MA 253 Exam 3 Review Partial Solutions

1. Orthogonality

(1) What does it mean for a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ to be orthonormal? How can you check if they are orthonormal?

Solution: If $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ if $i \neq j$ and $\mathbf{v}_i \cdot \mathbf{v}_j = 1$ if $i = j$. You can make them the columns of a matrix Q, and then check if $Q^T Q = I$.

(2) Let $A = \begin{pmatrix} | & | & \cdots & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \\ | & | & \cdots & | \end{pmatrix}$ be an orthonormal $n \times k$ matrix.

(a) Explain why $A^T A = I_k$.

Solution: The entries of the matrix $A^T A$ are formed by taking a row of A^T and dotting it a column of A. Since the columns of A (and hence the rows of A^T) are orthonormal, these dot products produce 1s down the diagonal and 0s off the diagonal.

(b) If $n = k$, explain why also $AA^T = I_n$

Solution: If $n = k$, then A is a square matrix. Since A is orthogonal, $A^T A = I_n$. We have $\det(A^T A) = \det(A^T)\det(A) = 1$, by properties of the determinant. Thus, $\det(A) \neq 0$ and so A is invertible. The inverse must be A^T since a square invertible matrix does not have different left and right inverses. Hence, $AA^T = I$.

(c) Explain why, if $n \neq k$, then AA^T is the matrix for projecting vectors onto the subspace spanned by the columns of A.

Solution: Here's a slightly different proof from what we did in class:

Choose an orthonormal basis $\mathbf{a}_1, \ldots, \mathbf{a}_k$ for the image of A and an orthonormal basis $\mathbf{b}_1 \ldots \mathbf{b}_{n-k}$ for the orthogonal complement of the image of A. Since $(\mathrm{im} A)^\perp = \ker A^T$, the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_{n-k}$ are a basis for $\ker A^T$. Write an arbitary x in this basis for $\mathbb{R}^n$ as:

$$\mathbf{x} = \alpha_1 \vec{a}_1 + \ldots + \alpha_k \mathbf{a}_k + \beta_1 \mathbf{b}_1 + \ldots + \beta_{n-k} \mathbf{b}_{n-k}.$$

Applying AA^T to both sides we get:

$$\begin{aligned} AA^T\mathbf{x} &= \alpha_1 AA^T \mathbf{a}_1 + \ldots + \beta_{n-k} AA^t \mathbf{b}_{n-k} \\ &= \alpha_1 AA^T \mathbf{a}_1 + \ldots + \alpha_k AA^T \mathbf{a}_k \end{aligned}$$

since each of the $\mathbf{b}_i$ are in the kernel of A^T.

Since each $\mathbf{a}_i \in \operatorname{im} A$, we have $\mathbf{a}_i = A\mathbf{c}_i$ for some $\mathbf{c}_i \in \mathbb{R}^k$. Thus,

$$\begin{aligned} AA^T\mathbf{x} &= \alpha_1 AA^T\mathbf{a}_1 + \ldots + \alpha_k AA^T\mathbf{a}_k \\ &= \alpha_1 AA^T A\mathbf{c}_i + \ldots + \alpha_k AA^T A\mathbf{c}_k \end{aligned}$$

Recalling that $A^T A = I_k$ and that $A\mathbf{c}_i = \mathbf{a}_i$, we have

$$AA^T\mathbf{x} = \alpha_1\mathbf{a}_1 + \ldots + \alpha_k\mathbf{a}_k$$

which is what we mean by orthogonal projection onto the subspace spanned by the columns of $\mathbf{a}$.

(3) Explain all details of the Gram-Schmidt process and be able to apply to at least 3 vectors.

(4) Explain how the Gram-Schmidt process gives rise to a QR-factorization of a matrix.

Solution: The basis produced by GS is orthonormal, when put into the columns of a matrix we get Q. The original basis can be written in terms of this orthonormal basis; those equations give rise to the triangular matrix R.

(5) Explain how to view QR-factorization as a change of basis.

(6) What is the orthogonal complement $V^\perp$ of a subspace $V \subset \mathbb{R}^n$?

(7) Show that $(V^\perp)^\perp = V$.

Solution: We can form a basis $\mathscr{B}$ for $\mathbb{R}^n$ by taking a basis for V and a basis for $V^\perp$. An arbitrary vector of $\mathbb{R}^n$ can be written as a linear combination of a vector in V and a vector in $V^\perp$. Recall that $(V^\perp)^\perp$ is the set of all vectors in $\mathbb{R}^n$ orthogonal to the vectors of $V^\perp$. Such a vector when written in the basis $\mathscr{B}$ cannot have any terms from the basis for $V^\perp$ appearing (examine the dot product). Thus it must lie in V. So $(V^\perp)^\perp \subset V$. Similarly, if $v \in V$ then, by definition of $V^\perp$, for every $\mathbf{x} \in V^\perp$, $\mathbf{v} \cdot \mathbf{x} = 0$. Thus, $\mathbf{v} \in (V^\perp)^\perp$. Hence, $V \subset (V^\perp)^\perp$. Consequently, $V = (V^\perp)^\perp$

(8) Explain how the attempt to fit a curve to "most" data sets gives rise to an inconsistent linear system. Explain why if it gives rise to a consistent linear system you are probably "over-fitting".

(9) Explain the least-squares process in terms of projecting points onto subspaces.

(10) Prove that if A is an $n \times m$ matrix, then $(\operatorname{im} A)^\perp = \ker A^T$.

Solution: Let $\mathbf{x} \in (\operatorname{im} A)^\perp$. So for every $\mathbf{a} \in \mathbb{R}^n$, $(A\mathbf{a}) \cdot \mathbf{x} = 0$. Hence, $\mathbf{a}^T A^T \mathbf{x} = 0$. Since this is true for every $\mathbf{a}$, we must have $A^T\mathbf{x} = 0$. Thus, $(\operatorname{im} A)^\perp \subset \ker A^T$. Now let $\mathbf{x} \in \ker A^T$. Then $A^T\mathbf{x} = \mathbf{0}$. Thus, if $\mathbf{a}$ is a column of A, we have $\mathbf{a} \cdot \mathbf{x} = 0$. The columns of A span $\operatorname{im} A$, so $\mathbf{x}$ is perpindicular to every vector in the image of A. Hence, $\mathbf{x} \in (\operatorname{im} A)^\perp$. Thus, $\ker A^T = (\operatorname{im} A)^\perp$.

(11) Prove that if A is an $n \times m$ matrix, then $\ker A^T A = \ker A$.

Solution: If $A\mathbf{v} = \mathbf{0}$, then $A^T A\mathbf{v} = A^T\mathbf{0} = \mathbf{0}$. So $\ker A \subset \ker A^T A$. Now suppose that $\mathbf{x} \in \ker A^T A$. Thus,

$$A^T A\mathbf{x} = \mathbf{0}$$

This means that $A\mathbf{x} \in \ker A^T$. But $\ker A^T = (\operatorname{im} A)^\perp$, so $\mathbf{Ax}$ is perpindicular to every vector of $\operatorname{im} A$, but it itself is in $\operatorname{im} A$ so the only way this is possible, is if $A\mathbf{x} = 0$. That is, $\mathbf{x} \in \ker A$. Hence, $\ker A^T A = \ker A$.

(12) Suppose that A is an $n \times m$ matrix. Explain why the least squares approximate solution to $A\mathbf{x} = \mathbf{b}$ is given by

$$\mathbf{x}^* = (A^T A)^{-1} A^T \mathbf{b} = \mathbf{x}^*.$$

(13) Find a quadratic curve $g(x) = a + bx + cx^2$ which is as close as possible (in the least-squares sense) to a curve which passes through the points

$$(0,0), (4,5), (8,2), (16,0)$$

(14) Define "singular value of an $n \times m$ matrix A".

(15) Let $q(\mathbf{x}) = ||A\mathbf{x}||^2$. Use q to prove that the eigenvalues of $A^T A$ are nonnegative.

Solution: Let λ be an eigenvalue of $A^T A$ and let $\mathbf{v}$ be the corresponding eigenvector. Then

$$q(\mathbf{v}) = ||A\mathbf{v}||^2 \geq 0.$$

On the other hand,

$$q(\mathbf{v}) = (A\mathbf{v}) \cdot (A\mathbf{v}) = \mathbf{v}^T A^T A\mathbf{v} = \lambda ||\mathbf{v}||^2$$

So we must have $\lambda \geq 0$.

(16) Prove that if $\mathbf{v}$ and $\mathbf{w}$ are orthogonal eigenvectors for $A^T A$, then the vectors $A\mathbf{v}$ and $A\mathbf{w}$ are orthogonal.

Solution: We have,

$$(A\mathbf{v}) \cdot (A\mathbf{w}) = \mathbf{v}^T A^T A\mathbf{w} = \lambda \mathbf{v}^T \mathbf{w} = \lambda \mathbf{v} \cdot \mathbf{w} = 0$$

where λ is the eigenvalue associated to $\mathbf{w}$.

(17) State the spectral theorem.

(18) State the Singular Value Decomposition Theorem and explain the statement (but not why it is true) in terms of change of bases.

(19) Explain how Singular Value Decomposition can be used to efficiently compute powers of a matrix A.

2. Linear Spaces

(1) Know the informal definition of a linear space. (You do not, however, have to know all the axioms for a linear space)

(2) Explain why the following are linear spaces:

Partial Solution: All of the following sets are closed under addition and scalar multiplication (check!).

(a) polynomials with real coefficients

(b) $n \times m$ matrices with real entries

(c) functions from any given set X to $\mathbb{R}$

(d) Solutions to a linear homogenous differential equation like $f''(x) - f(x) = 0$.

(e) Sequences of real numbers

(f) $n \times n$ matrices with trace equal to 0

(g) $n \times n$ matrices with entries summing to 0

(3) Know the precise definition of a linear transformation $T : X \to Y$ between two linear spaces.

(4) Know the precise definition of an isomorphism between two linear spaces.

(5) Explain why every finite dimensional linear space is isomorphic to $\mathbb{R}^n$ for some n.

(6) Explain why the following are linear transformations (between what spaces??). What are their images? What are their kernels?

Partial Solution: Each of the following preserves linear combinations. To calculate the kernel, figure out what goes to 0. To figure out the image, ask what sorts of objects can come out of the linear transformation.

(a) The derivative

(b) Taking a limit

(c) the integral $\int_0^x$.

(d) the shift operator on sequences of real numbers

$$a_1, a_2, a_3, \ldots \mapsto a_2, a_3, a_4, \ldots$$

(7) Let G be a planar directed graph with finitely many vertices and edges. Let V be the vector space of functions from the vertices of G to $\mathbb{R}$. Let E be the vector space of functions from the edges of G to $\mathbb{R}$. Let F be the vector space of functions from the faces of G to $\mathbb{R}$. For a function $f \in V$, let ∇f

be the function which assigns to an edge e having head v and tail w the number $f(v) - f(w)$. For a function $g \in E$, let $c(g)$ be the function which assigns to a face P of G, the sum

$$\pm g(e_1) \pm g(e_2) \pm \ldots \pm g(e_n)$$

where $e_1, \ldots, e_n$ are the edges of P in counterclockwise order and the sign $\pm$ is determined by whether we go with the arrow on the edge or against the arrow on the edge as we move counterclockwise around P.

Prove that $\operatorname{im} \nabla \subset \ker c$.

Solution: Let e_1, e_2, and e_3 be edges going (partway) around a face in counter-clockwise order. Let their endpoints be

$$v_1, v_2, v_3$$

in order. Then if $f \in V$, we have $\nabla f(e_1) = \pm(f(v_2) - f(v_1))$ (with a $+$ if e_1 is oriented counterclockwise around the face and $-$ otherwise). Also, $f(e_2) = \pm(f(v_3) - f(v_2))$ (with same sign convention.). Thus, when we apply c to ∇f we have $\pm(f(v_2) - f(v_1) + f(v_3) - f(v_2)) = \pm(f(v_3) - f(v_2))$ (keeping careful track of the signs if e_1 or e_2 does not point counterclockwise around the face. Thus, there is no term $f(v_1)$ in $c(\nabla f)$. But v_1 was an arbitrary vertex in the boundary of the face and the same would be true of any other vertex, so $\nabla \circ c = 0$ as desired.