Homework #3 Solutions

Thayer Anderson, Davis Lazowski, Handong Park, Rohil Prasad
Eric Peterson

1 For submission to Thayer Anderson

Problem 1.1. Let $g_1, g_2 \colon \mathbb{R}^2 \to \mathbb{R}$ both be continuous functions. Define $f \colon \mathbb{R}^2 \to \mathbb{R}$ by the following integral:

$$f(x, y) = \int_0^x g_1(t, 0)\, \mathrm{d}t + \int_0^y g_2(x, t)\, \mathrm{d}t.$$

(You can think of f as measuring the total charge for moving from $(0, 0)$ to (x, y) along a rectilinear path.)

1. Demonstrate $\frac{\partial f}{\partial y}(x, y) = g_2(x, y)$. (You'll want to invoke the single-variable fundamental theorem of calculus, which we have not yet proven. Go ahead.)
2. How can the definition of f be modified so that $\frac{\partial f}{\partial x}(x, y) = g_1(x, y)$? Can you get both identities to work at once?
3. Find a function f with $\frac{\partial f}{\partial x} = x$ and $\frac{\partial f}{\partial y} = y$. Now find a function f with $\frac{\partial f}{\partial x} = y$ and $\frac{\partial f}{\partial y} = x$.
4. Explain why you don't expect to be able to find a function f with $\frac{\partial f}{\partial x} = y$ and $\frac{\partial f}{\partial y} = -x$.

Solution.

1. Certainly we have

$$\frac{\partial}{\partial y}\left(\int_0^x g_1(t, 0)\, \mathrm{d}t + \int_0^y g_2(x, t)\, \mathrm{d}t\right) = \frac{\partial}{\partial y}\left(\int_0^y g_2(x, t)\, \mathrm{d}t\right),$$

since the first term has no dependence on y. To evaluate the remaining term, we directly appeal to the single-variable fundamental theorem of calculus: the derivative of the integral of a continuous function recovers the original function. Hence,

$$\frac{\partial}{\partial y}\left(\int_0^y g_2(x, t)\, \mathrm{d}t\right) = g_2(x, y).$$

2. You can always integrate along the "other" rectilinear path:

$$f_{\text{modified}}(x, y) = \int_0^y g_2(0, t)\, \mathrm{d}t + \int_0^x g_1(t, y)\, \mathrm{d}t.$$

You certainly can't get both to work in general; if g_1 and g_2 happen to be continuously differentiable themselves, then if both identities are true then it must certainly also be the case that

$$\frac{\partial}{\partial x} g_2 = \frac{\partial^2}{\partial x \partial y} f = \frac{\partial^2}{\partial y \partial x} f = \frac{\partial}{\partial y} g_1,$$

because the mixed partials of f would then commute. This puts a serious interdependence on g_1 and g_2, which is typically not satisfied, and so no such function f could possibly exist.

3. $f = \frac{1}{2}(x^2 + y^2)$ works for the first condition, which you can get by setting $g_1(t, 0) = t$ and $g_2(x, t) = t$. For the second, we can use $f = x \cdot y$, which you can get by setting $g_1 = 0$ and $g_2(x, t) = x$.

4. Again, mixed partials are supposed to commute for nice functions, but $\frac{\partial^2}{\partial x \partial y} f = \frac{\partial}{\partial x}(-x) = -1$ while $\frac{\partial^2}{\partial y \partial x} f = \frac{\partial}{\partial y}(y) = 1$ shows that that commutation does not hold here. (ECP)

Problem 1.2. For $f\colon \mathbb{R}^2 \to \mathbb{R}$, show that if $\frac{\partial f}{\partial y} = 0$ then f is independent of the second variable — i.e., $f(x, y_1) = f(x, y_2)$ for any $y_1, y_2 \in \mathbb{R}$. If both $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$, show that f is a constant function.

Solution. This is entirely a consequence of the mean value theorem. If an everywhere differentiable function were not constant, then the nonzero slope of the secant line passing through two points witnessing the nonconstancy guarantee (by the MVT) a point between where the derivative matches the secant slope. Since $\partial/\partial y$ is computed by restricting to a particular value $x = a$ and considering the function as a single-variable function, the MVT applies and shows that f cannot depend on the second variable. If $\partial f/\partial x$ is *also* zero, then we can always build a rectilinear path between any pair of points (x_0, y_0) to (x_1, y_1) by

$$(x_0, y_0) \to (x_0, y_1) \to (x_1, y_1).$$

Since the function is independent of the second variable, its value on the first two points is the same. Since the function is also independent of the first variable, its value on the last two points on the same. By transitivity, its values agree on the first and third points. (ECP)

2 For submission to Davis Lazowski

Problem 2.1. 1. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function, and assume $f'(a) \neq 0$ for all $a \in \mathbb{R}$. Show that f is injective.

2. Now consider $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, described by the formula

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}.$$

Show that $\det(D_{(a,b)} f) \neq 0$ for all $(a, b) \in \mathbb{R}^2$, and yet f is *not* injective.[1]

Solution. 1. Otherwise, there exists a, b: $f(a) = f(b)$. By the mean value theorem then there exists c: $f'(c) = \frac{f(b)-f(a)}{b-a} = 0$. Therefore contradiction.

2. The Jacobian is generally

$$\begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix} = e^x \begin{pmatrix} \cos y & -\sin y \\ \sin y & \cos y \end{pmatrix}$$

So this is e^x times the rotation matrix $R(y)$, which is invertible. Therefore the determinant is nonzero everywhere with these two components. However $f(x, y) = f(x, 2\pi + y)$. (DL)

Problem 2.2. The statement of the inverse function theorem for a function f requires that the derivative f' be continuous. Consider the following function:

$$f(x) = \begin{cases} \frac{x}{2} + x^2 \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

1. Check that f is differentiable everywhere and that its derivative fails to be continuous at 0.

[1] The inspiration for this function comes from the complex exponential: $e^{a+bi} = e^a \cdot (\cos b + i \sin b)$.

2. Check that the conclusion of the inverse function theorem fails at 0.

Solution. 1. Away from zero, the function is differentiable with the chain rule and product rule.

For $x > 0$, we find that $f'(x) = \frac{1}{2} + 2x \sin \frac{1}{x} - \cos \frac{1}{x}$.

At 0,

$$\frac{h}{2} + \frac{h^2}{2} \sin \frac{1}{h} = 0 + f'(0)h + \varepsilon(0, h)$$

We can check that $f'(0) = \frac{1}{2}$ satisfies the definition. Then $\varepsilon(0, h) = \frac{h^2}{2} \sin \frac{1}{h}$, and $h \sin \frac{1}{h} \to 0$ as $h \to 0$. But certainly $\lim_{x\to 0} 2x \sin \frac{1}{x} - \cos \frac{1}{x}$ is not 0; in fact, the limit of the second term does not exist. Therefore the function is not continuous around zero.

2. In particular, the inverse function theorem requires that f be invertible on some open neighbourhood of 0.

Restrict f to $(-\varepsilon, 0) \cup (0, \varepsilon)$. We want to show that there are $x, x' \in (-\varepsilon, \varepsilon)$ such that $f(x) = f(x')$. Equivalently, we can show that $f'(x'') = 0$ for some $x'' \in (-\varepsilon, 0) \cup (0, \varepsilon)$.

By continuity of the derivative away from zero, it's again enough to show that there are a, b in $(0, \varepsilon)$ so that $f'(a) > 0$, $f'(b) < 0$. Then between them there must be a point which evaluates to zero.

We can choose $a = \frac{1}{n\pi/2}$, where n is an odd integer and $\sin n\pi/2 = 1$. By choosing sufficiently large n, then $0 < a < \varepsilon$, and

$$f'(a) = \frac{1}{2} + 2a > 0$$

We can choose $b = \frac{1}{2m\pi}$, where m is an integer , so that $\cos 2m\pi = 1$. Then

$$f'(b) = \frac{1}{2} - 1 = -\frac{1}{2}$$

And b can be as small as we like by choosing large m. (DL)

3 For submission to Handong Park

Problem 3.1. Consider the domain

$$A = \{(x, y) \in \mathbb{R}^2 \mid x < 0\} \cup \{(x, y) \in \mathbb{R}^2 \mid x \geq 0 \text{ and } y \neq 0\} = \mathbb{R}^2 \setminus ([0, \infty) \times \{0\}).$$

1. Let $f\colon A \to \mathbb{R}$ be a differentiable function with $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$. Show f is constant.[2]

2. Find a differentiable function $f\colon A \to \mathbb{R}$ satisfying $\frac{\partial f}{\partial y} = 0$ but which is *not* independent of y.

Solution. 1. Take some $f : A \to \mathbb{R}$ satisfying the conditions, and take (x_1, y_1) and (x_2, y_2) in A. If $x_1 = -1$, we have that $f(-1, y_1) = f(x_1, y_1)$. Else, by the Mean Value Theorem, we know that for $x_1 \neq -1$, there exists a between x_1 and -1 such that

$$\frac{\partial f}{\partial x}(a, y_1) = \frac{f(x_1, y_1) - f(-1, y_1)}{x_1 - (-1))} \to f(x_1, y_1) = f(-1, y_1)$$

Use similar reasoning to show that if $y_1 = y_2$, we have $f(-1, y_1) = f(-1, y_2)$, and if $y_1 \neq y_2$, we have by the Mean Value Theorem that there is some b between y_1 and y_2 such that

$$\frac{\partial f}{\partial y}(-1, b) = \frac{f(-1, y_1) - f(-1, y_2)}{y_1 - y_2} \to f(-1, y_1) = f(-1, y_2)$$

All together, we find that $f(x_1, y_1) = f(-1, y_1) = f(-1, y_2) = f(x_2, y_2)$, showing that f is constant.

[2]Note that this is *not* a repeat of Problem 1.2, since the domain is different.

2. There are many examples that work. One example is

$$f(x) = \begin{cases} -x^2 & (x, y) \in (-\infty, 0) \times \mathbb{R} \cup [0, \infty) \times (-\infty, 0) \\ x^2 & \text{else} \end{cases}$$

The partial derivative with respect to y is certainly 0, yet the derivative depends on y due to the function's definition, although continuity and differentiability are indeed satisfied. (HP)

Problem 3.2. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable. Given a point $a \in \mathbb{R}^n$ and a direction $v \in T_a\mathbb{R}^n$, the associated *directional derivative* of f is defined by

$$\lim_{t \to 0} \frac{f(a + tv) - f(a)}{t},$$

which we will temporarily denote by $\mathbb{D}_a^v f$.

1. Choosing $v = e_i$ to be a standard basis vector, show $\mathbb{D}_a^{e_i} f = \frac{\partial f}{\partial x_i}$.
2. Choosing a scalar $k \in \mathbb{R}$, show $\mathbb{D}_a^{kv} f = k \cdot \mathbb{D}_a^v f$.
3. Now suppose that f is a differentiable function. Show that $\mathbb{D}_a^v f = (D_a f)(v)$, so that we have not invented a new notion of derivative. Conclude that $\mathbb{D}_a^{v+w} f = \mathbb{D}_a^v f + \mathbb{D}_a^w f$.

Solution. 1. Let e_i be the standard basis vector in question, then it is just

$$e_i = (0, 0, ..., 0, 1, 0, ..., 0, 0)$$

(While this could probably be written vertically, we can write it in transpose form to make notation easier.)
Then, we just have that by definition

$$\begin{aligned} \lim_{t \to 0} \frac{f(a + te_i) - f(a)}{t} &= \lim_{t \to 0} \frac{f(a_1, ..., a_{i-1}, a_i + t, a_{i+1}, ..., a_n) - f(a_1, ..., a_n)}{t} \\ &= \frac{\partial f}{\partial i} \end{aligned}$$

2. Let $u = tk$, then we have $t = \frac{u}{k}$. Since u is also a scalar that tends to zero as t tends to zero, we get

$$\lim_{t \to 0} \frac{f(a + uv) - f(a)}{\frac{u}{k}} = \lim_{u \to 0} \frac{k(f(a + uv) - f(a))}{u} = k \cdot \mathbb{D}_a^v$$

3. Since f is differentiable, there exists the derivative $D_a f$ such that

$$\lim_{h \to 0} \frac{|f(a + h) - f(a) - (D_a f)(h)|}{||h||} = 0$$

Choose some $v \in T_a\mathbb{R}^n$. If $v = 0$, we know that $D_a f(v) = 0$ by linearity, and we have that

$$\mathbb{D}_a^v f = \lim_{t \to 0} \frac{f(a + v) - f(a)}{t} = \lim_{t \to 0} \frac{f(a) - f(a)}{t} = 0$$

so that $D_a f(v) = \mathbb{D}_a^v f$ in this case.
If $v \neq 0$, we have that

$$\begin{aligned} \lim_{h \to 0} \frac{|f(a + h) - f(a) - (D_a f)(h)|}{||h||} = 0 &\to \lim_{t \to 0} \frac{f(a + tv) - f(a) - (D_a f)(tv)}{t||v||} = 0 \\ &\to \lim_{t \to 0} \left| \frac{f(a + tv) - f(a)}{t} - (D_a f)(v) \right| \frac{1}{||v||} = 0 \end{aligned}$$

Since $||v||$ is nonzero, and the whole limit is zero, we get that $D_a f(v)$ must equal the left term, which is the definition of $\mathbb{D}_a^v f$, so we must have that $D_a f(v) = \mathbb{D}_a^v f$.
Then, from what we've proved, we can immediately state by linearity of $D_a f$ that

$$\mathbb{D}_a^{v+w} f = (D_a f)(v+w) = (D_a f)(v) + (D_a f)(w) = \mathbb{D}_a^v f + \mathbb{D}_a^w f$$

and we are done. (HP)

4 For submission to Rohil Prasad

Problem 4.1. Let $A \subseteq \mathbb{R}^n$ be an open set, and let $f : A \to \mathbb{R}^n$ be a continuously differentiable injective function with $\det D_a f \neq 0$ for all $a \in A$.

1. Show that the image $f(A) \subseteq \mathbb{R}^n$ is an open set.
2. Show that $f^{-1} : f(A) \to A$ is differentiable.
3. Show that for any open $B \subseteq A$, $f(B)$ is also an open set.

Solution.

1. Since the differential has nonzero determinant everywhere, for all $a \in A$ there exists an open neighborhood V around $f(a)$ contained inside $f(A)$ and a continuously differentiable map $g : V \to f^{-1}(V) \subset A$ such that $fg = \mathrm{id}$ and $gf|_V = \mathrm{id}$.

   Therefore, every point in $f(A)$ has an open neighborhood around it that is also contained in $f(A)$. By definition, this means $f(A)$ is open.

2. By the inverse function theorem, we can produce open subsets $V, V' \subseteq f(A)$ and continuously differentiable maps $g : V \to f^{-1}(V)$, $g' : V' \to f^{-1}(V')$ that are inverse to f in the way described above.

   Then, for any $x \in V \cap V'$, we must have $f(g(x)) = f(g'(x))$ by the inverse property, so it follows by injectivity of f that $g(x) = g'(x)$.

   Thus, for all $a \in A$ we can use the inverse function theorem to construct a neighborhood V_a and an inverse $g_a : V_a \to f^{-1}(V_a)$. It follows that we get an inverse map $f^{-1} : f(A) \to A$ by setting $f^{-1}(a) = g_a(a)$.

   As we have checked above, f^{-1} is well-defined and furhermore, since g_a is continuously differentiable at a and f^{-1} agrees with g_a in an open set around a, it follows that f^{-1} is also continuously differentiable at a for every $a \in A$.

3. Since $f^{-1} : f(A) \to A$ is continuous as we proved above, it pulls back open sets to open sets. Since $f^{-1}(f(B)) = B$, and B is open, it follows that $f(B)$ is open as well. (RP)

Problem 4.2. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuously differentiable function. Show that f *cannot* be injective. Now generalize this to the case of $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m < n$.

Solution. Assume for the sake of contradiction that $f : \mathbb{R}^2 \to \mathbb{R}$ is injective.

Then, it follows that $\partial_x f(x, y)$ is not identically zero for any y, since otherwise we would have f is constant along some fixed line in $\mathbb{R}^2$ and would not be injective.

Now define $g(x, y) : \mathbb{R}^2 \to \mathbb{R}^2$ by $g(x, y) = (f(x, y), y)$. Then we calculate the Jacobian of g at (x, y) to be

$$\begin{bmatrix} \partial_x f(x,y) & \partial_y f(x,y) \\ 0 & 1 \end{bmatrix}.$$

The determinant of this is just $\partial_x f(x, y)$. By our reasoning above, there exists (x, y) such that this is nonzero, and so we can apply the inverse function theorem to get a neighborhood $V \subset \mathbb{R}^2$ of (x, y) and a map $h : V \to g^{-1}(V)$ that is continuously differentiable and inverse to g where defined.

Now since V is open, we can find some y' sufficiently close to (but not equal to) y such that $(f(x,y),y') \in V$. Set $(x',y') = h(f(x,y),y')$. Since h is an inverse, we have $g(x',y') = (f(x,y),y')$. However, by definition of g, $g(x',y') = (f(x',y'),y')$. Therefore, we require $f(x',y') = f(x,y)$ and arrive at a contradiction.

The case for $f : \mathbb{R}^n \to \mathbb{R}^m$ proceeds identically in the case where there is some point $p \in \mathbb{R}^n$ such that $D_p f$ has full rank.

Again consider $g(x,y) : \mathbb{R}^n \to \mathbb{R}^n$ given by $g(x,y) = (f(x,y),y)$, where $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^{n-m}$.

Then the Jacobian of g at p looks like

$$D_p g = \begin{pmatrix} \multicolumn{2}{c}{D_p f} \\ \hline 0 & I_m \end{pmatrix}$$

and since $D_p f$ is of rank m, it is clear by using row reduction or some other method that $D_p g$ has rank n. Then, we apply the inverse function theorem like in the other case and we are done.

Now consider the case where $D_p f$ has rank less than m for every $p \in \mathbb{R}^n$. We will present a proof here different from the one in class that f cannot be injective in this case.

We will make the following slightly more general claim:

Claim: A continuously differentiable map $f : U \to V$ of open sets $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^m$ with $n > m$ and no differential of rank m cannot be injective.

We will show this by induction. The base case, $m = 1$, is immediate. In this case, the differential having rank less than m implies that all of the partial derivatives vanish, which implies that f is constant and therefore cannot be injective.

Let $0 < k < m$ be the maximal rank attained by the differentials of f, and assume for the sake of contradiction that f is injective. Now assume that a point $(x,y) \in U$ achieves that rank. Then, there exist invertible matrices L_1, L_2 such that $L_1 \cdot D_{(x,y)} f \cdot L_2$ is equal to the projection matrix sending $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_k, 0, \ldots, 0)$ (the matrices encode row and column operations resp. on $D_{(x,y)} f$).

Since $L_1 = D_p L_1, L_2 = D_p L_2$ for every point p, it follows by the chain rule that $D_{(x,y)}(L_1 \circ f \circ L_2)$ is the projection matrix described above. Since f is injective iff $L_1 \circ f \circ L_2$ is by invertibility of L_1, L_2, we can just assume without loss of generality that $D_{(x,y)} f$ is this projection matrix.

Now, compose f with the projection $\pi_k : \mathbb{R}^m \to \mathbb{R}^k$ onto the first k coordinates. Then we have π_k is linear, so it follows that $D_{(x,y)} \pi_k f$ is projection from $\mathbb{R}^n$ onto the first k coordinates, which has full rank. Then, set $g = \pi_k f - \pi_k f(x,y)$, so $g(x,y) = 0$ and since it differs by a constant, it has the same differentials as $\pi_k f$.

Therefore, we can apply the implicit function theorem to produce a continuous, differentiable function $h : A \to B$ for $A \subseteq \mathbb{R}^{n-k}, B \subseteq \mathbb{R}^k$ open such that $g(x,h(x)) = 0$ for any $x \in A$.

Now set $\pi_{m-k} : \mathbb{R}^m \to \mathbb{R}^{m-k}$ to be the projection onto the *last* $m-k$ coordinates. We define $\varphi : A \to \mathbb{R}^{m-k}$ by $\varphi(x) = \pi_{m-k} g(x,h(x))$. If f is injective, then by definition we must have g is injective as well, so it follows φ is injective. Furthermore, since everything is continuously differentiable it follows that φ is continuously differentiable as well.

Therefore, we have produced a continuously differentiable, injective map from an open set of $\mathbb{R}^{n-k}$ to an open set of $\mathbb{R}^{m-k}$. Since these are respectively less than n, m we can apply our inductive hypothesis to arrive at a contradiction and show that f is not injective. (RP)

Solution. For completeness's sake, I will also give the version of the solution from class. Our solution to the first part is identical to Rohil's, and we additionally note that it applies essentially without modification to functions of the form $\mathbb{R}^n \to \mathbb{R}$ (i.e., in the case $m = 1$). We now seek to inductively reduce the general case to the specific case of $m = 1$.

Suppose, for technical reasons, that f is merely defined on an open subset $W \subseteq \mathbb{R}^n$. Consider its first component, f_1. If this component is a constant function, then it cannot affect an injectivity argument: a pair of points $x, x' \in W$ with $(f_2, \ldots, f_m)(x) = (f_2, \ldots, f_m)(x')$ must also have $f(x) = f(x')$, since $f_1(x) = f_1(x')$ automatically. Hence, if f_1 is a constant function, we may reduce to the case of $(f_2, \ldots, f_m) \colon W \to \mathbb{R}^{m-1}$.

Now consider instead the more interesting case where f_1 is *not* a constant function. Starting just with a pair of points $(x_1, \ldots, x_n)$ and $(x'_1, \ldots, x'_n)$ with different outputs, we may compare the intermediate pairs where we change just one coordinate, like

$$(x_1, \ldots, x_j, x'_{j+1}, \ldots, x'_n) \quad \text{vs.} \quad (x_1, \ldots, x_{j-1}, x'_j, \ldots, x'_n),$$

and at some index j there must be a change in the output of f_1. Without loss of generality, we may as well re-order the inputs so that $j = 1$. We can even find an open set on which $\frac{\partial f_1}{\partial x_1} \neq 0$, on which we can apply the inverse function theorem to

$$g(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), x_2, \ldots, x_n)$$

to get a local inverse $(s_1, \ldots, s_n)$ with the property that the composition $f \circ s = (x_1, r_2, \ldots, r_n)$ untangles the first coordinate (at the cost of further tangling the others). Note because s is invertible, the failure of $f \circ s$ to be injective is identical to the failure of f to be injective, so we may study $f \circ s$ instead.

We now come to the dimension reduction. Select a particular value $x_1 = a_1$ for the first coordinate so that the hyperplane determined by this equation cuts out a nonempty relative open set from the domain of s. This has the delightful effect of putting us in the first situation: after restricting to $x_1 = a_1$, the first coordinate becomes a constant function, and so we discard not only a dimension from our domain (by passing to the restriction $x_1 = a_1$) but also a dimension from our codomain (by ignoring the first coordinate, which no longer contains any information). (ECP)