Optimization in Machine Learning **Universität Mannheim**
FSS 2023 Prof. Simon Weißmann, Felix Benning

Sheet 1

For the exercise class on the 02.03.2023.
Hand in your solutions by 12:00 in the exercise on Thursday 02.03.2023.

Exercise 1 (Convex Examples). **(2 Points)**

Prove the following functions are convex

(i) affine linear functions, i.e. $f(x) = a^T x + c$ for $a \in \mathbb{R}^d$, $c \in \mathbb{R}$, (0.5 pts)

(ii) norms, i.e. $x \mapsto \|x\|$, (0.5 pts)

(iii) sums of convex functions f_k, i.e. $f(x) = \sum_{k=1}^n f_k(x)$, (0.5 pts)

(iv) $F(x) := \sup_{f \in \mathcal{F}} f(x)$ for a set of convex functions $\mathcal{F}$. (0.5 pts)

Exercise 2 (Finite Jensen). **(2 Points)**

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

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

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 (Strict & Strong Convexity). **(4 Points)**

Prove the following statements

(i) μ-strong convexity implies strict convexity. (1 pt)

(ii) For twice differentiable f, the following are equivalent (3 pts)

(a) $\nabla^2 f(x) \succeq \mu \mathbb{I}$
(b) $z^T \nabla^2 f(x) z \geq \mu \|z\|^2$
(c) f is μ-strongly convex

where $\mathbb{I}$ is the identity matrix and

$$A \succeq B :\iff A - B \text{ is (weakly) positive definite.}$$

Exercise 4 (Convexity and Minima). **(3 Points)**

Prove the following statements

(i) If f is convex, then every local minimum is also a global minimum. (1 pt)

(ii) If f is strictly convex, then there exists exactly one minimum. (1 pt)

(iii) If f convex and differentiable and $\nabla f(x^*) = 0$, then x^* is a minimum. (1 pt)

Exercise 5 (Directional Minima). **(3 Points)**

Let f be some differentiable function. For every direction $d \in \mathbb{R}^n$ define

$$g_d(\alpha) := f(x^* + \alpha d).$$

Assume that for every d, g_d is minimized by $\alpha = 0$. Prove that

(i) We have necessarily $\nabla f(x^*) = 0$. (1 pt)

(ii) $f(x^*)$ is *not* necessarily a minimum of f. (2 pts)

Hint. *Let* $0 < p < q$ *and define*

$$f(y, z) := (z - py^2)(z - qy^2)$$

consider $x^* = (0, 0)$ *and prove that* $f(y, my^2) < 0$ *for* $p < m < q$.

Exercise 6 (Bregman Divergence). **(10 Points)**

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

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

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

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

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

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

What is the minimizer? (2 pts)

Hint. *Minimize the upper bound* (1). *For this, 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 (2 pts)

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

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 (1 pt)

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

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. (1 pt)

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 (4 pts)

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

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*

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

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

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

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.