

Solution Sheet 2

For the exercise class on the 15.10.2026 at 12:00.

Exercises from the in-class test

Exercise 1 (Strict & Strong Convexity).

(0 Points)

Prove the following statements

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

Solution. Assuming $x \neq y$ we have

$$0 < \frac{\mu}{2} \|x - y\|^2 \stackrel{\mu\text{-strong conv.}}{\leq} f(x) - f(y) - \langle \nabla f(y), x - y \rangle$$

But after moving the negative parts to the left we are left with a condition which is equivalent to strict convexity by Proposition A.1.8 in the lecture. $\square$

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

- (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

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

Solution. Let us get (a) $\Leftrightarrow$ (b) out of the way first:

$$z^T \nabla^2 f(x) z - \mu \|z\|^2 = \langle z, \nabla^2 f(x) z \rangle - \langle z, \mu \mathbb{I} z \rangle = z^T (\nabla^2 f(x) - \mu \mathbb{I}) z$$

If we have (a), then the rightmost side is positive and thus we also have (b). For the other direction we simply start from the left.

For (c) $\Rightarrow$ (a) let

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

because f is μ strongly convex

$$\begin{aligned} g(y) &= f(y) - \frac{\mu}{2} \|y\|^2 \\ &\stackrel{f \text{ } \mu\text{-strong conv.}}{\geq} f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2 - \frac{\mu}{2} \|y\|^2 \\ &= \underbrace{f(x) - \frac{\mu}{2} \|x\|^2}_{=g(x)} + \frac{\mu}{2} \|x\|^2 + \underbrace{\langle \nabla f(x) - \mu x, y - x \rangle}_{\nabla g(x)} + \mu \langle x, y - x \rangle - \mu \langle x, y \rangle + \frac{\mu}{2} \|x\|^2 \\ &= g(x) + \langle \nabla g(x), y - x \rangle \end{aligned}$$

g is still convex. So we know by the lecture that

$$\nabla^2 g(x) = \nabla^2 f(x) - \mu \mathbb{I}$$

is positive definite.

On the other hand for (a) $\Rightarrow$ (c): if $\nabla^2 f \succeq \mu \mathbb{I}$, we know that $\nabla^2 g$ is positive definite and thus that g is convex by the lecture. Reusing the equations above except for the inequality we obtain

$$\begin{aligned} f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2 - \frac{\mu}{2} \|y\|^2 &= g(x) + \langle \nabla g(x), y - x \rangle \\ &\stackrel{g \text{ conv.}}{\leq} g(y) \\ &= f(y) - \frac{\mu}{2} \|y\|^2 \end{aligned}$$

Adding $\frac{\mu}{2} \|y\|^2$ on both sides results in the definition of μ -strong convexity of f . $\square$

Exercise 2 (Directional Minima).

(0 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 $\nabla f(x^*) = 0$.

Solution. If we select a standard basis vector as the direction, then

$$0 = g'_{e_i}(\alpha) \Big|_{\alpha=0} = \langle \nabla f(x^* + \alpha e_i), e_i \rangle \Big|_{\alpha=0} = \langle \nabla f(x^*), e_i \rangle$$

So the i -th entry of $\nabla f(x^*)$ is zero. As i was arbitrary, we are done. $\square$

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

Solution. Let $0 < p < q$ and define

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

then for $p < m < q$ we have for $y \neq 0$

$$f(y, my^2) = (my^2 - py^2)(my^2 - qy^2) = \underbrace{y^2}_{>0} \underbrace{(m-p)}_{>0} \underbrace{(m-q)}_{<0} < 0.$$

This means that $f(0, 0) = 0$ is not a minimum since we can get arbitrarily close to $x^* = (0, 0)$ with (y, my^2) with smaller and smaller y .

Let $d = (d_1, d_2)$ be an arbitrary direction. Then

$$\begin{aligned} g_d(\alpha) &= (\alpha d_2 - p(\alpha d_1)^2)(\alpha d_2 - q(\alpha d_1)^2) \\ &= \alpha^2 (d_2 - p\alpha d_1^2)(d_2 - q\alpha d_1^2) \\ &= \alpha^2 (d_2^2 - \alpha d_2(p+q)d_1^2 + \alpha^2 pqd_1^2) \end{aligned}$$

Thus $g'(0) = 0$ and $g''(0) = 2d_2^2 > 0$ which implies g has a minimum in zero. $\square$

Exercise 3 (Convergence to Stationary Point).**(0 Points)**

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a continuously differentiable function.

(i) Let $(x_k)_{k \in \mathbb{N}}$ be defined by gradient descent

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad x_0 \in \mathbb{R}^d$$

with diminishing step size $\alpha_k > 0$ such that $\sum_{k=1}^{\infty} \alpha_k = \infty$. Suppose that $(x_k)_{k \in \mathbb{N}}$ converges to some $x_* \in \mathbb{R}^d$. Prove that x_* is a stationary point of f , i.e. $\nabla f(x_*) = 0$.

Solution. For any $\epsilon > 0$ there exists $n \geq 0$ such that for all $i, j \geq n$ we have by Cauchy-Schwarz and convergence of $\nabla f(x_i)$ to $\nabla f(x_*)$ due to continuity of ∇f

$$\begin{aligned} & \langle \nabla f(x_i), \nabla f(x_j) \rangle \\ &= \|\nabla f(x_*)\|^2 + \langle \nabla f(x_i) - \nabla f(x_*), \nabla f(x_*) \rangle + \langle \nabla f(x_i), \nabla f(x_j) - \nabla f(x_*) \rangle \\ &\stackrel{\text{C.S.}}{\geq} \|\nabla f(x_*)\|^2 - \underbrace{\|\nabla f(x_i) - \nabla f(x_*)\|}_{\leq \epsilon} \|\nabla f(x_*)\| - \underbrace{\|\nabla f(x_i)\|}_{\leq \|\nabla f(x_*)\| + \epsilon} \underbrace{\|\nabla f(x_j) - \nabla f(x_*)\|}_{\leq \epsilon} \\ &\geq \|\nabla f(x_*)\|^2 - 2\epsilon \|\nabla f(x_*)\| - \epsilon^2 =: p(\epsilon) \end{aligned}$$

This results in

$$\begin{aligned} \|x_n - x_m\|^2 &= \left\| \sum_{k=n}^{m-1} \alpha_k \nabla f(x_k) \right\|^2 = \sum_{i,j=n}^{m-1} \alpha_i \alpha_j \langle \nabla f(x_i), \nabla f(x_j) \rangle \\ &\geq \left(\sum_{k=n}^{m-1} \alpha_k \right)^2 p(\epsilon) \end{aligned}$$

Taking the limit over m results in

$$\infty > \|x_n - x_*\|^2 \geq \underbrace{\left(\sum_{k=n}^{\infty} \alpha_k \right)^2}_{=\infty} p(\epsilon).$$

So we necessarily need $p(\epsilon) \leq 0$. But as ϵ was arbitrary, we have

$$0 \leq \|\nabla f(x_*)\|^2 = \lim_{\epsilon \rightarrow 0} p(\epsilon) \leq 0. \quad \square$$

(ii) Assume that f is also L -smooth. Prove for x_n generated by gradient descent with constant step size $\alpha \in (0, \frac{2}{L})$ we have

$$\sum_{k=n}^m \|\nabla f(x_k)\|^2 \leq \frac{f(x_n) - f(x_m)}{\alpha(1 - \frac{L}{2}\alpha)} \leq \frac{f(x_n) - \min_x f(x)}{\alpha(1 - \frac{L}{2}\alpha)}$$

for any $n, m \in \mathbb{N}$. Deduce for the case $\min_x f(x) > -\infty$, that we have

$$\min_{k \leq n} \|\nabla f(x_k)\|^2 \in o(1/n).$$

Solution. By L -smoothness of f , we have

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), \overbrace{x_{k+1} - x_k}^{-\alpha \nabla f(x_k)} \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2 \\ &= f(x_k) - (\alpha - \frac{L}{2} \alpha^2) \|\nabla f(x_k)\|^2 \end{aligned}$$

and therefore

$$\sum_{k=n}^m \|\nabla f(x_k)\|^2 \leq \sum_{k=n}^m \frac{f(x_k) - f(x_{k+1})}{\alpha(1 - \frac{L}{2} \alpha)} \stackrel{\text{telescope}}{=} \frac{f(x_n) - f(x_m)}{\alpha(1 - \frac{L}{2} \alpha)}.$$

Now $a_n := \min_{k \leq n} \|\nabla f(x_k)\|^2$ is non-increasing, therefore

$$na_{2n} \leq \sum_{k=n}^{2n} a_k \leq \sum_{k=n}^{\infty} a_k \rightarrow 0 \quad (n \rightarrow \infty).$$

Thus $a_{2n} \in o(1/n)$. And we can simply bound the odd elements of the sequence

$$a_{2n+1} \leq a_{2n} \in o(1/n). \quad \square$$

Exercise 4 (Sub-gradients).

(0 Points)

Let $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ be convex functions.

- (i) Prove that $\partial f(x)$ is a convex set for any $x \in \mathbb{R}^d$.

Solution. Let $g_1, g_2 \in \partial f(x)$. Then for any $\lambda \in [0, 1]$ and any $y \in \mathbb{R}^d$

$$\begin{aligned} f(y) &= \lambda f(y) + (1 - \lambda) f(y) \\ &\stackrel{g_1, g_2 \in \partial f(x)}{\geq} \lambda \left(f(x) + \langle g_1, y - x \rangle \right) + (1 - \lambda) \left(f(x) + \langle g_2, y - x \rangle \right) \\ &= f(x) + \langle \lambda g_1 + (1 - \lambda) g_2, y - x \rangle \end{aligned}$$

thus $\lambda g_1 + (1 - \lambda) g_2 \in \partial f(x)$ by definition. $\square$

- (ii) For $h(x) = f(Ax + b)$ prove $\partial h(x) \supseteq A^T \partial f(Ax + b)$. Prove equality for invertible A .

Solution. Let $g_x \in \partial f(x)$ i.e. $g_{Ax+b} \in \partial f(Ax + b)$. Then $A^T g_{Ax+b} \in \partial h(x)$ because

$$\begin{aligned} h(y) &= f(Ay + b) \geq f(Ax + b) + \langle g_{Ax+b}, (Ay + b) - (Ax + b) \rangle \\ &= h(x) + \langle A^T g_{Ax+b}, y - x \rangle. \end{aligned}$$

If A is invertible, we have $f(x) = h(A^{-1}x - A^{-1}b)$ so by the previous statement with $\tilde{A} = A^{-1}$ and $\tilde{b} = -A^{-1}b$, we get the other direction. $\square$

Exercise 5 (PL-Inequality).

(0 Points)

Assume $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is L -smooth and satisfies the Polyak-Łojasiewicz inequality

$$\|\nabla f(x)\|^2 \geq 2c(f(x) - f_*) \quad (\text{PL})$$

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

(i) Prove that gradient descent with fixed step size $\alpha_k = \frac{1}{L}$ converges linearly in the sense

$$f(x_k) - f_* \leq (1 - \frac{c}{L})^k (f(x_0) - f_*).$$

Solution. By L -smoothness and the descent lemma, we have

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|^2 \stackrel{\text{(PL)}}{\leq} f(x_k) - \frac{c}{L} (f(x_k) - f_*).$$

Subtracting f_* from both sides, we get

$$f(x_{k+1}) - f_* \leq (1 - \frac{c}{L})(f(x_k) - f_*) \quad \square$$

(ii) Prove that μ -strong-convexity and L -smoothness imply the PL-inequality.

Solution. Recall by the solution of sheet 1, exercise 6 (iii), and strong convexity we have

$$\begin{aligned} \mu \|x - y\|^2 &\leq D_f^{(B)}(x, y) + D_f^{(B)}(y, x) \\ &= \langle \nabla f(x) - \nabla f(y), x - y \rangle \\ &\stackrel{\text{C.S.}}{\leq} \|\nabla f(x) - \nabla f(y)\| \|x - y\| \end{aligned}$$

and therefore

$$\mu \|x - y\| \leq \|\nabla f(x) - \nabla f(y)\|. \quad (1)$$

Finally we know by L -smoothness and $\nabla f(x_*) = 0$ where x_* is the minimum

$$f(x) - f(x_*) \stackrel{\nabla f(x_*)=0}{=} D_f^{(B)}(x, x_*) \stackrel{L\text{-smooth}}{\leq} \frac{L}{2} \|x - x_*\|^2 \stackrel{(1)}{\leq} \frac{L}{2\mu} \|\nabla f(x) - \underbrace{\nabla f(x_*)}_{=0}\|^2. \quad \square$$