Math 291-2: Final Exam Solutions
Northwestern University, Winter 2016

1. Determine whether each of the following statements is true or false. If it is true, explain why; if it is false, give a counterexample.

(a) If A and B are diagonalizable 2×2 matrices, then $A + B$ is diagonalizable.

(b) There is no 3×3 matrix which transforms a sphere of radius 1 into a sphere of radius 3 and at the same time transforms a cube with edges of length 2 into one with edges of length 1.

(c) The limit $\lim_{(x,y)\to(0,0)} \frac{x^4y^4}{(x^2+y^4)^3}$ does not exist.

Solution. (a) This is false. For instance, $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ are both diagonalizable, but $A + B = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is not. (A point of clarification: some of attempted solutions claimed that the zero matrix was not diagonalizable, but the zero matrix is diagonal and any diagonal matrix is diagonalizable!)

(b) This is true. If A was going to be such a matrix A, then we would need to have

$$\text{volume of output sphere } = |\det A|(\text{volume of input sphere})$$

based on the expansion factor interpretation of $|\det A|$, so in particular this would imply that $|\det A| > 1$. But at the same time a larger cube is being transformed into a smaller cube, which would require that $|\det A| < 1$, which is not compatible with the first inequality.

(c) This is true. Approaching the origin along $x = 0$ gives 0 as the value for the limit, while approaching along $x = y^2$ gives $\frac{1}{8}$ since when $x = y^2$ we have:

$$\frac{x^4y^4}{(x^2+y^4)^3} = \frac{y^{12}}{(2y^4)^3} = \frac{y^{12}}{8y^{12}}.$$

Since approaching $(0,0)$ along different directions gives different values for the limit, the limit in question does not exist. $\square$

2. Suppose Q is an $n \times n$ matrix such that

$$Q\mathbf{x} \cdot Q\mathbf{y} = 4(\mathbf{x} \cdot \mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

Show that $\det Q = \pm 2^n$. Hint: Figure out how to relate Q to something which is orthogonal.

Proof 1. The given equality gives

$$\left(\frac{1}{2}Q\mathbf{x}\right) \cdot \left(\frac{1}{2}Q\mathbf{y}\right) = \mathbf{x} \cdot \mathbf{y}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. This says that $\frac{1}{2}Q$ preserves dot products, so $\frac{1}{2}Q$ is orthogonal and hence

$$\det\left(\frac{1}{2}Q\right) = \pm 1$$

since the determinant of an orthogonal matrix is ± 1. On the other hand,

$$\det\left(\frac{1}{2}Q\right) = \left(\frac{1}{2}\right)^n \det Q$$

since $\frac{1}{2}Q$ is obtained by multiplying each of the n rows of Q by $\frac{1}{2}$, and by multilinearity of the determinant, each such row operation scales the determinant by that same amount. Thus

$$\frac{1}{2^n}Q = \pm 1,$$

so $\det Q = \pm 2^n$ as claimed. $\square$

Proof 2. Using properties of transposes, we have

$$\mathbf{x} \cdot 4\mathbf{y} = 4(\mathbf{x} \cdot \mathbf{y}) = Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot Q^T Q\mathbf{y}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. This implies that $Q^T Q\mathbf{y} = 4\mathbf{y}$ for any $\mathbf{y}$ (since in general, $\mathbf{x} \cdot \mathbf{v} = \mathbf{x} \cdot \mathbf{w}$ for all $\mathbf{x}$ implies $\mathbf{v} = \mathbf{w}$), so $Q^T Q = 4I$. Thus

$$(\det Q)^2 = (\det Q^T)(\det Q) = \det(Q^T Q) = \det(4I) = 4^n,$$

where $\det(4I) = 4^n$ since $4I$ is a diagonal matrix consisting of n fours down the diagonal. Hence $\det Q = \pm 2^n$ as claimed. $\square$

Proof 3. The given equality gives

$$\|Q\mathbf{x}\|^2 = Q\mathbf{x} \cdot Q\mathbf{x} = 4(\mathbf{x} \cdot \mathbf{x}) = 4\|\mathbf{x}\|^2$$

for any $\mathbf{x} \in \mathbb{R}^n$, so $\|Q\mathbf{x}\| = 2\|\mathbf{x}\|$ for any $\mathbf{x}$. Also, for any nonzero $\mathbf{x}, \mathbf{y}$ we have:

$$\frac{Q\mathbf{x} \cdot Q\mathbf{y}}{\|Q\mathbf{x}\| \|Q\mathbf{y}\|} = \frac{4(\mathbf{x} \cdot \mathbf{y})}{2\|\mathbf{x}\| 2\|\mathbf{y}\|} = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}.$$

The first fraction is cosine of the angle between $Q\mathbf{x}$ and $Q\mathbf{y}$ and the second is cosine of the angle between $\mathbf{x}$ and $\mathbf{y}$, so this equality implies that these angles must be the same and hence Q preserves angles. Now, take the parallelopiped in $\mathbb{R}^n$ formed by the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Applying Q scales the length of each edge of this parallelopiped by 2 and leaves all right angles as right angles, so the resulting parallelopiped has volume given by the products of the lengths of the edges, which 2^n since each such length is 2 times the corresponding length of 1 in the original parallelopiped. Thus the expansion factor $|\det Q|$ is 2^n, so $\det Q = \pm 2^n$ as claimed. $\square$

3. For $n \geq 2$, let $T : P_n(\mathbb{R}) \to P_n(\mathbb{R})$ be the linear transformation defined by

$$T(p(x)) = x^2 p''(x).$$

That is, T sends a polynomial $p(x)$ to $x^2 p''(x)$. Determine all eigenvalues and eigenvectors of T.

Solution. The $n \geq 2$ stipulation is actually irrelevant here—it was leftover from an earlier version of this problem where I asked to show that T was not diagonalizable, which requires $n \geq 2$.

First, we have $T(1) = 0$ and $T(x) = 0$, so any constant polynomial and any polynomial of degree 1 is an eigenvector with eigenvalue 0. For any $2 \leq k \leq n$, we have

$$T(x^k) = x^2[k(k-1)x^{k-2}] = k(k-1)x^k,$$

so x^k—and any of its scalar multiples—is an eigenvector of T with eigenvalue $k(k-1)$. So far this gives that anything of the form cx^k for $0 \leq k \leq n$ and $c \in \mathbb{R}$ is an eigenvector with eigenvalue

$k(k-1)$, but since among these is the basis $1, x, \dots, x^n$ of $P_n(\mathbb{R})$, there can be no more eigenvectors. Hence we have found all eigenvalues and eigenvectors.

Alternatively, denote an arbitrary $p(x) \in P_n(\mathbb{R})$ by

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n.$$

In order for this to be an eigenvector of T, we need:

$$2a_2x^2 + 3(2)a_3x^3 + \cdots + \cdots + n(n-1)a_n^n = \lambda(a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_nx^n),$$

since left side is $T(p(x)) = x^2p''(x)$ and the right side is $\lambda p(x)$. Comparing coefficients on both sides shows that such an equality holds only when

$$a_2 = \cdots = a_n = 0 \text{ and } \lambda = 0, \text{ or } a_i = 0 \text{ for } i \neq k \text{ and } \lambda = k(k-1),$$

where k in the latter case is between 2 and n. Thus we get the same eigenvalues and eigenvectors as in the first method. □

4. Let A be an $n \times n$ symmetric matrix. Show that all eigenvalues of A are positive if and only if $\mathbf{x}^T A\mathbf{x} > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$.

Proof. This was a discussion problem. Orthogonally diagonalize A to get an orthonormal eigenbasis $\mathbf{u}_1, \dots, \mathbf{u}_n$ of $\mathbb{R}^n$. Denote by $c_1, \dots, c_n$ the coordinates corresponding to this basis and by $\lambda_1, \dots, \lambda_n$ the associated eigenvalues. Then we can rewrite $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ as

$$q(\mathbf{c}) = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2$$

relative to this basis. If all the eigenvalues of A are positive, then $q(\mathbf{c}) > 0$ for all $\mathbf{c} \neq \mathbf{0}$ since each term in the rewritten expression for q is nonnegative and at least one will be strictly positive. Conversely, if $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$, then in particular

$$q(\mathbf{u}_i) = \lambda_i > 0$$

since the coordinates for $\mathbf{u}_i$ are $c_i = 1$ and $c_j = 0$ for $j \neq i$. Hence all the eigenvalues of A are positive. □

5. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ has the property that for any $\mathbf{x} \in \mathbb{R}^n$, $\|f(\mathbf{x})\| \leq \|\mathbf{x}\|$ and

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

Show that f is the zero function.

Proof. This was also a discussion problem, only with an extra assumption. The given property implies that $Df(\mathbf{a}) = 0$ at every $a \in \mathbb{R}^n$. Indeed, writing the given limit as

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

where 0 in the numerator denotes the zero matrix, shows that f is differentiable at $\mathbf{a}$ with 0 being the matrix satisfying the required property of the derivative of f at $\mathbf{a}$. Hence $Df(\mathbf{a}) = 0$ as claimed. Thus, all entries of $Df(\mathbf{a})$ are zero, so all partial derivatives of all components of f are 0 at $\mathbf{a}$. Since this is true for any $\mathbf{a} \in \mathbb{R}^n$, f is constant. Since $\|f(\mathbf{0})\| \leq \|\mathbf{0}\| = 0$, we must have $f(\mathbf{0}) = \mathbf{0}$, so the constant which f equals must be zero as claimed. □

6. Suppose $g : \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^{m+n} \to \mathbb{R}^m$ are each differentiable and that

$$F(g(\mathbf{x}), \mathbf{x}) = \mathbf{0} \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

Write the Jacobian matrix of F at a point $\mathbf{y} \in \mathbb{R}^{m+n}$ as

$$DF(\mathbf{y}) = \begin{pmatrix} A(\mathbf{y}) & B(\mathbf{y}) \end{pmatrix}$$

where $A(\mathbf{y})$ is the $m \times m$ matrix consisting of the first m columns of $DF(\mathbf{y})$ and B is the $m \times n$ matrix consisting of the last n columns. If $\det A(g(\mathbf{x}), \mathbf{x}) \neq 0$ for each $\mathbf{x} \in \mathbb{R}^n$, show that

$$Dg(\mathbf{x}) = -A(g(\mathbf{x}), \mathbf{x})^{-1} B(g(\mathbf{x}), \mathbf{x}).$$

Hint: Express the function $\mathbf{x} \to F(g(\mathbf{x}), \mathbf{x})$ as a composition of differentiable functions.

What's the point? This problem seems to have come out of nowhere, but it is actually a part of the statement of what is known as the *Implicit Function Theorem*, which is one of the most important theorems in analysis. The setup is as follows. Say we have some equation of the form

$$F(y_1, \ldots, y_m, x_1, \ldots, x_n) = 0.$$

The Implicit Function Theorem says that under the determinant assumption given in the problem, the variables $y_1, \ldots, y_m$ can be solved for in terms of the variables $x_1, \ldots, x_n$, or more concretely, the given equation implicitly defines $\mathbf{y} = (y_1, \ldots, y_n)$ as a function of $\mathbf{x} = (x_1, \ldots, x_n)$, meaning there is some differentiable function g such that $\mathbf{y} = g(\mathbf{x})$. (Think of how the equation of a circle $x^2 + y^2 = 1$ implicitly defines y as a function of x via $y = \pm\sqrt{1 - x^2}$. This is also analogous to how when solving a system of linear equations $A\mathbf{x} = \mathbf{b}$, the "non-free variables" can be expressed in terms of the "free" variables; indeed, the Implicit Function Theorem is *precisely* the non-linear analog of this fact about systems of linear equations.) The point of this problem is that it describes the Jacobian matrix of the "implicit" function g in terms of the Jacobian matrix of F. Note, however, that you do not need to know anything about this theorem in order to solve this problem, which is just an application of the chain rule.

Proof. Consider the function $h : \mathbb{R}^n \to \mathbb{R}^{m+n}$ defined by $h(\mathbf{x}) = (g(\mathbf{x}), \mathbf{x})$. The composition $F \circ h$ is the function $\mathbf{x} \mapsto F(g(\mathbf{x}), \mathbf{x})$, which we are assuming in the problem to be identically zero. Hence $D(F \circ h)(\mathbf{x}) = 0$ at all $\mathbf{x} \in \mathbb{R}^n$. On the other hand, the chain rule gives:

$$D(F \circ h)(\mathbf{x}) = DF(h(\mathbf{x}))Dh(\mathbf{x}) = \begin{pmatrix} A(g(\mathbf{x}), \mathbf{x}) & B(g(\mathbf{x}), \mathbf{x}) \end{pmatrix} \begin{pmatrix} Dg(\mathbf{x}) \\ I \end{pmatrix},$$

where $Dh(\mathbf{x}) = \left(\begin{smallmatrix} Dg(\mathbf{x}) \\ I \end{smallmatrix}\right)$ is the $(m+n) \times n$ matrix whose upper portion is the $m \times n$ matrix $Dg(\mathbf{x})$ and lower portion the $n \times n$ identity matrix I. (This is found by computing the partial derivatives of $h(x_1, \ldots, x_n) = (g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n), x_1, \ldots, x_n)$, where $g_1, \ldots, g_m$ are the components of g; the partial derivatives of the first m components contribute to Dg and the partial derivatives of the final n components are either 0's or 1's, which is what gives the identity matrix portion.)

Thus we have:

$$0 = \begin{pmatrix} A(g(\mathbf{x}), \mathbf{x}) & B(g(\mathbf{x}), \mathbf{x}) \end{pmatrix} \begin{pmatrix} Dg(\mathbf{x}) \\ I \end{pmatrix} = A(g(\mathbf{x}), \mathbf{x})Dg(\mathbf{x}) + B(g(\mathbf{x}), \mathbf{x}).$$

Since $\det A(g(\mathbf{x}), \mathbf{x}) \neq 0$, $A(g(\mathbf{x}), \mathbf{x})$ is invertible, and solving for $Dg(\mathbf{x})$ gives

$$Dg(\mathbf{x}) = -A(g(\mathbf{x}), \mathbf{x})^{-1} B(g(\mathbf{x}), \mathbf{x})$$

as claimed. (This was certainly the toughest problem, but illustrates the power of being able to express derivatives in terms of Jacobian matrices, in particular when it comes to working out some chain rule application.) □

7. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is C^2. Fix a unit vector $\mathbf{u} \in \mathbb{R}^n$ and define the differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ by $g(\mathbf{x}) = D_{\mathbf{u}} f(\mathbf{x})$. Show that the maximal directional derivative of g at $\mathbf{x} \in \mathbb{R}^n$ occurs in the direction of $D^2 f(\mathbf{x})\mathbf{u}$, which is the product of the Hessian matrix $D^2 f(\mathbf{x})$ and the vector $\mathbf{u}$.

What's the point? Again, this problem seems to have come out of nowhere. But actually, this problem is related to the notion of a *second-order directional derivative*, which is analogous to second-order partial derivatives only that we differentiate in arbitrary directions. Given unit vectors $\mathbf{u}$ and $\mathbf{v}$, the second-order directional derivative $D_{\mathbf{v}} D_{\mathbf{u}} f(\mathbf{x})$ measures how the directional derivative in the $\mathbf{u}$-direction changes as you move in the $\mathbf{v}$-direction. It turns out that if f is C^2, this second-order directional derivative is given by

$$D_{\mathbf{v}} D_{\mathbf{u}} f(\mathbf{x}) = \mathbf{v}^T D^2 f(\mathbf{x}) \mathbf{u}.$$

Thus, just as the Jacobian matrix $Df(\mathbf{x})$ encodes information about all possible directional derivatives, the Hessian matrix $D^2 f(\mathbf{x})$ encodes information about all possible second-order directional derivatives. Note, however, that you do not need to know anything about second-order directional derivatives in order to solve this problem.

Proof. Since f is differentiable, $D_{\mathbf{u}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{u}$ for any $\mathbf{x} \in \mathbb{R}^n$. Thus g is explicitly given by

$$g(\mathbf{x}) = D_{\mathbf{u}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{u} = \frac{\partial f}{\partial x_1}(\mathbf{x}) u_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}) u_n$$

where $\mathbf{u} = (u_1, \ldots, u_n)$. The direction of the maximal directional derivative of g at $\mathbf{x} \in \mathbb{R}^n$ is given by $\nabla g(\mathbf{x})$, so we must compute this gradient.

Differentiating the explicit expression for g given above with respect to each of $x_1, \ldots, x_n$ shows that:

$$\nabla g(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) u_1 + \cdots + \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) u_n \\ \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) u_1 + \cdots + \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) u_n \end{pmatrix},$$

where the i-th entry is the derivative of $\frac{\partial f}{\partial x_1}(\mathbf{x}) u_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}) u_n$ with respect to x_i. But this expression can be written as

$$\nabla g(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

which is the transpose of $D^2 f(\mathbf{x})$ times $\mathbf{u}$. Since f is C^2, $D^2 f(\mathbf{x})$ is symmetric so we get

$$\nabla g(\mathbf{x}) = (D^2 f(\mathbf{x}))^T \mathbf{u} = D^2 f(\mathbf{x}) \mathbf{u}$$

as the direction of the maximal directional derivative at $\mathbf{x}$ as claimed. □