Math 350 Problem Set 2 (Part I) Solutions

1. Mean Value Theorems:

   (a) (10pts) Use the Mean Value Theorem from single variable calculus to prove the Mean Value Theorem below for scalar functions of several variables. A **convex set** $A \subset \mathbb{R}^n$ is a set such that, for any two points in A, the line segment between them is also in A:

   $$\mathbf{x}, \mathbf{y} \in A \implies \{(1-t)\mathbf{x} + t\mathbf{y} \mid 0 \le t \le 1\} \subset A$$

   Theorem (MVT). *If $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable on a convex set A, for any pair of points $\mathbf{x}, \mathbf{y} \in A$, we have*

   $$f(\mathbf{x}) - f(\mathbf{y}) = \nabla f(\mathbf{z}) \cdot (\mathbf{x} - \mathbf{y}) = \mathbf{D}f(\mathbf{z})(\mathbf{x} - \mathbf{y})$$

   for some $\mathbf{z} \in \{(1-t)\mathbf{x} + t\mathbf{y} \mid 0 \le t \le 1\} \subset A$.

   Proof. Let $\mathbf{c} : \mathbb{R} \to \mathbb{R}^n$ be the curve $t \mapsto (1-t)\mathbf{x} + t\mathbf{y} = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$, so the composition $f \circ \mathbf{c} : \mathbb{R} \to \mathbb{R}$ is a single variable function given by

   $$t \mapsto f((1-t)\mathbf{x} + t\mathbf{y}).$$

   By the Mean Value Theorem in single variable calculus,

   $$(f \circ \mathbf{c})(0) - (f \circ \mathbf{c})(1) = \frac{d(f \circ \mathbf{c})}{dt}(t')(0-1), \qquad \text{for some } t' \in [0,1].$$

   However, evaluating this expression and using the chain rule, we have

   $$f(\mathbf{x}) - f(\mathbf{y}) = \mathbf{D}f((c(t'))\mathbf{D}\mathbf{c}(t')(0-1) = -\mathbf{D}f(\mathbf{z})\mathbf{c}'(t') = -\mathbf{D}f(\mathbf{z})(\mathbf{y} - \mathbf{x}) = \mathbf{D}f(\mathbf{z})(\mathbf{x} - \mathbf{y})$$

   where $z = (1-t')\mathbf{x} + t'\mathbf{y} = \mathbf{c}(t')$. ☐

   (Sorry about the minus signs. It would have been clearer to write the line between $\mathbf{x}$ and $\mathbf{y}$ as $\{t\mathbf{x} + (1-t)\mathbf{y} \mid 0 \le t \le 1\}$).

   (b) (10pts) Why is the analogous statement (using the $\mathbf{D}f$ form of the right hand side) false in general for a vector valued function $f : \mathbb{R}^n \to \mathbb{R}^m$ when $m > 1$?

   Since $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$, where the $f_i(\mathbf{x})$ are scalar functions, we know from part (a) that there exist points $\mathbf{z}_1, \ldots, \mathbf{z}_m$ on the line segment between $\mathbf{x}$ and $\mathbf{y}$ such that

   $$f_i(\mathbf{x}) - f_i(\mathbf{y}) = \mathbf{D}f_i(\mathbf{z}_i)(\mathbf{x} - \mathbf{y}),$$

   and certainly we have that

   $$f(\mathbf{x}) - f(\mathbf{y}) = \begin{bmatrix} f_1(\mathbf{x}) - f_1(\mathbf{y}) \\ \vdots \\ f_m(\mathbf{x}) - f_m(\mathbf{y}) \end{bmatrix} = \begin{bmatrix} \mathbf{D}f_1(\mathbf{z}_1) \\ \vdots \\ \mathbf{D}f_m(\mathbf{z}_m) \end{bmatrix} (\mathbf{x} - \mathbf{y})$$

   where the $\mathbf{D}f_i(\mathbf{z}_i)$ (which are $1 \times n$ matrices) provide the rows for a $m \times n$ matrix. However, we cannot guarantee that $\mathbf{z}_1 = \mathbf{z}_2 = \cdots = \mathbf{z}_m$ in general, so this matrix is not generally equal to $\mathbf{D}f(\mathbf{z})$ for any $\mathbf{z}$.

2. (5pts) What is the derivative of a constant function $f : \mathbb{R}^n \to \mathbb{R}^m$, (so $(x_1, \ldots, x_n) \mapsto (f_1, \ldots, f_m)$, where each $f_i \in \mathbb{R}$ is independent of $\mathbf{x}$)? Prove your answer using the definition of the derivative; i.e. that the derivative is the unique linear function $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{T}(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0$$

The derivative of such a function is identically 0 for all $\mathbf{x}$; that is,

$$\mathbf{D}f(\mathbf{x}) = \mathbf{0}$$

where $\mathbf{0} : \mathbb{R}^n \to \mathbb{R}^m$ is the linear function $\mathbf{x} \mapsto (0, \ldots, 0)$ for all x, which is represented by the zero matrix

$$\mathbf{0} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}.$$

Indeed, we have

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{0}(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = \lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|0 - \mathbf{0}(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = \lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|0-0\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0.$$

3. (5pts) What is the derivative of a linear function $f : \mathbb{R}^n \to \mathbb{R}^m$, (so $f(a\mathbf{x} + b\mathbf{y}) = af(\mathbf{x}) + bf(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $a, b \in \mathbb{R}$)? Prove your answer using the definition of the derivative.

   The derivative of f at $\mathbf{x}$ is supposed to be the "best linear approximation to f" at $\mathbf{x}$. Since f is already linear, it is best approximated by f itself. Thus, $\mathbf{D}f(\mathbf{x}) = f$ for all $\mathbf{x}$. Indeed,

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - f(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = \lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}-\mathbf{x}_0) - f(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0$$

   using the property of linearity that $f(\mathbf{x}) - f(\mathbf{x}_0) = f(\mathbf{x}-\mathbf{x}_0)$.

4. (10pts) A function $f : A \subset \mathbb{R}^n \to B \subset \mathbb{R}^n$ is said to be **invertible** if there exists a function (called the inverse) $f^{-1} : B \subset \mathbb{R}^n \to A \subset \mathbb{R}^n$ such that

$$f^{-1} \circ f = \mathrm{Id} : A \to A \quad \text{and} \quad f \circ f^{-1} = \mathrm{Id} : B \to B,$$

   where Id is the **identity function** which maps each point to itself,

$$\mathrm{Id} : \mathbb{R}^n \to \mathbb{R}^n, \quad \mathbf{x} \mapsto \mathbf{x}.$$

   (Note that the dimension of the domain and range must be the same, and note that $f^{-1}(\mathbf{x})$ does not mean $1/f(\mathbf{x})$ unless $n = 1$, since division does not make sense for $n > 1$.)

   Show that if $f : A \subset \mathbb{R}^n \to B \subset \mathbb{R}^n$ is invertible, and if f is differentiable at $\mathbf{x}_0 \in A$, then f^{-1} is differentiable at $\mathbf{y}_0 = f(\mathbf{x}_0)$ with derivative

$$\mathbf{D}\left(f^{-1}\right)(\mathbf{y}_0) = \left(\mathbf{D}f(\mathbf{x}_0)\right)^{-1}.$$

   (Hint: use the chain rule and your result from problem 3).

   Since $(f^{-1} \circ f) = \mathrm{Id}$ is differentiable at $\mathbf{x}_0$, it follows that f^{-1} is differentiable at $\mathbf{y}_0$. (It's OK if you did not make this argument, but rather assumed that f^{-1} was differentiable at $\mathbf{y}_0$). What's left is to show that $\mathbf{D}(f^{-1})(\mathbf{y}_0)$ is an inverse to $\mathbf{D}f(\mathbf{x}_0)$, i.e. that

$$A\,\mathbf{D}f(\mathbf{x}_0) = \mathrm{Id}, \qquad \mathbf{D}f(\mathbf{x}_0)\,A = \mathrm{Id},$$

where $A = \mathbf{D}(f^{-1})(\mathbf{y}_0)$.

But using the fact that Id is linear, so $\mathbf{D}(\mathrm{Id})(\mathbf{x}) = \mathrm{Id}$, and the chain rule, we have

$$\mathrm{Id} = \mathbf{D}(\mathrm{Id})(\mathbf{x}_0) = \mathbf{D}(f^{-1} \circ f)(\mathbf{x}_0) = \mathbf{D}(f^{-1})(\mathbf{y}_0)\,\mathbf{D}f(\mathbf{x}_0)$$

and

$$\mathrm{Id} = \mathbf{D}(\mathrm{Id})(\mathbf{y}_0) = \mathbf{D}(f \circ f^{-1})(\mathbf{y}_0) = \mathbf{D}f(\mathbf{x}_0)\,\mathbf{D}(f^{-1})(\mathbf{y}_0),$$

so by definition of invertibility, we have

$$\mathbf{D}(f^{-1})(\mathbf{y}_0) = (\mathbf{D}f(\mathbf{x}_0))^{-1}.$$

5. (10pts) In class (and in Ch. 2.6 in the book), we discussed the product rule

$$\mathbf{D}(fg)(\mathbf{x}_0) = g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0) + f(\mathbf{x}_0)\mathbf{D}g(\mathbf{x}_0)$$

for scalar functions $f, g : \mathbb{R}^n \to \mathbb{R}$, where the product $fg : \mathbb{R}^n \to \mathbb{R}, \quad \mathbf{x} \mapsto f(\mathbf{x})g(\mathbf{x})$ is ordinary multiplication in $\mathbb{R}$.

For vector valued functions $f, g : \mathbb{R}^n \to \mathbb{R}^m$, we can form the dot product function (note that it is scalar valued!)

$$f \cdot g : \mathbb{R}^n \to \mathbb{R}, \quad \mathbf{x} \mapsto f(\mathbf{x}) \cdot g(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{x})g_i(\mathbf{x})$$

where the product is the dot product in $\mathbb{R}^m$. Formulate and prove a product rule for the dot product. You may use the scalar version of the product rule in your proof, and any other properties of derivatives discussed in class. (Hint: in formulating your result in a concise manner, you may be interested in the **transpose** operation on matrices $A \mapsto A^{\mathrm{T}}$. If A is an $r \times s$ matrix, A^{T} is a $s \times r$ matrix with $(A^{\mathrm{T}})_{ij} = A_{ji}$, that is, the rows and columns are swapped with one another.)

Since $f \cdot g$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$, we expect $\mathbf{D}(f \cdot g)(\mathbf{x})$ to be a linear map from $\mathbb{R}^n$ to $\mathbb{R}$, that is, a $1 \times n$ matrix. Using the sum and product rules, we have

$$\mathbf{D}(f \cdot g)(\mathbf{x}) = \mathbf{D}\left(\sum_{i=1}^{m} f_i(\mathbf{x})g_i(\mathbf{x})\right) = \sum_{i=1}^{m} \left(f_i(\mathbf{x})\mathbf{D}g_i(\mathbf{x}) + g_i(\mathbf{x})\mathbf{D}f_i(\mathbf{x})\right).$$

Each term in the sum is a scalar, such as $f_i(\mathbf{x})$, times a $1 \times n$ matrix, such as $\mathbf{D}g_i(\mathbf{x})$, so the result is indeed a $1 \times n$ matrix. We can write this more economically by noting that

$$[\mathbf{D}g_i(\mathbf{x})]_{1j} = [\mathbf{D}g(\mathbf{x})]_{ij}$$

that is, the entry in the jth column of the $1 \times n$ matrix $\mathbf{D}g_i(\mathbf{x})$ (which is $\frac{\partial g_i(\mathbf{x})}{\partial x_j}$), is equal to the entry in the ith row, jth column entry of the $m \times n$ matrix $\mathbf{D}g(\mathbf{x})$.

Thus, using the formula $[AB]_{ik} = \sum_j A_{ij}B_{jk}$ for matrix multiplication, we have

$$[\mathbf{D}(f \cdot g)(\mathbf{x})]_{1j} = \sum_{i=1}^{m} \left(f_i(\mathbf{x})\,[\mathbf{D}g(\mathbf{x})]_{ij} + g_i(\mathbf{x})\,[\mathbf{D}f(\mathbf{x})]_{ij}\right) = \left[f(\mathbf{x})^{\mathrm{T}}\,\mathbf{D}g(\mathbf{x}) + g(\mathbf{x})^{\mathrm{T}}\,\mathbf{D}f(\mathbf{x})\right]_{1j},$$

for the entry in each column, and so

$$\mathbf{D}(f \cdot g)(\mathbf{x}) = f(\mathbf{x})^{\mathrm{T}}\,\mathbf{D}g(\mathbf{x}) + g(\mathbf{x})^{\mathrm{T}}\,\mathbf{D}f(\mathbf{x})$$

Note that $f^{\mathrm{T}}(\mathbf{x})$ is a $1 \times m$ matrix and $\mathbf{D}g(\mathbf{x})$ is an $m \times n$ matrix, (similarly for $g^{\mathrm{T}}(\mathbf{x})$ and $\mathbf{D}f(\mathbf{x})$) so the result is a $1 \times n$ matrix, as expected.

This can equivalently be written as

$$\mathbf{D}(f \cdot g)(\mathbf{x}) = (\mathbf{D}g(\mathbf{x}))^{\mathrm{T}}\ f(\mathbf{x}) + (\mathbf{D}f(\mathbf{x}))^{\mathrm{T}}\ g(\mathbf{x}).$$