Eric Gelphman *April 18, 2020*

Optimization Algorithms Problem Set 1 Solutions

Problem 1

Let $x, y \in \mathbb{R}$ Define the following distance functions on $\mathbb{R}$:

1. $d_1(x,y) = (x-y)^2$
2. $d_2(x,y) = \sqrt{|x-y|}$
3. $d_3(x,y) = |x^2 - y^2|$
4. $d_4(x,y) = |x - 2y|$
5. $d_5(x,y) = \frac{|x-y|}{1+|x-y|}$

For each distance function, determine whether or not it is a valid metric on $\mathbb{R}$.

Solution:

1. Clearly, $d_1(x,y)$ is > 0 if $x \neq y$ and is 0 if and only if $x = y$. Additionally, it is clear that $d_1(x,y) = d_1(y,x)$.
   $d_1(x,y)$ does not satisfy the triangle inequality. Let $x = 3, y = -6$, and $z = 1 \in \mathbb{R}$:
   $d_1(x,y) = (3-(-6))^2 = 81 > d_1(x,z) + d_1(z,y) = (3-1)^2 + (1-(-6))^2 = 53$
   Thus, $d_1(x,y)$ is not a valid metric on $\mathbb{R}$

2. Clearly, $d_2(x,y)$ is > 0 if $x \neq y$ and is 0 if and only if $x = y$. Additionally, it is clear that $d_2(x,y) = d_2(y,x)$.
   Consider the "standard" metric on $\mathbb{R}$: $d(x,y) = |x-y|$ for $x, y \in \mathbb{R}$. We know that this metric satisfies the triangle inequality, so for $x, y, z \in \mathbb{R}$:
   $|x-y| \leq |x-z| + |z-y| \leq |x-z| + \sqrt{|x-z||z-y|} + |z-y|$ as $\sqrt{|x-z||z-y|} \geq 0$
   $\iff (\sqrt{|x-y|})^2 \leq (\sqrt{|x-y|} + \sqrt{|z-y|})^2$
   $\iff \sqrt{|x-y|} \leq \sqrt{|x-y|} + \sqrt{|z-y|}$
   Therefore, d_2 satisfies the triangle inequality. Thus, it defines a valid metric on $\mathbb{R}$

3. This is not a valid metric on $\mathbb{R}$. Suppose $x = -y$. Clearly, $y \neq x$ but $d_3(x,y) = |x^2 - y^2| = |y^2 - y^2| = 0$, which violates one of the metric axioms.

4. $d_4(x,y)$ is not symmetric: Let $x = 3, y = 1 \in \mathbb{R}$:
   $d_4(x,y) = |3 - 2(1)| = 1 \neq 5 = |1 - 2(3)| = d_4(y,x)$
   Thus, d_4 is not a valid metric on $\mathbb{R}$

5. Clearly, $d_5(x,y)$ is > 0 if $x \neq y$ and is 0 if and only if $x = y$. Additionally, it is clear that $d_5(x,y) = d_5(y,x)$.
   Let $x, y, z \in \mathbb{R}$ and let $a = |x-y|, b = |x-z|, c = |z-y|$. Since the "standard" metric on $\mathbb{R}$ satisfies the triangle inequality, it follows that
   $a \leq b + c$
   $\iff a \leq b + bc + abc + c + bc$ as $a, b, c \geq 0$
   $\iff a + ac + ab + abc \leq b + bc + ab + abc + c + bc + ac + abc$ by adding $ac + ab + abc$ to

both sides
$\iff a(1+c+b+bc) \leq b(1+c+a+ac) + c(1+b+a+ab)$
$\iff a(1+b)(1+c) \leq b(1+a)(1+c) + c(1+a)(1+b)$
$\iff \frac{a}{1+a} \leq \frac{b}{1+b}\frac{c}{1+c}$ by dividing both sides by $(1+a)(1+b)(1+c)$
$\iff d_5(x,y) \leq d_5(x,z) + d_5(z,y)$
Thus, d_5 is a valid metric on $\mathbb{R}$

QED

Problem 2

The statement of Taylor's Theorem, FONC, SONC, SOSC all begin with "Let $E \subseteq \mathbb{R}$ be open, let $E \subseteq \mathbb{R}^n$ be open," and such. Can you think of a reason for this?

Solution: The reason for this is that, if E is not open, we would not be able to find an open neighborhood/ball that lies entirely within that set. This is necessary for the definition of a limit of a function at a point, i.e. $\lim_{x \to a} f(x)$, is given by $\forall \epsilon > 0, \exists \delta > 0$ such that $\forall x$ such that $0 < |x-a| < \delta, |f(x) - f(a)| < \epsilon$. $B(a,\delta) \setminus \{a\}$ and $B(f(x), \epsilon)$ are both open sets. In a Taylor series, derivatives are involved, which is an operation involving limits.

See Jupyter notebook Solution-to-Exercise-1.3.ipynb for Problem 3

Problem 4

Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

. Let $t = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$

1. Calculate the gradient of f at t $\nabla f(t)$
2. Calculate the Hessian of f at t $\nabla^2 f(t)$
3. Is FONC satisfied at t?
4. is SONC satisfied at t?
5. is SOSC satisfied at t?

Solution:

1. $\nabla f(t) = \begin{bmatrix} \frac{\partial f(t)}{\partial x_1} & \frac{\partial f(t)}{\partial x_2} \end{bmatrix}$. Using standard differentiation rules from calculus, we can compute the first-order partials as:

$$\frac{\partial f}{\partial x_1} = 2(100)(x_2 - x_1^2)(-2x_1) + 2(1-x_1)(-1) = -400x_1(x_2 - x_1^2) - 2(1-x_1)$$

$$\frac{\partial f}{\partial x_2} = 200(x_2 - x_1^2)$$

So,

$$\nabla f(t) = \begin{bmatrix} -400(1)(1-1^2) - 2(1-1) & 200(1-1^2) \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

2. $\nabla^2 f(t) = \begin{bmatrix} \frac{\partial^2 f(t)}{\partial x_1^2} & \frac{\partial^2 f(t)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(t)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(t)}{\partial y^2} \end{bmatrix}$. Since f is a polynomial in 2 variables, it is a C^2 function on $\mathbb{R}^2$, so by Clairaut's Theorem it follows that $\frac{\partial^2 f(t)}{\partial x_1 \partial x_2} = \frac{\partial^2 f(t)}{\partial x_2 \partial x_1}$.

$$\frac{\partial^2 f(t)}{\partial x_1^2} = \frac{\partial}{\partial x_1}[\frac{\partial f}{\partial x_1}] = -400x_1(-2x_1) - 400(x_2 - x_1^2) + 2 = 1200x_1^2 - 400x_2 + 2$$

$$\frac{\partial^2 f(t)}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}[\frac{\partial f}{\partial x_2}] = -400x_1$$

$$\frac{\partial^2 f(t)}{\partial x_2^2} = \frac{\partial}{\partial x_2}[\frac{\partial f}{\partial x_2}] = 200$$

So,

$$\nabla^2 f(t) = \begin{bmatrix} 1200(1)^2 - 400(1) + 2 & -400(1) \\ -400(1) & 200 \end{bmatrix} = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

3. The first order necessary condition is satisfied at t because both the x_1- and x_2 components of ∇f are 0 at t.

4. We will prove that the Hessian matrix $\nabla^2 f(t)$ is positive definite. Clearly, the matrix is symmetric.
An equivalent condition for a matrix being positive semidefinite is that its eigenvalues are greater than or equal to 0. We will prove this below: Let $x \in \mathbb{R}^n$ be an eigenvector of the matrix $A \in \mathbb{R}^{n \times n}$ with corresponding eigenvalue λ:

$$x^T A x = x^T(\lambda x) = \lambda \|x\|^2 \geq 0 \iff \lambda \geq 0$$

We will now compute the eigenvalues of the matrix $\nabla^2 f(t)$. The characteristic polynomial of this matrix, found by taking the determinant of the matrix $\nabla^2 f(t) - \lambda I_2$ where I_2 is the 2 x 2 identity matrix, is

$$p(\lambda) = \lambda^2 - 1002\lambda + 400$$

The roots of this polynomial are given by the quadratic equation:

$$\lambda = \frac{1002 \pm \sqrt{(-1002)^2 - 4(400)(1)}}{2(1)}$$
$$\lambda = 501 \pm \frac{1}{2}\sqrt{1002404}$$
$$\lambda = 501 \pm 500.60064 \geq 0$$

Thus, the matrix $\nabla^2 f(t)$ is positive semidefinite and the second order necessary condition is satisfied at t.

5. The second order sufficient condition is satisifed at t as both eigenvalues of $\nabla^2 f(t)$ are strictly positive. It then follows from the proof above that $\nabla^2 f(t)$ is positive definite and thus t is a local minimizer of f.

Problem 5

Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x) = \|x\|^2$. Let $\{x_k\}_{k=0}^{\infty}$ be the vector-valued sequence given by

$$x_k = (1 + \frac{1}{2^k}) \begin{bmatrix} \cos k \\ \sin k \end{bmatrix}$$

Prove that $f(x_k)$ is a strictly decreasing sequence and that every point on the unit circle is a limit point of $\{x_k\}$.

<u>Solution</u>: A limit point of a sequence is any one of its subsequential limits. What the hint is saying is that, given arbitrary $\theta \in [0, 2\pi]$ $\exists$ a subsequence of $\{\xi_k\}$, given by $\{\xi_{k_m}\}$ such that $\{\xi_{k_m}\} \to \theta \text{as} k_m \to \infty$.

$f(x_k)$ is a strictly decreasing sequence: Proof is by induction on k.
Base Case: k = 0: $f(x_0) = 4$.
$f(x_1) = \|x_1\|^2 = ((1 + \frac{1}{2}) \cos 1)^2 + ((1 + \frac{1}{2}) \sin 1)^2 = (\frac{3}{2})^2 (\cos^2 1 + \sin^2 1)^2 = \frac{9}{4}(1) = \frac{9}{4} < f(x_0)$
Thus, the base case is verified
Induction step: Let $m \in \mathbb{N}$ and suppose $f(x_m) < f(x_{m-1})$

$$\begin{aligned}
f(x_{m+1}) &= \|x_{m+1}\|^2 = (1 + \frac{1}{2^{m+1}})^2 (\cos^2(m+1) + \sin^2(m+1)) \\
f(x_{m+1}) &= (1 + \frac{1}{2^{m+1}})^2 \\
&< (1 + \frac{1}{2^m})^2 \\
&= (1 + \frac{1}{2^m})^2 (\cos^2(m) + \sin^2(m)) \\
&= f(x_m)
\end{aligned}$$

Thus, the induction step follows.

Every point on the unit circle is a limit point of $\{x_k\}$: Let $\theta \in [0, 2\pi]$ and let $x^* = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$
Let $\epsilon > 0$ and let $\{\xi_{k_m}\}$ denote the subsequence of $\{\xi_k\}$ given by the hint that converges to θ. Using one of the definitions of continuity, since both $\sin x, \cos x$ are continuous functions, it follows that $\exists N_1, N_2 \in \mathbb{N}$ and $\delta_1, \delta_2 > 0$ such that $\forall k_m \geq N_1$ we have $0 < \|x_{k_m} - \xi_{k_m}\| < \delta_1$, and it follows that $|\cos k_m - \cos\theta| < \frac{\epsilon}{\sqrt{8}}$, and $\forall k_m \geq N_2$ we have $0 < \|x_{k_m} - \xi_{k_m}\| < \delta_2$, and it follows that $|\sin k_m - \sin\theta| < \frac{\epsilon}{\sqrt{8}}$.
$\{\frac{1}{2^k}\}$ is a sequence that converges to 0, so $\exists N_3 \in \mathbb{N}$ such that $\forall k \geq N_3, |\frac{1}{2^k}| < \frac{\epsilon}{2}$.

See Next Page

Let $N = \max(N_1, N_2, N_3)$. Then, $\forall k_m \geq N$,

$$\begin{aligned}
\left\| x_{k_m} - x^* \right\| &= \left\| x_{k_m} - \begin{bmatrix} \cos k_m \\ \sin k_m \end{bmatrix} + \begin{bmatrix} \cos k_m \\ \sin k_m \end{bmatrix} - x^* \right\| \\
&= \left\| (1 + \frac{1}{2^{k_m}}) \begin{bmatrix} \cos x_{k_m} \\ \sin x_{k_m} \end{bmatrix} - \begin{bmatrix} \cos k_m \\ \sin k_m \end{bmatrix} + \begin{bmatrix} \cos k_m \\ \sin k_m \end{bmatrix} - x^* \right\| \\
&\leq \left\| \begin{bmatrix} \cos x_{k_m} \\ \sin x_{k_m} \end{bmatrix} (1 + \frac{1}{2^{k_m}} - 1) \right\| + \left\| \begin{bmatrix} \cos k_m \\ \sin k_m \end{bmatrix} - x^* \right\| \\
&= \sqrt{(\cos^2 k_m + \sin^2 k_m)} |\frac{1}{2^{k_m}}| + \sqrt{(\cos k_m - \cos\theta)^2 + (\sin k_m - \sin\theta)^2} \\
&< (1)(\frac{\epsilon}{2}) + \sqrt{\frac{\epsilon^2}{4}} \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon
\end{aligned}$$

QED