Practice Math 25b Midterm #1.1 Solutions

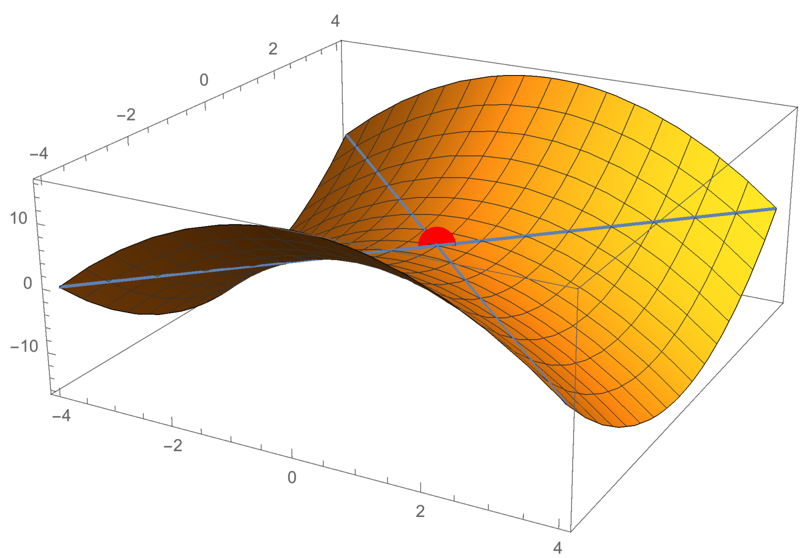
Eric Peterson

Problem 1. 1. Sketch the graph of $z = f(x, y) = (y - x)(x + y)$.

2. Label the points a where $D_a f$ changes rank.

3. Describe the geometry of the level set $z = 0$.

Solution. 1. If you were to generate this picture by hand, you would draw a series of cross-sections. For instance, start by specializing to $x = 0$ and drawing the graph just as y varies in the xz–plane, then specialize to $y = 0$ and draw the graph just as x varies in the yz–plane. Now build some parallel graphs by picking other x and y values and drawing their graphs in the respective parallel–to–xz and parallel–to–yz planes. In the end, you'll draw something like the following Mathematica-generated figure, where the curved grid lines are exactly the parabolas you've drawn by specializing to your various x and y values.



2. We calculate $D_a f = \begin{pmatrix} -2x & 2y \end{pmatrix}$. This is typically rank 1, except when x and y are simultaneously equal to zero, in which case it has rank 0. We've labeled this point with a red dot in the center of the graph, and it forms a *saddle point*.

3. The set $z = 0$ is the set $(y - x)(x + y) = 0$, which is the set where either of $y - x = 0$ or $y + x = 0$ is true. In the former case, $x = y$ traces out a line in the xy–plane, and in the second case $x = -y$ traces out a perpendicular line in the xy–plane. We've drawn these into our picture as the gray-blue lines. (ECP)

Problem 2. 1. Identify $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$, the set of 2×2 matrices, with $\mathbb{R}^4$ in the usual way. For any $a \in \mathbb{R}$, consider the matrix

$$A = \begin{pmatrix} a & a \\ a & a \end{pmatrix},$$

as well as its sequence of powers (A^n). For which values a does this sequence converge?

2. Describe a generalization of your answer to $m \times m$ matrices.

Solution. 1. We claim $A^n = \begin{pmatrix} 2^{n-1}a^n & 2^{n-1}a^n \\ 2^{n-1}a^n & 2^{n-1}a^n \end{pmatrix}$. This has the right form at $n = 1$, hence we can check the claim inductively by calculating

$$\begin{aligned} A^{n+1} &= A^n A \\ &= \begin{pmatrix} 2^{n-1}a^n & 2^{n-1}a^n \\ 2^{n-1}a^n & 2^{n-1}a^n \end{pmatrix} \cdot \begin{pmatrix} a & a \\ a & a \end{pmatrix} \\ &= \begin{pmatrix} 2^{n-1}a^n \cdot a + 2^{n-1}a^n \cdot a & 2^{n-1}a^n \cdot a + 2^{n-1}a^n \cdot a \\ 2^{n-1}a^n \cdot a + 2^{n-1}a^n \cdot a & 2^{n-1}a^n \cdot a + 2^{n-1}a^n \cdot a \end{pmatrix} \\ &= \begin{pmatrix} 2^n a^{n+1} & 2^n a^{n+1} \\ 2^n a^{n+1} & 2^n a^{n+1} \end{pmatrix}. \end{aligned}$$

The question is thus equivalent to calculating those values a for which the sequence $(2^{n-1}a^n)_n$ converges, and this happens exactly in the range $-1/2 < a \leq 1/2$.

2. For a general m, we claim $A^n = (m^{n-1}a^n)_{i,j}$. Again, this can be checked inductively, using the summation formula for matrix multiplication:

$$A^{n+1}_{ij} = (A^n \cdot A)_{ij} = \sum_{k=1}^{m} (A^n)_{ik} \cdot A_{kj} = \sum_{k=1}^{m} m^{n-1}a^n \cdot a = m^n a^{n+1}.$$

Accordingly, this sequence of matrix powers converges when $-1/m < a \leq 1/m$. (ECP)

Problem 3. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function, and let $p \in \mathbb{R}^n$ be any point in its domain. Show that there exist invertible linear maps $L_n\colon \mathbb{R}^n \to \mathbb{R}^n$ and $L_m\colon \mathbb{R}^m \to \mathbb{R}^m$ such that $D_p(L_m \circ f \circ L_n)$ has the block form

$$D_p(L_m \circ f \circ L_n) = \left(\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right)$$

(where, perhaps, any one of these blocks may be missing entirely).

Solution. This is a consequence of Gaussian elimination. The presentation of $D_p f$ as a matrix of partial derivatives admits some pair of strings of elementary matrices $(E_1, \ldots, E_j)$ and $(F_1, \ldots, F_k)$ such that the product

$$E_1 \cdots E_j \cdot (D_p f) \cdot F_1 \cdots F_k$$

has the form requested in the problem statement. However, we also have that $D_a L = L$ for a linear operator L and *any* point a, so we may take $L_m = E_1 \cdots E_j$ and $L_n = F_1 \cdots F_k$ and apply the chain rule. (ECP)

Problem 4. If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ are both differentiable at $a \in \mathbb{R}^n$, prove that the mapping $z\colon \mathbb{R}^n \to \mathbb{R}^{m+k}$ given by $z(v) = (f(v), g(v))$ is also differentiable.

Solution. We claim $D_a z = (D_a f, D_a g)$ is the direct sum of the two individual derivatives. To see this, we write out the difference equation:

$$\lim_{h \to 0} \frac{\|z(a+h) - (z(a) + (D_a z)(h))\|}{\|h\|} = \lim_{(h_f, h_g) \to 0} \frac{\left\| \begin{pmatrix} f(a+h_f) - (f(a) + (D_a f)(h_f)) \\ g(a+h_g) - (g(a) + (D_a g)(h_g)) \end{pmatrix} \right\|}{\|h\|}.$$

Applying the triangle inequality to separate the two limits, we have

$$\begin{aligned} \lim_{(h_f, h_g) \to 0} \frac{\left\| \begin{pmatrix} f(a+h_f) - (f(a) + (D_a f)(h_f)) \\ g(a+h_g) - (g(a) + (D_a g)(h_g)) \end{pmatrix} \right\|}{\|h\|} &\le \lim_{(h_f, h_g) \to 0} \frac{\|f(a+h_f) - (f(a) + (D_a f)(h_f))\|}{\|(h_f, h_g)\|} \\ &\quad + \lim_{(h_f, h_g) \to 0} \frac{\|g(a+h_g) - (g(a) + (D_a g)(h_g))\|}{\|(h_f, h_g)\|}. \end{aligned}$$

Consider just the first of these terms, as the reasoning for the other is identical. Observing that $\|h\| \ge \|h_f\|$, again by the triangle inequality, we have

$$\lim_{(h_f, h_g) \to 0} \frac{\|f(a+h_f) - (f(a) + (D_a f)(h_f))\|}{\|(h_f, h_g)\|} \le \lim_{(h_f, h_g) \to 0} \frac{\|f(a+h_f) - (f(a) + (D_a f)(h_f))\|}{\|h_f\|} = 0.$$

It follows that the original term is zero. Since the original entire limit is zero, z is differentiable with the claimed derivative. (ECP)

Problem 5. On a space of operators $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, we can define the *operator norm* by the amount that an operator stretches the unit sphere:

$$\|A\|_\infty = \max\{\|Av\| : \|v\| = 1\}.$$

In addition to all the usual properties of a norm, this has the additional feature that $\|AB\|_\infty \leq \|A\|_\infty \cdot \|B\|_\infty$.

1. Recite what it means for a function

$$f\colon \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mathcal{L}(\mathbb{R}^k, \mathbb{R}^\ell)$$

to be differentiable at a point $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

2. Consider the specific function

$$\begin{gathered} f\colon \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m), \\ f(A) = AA^*, \end{gathered}$$

where "A^*" denotes the transpose of A. Using the operator norm, show that f is everywhere differentiable and compute its derivative $D_A f$.

Solution. 1. The function f is differentiable at $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with derivative $D_A f$ when the following limit equation is satisfied:

$$\lim_{H \to 0} \frac{\|f(A + H) - (f(A) + (D_a f)(H))\|}{\|H\|} = 0.$$

2. We make a guess at the derivative by computing the following difference:

$$\begin{aligned} f(A + H) - f(A) &= (A + H)(A + H)^* - AA^* \\ &= AA^* + HA^* + AH^* + HH^* - AA^* \\ &= (HA^* + AH^*) + HH^*. \end{aligned}$$

We purport that the linear part of the difference gives $D_A f = HA^* + AH^*$, which we verify by checking the limit equation:

$$\begin{aligned} 0 \leq \lim_{H \to 0} \frac{\|f(A + H) - (f(A) + (D_a f)(H))\|}{\|H\|} &= \lim_{H \to 0} \frac{\|HH^*\|}{\|H\|} \\ &\leq \lim_{H \to 0} \frac{\|H\| \cdot \|H^*\|}{\|H\|} \\ &= \lim_{H \to 0} \|H^*\| = 0. \end{aligned} \qquad \text{(ECP)}$$

Problem 6. Consider the squaring operator

$$\begin{gathered} S\colon \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2) \to \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2), \\ S(A) = A^2. \end{gathered}$$

1. Recall the notion of canonical decomposition from last semester: after complexifying such an A, we are guaranteed that A admits one of two presentations:

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$$

for some $\lambda, \mu \in \mathbb{C}$. Use this to conclude that S has an inverse function f satisfying $f(I) = I$.

2. Now consider the 25b approach to this problem. State the inverse function theorem and show that S satisfies its hypotheses at I.

3. Consider the two approaches. How do they compare? What about the 25b approach is *stronger* and what about it is *weaker*?

Solution. 1. Given such a presentation of such a matrix A, we manually construct a square root.

(a) In the case $A = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$, we take $\sqrt{A} = \begin{pmatrix} \sqrt{\lambda} & 0 \\ 0 & \sqrt{\mu} \end{pmatrix}$.

(b) In the case $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, we write $A = \lambda I + N$, where $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ satisfies $N^2 = 0$. The Taylor series $\sqrt{\lambda I + N} = \sqrt{\lambda} I + \frac{1}{2\sqrt{\lambda}} N + \cdots$ gives

$$\sqrt{A} = \begin{pmatrix} \sqrt{\lambda} & \frac{1}{2\sqrt{\lambda}} \\ 0 & \sqrt{\lambda} \end{pmatrix}.$$

2. The inverse function theorem claims that a function f which is continuously differentiable in an open neighborhood of a point a and which has $D_a f$ invertible has a differentiable local inverse: there exist open sets U and V of the domain and codomain respectively, as well as a differentiable function $f^{-1}\colon V \to U$ satisfying $f^{-1} f = \mathrm{id}_U$, $f f^{-1} = \mathrm{id}_V$, and $D_{f(a)} f^{-1} = -(D_a f)^{-1}$. We apply this to the case of the operator S: because the entries of $S(A)$ are polynomial in the entries of A, S is certainly a continuously differentiable function. We also have $D_A S = AH + HA$, which at the point $A = I$ becomes $D_I S = 2H$, so that $D_I S$ is an invertible operator. The inverse function theorem thus applies.

3. The main advantage of the first approach is that it gives a globally defined square root function. However, it is guaranteed essentially *no* good properties: the definition at every point requires a choice of basis, and there is no reason to think we can make a consistent choice of basis, so square roots even of very nearby operators may well end up looking very different. Meanwhile, while the 25b inverse is only defined on a neighborhood of the identity operator, it is guaranteed to be not only continuous but even differentiable. (ECP)

Practice Math 25b Midterm #1.2 Solutions

Eric Peterson

Problem 1. Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be functions, and define $h(x) = \max\{f(x), g(x)\}$.

1. Suppose that f and g are continuous. Show that h is then continuous.
2. Suppose that f and g are differentiable. Show that h need not be differentiable.

Solution. Consider a point $a \in \mathbb{R}$. If $f(a) < g(a)$, then the same is true on a local neighborhood U of a, and we have $h|_U = g|_U$. Similarly, if $g(a) < f(a)$, then this inequality holds on a local neighborhood U of a, where we have $h|_U = f|_U$. In either case, the continuity of g and of f immediately give the continuity of h, since continuity is a local property.

In the remaining case that $f(a) = g(a)$, we have more work to do. Given an $\varepsilon > 0$, select δ_f and δ_g satisfying the following sentences:

$$|x - a| < \delta_f \implies |f(x) - f(a)| < \varepsilon,$$
$$|x - a| < \delta_g \implies |g(x) - g(a)| < \varepsilon.$$

Set $\delta = \min\{\delta_f, \delta_g\}$, so that $|x - a| < \delta$ gives the simultaneous conclusions $|f(x) - f(a)| < \varepsilon$ and $|g(x) - g(a)| < \varepsilon$. Since $h(a) = f(a) = g(a)$ and either $h(x) = f(x)$ or $h(x) = g(x)$ for any point x, we are thus guaranteed $|h(x) - h(a)| < \varepsilon$. (ECP)

Problem 2. Suppose that for the distance between any two points in $\mathbb{R}$, instead of using our usual Euclidean metric, we instead use the following distance function $d\colon \mathbb{R}^2 \to \mathbb{R}$, known as the **discrete metric**:

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{otherwise} \end{cases}$$

1. Under this new distance function, what are the open and closed sets?
2. Prove that $[0, 1]$ would *not* be compact if we defined distances in this way.

Solution. 1. *Every* set is open. Recall that a set U is open when for every point $u \in U$ one can find a radius $\varepsilon > 0$ such that $|x - u| < \varepsilon$ forces $x \in U$ as well. Fix $\varepsilon = 1/2$: then the only point x satisfying $|x - u| < 1/2$ is $x = u$ itself, which is indeed a member of U. Hence, the condition is satisfied at every point u for $\varepsilon = 1/2$. (Ignore the business about $\mathbb{R}$, which is a red herring, and draw a unit-length tetrahedron if you'd like help visualizing this.)

Similarly, *every* set is closed, because every set is the complement of some other set, and that other set is guaranteed to be open.

2. The sets $U_u = \{u\}$ parameterized over all points $u \in [0, 1]$ give an open cover of $[0, 1]$. Since each point belongs to a unique member of the cover, it is not possible to reduce the cover *at all*, nevermind to a finite subcover. (ECP)

Problem 3. Suppose $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a differentiable function satisfying

$$f\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad D_0 f = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Can there be a continuously differentiable function $g\colon \mathbb{R}^3 \to \mathbb{R}^3$ satisfying

$$g\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad f \circ g\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} z \\ x \\ y \end{pmatrix}?$$

Solution. The main observation is that $D_0 f$ is singular: writing c_j for the j^{th} column, $c_3 - c_1 = 2(c_2 - c_1)$ gives a nontrivial dependence among the output vectors. However, $f \circ g$ is simultaneously claimed to be too nice of an expression, so that we can calculate

$$D_{(1,1,1)}(f \circ g) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

which has $\det D_{(1,1,1)}(f \circ g) = 1$. This violates the chain rule, which would claim: $\det D_{(1,1,1)}(f \circ g) = \det(D_{(0,0,0)} f \circ D_{(1,1,1)} g) = \det D_{(0,0,0)} f \cdot \det D_{(1,1,1)} g$. The left-hand side is 1, while the right-hand side has a factor of zero. (ECP)

Problem 4. 1. Show that the mapping

$$f_3\colon \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n),$$
$$f_3(A) = A^3$$

is differentiable at an arbitrary matrix A and compute its derivative. (Hint: *don't* write out an arbitrary matrix and try to compute the matrix of partial derivatives of the function. Proceed directly to the definition.)

2. Analogously, define a function f_k by $f_k(A) = A^k$. Show that f_k is differentiable for each $k \geq 0$ and calculate its derivative at A.

Solution. 1. As instructed, we proceed to the definition:

$$\lim_{H\to 0} \frac{\|f_3(A+H) - (f_3(A) + (D_A f_3)(H))\|}{\|H\|} = \lim_{H\to 0} \frac{\left\| \begin{matrix} A^3 + AAH + AHA + HAA + \\ +AHH + HAH + HHA + \\ +H^3 - A^3 - (D_A f_3)(H) \end{matrix} \right\|}{\|H\|}.$$

We are thus moved to set $(D_A f_3)(H) = AAH + AHA + HAA$ to account for the linear terms. Performing cancellation yields

$$\cdots = \lim_{H\to 0} \frac{\|AHH + HAH + HHA + H^3\|}{\|H\|}.$$

At this point we can use the triangle inequality to bound this limit:

$$\cdots \leq \lim_{H\to 0} \left(3 \cdot \frac{\|A\| \cdot \|H\|^2}{\|H\|} + \frac{\|H\|^3}{\|H\|} \right) = 0.$$

Our guess for the derivative therefore works.

2. This is an elaboration of the method above. We set the derivative right off the bat:

$$(D_A f_k)(H) = \sum_{j=1}^{k} A^{j-1} H A^{k-j}.$$

The numerator of the difference fraction then becomes

$$\|f_k(A+H) - (f_k(A) + (D_A f_k)(H))\| \leq \sum_{n=2}^{k} \binom{k}{n} \|A\|^{k-n} \|H\|^n.$$

After dividing through by $\|H\|$, each term still has at least one factor of $\|H\|$ still in it and no other confounding dependence on H, hence each term tends to 0 as H tends to 0. (ECP)

Problem 5. Let $W := \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ be the space of 2×2 matrices, and let $U \subseteq W$ be the subset of 2×2 matrices A such that $A - I$ is invertible.

1. Consider the mapping

$$\begin{aligned} &f\colon U \to W, \\ &f(A) = (A^2 - I)(A - I)^{-1}. \end{aligned}$$

For $A \in U$, does $\lim_{A\to I} f(A)$ exist? If so, what is the limit?

2. Now let $J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and consider the set V of matrices B such that $B - J$ is invertible. Consider the mapping

$$\begin{aligned} &g\colon V \to W, \\ &g(B) = (B^2 - J^2)(B - J)^{-1}. \end{aligned}$$

Does $\lim_{B \to J} g(B)$ exist? If so, what is the limit?

Solution. 1. We can perform a factorization $A^2 - I = (A + I)(A - I)$, so that

$$(A^2 - I)(A - I)^{-1} = (A + I)(A - I)(A - I)^{-1} = A + I.$$

This is a continuous function of A, and its limit at I is $2I$.

2. Factorization here does not exactly work:

$$(B + J)(B - J) = B^2 - J^2 + JB - BJ.$$

Nonetheless, we can plug this in:

$$\begin{aligned} (B^2 - J^2)(B - J)^{-1} &= ((B + J)(B - J) - JB + BJ)(B - J)^{-1} \\ &= (B + J) + (BJ - JB)(B - J)^{-1}. \end{aligned}$$

The second term is the troubling one, so we focus on it. The first factor measures the failure of B and J to commute, while the second term measures their difference. From here the computation gets a little brutal, but it is not clever. Here is an example sequence of B operators, picked at random:

$$\begin{aligned} B_n &= \begin{pmatrix} 1 & 1/n \\ 1/n & -1 \end{pmatrix}, & B_n J - J B_n &= \begin{pmatrix} 0 & -2/n \\ 2/n & 0 \end{pmatrix}, \\ (B_n - J)^{-1} &= \begin{pmatrix} 0 & -n \\ -n & 0 \end{pmatrix}, & (B_n J - J B_n)(B_n - J)^{-1} &= \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}. \end{aligned}$$

In particular, the limit of the product is easy to calculate, since the sequence is constant. Meanwhile, here is a second sequence, also picked at random:

$$\begin{aligned} B_n &= \begin{pmatrix} 1 + 1/n & 1/n \\ 1/n & -1 \end{pmatrix}, & B_n J - J B_n &= \begin{pmatrix} 0 & -2/n \\ 2/n & 0 \end{pmatrix}, \\ (B_n - J)^{-1} &= \begin{pmatrix} 0 & -n \\ -n & n \end{pmatrix}, & (B_n J - J B_n)(B_n - J)^{-1} &= \begin{pmatrix} 2 & -2 \\ 0 & -2 \end{pmatrix}. \end{aligned}$$

These two sequences show different limiting values by approaching along different trajectories from B to J, hence the multivariate limit cannot exist. (ECP)

Problem 6. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function of *one* variable.

1. State what the inverse function theorem claims in this context.

2. Prove the inverse function theorem in this context.

Solution. 1. Assume that f is continuously differentiable on an open neighborhood of a point $a \in \mathbb{R}$, where $f'(a) \neq 0$. There is then a (smaller) open neighborhood U on which f is injective, as well as an inverse function $g\colon f(U) \to U$ to f which is itself differentiable with derivative $g'(f(a)) = (f'(a))^{-1}$.

2. We have $f'(a) \neq 0$, and we can (by replacing f with $-f$ if necessary) assume that $f'(a) > 0$. Because f' is a continuous function, we even have $f'(x) > 0$ on a connected open neighborhood U of a. On this neighborhood, f is monotonically increasing, hence injective, hence has an inverse $g\colon f(U) \to U$. We calculate $g'(f(a))$ through its limit definition:

$$g'(f(a)) = \lim_{y \to f(a)} \frac{g(y) - g(f(a))}{y - f(a)}$$

By setting $x = g(y)$, we equivalently have

$$= \lim_{f(x) \to f(a)} \frac{x - a}{f(x) - f(a)}.$$

This is the reciprocal of the difference quotient defining the derivative of f at a. Because we have assumed that the original limit is nonzero, this limit exists and is the reciprocal of the original, finally giving

$$= (f'(a))^{-1}. \tag{ECP}$$

Math 25b Midterm #1 Solutions

Eric Peterson

Problem 1. Given a subset $S \subset \mathbb{R}$, we define the *interior* of S, denoted S°, to be the set of all points $x \in S$ such that there exists an open interval (a, b) containing x that lies within S.

1. Show that the interior of the closed interval $[a, b]$ is the open interval (a, b).
2. Calculate the interiors of $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{R} \setminus \mathbb{Q}$. (You should discover that while $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$, it is not the case that $\mathbb{R}^\circ = \mathbb{Q}^\circ \cup (\mathbb{R} \setminus \mathbb{Q})^\circ$.)

Solution.

1. For any point $a < t < b$, set $\varepsilon = \min\{t - a, b - t\}$, so that $(t - \varepsilon, t + \varepsilon) \subseteq [a, b]$ and t is an interior point. On the other hand, no matter what $\varepsilon > 0$ we select, $(a - \varepsilon, a + \varepsilon)$ contains the point $a - \varepsilon/2$, which is not in the interval, so a cannot be an interior point. Similarly, $(b - \varepsilon, b + \varepsilon)$ contains $b + \varepsilon/2$, which is not in the interval.
2. Every point is an interior point of $\mathbb{R}$. However, *no* point is an interior point of $\mathbb{Q}$: $(p/q - \varepsilon, p/q + \varepsilon)$ always contains an irrational point no matter what $\varepsilon > 0$ is chosen, hence is not a subset of $\mathbb{Q}$. Similarly, $(r - \varepsilon, r + \varepsilon)$ contains a rational point for any irrational r and $\varepsilon > 0$, hence is not a subset of $\mathbb{R} \setminus \mathbb{Q}$. So, we have

$$\mathbb{R}^\circ = \mathbb{R}, \qquad \mathbb{Q}^\circ = \varnothing, \qquad (\mathbb{R} \setminus \mathbb{Q})^\circ = \varnothing. \tag{ECP}$$

Problem 2.

1. Show that if a continuously differentiable function $f\colon \mathbb{R}^2 \to \mathbb{R}$ can be written as $f(x, y) = \varphi(x^2 + y^2)$ for some auxiliary continuously differentiable function $\varphi(s)\colon \mathbb{R} \to \mathbb{R}$, then f satisfies

$$x \cdot \frac{\partial f}{\partial y} - y \cdot \frac{\partial f}{\partial x} = 0.$$

2. Now show the converse: if a continuously differentiable f satisfies that same identity, then there is necessarily a continuously differentiable $\varphi\colon \mathbb{R} \to \mathbb{R}$ with $f(x, y) = \varphi(x^2 + y^2)$. (Hint: consider the polar coordinate system (r, θ), related to the standard coordinate system by $x = r\cos\theta$, $y = r\sin\theta$. Try thinking of f as a function of r and θ instead. What is its partial derivative with respect to θ?)

Solution. 1. Writing $\varphi(s)$ as a function of s, we use the chain rule:

$$\frac{\partial f}{\partial x} = \frac{\partial \varphi}{\partial s} \cdot \frac{\partial s}{\partial x} = \varphi'(x^2+y^2) \cdot 2x,$$
$$\frac{\partial f}{\partial y} = \frac{\partial \varphi}{\partial s} \cdot \frac{\partial s}{\partial y} = \varphi'(x^2+y^2) \cdot 2y.$$

It's true that these agree after multiplying by y and x respectively, and hence their difference vanishes.

2. As instructed by the hint, we calculate the partial derivative of f with respect to θ:

$$\begin{aligned}\frac{\partial f}{\partial \theta} &= \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial \theta} \\ &= \frac{\partial f}{\partial x} \cdot (-r\sin\theta) + \frac{\partial f}{\partial y} \cdot (r\cos\theta) \\ &= \frac{\partial f}{\partial x} \cdot (-y) + \frac{\partial f}{\partial y} \cdot x.\end{aligned}$$

Our assumption is that this difference vanishes, hence $\frac{\partial f}{\partial \theta}$ vanishes, and hence f is independent of θ. We can thus write f solely as a function of r, i.e., there exists a single-variable function φ with $f(x, y) = f(r, \theta) = \varphi(\theta)$. (ECP)

Problem 3. Consider the following function:

$$f\colon \mathbb{R}^2 \to \mathbb{R},$$
$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{otherwise.} \end{cases}$$

1. Show that both partial derivatives exist everywhere.

2. Where is f differentiable?

Solution. 1. When $x = 0$ or $y = 0$, this is easy:

$$\frac{\partial f}{\partial x} = \frac{(x^2+y^2)\cdot y - xy \cdot 2x}{(x^2+y^2)^2},$$
$$\frac{\partial f}{\partial y} = \frac{(x^2+y^2)\cdot x - xy \cdot 2y}{(x^2+y^2)^2}.$$

In fact, when $x = 0$ or $y = 0$, this is also easy, since in this case $f = 0$ is the constant function. Hence, the derivative vanishes.

2. Away from the point $(0, 0)$, the partial derivatives are both continuous, and hence f is differentiable with derivative expressed by the matrix of partials. At the origin, however, the function is not differentiable, or even continuous: approaching along $x = y = t$ we have $f(t, t) = t^2/(t^2 + t^2) = 1/2$, which does not agree with approaching along $y = 0$ or $x = 0$. (ECP)

Problem 4. 1. On the homework, you considered the directional derivative

$$\mathbb{D}_I^H(\det) = \lim_{t\to 0} \frac{\det(I+tH) - \det(I)}{t}.$$

Use the permutation formula for the determinant to compute $\det(I+tH)$ and show that this directional derivative is $\operatorname{tr}(H)$.

2. On your homework, you also showed the very complicated identity

$$(D_A \det)(H) = \sum_{j=1}^{n} \det\left(a_1|\cdots|a_{j-1}|h_j|a_{j+1}|\cdots|a_n\right),$$

where A is an $n\times n$ matrix, $H \in T_A\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is a matrix of displacement values, and det is considered as a differentiable function $\det\colon \mathcal{L}(\mathbb{R}^n,\mathbb{R}^n) \to \mathbb{R}$. Set $A = I$ to be the identity matrix, and conclude also from *this* homework problem that $D_I \det = \operatorname{tr}$.

3. Now assume that A is an arbitrary invertible matrix. Use $(A+H) = A(I+A^{-1}H)$ and the definition of the derivative to calculate $D_A \det$ in terms of a trace.

Solution. 1. Start by writing down the permutation formula:

$$\det A = \sum_{P \text{ a permutation of } \{1,\dots,n\}} (-1)^{\text{disorders}(P)} \prod_j a_{j,P(j)}.$$

The idea is to compute $\det(I+tH)$ as a polynomial in t by considering which terms in this product expansion involve *no* ts or *one* t for $A = I + tH$.

- For no ts, we look only for terms in the product expansion that come from I. Only a single term is nonzero: the identity permutation contributes a 1.
- For one t, we allow only one term in the product expansion to have something to do with tH. However, in order to get something nonzero, all the other factors have to select entries from the diagonal of I, hence the permutation must again be the identity and we have $\sum_j t(1\cdots 1\cdot h_{jj}\cdot 1\cdots 1) = t\operatorname{tr}(H)$.

All the other terms have at least two factors of t in them, which we divide out (leaving something that's still polynomial, hence continuous, in t). Hence, we have

$$\mathbb{D}_I^H(\det) = \lim_{t\to 0}\frac{\det(I+tH)-\det(I)}{t} = \lim_{t\to 0}\frac{t\operatorname{tr}(H) + t^2(\cdots)}{t} = \operatorname{tr} H.$$

2. Just make the appropriate substitutions:

$$(D_I \det)(H) = \sum_{j=1}^{n} \det(I_1|\cdots|I_{j-1}|h_j|I_{j+1}|\cdots|I_n) = \sum_{j=1}^{n} h_{jj} = \operatorname{tr} H.$$

3. To start, take the difference:

$$\det(A+H) - (\det A + (D_A \det)H) =$$

$$= \det A \cdot (\det(I + A^{-1}H) - (\det I + (\det A)^{-1} \cdot (D_A \det)(H))).$$

Based on the previous answer, this suggests

$$(\det A)^{-1} \cdot (D_A \det)(H) = (D_I \det)(A^{-1}H) = \operatorname{tr}(A^{-1}H),$$

and solving for $D_A \det$ gives

$$(D_A \det)(H) = (\det A) \cdot \operatorname{tr}(A^{-1}H).$$

Substituting all this back in, we check the limit property:

$$\begin{aligned} &\lim_{H\to 0} \frac{\det(A+H) - (\det A + (D_A \det)(H))}{\|H\|} \\ &\qquad = \lim_{H\to 0} \frac{\det(A+H) - (\det A + (\det A)\operatorname{tr}(A^{-1}H))}{\|H\|} \\ &\qquad = \det A \cdot \lim_{H\to 0} \frac{\det(I + A^{-1}H) - (\det I + (D_I \det)(A^{-1}H))}{\|H\|}. \end{aligned}$$

Making the substitution $J = A^{-1}H$ makes this more readable:

$$= \det A \lim_{H\to 0} \frac{\det(I+J) - (\det I + (D_I \det)(J))}{\|AJ\|.}$$

Taking the minimum singular value s_{min} of A, which is guaranteed to be nonzero, and noting that $H \to 0$ if and only if $J \to 0$, we have

$$\leq \frac{\det A}{s_{min}} \lim_{J\to 0} \frac{\det(I+J) - (\det I + (D_I \det)(J))}{\|J\|}.$$

The limit goes to zero, hence the original limit goes to zero, verifying that our guess derivative is correct. (ECP)

Problem 5. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable. Assume that there exists $0 < B < 1$ such that $|f'(x)| < B$ for all x. Prove that there exists an $a \in \mathbb{R}$ such that $f(a) = a$.

Solution. If $f(x) = x$ for all choices of x, we are certainly done. Otherwise, we may assume that there exists some choice of x_0 with $f(x_0) \neq x_0$ — the more interesting case. Replacing f with $-f$ if necessary, we may assume that $f(x_0) > x_0$. The graph of f is constrained to live in the right-ward facing cone

$$\{(x, y) : f(x_0) - B|x - x_0| < y < f(x_0) + B|x - x_0|\}$$

If the graph escaped this cone, then the mean value theorem for derivatives would guarantee a point on the graph of f whose derivative exceeds the bound B. However, all of the points y in this range eventually satisfy $y < x$: namely, once $(x - x_0) > (f(x_0) - x_0)/B$. Select some x_1 satisfying this inequality. Since there must exist a point y_1 with (x_1, y_1) in the graph of f, since it must satisfy $f(x_1) - x_1 < 0$, and since $f(x_0) - x > 0$, we use the intermediate value theorem for continuous functions to conclude that somewhere between these points there must be an $x_0 < x < x_1$ with $f(x) - x = 0$. (ECP)

Problem 6. Consider the system of equations

$$\begin{aligned} x + y + \sin(xy) &= h, \\ \sin(x^2 + y) &= 2h. \end{aligned}$$

Does this system have a solution for sufficiently small values $h \in \mathbb{R}$?

Solution. Write the left-hand side as a function:

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + y + \sin(xy) \\ \sin(x^2 + y) \end{pmatrix}.$$

This function is continuously differentiable near the origin, and at the origin we see that

$$D_0 f = \begin{pmatrix} 1 + y\cos(xy) & 1 + x\cos(xy) \\ 2x\cos(x^2 + y) & \cos(x^2 + y) \end{pmatrix}\Bigg|_{(x,y)=(0,0)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is an invertible operator. Hence, the inverse function theorem applies, so that f admits an inverse near the origin. As a special case, the domain of the local inverse includes values like $(h, 2h)$ for $h \ll 1$, and hence solutions are guaranteed to exist. (ECP)