FAKULTÄT FÜR WIRTSCHAFTSINFORMATIK
UND WIRTSCHAFTSMATHEMATIK

UNIVERSITÄT MANNHEIM

Prof. Dr. Leif Döring
Benedikt Wille

Reinforcement Learning

10. Exercise Sheet - Solutions

13.05.2025

1. (μ-strong) convexity

a) Prove that all local minima of convex functions must be global minima. In particular, this means that all global minima have same height.

Solution:
Let x^* be a local minimum of a convex function f and assume there exists x_0 such that $f(x_0) < f(x^*)$. By convexity for every $t \in (0,1]$ we then have

$$f\big((1-t)x^* + tx_0\big) \;\leq\; (1-t)f(x^*) + tf(x_0) \;<\; (1-t)f(x^*) + tf(x^*) \;=\; f(x^*).$$

Then for every ϵ-ball around x^* there exists a t_0 such that $\|(1-t_0)x^* + t_0x_0 - x^*\| < \epsilon$ but $f((1-t)x^* + t_0x_0) < f(x^*)$, which is a contradiction. Thus, x^* must be a global minimum.

b) Show that a μ-strongly convex function has a unique global minimum.

Solution:
Let $x_0 = 0$ be fixed. From the μ-strong convexity and we know that

$$f(y) \;\geq\; f(0) + y^T\nabla f(0) + \frac{\mu}{2}\|y\|^2 \;\geq\; f(0) + c\|y\|_1 + \frac{\mu}{2}\|y\|_1^2,$$

where we used wlog. the 1-norm because of equivalence of norms and $c = \min_i\big(\nabla f(0)\big)_i$. Therefore we know that there exists an $R > 0$ such that for $\|y\| > R$ it holds $f(y) \geq f(0)$. Thus, if a minimum exists, it must lie in the closure of $B(0,R)$, which is compact. Since we assumed μ-strongly convex functions to be differentiable this yields a unique global minimum. Note, that with a more general definition of μ-strong convexity that involves sub-gradients this proof does not work!

c) Let $f : \mathbb{R}^d \to \mathbb{R}$ with $f(x) = x^TAx$ for a matrix $A \in \mathbb{R}^{d\times d}$. Show that f is L-smooth and μ-strongly convex, where L is twice the largest and μ twice the smallest singular value.

Solution:
Due to the equivalence of norms, the definition of the operator norm, and the fact that $\nabla f(x) = (A^T + A)x$ it holds

$$\frac{\|\nabla f(x) - \nabla f(y)\|_2}{\|x-y\|_2} \;=\; \frac{\|(A^T+A)(x-y)\|_2}{\|x-y\|_2} \leq 2\|A\|_2 =: L$$

Because the 2-norm matches the largest singular value the first part of the assertion is proved. With the singular value decomposition and the fact that orthogonal matrices do not change the length of a vector we obtain

$$y^TAy \;=\; y^TO\Sigma V^*y \;\geq\; \frac{\mu}{2}y^TOV^*y \;=\; \frac{\mu}{2}\|y\|_2^2,$$

where $\mu = \min_i \Sigma_{ii}$, i.e. the smallest singular value. This immediately yields

$$\begin{aligned} f(x+y) &= (x+y)^T A(x+y) = x^T Ax + x^T Ay + y^T Ax + y^T Ay \\ &\geq x^T Ax + y^T (A + A^T)x + \frac{\mu}{2}\|y\|_2^2 = f(x) + y^T \nabla f(x) + \frac{\mu}{2}\|y\|_2^2 \end{aligned}$$

and with $y = z - x$ the μ-strong convexity.

2. PL-condition

a) Prove that μ-strong convexity implies the PL-condition (4.5), i.e.

$$\|\nabla f(x)\|^2 \geq 2r(f(x) - f_*) \quad \forall x \in \mathbb{R}^d \tag{1}$$

for $r = \mu$ and $f_* = \min_{x \in \mathbb{R}^d} f(x) > -\infty$.

Solution:

Recall by the definition of μ-strong convexity, that

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2}\|y - x\|^2.$$

In the first exercise we showed that there exists a global minimum with value $f_* > -\infty$. Minimizing both sides of the equation still fulfills the inequality:

$$\min_y f(y) = f_*,$$

and

$$\nabla f(x) + \mu(y - x) \stackrel{!}{=} 0 \quad \Leftrightarrow \quad y = x - \frac{1}{\mu}\nabla f(x).$$

As the second derivative wrt y of the RHS is $\mu > 0$ we see that this is a minimum. Plugging in both minima we obtain that

$$\begin{aligned} f_* &\geq f(x) + \nabla f(x)^T (x - \frac{1}{\mu}\nabla f(x) - x) + \frac{\mu}{2}\|x - \frac{1}{\mu}\nabla f(x) - x\|^2 \\ &= f(x) - \frac{1}{\mu}\|\nabla f(x)\|^2 + \frac{1}{2\mu}\|\nabla f(x)\|^2 \\ &= f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2. \end{aligned}$$

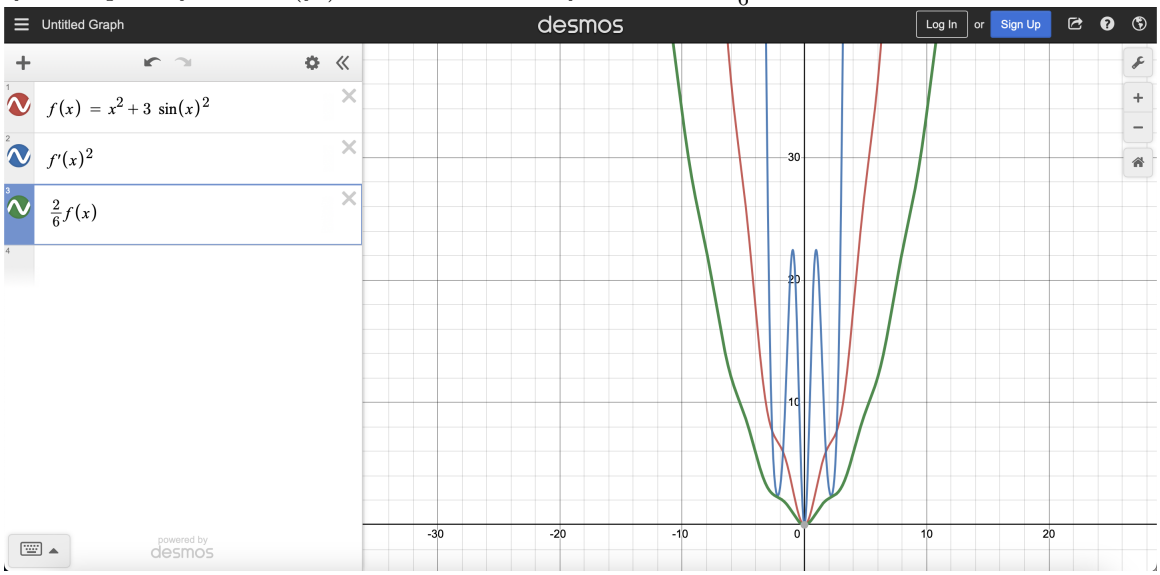
Rearranging results yields

$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f_*).$$

b) Show that $f(x) = x^2 + 3\sin^2(x)$ satisfies the PL-condition and prove that f is not convex. Plot the function to see why gradient descent converges.
Hint: The plot can also help to find the parameter r of the PL-condition.

Solution:

If we plot f and $(f')^2$ we see that for $r = \frac{1}{6}$ we have the PL-condition.



We argue why large x are not a problem: we have $f'(x) = 2(x + 3\sin(x)\cos(x))$ and therefore

$$\begin{aligned} f'(x)^2 &= 4(x + 3\sin(x)\underbrace{\cos(x)}_{\in[-1,1]})^2 \\ &\overset{|x|\geq 3}{\geq} 4(|x| - 3)^2 \\ &= 4x^2 - 24|x| + 36 \\ &= \frac{1}{3}x^2 + \underbrace{\frac{11}{3}x^2 - 3(8|x|)}_{\geq 0,\ for\ x \geq 8} + 36 \\ &\geq \frac{1}{3}(x^2 + 3\sin(x)^2), \end{aligned}$$

for $x \geq 8$. $x \leq 8$ is clear from the plot.

Further, f is not convex, as

$$f(\frac{1}{2}\pi + \frac{1}{2}0) = \frac{\pi^2}{4} + 3 > \frac{1}{2}f(\pi) + \frac{1}{2}f(0).$$

3. Stochastic gradient descent

In the lecture we proved convergence of SGD to stationary points if the function is L-smooth and bounded from below and we use stepsizes fulfilling the Robbins-Monro conditions. Consider the setting from the theorem of the lecture and additionally assume μ-strong convexity. Prove that $\|X_n - x_*\| \to 0$ almost surely.

Solution:

We proved above that the PL inequality is satisfied for μ-strongly convex functions. This implies

$$\frac{\mu}{2}\|X_n - x_*\|^2 \leq F(X_n) - F(x_*) - \underbrace{\langle \nabla F(x_*)}_{=0}, X_n - x_* \rangle \overset{PL}{\leq} \frac{1}{2\mu}\|\nabla F(X_n)\|^2 \to 0.$$

The sandwich theorem now implies $F(X_n) \to F(x_*)$ almost surely.

4. Fully deterministic game with non-convex value function

Imagine a game where you have two states s_1 and s_2 and two actions a_1 and a_2 in each of the states. The game is fully deterministic and after each turn switches, regardless of the action, between s_1 s_2, giving, regardless of the state, a reward of 1 in a_1 and -1 in a_2 and $\gamma \in (0, 1)$. Assume a "learner", that decides based on a one-dimensional linear softmax with features $\phi(s, a) \in \mathbb{R}$. Show that there exist (non-trivial) values of the features such that the value function of this game is non-convex in the parameter θ of the softmax family.

Solution:
The game dynamics imply $p(s_i; s_j, a) = 1 - \delta_{ij}$ for all actions a and $r(s, a_i) = \delta_1(i) - \delta_2(i)$ for all states s. Because the policies are measures we further know $\sum_a \pi_\theta(a; s) = 1$ for all states s. Additionally recall the solution to the Bellman expectation operator in vector notation:

$$V^\pi \;=\; (1 - \gamma P^\pi)^{-1} r^\pi,$$

where

$$P^\pi \;=\; \left(\sum_a \pi(a; s) p(s'; s, a) \right)_{(s,s')} \quad \text{and}$$

$$r^\pi \;=\; \left(\sum_a \pi(a; s) r(s, a) \right)_s .$$

With the observations above, we can deduce

$$P^\pi \;=\; \begin{pmatrix} 0 & \sum_a \pi_\theta(a, s_1) \\ \sum_a \pi_\theta(a, s_2) & 0 \end{pmatrix} \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{and } r^\pi \;=\; \begin{pmatrix} \pi_\theta(a_1, s_1) - \pi_\theta(a_2, s_1) \\ \pi_\theta(a_1, s_2) - \pi_\theta(a_2, s_2) \end{pmatrix}.$$

Now we choose the features $\phi(a_i, s) = \delta_1(i) - \delta_2(i)$ for all states s. The inverse of $1 - \gamma P^\pi$ only changes constants and does not depend on the parameter so an analysis of r^π suffices. We calculate

$$r^\pi \;=\; \begin{pmatrix} \frac{\exp(\theta) - \exp(-\theta)}{\exp(\theta) + \exp(-\theta)} \\ \frac{\exp(\theta) - \exp(-\theta)}{\exp(\theta) + \exp(-\theta)} \end{pmatrix},$$

which is non-convex in θ.

5. Non-convex, L-smooth value function

Imagine a game with a starting state, two states s_0 and s_1 as well as a terminal state T. A player in the starting state may decide to transfer to either state and receive rewards R_0 and R_1 accordingly, after which he is transferred to the terminal state. Consider the (differentiable parameterised) family of stationary policies $\{\pi^\theta\}_{\theta \in [0,1]}$ with

$$\pi_\theta = \theta^2 \delta_{\to s_0} + (1 - \theta^2) \delta_{\to s_1}$$

that transfers to s_0 with probability θ^2 and to s_1 with probability $1 - \theta^2$.

a) Assume that R_0 and R_1 are chosen uniformly and independently according to a continuous probability distribution. Show that the value function is non-convex with a probability of at least $\frac{1}{2}$.

Solution:

The value function in all states other than the starting state is zero because we can only receive rewards in the after the first decision, so we drop the dependence on the state. The value function of this one-step MDP in dependence on the parameter θ can be directly computed for fixed R_0 and R_1:

$$J(\theta) \;=\; \mathbb{E}[R] \;=\; R_0\theta^2 + R_1(1-\theta^2).$$

This function is definitely non-convex whenever for $\lambda \in (0,1)$

$$\begin{aligned} J\big((1-\lambda)0+\lambda 1\big) \;&>\; (1-\lambda)J(0)+\lambda J(1) \\ \Leftrightarrow \lambda^2 R_0 + (1-\lambda^2)R_1 \;&>\; (1-\lambda)R_1 + \lambda R_0 \\ \Leftrightarrow (1-\lambda)(R_1-R_0) \;&>\; 0 \\ \Leftrightarrow R_1 \;&>\; R_0 \end{aligned}$$

If we choose R_1 and R_2 uniformly and independently from a continuous probability distribution it holds

$$\mathbb{P}(\Omega) \;=\; \mathbb{P}(R_1 = R_0) + \mathbb{P}(R_1 > R_0) + \mathbb{P}(R_1 < R_0) \;=\; 2\mathbb{P}(R_1 > R_0)$$

and therefore the probability of having $R_1 > R_0$ (in which case J is non-convex) is exactly $\frac{1}{2}$, which proves the assertion.

b) Show that regardless of the values of R_0 and R_1 the value function is L-smooth, so that the gradient descent with diminishing stepsizes from the lecture converges.

Solution:

The gradient of J is given by $\nabla J(\theta) = 2\theta R_0 - 2\theta R_1$ and thus direct computation reveals

$$\|\nabla J(\theta_1) - \nabla J(\theta_2)\| \;=\; \|2(R_0-R_1)(\theta_1-\theta_2)\| = L\|\theta_1-\theta_2\|$$

with $L = \|2(R_0-R_1)\|$ for $R_0 \neq R_1$, whereas for $R_0 = R_1$ it is trivially L-smooth for all values of L. Since Theorem 5.4.1 only needs an L-smooth function, gradient descent with diminishing stepsizes converges.

c) Show that, regardless of the values of R_0 and R_1, if you restrict the parameter θ to a subset $[\theta_0, \theta_1] \subset [0,1]$ it fulfills the PL-condition and find a constant stepsize, such that gradient descent with constant stepsize converges to a global minimum.

Solution:

First assume $R_0 = R_1 = R$. Then

$$\|\nabla J(\theta)\|^2 \;=\; \|2\theta(R_0-R_1)\|^2 \;=\; 0 \text{ and } J(\theta) \;=\; R \;=\; \min_{\theta} J(\theta)$$

and thus the PL-condition is fulfilled on $[0,1]$ *for any* r*. Now assume that* $R_0 > R_1$ *and* $\theta \in [\theta_0, \theta_1]$*. Then*

$$\|\nabla J(\theta)\|^2 \;=\; 2\theta(R_1 - R_0)^2 \text{ and } \min_\theta J(\theta) \;=\; R_0\theta_1^2 + R_1(1-\theta_1^2)$$

Then we see that the PL-condition is fulfilled for $r = \frac{\theta_0^2}{\theta_1^2-\theta_0^2}\|2(R_0-R_1)\|$*:*

$$\begin{aligned}
\|\nabla J(\theta)\|^2 \;&\geq\; 2r(J(\theta) - J_*) \\
\Leftrightarrow 4\theta^2(R_0-R_1)^2 \;&\geq\; 2r\big(R_0\theta^2 + R_1(1-\theta^2) - R_0\theta_1^2 - R_1(1-\theta_1^2)\big) \\
\Leftrightarrow 4\theta^2(R_0-R_1)^2 \;&\geq\; 2r(R_1-R_0)(\theta_1^2-\theta^2) \\
\Leftrightarrow \frac{\theta^2}{\theta_1^2-\theta^2}\|2(R_0-R_1)\| \;&\geq\; \frac{\theta_0^2}{\theta_1^2-\theta_0^2}\|2(R_0-R_1)\| \;=\; r,
\end{aligned}$$

where switching to the last line we exclude the case $\theta = \theta_1$ *since in this case the inequality holds trivially. In the case of* $R_1 > R_0$ *the minimum of* J *is attained at* θ_0 *and the condition for the inequality similarly becomes*

$$\frac{\theta^2}{\theta^2-\theta_0^2}\|2(R_0-R_1)\| \;\geq\; \frac{\theta_0^2}{\theta_1^2-\theta_0^2}\|2(R_0-R_1)\| \;=\; r,$$

where analogously the case $\theta = \theta_0$ *is excluded due to trivially fulfilling the inequality in the line before. In the case* $R_0 = R_1$ *the function* J *is* L*-smooth with any chosen* L *and the PL-condition holds for any* r*, so by Theorem 5.3.1 convergence of the gradient descent holds with any constant stepsize. In order to obtain convergence with constant stepsize* $\alpha = \frac{1}{L} = \frac{1}{\|2(R_0-R_1)\|}$ *first we need that* $L > r$*. This is equivalent to*

$$1 \;>\; \frac{\theta_0^2}{\theta_1^2-\theta_0^2} \;\Leftrightarrow\; \theta_1^2 - \theta_0^2 \;>\; \theta_0^2 \;\Leftrightarrow\; \theta_1^2 - 2\theta_0^2 > 0.$$

Finally, we also need to have $r > 0$*, which is the case whenever additionally* $\theta_0 > 0$ *holds. Thus we can set* $\theta_1 = 1$ *and obtain convergence with constant stepsize* α *if we restrict the parameter* θ *to* $[\theta_0, 1]$ *for some* $\theta_0 \in \big(0, \frac{1}{4}\big)$*.*