Homework 3

Toffalini Leonardo

> **Exercise 1**
>
> Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function which is G-Lipschitz continuous, then $\|\nabla f(x)\| \leq G$ for all $x \in \mathbb{R}^n$.

Solution: By the first order characterization of convex functions:

$$f(x) + \nabla f(x) \cdot (y - x) \leq f(y).$$

Since f is G-Lipschitz:

$$\|f(y) - f(x)\| \leq G \cdot \|y - x\|.$$

Reordering the FOC we get the following:

$$\begin{aligned}
\nabla f(x) \cdot (y - x) &\leq f(y) - f(x) \\
\|\nabla f(x) \cdot (y - x)\| &\leq \|f(y) - f(x)\| \\
\|\nabla f(x)\| \cdot \|(y - x)\| &\leq \|f(y) - f(x)\| \\
\|\nabla f(x)\| &\leq \frac{\|f(y) - f(x)\|}{\|(y - x)\|} \leq G
\end{aligned}$$

□

> **Exercise 2**
>
> Give an example of a function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\|\nabla f(x)\|_2 \leq 1$ for all $x \in \mathbb{R}^n$ but the Lipschitz constant of its gradient is unbounded.

Solution: Basically we have to give a function F that is lies in the $[-1, 1]$ strip, but is not Lipschitz continuous, then find a function f for which $\nabla f = F$.

Let's just restrict ourselves to function of the form $F : \mathbb{R} \to \mathbb{R}$. A classic example of a bounded function that is not Lipschitz is $F(x) = \sin(1/x)$. Clearly $F(x) \in [0, 1]$ for all $x \in \mathbb{R}$. However near 0 it is not Lipschitz.

Now we just have to give a function such that $f'(x) = F(x)$, so just solve the following differential equation:

$$\begin{aligned}
f'(x) &= \sin\left(\frac{1}{x}\right) \\
f(x) &= \int_0^x \sin\left(\frac{1}{t}\right) \mathrm{d}t
\end{aligned}$$

By the Newton–Leibniz formula $f'(x) = \sin(1/x)$ if we define $\sin(1/t) = 0$ for $t = 0$. □

Exercise 3

Let $f(x) = \frac{1}{2}x^2$. Run a gradient descent with step size $\alpha \in (0, 2)$, that is $x_{t+1} = x_t - \alpha \nabla f(x_t)$.

1. Write the iteration explicitly.
2. For which values of α does the sequence converge to 0?

Solution: Let $x_0 = 2$ and $\alpha = 0.5$.

$$\begin{aligned}
x_1 &= x_0 - \alpha \nabla f(x_0) \\
x_1 &= 2 - 0.5 \cdot \nabla\left(\frac{1}{2}x_0^2\right) \\
x_1 &= 2 - 0.5 \cdot 2 \\
x_1 &= 1
\end{aligned}$$

For all values of $\alpha \in (0, 2)$ the method converges to $x^* = 0$. Since our only problem would arise if we have such a large step size that we end up higher on the other side of the parabola than where we started. Because, of course, we can only end up farther from the optimum on the otherside, since we always step towards the optimum.

This situation could only arise if $|x_k - \alpha x_k| > |x_k|$, which is only true when $\alpha > 2$ since α is positive. $\square$

Exercise 4

Suppose f is μ-strongly convex with L-Lipschitz continuous gradient. Prove that gradient descent with step size $\alpha \in (0, 2/L)$ satisfies the linear convergence bound

$$\left\|x_{k+1} - x^*\right\|^2 \leq (1 - \alpha\mu)\|x_k - x^*\|^2.$$

What is the optimal choice of α?

Solution: Let us denote $x_k - x^* = A$ and $\nabla f(x_k) = G$ for easier notation.

We can rewrite the left hand side of the desired inequality by substituing for the definition of how we got x_{k+1} with a single gradient step:

$$\left\|x_{k+1} - x^*\right\|^2 = \|x_k - \alpha\nabla f(x_k) - x^*\|^2 = \|A - \alpha G\|^2.$$

We can expand the norm square with a dot product as follows:

$$\|A - \alpha G\|^2 = \|A\|^2 - 2\alpha(A \cdot G) + \alpha^2\|G\|^2.$$

Now we just need to bound $A \cdot G$ by some values, for which we state two results:

$$G \cdot A \geq \mu\|A\|^2, \tag{1}$$

$$G \cdot A \geq \frac{1}{L}\|G\|^2. \tag{2}$$

With the above two results we can bound the above equation as follows:

$$\begin{aligned} \|A - \alpha G\|^2 &= \|A\|^2 - \alpha(A \cdot G) - \alpha(A \cdot G) + \alpha^2\|G\|^2 \\ &\leq \|A\|^2 - \alpha\mu\|A\|^2 - \frac{\alpha}{L}\|G\|^2 + \alpha^2\|G\|^2 \\ &= (1 - \alpha\mu)\|A\|^2 + \alpha\left(\alpha - \frac{1}{L}\right)\|G\|^2. \end{aligned}$$

For $\alpha \in \left(0, \frac{1}{L}\right)$

$$\alpha\left(\alpha - \frac{1}{L}\right)\|G\|^2 < 0,$$

so if we subtract it from the right hand side we get something that is greater or equal:

$$\|A - \alpha G\|^2 \leq (1 - \alpha\mu)\|A\|^2,$$

which is exactly what we were hoping to prove. □

Exercise 5

Let $f : \mathbb{R}^n \to \mathbb{R}$ be (possibly nonconvex) differentiable with L-Lipschitz continuous gradient, and let $\{x_k\}$ be generated by gradient descent with constant step size $\alpha = 1/L$. Show that after K iterations

$$\min_{0 \le k \le K-1} \|\nabla f(x_k)\|^2 \le \frac{2L(f(x_0) - \inf f)}{K}.$$

Solution: Lemma 6.3 from the Vishnoi Convex Optimization book states the following:

$$f(y) - f(x) - \nabla f(x) \cdot (y - x) \le \frac{L}{2}\|x - y\|^2,$$

for a function with an L-Lipschitz gradient. Notice, that the lemma does not rely on f being convex.

We can rewrite it as follows:

$$f(y) \le f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|y - x\|^2.$$

Let y be the result of a single gradient step from x, meaning $y = x - \alpha \nabla f(x)$, and plug this into the above inequality:

$$\begin{aligned} f(x - \alpha \nabla f(x)) &\le f(x) + \nabla f(x) \cdot \left(-\frac{1}{L}\nabla f(x)\right) + \frac{L}{2}\left\|-\frac{1}{L}\nabla f(x)\right\|^2 \\ &= f(x) - \frac{1}{L}\|\nabla f(x)\|^2 + \frac{1}{2L}\|\nabla f(x)\|^2 \\ &= f(x) - \frac{1}{2L}\|\nabla f(x)\|^2. \end{aligned}$$

Now, writing $x = x_k$ and thus $y = x_{k+1}$ we arrive at the following:

$$\|\nabla f(x_k)\|^2 \le 2L(f(x_k) - f(x_{k+1})).$$

With the gradient at a single point bounded by above, we are almost at the finish line, we just need to bound the minimum by above:

$$\min_{0 \le k \le K-1} \|\nabla f(x_k)\|^2 \le \frac{1}{K}\sum_{k=0}^{K-1} \le \frac{2L}{K}\sum_{k=0}^{K-1}(f(x_k) - f(x_{k+1})).$$

Notice, that the rightmost sum is telescoping and all that remains is the first and last elements:

$$= \frac{2L}{K}(f(x_0) - f(x_K)).$$

For the last step we just realize that $\inf f < f(x_K)$, so

$$\min_{0 \le k \le K-1} \|\nabla f(x_k)\|^2 \le \frac{2L}{K}(f(x_0) - \inf f).$$

$\square$

Exercise 6

Consider the quadratic function

$$f(x,y) = \frac{1}{2}(3x^2 + 2xy + 4y^2).$$

1. Write down its gradient and compute the Lipschitz constant L of the gradient.
2. Perform one step of gradient descent with step size $\alpha = 1/L$, starting from $(x_0, y_0) = (1,1)$.

Solution:

$$\nabla f(x,y) = \begin{bmatrix} 3x + 1y \\ 1x + 4y \end{bmatrix}$$

We need to find an $L > 0$ such that

$$\|\nabla f(x_1,y_1) - \nabla f(x_2,y_2)\| \leq L\|(x_1,y_1) - (x_2,y_2)\|.$$

Expanding the gradient we arrive at the following inequality that gives us L:

$$\begin{aligned} \left\| \begin{bmatrix} 3x_1 + 1y_1 \\ 1x_1 + 4y_1 \end{bmatrix} - \begin{bmatrix} 3x_2 + 1y_2 \\ 1x_2 + 4y_2 \end{bmatrix} \right\| &= \left\| \begin{bmatrix} 3(x_1 - x_2) + 1(y_1 - y_2) \\ 1(x_1 - x_2) + 4(y_1 - y_2) \end{bmatrix} \right\| \\ &= \left\| \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} x_1 - x_2 \\ y_1 - y_2 \end{bmatrix} \right\| \\ &\leq \left\| \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} \right\| \cdot \left\| \begin{bmatrix} x_1 - x_2 \\ y_1 - y_2 \end{bmatrix} \right\|. \end{aligned}$$

Thus $L = \left\| \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} \right\|$. For simplicity lets denote this matrix by A. We know that the 2-norm of a matrix is calculated as follows:

$$\|A\|_2 = \sqrt{\lambda_{\max}(AA^T)}.$$

This calculation is quite tiresome. However, if we are free to choose the norm, we can just choose the infinity norm, for which $L = \|A\|_\infty = 5$. □

Solution: Let $z = (x,y)$, thus $z_0 = (x_0, y_0) = (1,1)$.

$$\begin{aligned} z_1 &= z_0 - \nabla f(z_0) \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{5} \cdot \begin{bmatrix} 3+1 \\ 1+4 \end{bmatrix} = \begin{bmatrix} 1 - 0.8 \\ 1 - 1 \end{bmatrix} = \begin{bmatrix} -0.2 \\ 0 \end{bmatrix} \end{aligned}$$

□

Supplementary proofs

Proposition 0.1

If f is a μ-strongly convex function, then

$$(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \mu \|x - y\|^2.$$

Proof: Write out the definition of μ-strong convexity in both directions:

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

Summing these two inequalities we get the following:

$$\begin{aligned} f(x) + f(y) &\geq f(y) + f(x) + \nabla f(x) \cdot (y - x) + \nabla f(y) \cdot (x - y) + \mu\|x - y\|^2 \\ 0 &\geq (\nabla f(x) - \nabla f(y)) \cdot (y - x) + \mu\|x - y\|^2 \\ (\nabla f(x) - \nabla f(y)) \cdot (x - y) &\geq \mu\|x - y\|^2. \end{aligned}$$

With this we proved the proposition, and we can see that Equation 1 used in our solution can be achieved by the following variable assignment $x = x_k$ and $y = x^*$. □

Proposition 0.2 **Baillon–Haddad**

If f has an L-Lipschitz continuous gradient and is defined on a convex domain, then

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

Proof: The proof is quite involved and this is a somewhat well-known result so we just present the outline of the proof and how Equation 2 follows from this proposition. The following blog post has a great explanation of this result: https://samuelvaiter.com/a-first-look-at-convex-analysis/

1. $$(\nabla f(x) - \nabla f(y)) \cdot (x - y) \overset{\text{C.S.}}{\leq} \|\nabla f(x) - \nabla f(y)\| \cdot \|x - y\| \leq L\|x - y\|^2$$

2. From the previous we can prove the following:

$$f(y) \leq f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2$$

3. By the above result we can bound the difference of any function value to the optimum value:

$$\frac{1}{2L}\|\nabla f(z)\|^2 \leq f(z) - f(x^*) \leq \frac{L}{2}\|z - x^*\|^2,$$

where x^* is the minimizer of f.

4. $$f(y) - f(x) - \nabla f(x) \cdot (x - y) \geq \frac{1}{2L}\|\nabla f(y) - \nabla f(y)\|^2$$

5. Writing the previous inequality once for x and once for y and summing them up we arrive at our destination:

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

We can see that Equation 2 which we relied on in our solution is achieved by the variable substitution $x = x_k$ and $y = x^*$, since $\nabla f(x^*) = 0$ because it is the minimizer. $\square$