Math 52 Exam 2 Topic List

David Dumas

What will be covered

The second exam will cover the material from the reading and lectures before Monday, November 7. This corresponds approximately to chapters 1-5 from Strang. Material covered since the first exam (section 3.4 and later) is the primary focus of exam 2, but some dependence on foundational material from the earlier chapters is inevitable.

What to expect

The problems on the exam will closely resemble the problems from the homework. The distribution of different types of questions (e.g. calculations vs. theoretical exercises) will also mirror what you have seen on the homework assignments.

A typical question will ask you to apply the methods we have learned to a particular example (perhaps involving some unknown quantities) and draw conclusions about what the results mean.

How to prepare

In preparing for the exam, start by reviewing the reading and especially the "worked problems" at the end of each section. Strang includes a list of "conceptual questions" at the end of the book (starting on p. 546); ask yourself these questions after reviewing the reading, and make sure you can confidently answer them. For practice, try a few problems from the textbook, especially those similar to the problems assigned for homework. Solutions to many exercises can be found starting on p. 502.

Outline of topics

(1) The complete solution to $A\mathbf{x} = \mathbf{b}$ (§3.4)
- (a) The rank of an $m \times n$ matrix is the number of pivots. The rank r satisfies $r \leq m$ and $r \leq n$.
- (b) The RREF R has $m - r$ zero rows.
- (c) To decide whether or not $A\mathbf{x} = \mathbf{b}$ has any solutions, reduce the augmented matrix $[A\ \mathbf{b}]$ to $[R\ \mathbf{c}]$. There is a solution if and only if the last $m - r$ entries of $\mathbf{c}$ are zero (so the last $m - r$ equations become $0 = 0$).
- (d) To find a *particular solution* $\mathbf{x}_p$, set all $n - r$ free variables to zero and solve $R\mathbf{x} = \mathbf{c}$ by back-substitution.
- (e) The *general solution* is then $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$ where $\mathbf{x}_n \in N(A)$, i.e. $\mathbf{x}_n$ is a linear combination of the $(n - r)$ special solutions.
- (f) Special case: If $r = n$, then A has *full column rank*. There is exactly one solution if $\mathbf{b} \in C(A)$ and no solution otherwise. (In particular, if there is any solution at all, it is unique.)

(g) Special case: If $r = m$ then A has *full row rank*. There is exactly one solution if $m = n$, and infinitely many solutions if $m < n$. (In particular, there is at least one solution for any $\mathbf{b}$.)
(h) Special case: If $r = m = n$ then A is a *square invertible matrix*. There is exactly one solution for every $\mathbf{b}$. In fact, the solution is $\mathbf{x} = A^{-1}\mathbf{b}$.

(2) Linear Independence, Span, Basis, and Dimension (§3.5)
(a) Vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are *linearly independent* if

$$c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n \neq \mathbf{0}$$

except when $c_1 = c_2 = \cdots = c_n = 0$.
(b) The columns of a matrix A are linearly independent exactly when $\mathbf{x} = \mathbf{0}$ is the only solution to $A\mathbf{x} = 0$, i.e. $N(A) = \mathbf{0}$.
(c) The *span* of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is the vector space consisting of all linear combinations of these vectors.
(d) We say that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *span* V if V is equal to the span of $\mathbf{x}_1, \ldots, \mathbf{x}_n$.
(e) A *basis* of V is a set of vectors that span V and are linearly independent.
(f) Every basis of a vector space has the same number of elements. This number is the *dimension* of V, written $\dim V$.
(g) If $\dim V = d$, then any d linearly independent vectors form a basis of V.
(h) If $\dim V = d$, then $(d-1)$ or fewer vectors cannot span V, and $(d+1)$ or more vectors cannot be linearly independent.
(i) The columns of an $n \times n$ matrix A are a basis for $\mathbb{R}^n$ if and only if A is invertible.

(3) Bases and Dimensions of the Four Subspaces (§§3.5-3.6)
Let A be an $m \times n$ matrix of rank r.
(a) The row space $C(A^T)$ and the null space $N(A)$ are subspaces of $\mathbb{R}^n$.
(b) The column space $C(A)$ and the left null space $N(A^T)$ are subspaces of $\mathbb{R}^m$.
(c) To find bases for the four subspaces, start by reducing $[A\ I]$ to $[R\ E]$, where R is the RREF and $EA = R$.
(d) The row spaces of A and R are the same, $C(A^T) = C(R^T)$; a basis is given by the r nonzero rows of R, so $\dim C(A^T) = r$.
(e) The null spaces of A and R are the same, $N(A) = N(R)$; a basis is given by the $(n-r)$ special solutions so $\dim N(A) = (n-r)$.
(f) The r pivot columns of A give a basis for the column space $C(A)$, so $\dim C(A) = r$. (The column spaces of A and R are different, but they have the same dimension.)
(g) The last $m - r$ rows of E give a basis for the left null space $N(A^T)$, so $\dim N(A^T) = (m-r)$. (The left null spaces of A and R are different, but they have the same dimension.)

(4) Orthogonality of Vectors and Subspaces (§4.1)

(a) The product $\mathbf{x}^T\mathbf{y}$ is a real number, sometimes called the *dot product* of $\mathbf{x}$ and $\mathbf{y}$.
(b) Vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if $\mathbf{x}^T\mathbf{y} = 0$. In $\mathbb{R}^2$ and $\mathbb{R}^3$, this means the vectors are perpendicular in the usual sense.
(c) Alternately, $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x}+\mathbf{y}\|^2$, where $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$.
(d) If $\mathbf{x}$ is orthogonal to each of the vectors $\mathbf{y}_1, \dots, \mathbf{y}_m$, then $\mathbf{x}$ is orthogonal to every vector in the span of $\mathbf{y}_1, \dots, \mathbf{y}_m$.
(e) Subspaces V and W of $\mathbb{R}^n$ are *orthogonal* if every vector in V is orthogonal to every vector in W. Orthogonal subspaces have no vectors in common except $\mathbf{0}$.
(f) The *orthogonal complement* $V^\perp$ of a subspace V of $\mathbb{R}^n$ is the set of all vectors orthogonal to V. It is a subspace.
(g) We say V and W are orthogonal complements if $V^\perp = W$ (equivalently, $W^\perp = V$). If V and W are subspaces of $\mathbb{R}^n$ and are orthogonal complements, then their dimensions add up to n.

(5) Orthogonality of the Four Subspaces (§4.1)
(a) The null space $N(A)$ and row space $C(A^T)$ are orthogonal complements; this comes from the equation $A\mathbf{x} = \mathbf{0}$ satisfied by a vector $\mathbf{x} \in N(A)$.
(b) The left null space $N(A^T)$ and the column space $C(A)$ are also orthogonal complements. (This can be derived from the fact about the null space and row space, applied to A^T.)
(c) To find a basis for a subspace V of $\mathbb{R}^d$ or its orthogonal complement $V^\perp$, it is best to find a matrix A that has V as one of its four subspaces. For example:
- To find a basis for a subspace V, you can try to find a matrix A with V as its column space, and then take the pivot columns of A. You can also find a basis for the left null space of A, which is $C(A)^\perp = V^\perp$.
- Alternately, you can find a matrix B with V as its null space, then the special solutions of B give a basis for V. You can also find a basis for the row space of B, which is $N(A)^\perp = V^\perp$.

(6) Projections (§4.1)
(a) The projection of a vector $\mathbf{b}$ onto the line spanned by $\mathbf{a}$ is $\mathbf{p} = \mathbf{a}\frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}} = \mathbf{a}\frac{\mathbf{a}^T\mathbf{b}}{\|\mathbf{a}\|^2}$.
(b) The projection matrix $P = \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}\mathbf{a}^T$ multiplies $\mathbf{b}$ to produce $\mathbf{p}$. The matrix P has rank one.
(c) To project $\mathbf{b}$ onto the column space of a matrix A (which should have independent columns), we need to find $\mathbf{p} = A\hat{\mathbf{x}} \in C(A)$ so that $\mathbf{e} = \mathbf{b} - \mathbf{p}$ is orthogonal to $C(A)$. This leads to the equation

$$A^T A\hat{\mathbf{x}} = A^T b.$$

(d) If A has independent columns, A^TA is invertible.
(e) The matrix P that projects onto $C(A)$ is given by the formula

$$P = A(A^TA)^{-1}A^T,$$

so the projection of $\mathbf{b}$ onto $C(A)$ is $P\mathbf{b}$.
(f) Extreme cases: If P is the projection matrix onto V, then $P\mathbf{b} = \mathbf{b}$ if $\mathbf{b} \in V$ and $P\mathbf{b} = \mathbf{0}$ if $\mathbf{b} \in V^\perp$.
(g) Projection matrices are symmetric ($P = P^T$) and satisfy $P^2 = P$ (because projecting twice is the same as projecting once).
(h) If P is the projection onto V, then $I - P$ is the projection onto $V^\perp$.
(i) If V is a subspace of $\mathbb{R}^n$, then every vector $\mathbf{b} \in \mathbb{R}^n$ can be expressed as a sum $\mathbf{b} = \mathbf{b}_V + \mathbf{b}_{V^\perp}$ where $\mathbf{b}_V \in V$ and $\mathbf{b}_{V^\perp} \in V^\perp$. To find these vectors, let P be the projection matrix for V; then $\mathbf{b}_V = P\mathbf{b}$ and $\mathbf{b}_{V^\perp} = (I-P)\mathbf{b}$.

(7) Least Squares Approximation, Fitting Problems (§4.3)
(a) The best way to understand this application of orthogonality and projections is to work through an example, such as the one Strang describes on pp. 206 - 209.
(b) Fitting a line (or other model curve) with n parameters (e.g. slope D and intercept C) to a set of m data points $(b_1, \ldots, b_m)$ gives m equations in n variables, i.e. $A\mathbf{x} = \mathbf{b}$ where A is $m \times n$.
(c) Typically, m is much larger than n, so the system has no solution for most $\mathbf{b}$.
(d) The *normal equations* $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$ replace the original system with n equations in n variables. If A has independent columns, then the normal equations have a unique solution.
(e) The solution $\hat{\mathbf{x}}$ to the normal equations is the best approximate solution to $A\mathbf{x} = \mathbf{b}$, i.e. it makes $E = \|A\mathbf{x} - \mathbf{b}\|^2$ as small as possible. The number E is the sum of the squares of individual errors, so this kind of approximate solution is called "least squares".

(8) Orthonormal Sets and Orthogonal Matrices (§4.4)
(a) Vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ are *orthonormal* if they are mutually orthogonal and each has unit length. Equivalently $\mathbf{q}_i^T\mathbf{q}_j = 0$ when i and j are different, and $\mathbf{q}_i^T\mathbf{q}_i = 1$.
(b) Orthonormal vectors are linearly independent.
(c) Vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ are orthonormal if and only if $Q^TQ = I$, where Q is the $m \times n$ matrix with columns $\mathbf{q}_1, \ldots, \mathbf{q}_n$.
(d) Notation: Q always represents a matrix with orthonormal columns.
(e) If Q is square, it is called an *orthogonal matrix*. In this case $Q^T = Q^{-1}$.
(f) Then the vectors $\mathbf{x}$ and $Q\mathbf{x}$ have the same length ($\|\mathbf{x}\| = \|Q\mathbf{x}\|$).
(g) The projection onto $C(Q)$ has a particularly simple form: $P = QQ^T$. (If Q is square, then $P = I$ because the columns span $\mathbb{R}^n$.)

(9) The Gram-Schmidt Algorithm (§4.4)

(a) The Gram-Schmidt algorithm starts with independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and produces orthonormal vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ that span the same space.
(b) Gram-Schmidt is inductive; we explain how to do it for one vector, and then how to do it for n vectors if we already know how to do it for $n-1$ vectors.
(c) For one vector $\mathbf{a}_1$, Gram-Schmidt gives $\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$.
(d) To apply Gram-Schmidt to n vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, first apply it to the $(n-1)$ vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$ to obtain $\mathbf{q}_1, \ldots, \mathbf{q}_{n-1}$. Let V_{n-1} be the span of $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$, which is the same as the span of $\mathbf{q}_1, \ldots, \mathbf{q}_{n-1}$. Then $\mathbf{q}_n$ is given by the formula:

$$\begin{aligned} A_n &= \mathbf{a}_n - (\text{projection of } \mathbf{a}_n \text{ onto } V_{n-1}) \\ &= \mathbf{a}_n - (\mathbf{q}_1^T \mathbf{a}_n)\mathbf{q}_1 - (\mathbf{q}_2^T \mathbf{a}_n)\mathbf{q}_2 - \cdots - (\mathbf{q}_{n-1}^T \mathbf{a}_n)\mathbf{q}_{n-1} \\ \mathbf{q}_n &= \frac{A_n}{\|A_n\|} \end{aligned}$$

(e) Applying Gram-Schmidt to the columns of an $m \times n$ matrix A (which we assume are independent) produces an $m \times n$ matrix Q with orthonormal columns.
(f) The corresponding matrix factorization is $A = QR$ where $R = Q^T A$ is an $n \times n$ matrix. The entries of R are

$$R_{ij} = \mathbf{q}_i^T \mathbf{a}_j.$$

(g) The matrix R is upper triangular because *later* $\mathbf{q}$*s are orthogonal to earlier* $\mathbf{a}$*s*.

(10) Properties of the Determinant (§5.1)
(a) $\det(A) = |A| =$ the determinant of A (an $n \times n$ matrix)
(b) Three essential properties uniquely define the determinant:
  (i) $|I| = 1$
  (ii) The determinant *changes sign* when two rows are exchanged.
  (iii) The determinant is a *linear function* of any single row. (Which involves a *scalar multiplication* condition and a *row addition* condition.)
(c) Some additional properties of the determinant follow from the three above:
- If A has two rows that are equal, then $|A| = 0$.
- Subtracting a multiple of one row from another leaves $|A|$ unchanged.
- If A has a row of zeros, then $|A| = 0$.
- If A is triangular, then $|A| = a_{11}a_{22}\cdots a_{nn}$ (the product of the diagonal entries).
- $|A| = 0$ if and only if A is singular.
- $|AB| = |A|\,|B|$ (this is amazing!)

- $|A^T| = |A|$ (so we can substitute "column" for "row" in any other property)

(d) A permutation matrix P has determinant ± 1. If $|P| = 1$ then P is *even*, otherwise P is *odd*.
(e) It follows from these properties that $|A|$ is $\pm$ the product of the pivots if A is invertible (i.e. has n pivots). The sign is determines by the row exchanges: it is $+$ if the permutation of the rows is even, and $-$ if the permutation of the rows is odd.

(11) Formulas for the Determinant (§5.2)
(a) The "big formula" for the determinant expresses it as a sum of $n!$ terms, one for each permutation of $(1, \dots, n)$:

$$|A| = \sum_{P=(\alpha,\beta,\dots,\omega) \text{ permutation of } (1,\dots,n)} |P| \, a_{1\alpha} a_{2\beta} \cdots a_{n\omega}$$

(b) In the big formula, we have one term for each way to select a single entry of A from each row and each column. The term includes a $+$ or $-$ sign depending on whether the corresponding permutation of the columns is even or odd, respectively.
(c) The *minor* matrix M_{ij} has size $(n-1) \times (n-1)$, and is obtained from A by removing row i and column j.
(d) The *cofactor* C_{ij} the determinant of M_{ij}, up to sign:

$$C_{ij} = (-1)^{i+j} |M_{ij}|$$

(e) The signs of the cofactors follow a "checkerboard" pattern:

$$\begin{pmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{pmatrix}$$

(f) Choose a row i of A; then we get the *cofactor formula* for $|A|$ using that row:

$$|A| = a_{i1} C_{i1} + a_{i2} + \cdots + a_{in} C_{in}$$

This is essentially the "big formula", but we have separated the terms according to which entry of row i is present.
(g) A similar cofactor formula works for a column.
(h) The cofactor formula expresses an $n \times n$ determinant as a sum of $(n-1) \times (n-1)$ determinants. (It is "inductive".)

(12) Applications of the Determinant (§5.3)
(a) Cofactors give us a formula for entries of A^{-1}:

$$(A^{-1})_{ij} = \frac{1}{|A|} C_{ji}.$$

In other words,

$$A^{-1} = \frac{1}{|A|} C^T.$$

(b) The solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$, when A is invertible.

(c) Using the formula for A^{-1}, we can write the solution to $A\mathbf{x} = \mathbf{b}$ as

$$x_j = \frac{|B_j|}{|A|}$$

where B_j is the matrix obtained by replacing column j of A with $\mathbf{b}$. This is *Cramer's rule.*

(d) Cramer's rule is computationally inefficient.

(e) The determinant of A is the volume of a box (parallelepiped) in $\mathbb{R}^n$ formed from the rows of A.

(f) For $n = 2$ the volume interpretation means that the area of a *parallelogram* with vertices $(0, 0)$, (x_1, y_1), (x_2, y_2), and $(x_1 + x_2, y_1 + y_2)$ is the determinant:

$$\text{area} = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

(g) This leads to a nice formula for the area of a triangle in the plane with vertices (x_1, y_1), (x_2, y_2), (x_3, y_3):

$$\text{area} = \frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

Note:

- The references to sections of the textbook are only approximate. Some material appears in this outline in a slightly different order than in the textbook.
- This list is meant to highlight the most important topics, but it is not exhaustive. You are responsible for the material in the readings and lectures.