Numerical Optimization - Homework 1

Andy Tràn

October 12, 2022

Problem 1

1. Let $g(x) = \mathbf{x}^T A\mathbf{x}$, $h(x) = \mathbf{x} + \mathbf{b}$, then $f(x) = g(h(x))$. The gradient is $\nabla f(x) = h'(x)^T \nabla g(h(x)) = I \cdot (A + A^T)(\mathbf{x} + \mathbf{b}) = (A + A^T)(\mathbf{x} + \mathbf{b})$. $Hess_f(x) = A + A^T$

2. $f(\mathbf{x}) = \mathbf{b}^T(A\mathbf{x}\text{-}\mathbf{y}) = \mathbf{b}^T A\mathbf{x} - \mathbf{b}^T y$, thus $\nabla f(x) = A^T b$ and $Hess_f(x) = 0$ (zero matrix)

3. $f(x) = \|Ax - b\|^2 = (Ax - b)^T(Ax - b)$, let $h(x) = x^T x$, $g(x) = Ax - b$. We have $f(x) = h(g(x))$, thus $\nabla f(x) = g'(x)^T \cdot \nabla h(g(x)) = A^T \cdot 2(Ax - b)$ and $Hess_f(x) = 2A^T A$

4. We can employ the chain rule by writing down, $h(x) = \sqrt{(}x), x \in \mathbb{R}$, and $g(x) = \|Ax - b\|^2$, thus $f(x) = h(g(x))$. Calculating $\nabla f(x) = \nabla g(x) \cdot h'(g(x)) = \frac{2A^T(Ax-b)}{2\|Ax-b\|}$. Now the harder part comes:
   Let $\kappa(x) = \frac{1}{\|Ax-b\|}$ and $\tau(x) = A^T(Ax - b)$.We can write it as $\nabla f(x) = \tau(x)\kappa(x)$. Then we can employ the chain rule to calculate the derivative (Hessian). We can calculate these terms individually:

$$
\begin{aligned}
Hess_f(x) &= \tau(x)d\kappa(x) + d\theta(x)\kappa(x) \\
&= A^T(Ax - b) \cdot -\frac{(x^T A^T - b^T)A}{\|Ax - b\|^3} + A^T A \cdot \frac{1}{\|Ax - b\|} \\
&= \frac{A^T A}{\|Ax - b\|} - \frac{(A^T Ax - b)((x^T A^T - b^T)A)}{\|Ax - b\|^3}
\end{aligned}
$$

Problem 2 (Convex functions)

- Let's show $g_i(x) = |a_i^T x + b_i|$ is convex for any vector a_i and scalar b_i. Let those values be arb. but fix.

$$
\begin{aligned}
g_i(\lambda x + (1 - \lambda)y) &= |a_i^T(\lambda x + (1 - \lambda)y) + b_i| \\
&= |a_i^T \lambda x + a_i^T(1 - \lambda)y + b_i| \\
&= |a_i^T \lambda x + \lambda b_i + a_i^T(1 - \lambda)y + (1 - \lambda)b_i| \\
&\leq |a_i^T \lambda x + \lambda b_i| + |a_i^T(1 - \lambda)y + (1 - \lambda)b_i| \\
&= \lambda|a_i^T x + b_i| + (1 - \lambda)|a_i^T y + b_i| \\
&= \lambda g_i(x) + (1 - \lambda)g_i(y)
\end{aligned}
$$

  Then the pointwise maximum is also convex, due to the theorem from the lecture.

- Firstly, we need to show $g(x) = x^3$ is convex on $[0, \infty[$, which is rather straightforward since we see $g''(x) = 6x \geq 0 \forall x \in [0, \infty[$ (and strictly positive for $x \neq 0$).We also need to show $h(x) = \|Ax - b\|$ is convex, which it is from the proposition from the slide where all norms are convex and functions composed with affine functions are also convex (else prove it quickly using the definition of convexity and triangle inequality).

$$\begin{aligned} f(\lambda v + (1-\lambda)w) &= g(h(\lambda v + (1-\lambda)w)) \\ &\leq g(h(\lambda v) + (1-\lambda)w) \\ &\leq g(h(\lambda v)) + g(h((1-\lambda)w)) \\ &= f(\lambda v) + f((1-\lambda)w) \end{aligned}$$

The first inequality follows from the convexity of $h(x)$ and the second from convexity of $g(x)$.

- In the previous step (2) we proved $\|A\mathbf{x}+\mathbf{b}\|^3$ is convex, we need to prove $h(x) = e^x$ is convex as well for any x. The derivatives $h'(x) = h''(x) = e^x$ show us that $h'(x)$ is always increasing and $h''(x)$ shows us e^x is convex as well. Employing the composition rule from the lecture, $f(x) = h(\|A\mathbf{x}+\mathbf{b}\|^3)$ is convex as well due to the inner function also being non-decreasing.

- We can define new auxillary functions, namely $g_i(x) = a_i^T x$, where a_i is a vector where k values are set to 1 and all the other values are 0. We have $\binom{n}{k}$ of these vectors. We can redefine $f(x) = \max\limits_i g_i$. All there is left to show is that $g_i(x) = a_i^T x$ is convex as well. Namely, let $g_i(x)$ be arb. but fix

$$\begin{aligned} g_i(\lambda x + (1-\lambda)y) &= a_i^T(\lambda x + (1-\lambda)y) \\ &= a_i^T \lambda x + a_i^T(1-\lambda)y \\ &= \lambda g_i(x) + (1-\lambda)g_i(y) \end{aligned}$$

which concludes our proof combined with the proposition that a pointwise maximum is also convex.

Problem 3

1. **Proof:** let $S_f(t) = \{x \in dom(f) : f(x) \leq t\}$. We want to prove over the definition, namely $x, y \in S_f(t)$, then for $\theta \in [0,1] : \theta x + (1-\theta)y \in S_f(t)$
   Case Distinction:
   (a) Let $\theta = \{0, 1\}$, then it immediately follows $\theta x + (1-\theta)y = x \in S_f(t)$, conversely for $\theta = 0$
   (b) Let $\theta \in]0, 1[$ and case distinction over $x \leq y$ and $x > y$. We do the second case and the other one should be immediately clear.
   $$\theta x + (1-\theta)y < \theta x + (1-\theta)x = x.$$
   and as follows we have due to $x \leq y \Rightarrow \sqrt{x} \leq \sqrt{y}$ following $\sqrt{\theta x + (1-\theta)y} < \sqrt{x} \leq t$, which concludes our proof.

2. This can be easily proven over the definition. Let $f(x)$ be convex such that $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$ and let it's sublevel set $S_f(t) = \{x \in dom(f) : f(x) \leq t\}$. We want to show $x, y \in S_f(t)$, then for $\theta \in [0,1] : \theta x + (1-\theta)y \in S_f(t)$.
   We can show this as follows:
   $$\begin{aligned} f(\theta x + (1-\theta)y) &\leq \theta f(x) + (1-\theta)f(y) \\ &\leq \theta t + (1-\theta)t \\ &= t \end{aligned}$$
   Where the second inequality follows from $x \in S_f(t) \Rightarrow f(x) \leq t$

3. (a) We must show $f(x) = |x+b|^p$ is quasi convex for $p \in]0,1[$, since from the lecture we know it's convex for $p \geq 1$. Also the inequality $x \leq y \Rightarrow |x|^p \leq |y|^p$ for $p \in]0,1[$ (can be shown with the first derivative always being > 0, except for the point where the argument is 0, thus monotonically increasing) allows us to reuse the proof from (1). This altogether show its sublevel set is also convex.

   (b) This one was one of the harder proof, but nonetheless we can attempt it :) and didn't work for me over the defintion so we cheat a little bit here. We know from the lectures half-spaces $H = \{\mathbf{x} : \mathbf{a}^T\mathbf{x} \leq k\}$ are convex. So if we look at the sublevel set $S_f(t) = \{\mathbf{x} \in dom(f) : f(x) \leq t\} = \{\mathbf{x} : \frac{a^T x + b}{c^T x + d} \leq t\} = \{\mathbf{x} : a^T x + b \leq t(c^T x + d)\} = \{\mathbf{x} : (a^T - tc^T)x \leq dt - b\}$. Which is a halfspace, and thus convex.

4. (a) Analyzing the second derivative $f''(x) = p \cdot (p-1)|x+b|^{p-2}$, shows us that $f''(x) < 0 \forall x \neq 0$. At all those points it's concave down and we have showed it's not convex.

   (b) We can take an counterexample, let $(a,b,c,d) = (1,0,1,0)$ and $x = [u, \sqrt{u}]^T$, which gives $f(x) = \frac{u}{\sqrt{u}} = \sqrt{u}$, which is a concave function, given $u \geq 0$. This can be easily proven with going over the definition of a concave function $f(\theta x + (1-\theta)y) > \theta f(x) + (1-\theta) f(y)$

Problem 4

1.

$$\begin{aligned}
\frac{df(z)}{dz} &= \frac{(e^z + e^{-z})(e^z + e^{-z}) - (e^z - e^{-z})(e^z - e^{-z})}{(e^z + e^{-z})^2} \\
&= 1 - \frac{(e^z - e^{-z})^2}{(e^z + e^{-z})^2} \\
&= 1 - \tanh^2(z)
\end{aligned}$$

   and for the second derivative we can calculate:

$$\frac{d^2 f(z)}{dz^2} = -2[\tanh(z)]' \tanh(z) = -2(1 - \tanh^2(z)) \tanh(z).$$

2. Solution:

```python
import numpy as np
a1 = 1.0
a2 = -1.0
a3 = 0.5


def f_value(x):
    return np.tanh(x) # this is just a tan function with a single input

def g_value(x):
    return f_value(np.dot(np.array([a1, a2, a3]), x)) # we assume x is a vector of 3 variables
```

3. Solution:

```python
def g_gradient(x):
    d_f = 1 - (f_value(np.dot(np.array([a1, a2, a3]), x)))**2
    grad_g = np.array([[a1*d_f], [a2*d_f], [a3*d_f] ])
    return grad_g
```

4. Solution:

```
def g_Hessian(x):
    d_f2 = -2 * (1-f_value(np.dot(np.array([a1, a2, a3]), x))**2)*f_value(np.dot(np.array([a1, a2, a3]), x))
    h_f = np.matrix([[a1*a1 * d_f2, a1*a2 * d_f2, a1*a3 * d_f2],
                    [a2*a1*d_f2, a2*a2 * d_f2, a2*a3 * d_f2],
                    [a3*a1*d_f2, a3*a2 * d_f2, a3*a3 *d_f2]])
    return h_f
```