Prof. Dr. Leif Döring
André Ferdinand, Sara Klein

Reinforcement Learning

11. Solution Sheet

1. Baseline trick

Write down and proof the baseline gradient representation for infinite discounted MDPs.
Solution:
We aim to prove

$$\nabla J_s(\theta) = \sum_{s' \in \mathcal{S}} p^{\pi^\theta}(s') \sum_{a \in \mathcal{A}_{s'}} \nabla \pi^\theta(a; s')\big(Q^{\pi^\theta}(s', a) - b\big),$$

for some $b \in \mathbb{R}$*. By the finiteness of the state and action space we have that*

$$\begin{aligned}&\sum_{s' \in \mathcal{S}} p^{\pi^\theta}(s') \sum_{a \in \mathcal{A}_{s'}} \nabla \pi^\theta(a; s') b \\ &= b \sum_{s' \in \mathcal{S}} p^{\pi^\theta}(s') \nabla \underbrace{\sum_{a \in \mathcal{A}_{s'}} \pi^\theta(a; s')}_{=1} \\ &= 0.\end{aligned}$$

Hence, the claim follows from the policy gradient theorem for discounted MDPs (5.2.6) in the lecture.

2. PL-condition

a) Prove that μ-strong convexity implies the PL-condition (5.4.), 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*

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

Minimizing both sides of the equation still fulfills the inequality, so we will calculate the derivatives for both cases separately.

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

and

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

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

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

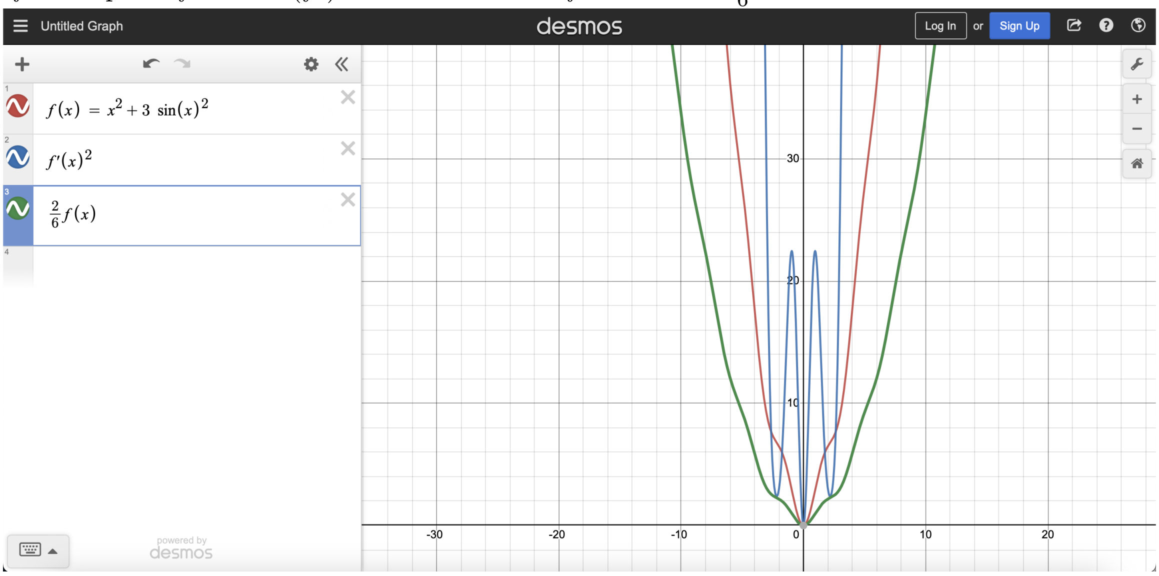
Rearranging results in

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

b) Show that $f(x) = x^2 + 3\sin^2(x)$ satisfies the PL-condition (1) 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. 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 L-smooth, strongly convex functions. This together with strong convexity implies

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

From part in the middle we get $F(X_n) \to F(x_*)$ for free.