A few caveats:

- This is only a brief summary of the topics on the test, and inevitably I've left some things out – make sure to study the notes as well!
- The notes for the first part are a bit thin, especially for things relating to elimination.
- You can find some old exams on the course site – these are the most important thing to look at.
- Note too the table of useful properties of eigenvalues on page 374. The glossary (557-565) and review questions (552-556) are also worth a look.
- Please let me know if you notice errors here (`johnl@math.mit.edu`)
- Review sheet subject to change, so check back for updated versions.

1 Things you definitely ought to know

- Decompositions:
  - $A = LU$, $PA = LU$, $A = LDU$
  - $A = QR$
  - $A = U\Sigma V^T$
  - Diagonalizable: $A = S\Lambda S^{-1}$
  - Jordan form: $A = SJS^{-1}$
  - A symmetric: $A = Q\Lambda Q^{-1}$.
- Types of matrices:
  - Triangular
  - Diagonal
  - Permutation
  - Symmetric: $A^T = A$. The eigenvalues are real, and the eigenvectors are orthogonal.
  - Skew-symmetric: $A^T = -A$. The eigenvalues are imaginary.
  - Positive definite: symmetric + any of the five conditions
  - Orthogonal: $Q^TQ = I$. The eigenvalues have $|\lambda| = 1$. The columns are any orthonormal set.
  - Projection $P = A(A^TA)^{-1}A^T$ (alternately, $P^2 = P$ and $P^T = P$)
  - Markov: Definition: column sums are 1, all entries positive. It follows that the largest eigenvalue is 1.
- Key computations:
  - Elimination (up to rref(A))
  - Compute the four fundamental subspaces of A
  - Inverse of a matrix
  - Gram-Schmidt orthonormalization
  - Solution to $d\mathbf{x}/dt = e^{At}\mathbf{x}$
  - Solution to $\mathbf{u}_{k+1} = A\mathbf{u}_k$
  - Eigenvectors and eigenvalues
- Properties of eigenstuff:
  - There's a table of this stuff on page 374. I'm not going to type it all out, but I'd recommend knowing just about everything on this list.
- Other goodies
  - Similar matrices

– Markov processes
– Analyzing graphs and networks

2 Linear equations

Hopefully by now you are pretty happy with matrix multiplication, so I won't spell out this basic stuff. One thing that's often forgotten that a way to think of multiplication is as $A[\mathbf{x}_1 \dots \mathbf{x}_n] = [A\mathbf{x}_1 \dots A\mathbf{x}_n]$; this can make some problems easier to think about.

The first big topic we covered was solving solving linear equations. When you have n linear equations in n variables, you can package them all up as a matrix equation $A\mathbf{x} = \mathbf{b}$. We have two ways of looking at this.

- Column picture: call the columns of A $\mathbf{c}_1, \dots, \mathbf{c}_n$. We're asking to write $\mathbf{b} = x_1\mathbf{c}_1 + \dots + x_n\mathbf{c}_n$; i.e. to express $\mathbf{b}$ as a combination of the columns of A.
- Row picture: call the rows of A $\mathbf{r}_1, \dots, \mathbf{r}_n$. The equation is equivalent to the n linear equations $\mathbf{r}_i \cdot \mathbf{x} = b_i$. Each of these defines a hyperplane in $\mathbb{R}^n$; the solutions are all vectors that lie in the intersections of all of these hyperplanes.

The systematic way to solve $A\mathbf{x} = \mathbf{b}$ is via elimination. Here I will only say what the general method is, and skip the special cases of it that we worked up to in the book. Bear in mind that you can just think of row operations as manipulating your original equations, so the solutions don't change; the strategy of elemination is just to substitute away the variables one at a time.

The important thing to remember is that $A\mathbf{x} = \mathbf{b}$ may have no solutions, a single solution, or infinitely many solutions. No solutions will happen when $\mathbf{b}$ is not in the column space of A. There will be a unique solution when $\mathbf{b}$ is in the column space of A, and A has no nullspace. When A has a nullspace, if there is a single solution then there are actually infinitely many: starting with any one solution, we can add vectors in the nullspace of A and so find many others. (Of course, it's possible that A has a nullspace but there are no solutions, because $\mathbf{b}$ isn't in the column space).

The general solution will be $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$, where $\mathbf{x}_p$ is a "particular solution". This means there are two things we have to do: find a particular solution $\mathbf{x}_p$, and describe all vectors $\mathbf{x}_n$ in the nullspace. Here's what to do. Write down the "augmented matrix" $[A \mid \mathbf{b}]$ and run elimination until A is in row reduced echelon form, call it R. This means that we have 1's in the pivots, and 0's above and below the pivots.

Identify the pivot columns of R (the ones with a pivot in them), and the free columns of R (the rest). Step 1 is to find the special solutions (a basis for the null space of A). There's going to be one special solution for each of the free variables. To find it, plug in 1 for a single free variable, and 0 for the rest. Then figure out what you need to set the pivot variables to be in order to get $\mathbf{0}$. Here's one way to write this down. If the r pivot columns are all to the left, then rref looks like

$$R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$$

Where I is an $r \times r$ identity matrix, F is who knows what, and the 0s are 0 matrices. The special solutions are then the columns of $N = \begin{bmatrix} -F \\ I \end{bmatrix}$. This is nothing more than what we said above; each column is a special solution, and the $-F$ is what you get when you solve for the pivot columns. If this way of writing things down seems confusing to you, you can

always just think of setting the free variables to 0's with one 1, and then solving for the pivot variables.

Here are some things you can read off after doing elimination on a matrix.

- Rank: this is just the number of pivots
- Determinant: you can multiply the pivots together.
- Null space: the nullspace of A is equal to the nullspace of R, which you can find as special solutions.
- Column space: the columns of A corresponding to the pivot columns of R are a basis for the column space of A.
- LU decomposition: keep track of the elimination matrices as you go.

Note: the ability to get the null space and the column space here allow you to compute all four fundamental subspaces: the row space is the column space of the transpose, and the left nullspace is the nullspace of the transpose.

Important: the column space of R is usually different from that of A (though the row spaces are the same). Elimination changes the column space, but not the row space. So when writing down a basis for the column space, be careful to use the columns of A rather than the columns of R.

Another big use of elimination is in inverting matrices. If A is a square matrix, its inverse is a matrix with $A^{-1}A = I$ and AA^{-1} (supposing such a matrix exists, which of course it doesn't have to). Here are some ways to check that a matrix is invertible, most using definitions that are later on in my summary:

- A has n pivots when we do elimination.
- A has independent columns (full column rank)
- There are no nonzero vectors $\mathbf{x}$ with $A\mathbf{x} = 0$.
- A has independent rows (full row rank)
- 0 is not an eigenvalue of A
- the determinant of A is nonzero

The inverse of AB is $B^{-1}A^{-1}$. To compute A^{-1}, we can use Gauss-Jordan elimination: make a $n \times 2n$ matrix with A on the left and I on the right. Run elimination on the whole thing, first to get the left half to be upper triangular, and then by back substitution so that the left half becomes the identity matrix. What's left on the right is A^{-1} (we have another way to compute this too, using Cramer's rule – this is less efficient in general, but helpful in many cases).

Don't forget about LU decomposition, which comes from elimination. At each step of elimination, you add a multiple of one row to another: so $A_1 = E_{ij}A_0$, where E_{ij} is the elimination matrix corresponding to the step. At the end, we get $U = E_{32}E_{31}E_{21}A$, and inverting this gives $A = (E_{21}^{-1}E_{31}^{-1}E_{32}^{-1})U$. This product of inverses of elimination matrices is lower-triangular, so we get a decomposition $A = LU$.

There are a couple variations on this theme. If elimination on A requires us to do a row swap, there's a permutation matrix thrown in the mix, and the product isn't lower triangular anymore. All we get is a factorization $PA = LU$. Also, note that L has 1s along the diagonal, while U doesn't. To make this more symmetric, we can pull these factors off into a diagonal matrix D, leaving U with 1s on the diagonal. This gives a factorization $A = LDU$. When A is symmetric, it will turn out that $U = L^T$, and so we get a factorization $A = LDL^T$.

3 The four fundamental subspaces

Suppose A is an $m \times n$ matrix, and the rank is r. The four fundamental subspaces are

1. The row space $C(A^T)$, which has dimension r
2. The column space $C(A)$, which has dimension r
3. The nullspace $N(A)$, which has dimension $n - r$
4. The left nullspace $N(A^T)$. which has dimension $m - r$

The row space and null space are orthogonal to each other, and the column space and left nullspace are orthogonal to each other.

4 Vector spaces other than $\mathbb{R}^n$

A few times in this class he have occasion to talk about vector spaces other than $\mathbb{R}^n$. A vector space just means a collection of things, such that it makes sense to add two of them together, or to multiply one by a scalar. The basic examples we have are

- $\mathbb{R}^n$
- $n \times n$ matrices
- polynomials of degree $\leq d$
- functions of a single variable

If you have a vector space, often it's interesting to think about the members of this vector space that have a certain property (e.g. "all vectors of length 1", "upper-triangular matrices"). Such a collection is called a *subset* of the vector space. A *subspace* is a special kind of subset: it's one where if you add two things in the subset together, or multiply one by a scalar, you get another thin gthat's in the subset. The first example is not a subspace, while the second one is.

5 Linear independence, basis, and all that

We say a bunch of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if the only combination of those vectors that gives $\mathbf{0}$ is the zero combination $0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n = \mathbf{0}$. The columns of a matrix A are linearly independent means that the only solution to $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. It's good to think about this and related notions just for vectors in $\mathbb{R}^3$. Three vectors in $\mathbb{R}^3$ are linearly independent if they don't all lie in a plane.

A set of vectors is said to span a space if every vector in the space can be written as a combination of those vectors. A set of vectors is said to be a basis for a space if they span the space, and they are linearly independent. The dimension of a vector space is defined to be the number of vectors in a basis. A common type of problem is the "find a basis for ..." problem. I think there are two basic strategies that are widely applicable (though now that we've seen a lot more examples, hopefully you're more comfortable with the "look at it and guess a basis" approach).

- Express the subspace as the nullspace of some matrix. We have a recipe for finding the nullspace of a matrix using special solutions, so if you can pull this off, you're done. Here are two examples:
  - The plane in $\mathbb{R}^3$ defined by $x + 2y + 3z = 0$. This is just the nullspace of the 1×3 matrix (1 2 3).
  - The set of vectors perpendicular to $(1, 1, 1)$ and $(1, 2, 3)$. This is just the nullspace

of $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}$.

- Write down what a general member of the subspace looks like, using letters for coefficients (but as few letters as you can get away with)
  - Upper-triangular 2×2 matrices. These look like $\left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right)$. Such a matrix can be written as $a\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) + b\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) + c\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Every upper-triangular matrix is a combination of those 3; thus they are a basis for the vector space.
  - Quadratic polynomials with $f(1) = 0$. These look like $ax^2 + bx + (-a - b)$, which is $a(x^2 - 1) + b(x - 1)$. A basis for this space is then $x^2 - 1$ and $x - 1$.

A lot of the stuff we talked about in this class works for these vector spaces other than $\mathbb{R}^n$. We can talk about projection matrices and orthonormal bases etc. We need a new version of the dot product; with vector spaces of functions, the analog is $\int f(x)g(x)\,dx$.

6 Networks

If G is a graph (a collection of nodes, with (directed) edges between some of the nodes), we associated to it an *incidence matrix* A. This matrix has one column for each node, and one row for each edge. For each edge in the graph, insert a row with a -1 in the column corresponding to the starting node, and $+1$ in the column corresponding to the ending node. The cool thing is that the various fundamental subspaces of A tell us things about graph G. A combination of rows which is 0 corresponds to a loop in the graph – thus loops come from vectors in the left nullspace of A. The nullspace of A consists of the single vector $(1, \ldots, 1)$.

7 Orthogonality

Two vectors are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. If V is a subspace, then the set of all vectors orthogonal to every vector $\mathbf{v}$ in V is a subspace too: this is the orthogonal complement of V, denoted $V^\perp$.

A matrix Q is said to orthogonal if its columns are an orthonormal set of vectors, meaning that $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ if $i \neq j$, where $\mathbf{v}_i \cdot \mathbf{v}_i = 1$ for every i. This implies that $Q^T Q = I$. Orthogonal matrices also satisfy $|Q\mathbf{x}| = |\mathbf{x}|$ for any vector $\mathbf{x}$.

Another useful property is that if you have an orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_n$ for a space, then for any vector, $\mathbf{x} = (\mathbf{x} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{x} \cdot \mathbf{q}_n)\mathbf{q}_n$. When your set is just $\mathbf{e}_1, \ldots, \mathbf{e}_n$ this just corresponds to writing out a vector as the sum of its components in the various coordinate directions. If $\mathbf{q}_1, \ldots, \mathbf{q}_k$ is a set of orthonormal vectors that isn't a basis (because they don't span the whole space), then $\mathbf{x} = (\mathbf{x} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{x} \cdot \mathbf{q}_k)\mathbf{q}_k$ is the projection of $\mathbf{x}$ onto the span of $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

8 Projection matrices and least squares

Suppose that V is a subspace of $\mathbb{R}^n$. If $\mathbf{x}$ is a vector which isn't contained in V, it's often useful to know the point of V that's the closest to $\mathbf{x}$. It turns out that there's a matrix P such that for any vector $\mathbf{x}$, the closest point in V to $\mathbf{x}$ is just $P\mathbf{x}$. P is the called a projection matrix. To get the matrix, you need to find a basis for V, form a matrix A by putting all your basis vectors together as the columns, and then compute $P = A(A^T A)^{-1} A^T$ (a formula well worth remembering). Note that $A^T A$ is a square matrix, so the inverse makes sense.

Two basic properties of projection matrices: $P^2 = P$ and $P^T = P$. If you choose your matrix $A = Q$ to be orthogonal (i.e. choose an orthonormal basis for V), this takes a

particularly simple form: $P = Q(Q^TQ)^{-1}Q^T = Q(I)Q^T = QQ^T$.

Suppose we're trying to solve $A\mathbf{x} = \mathbf{b}$. If $\mathbf{b}$ isn't in the column space of A, there can't be any solutions. The best we can do is to find the point $\mathbf{p}$ in $V = \text{col}(A)$ closest to $\mathbf{b}$ for which $A\mathbf{x} = \mathbf{b}$ does have a solution. A bit of work (explained in the book) shows that the closest $\mathbf{p}$ is the projection of $\mathbf{b}$ onto the column space of A: $\mathbf{p} = A(A^TA)^{-1}A^T\mathbf{b}$. Then we need to solve $A\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{b}$. Applying A^T to both sides we need to solve $A^TA\mathbf{x} = A^T\mathbf{b}$. This is a general guideline: if you want to solve $A\mathbf{x} = \mathbf{b}$, but there's no solution, the best-fit approximation will be the solution to $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$; this means $\mathbf{p} = A\hat{\mathbf{x}}$. The error $\mathbf{e} = \mathbf{b} - \mathbf{p}$ will always be orthogonal to V.

A typical setting where this method comes in handy is in trying to fit a curve of a specific type through a bunch of data points: say we have a bunch of pairs (x_1, y_1), ..., (x_n, y_n) and want to find the best-fit parabola $y = at^2 + bt + c$. To find a parabola going through all the points, we'd want a, b, and c to satisfy $y_i = ax_i^2 + bx_i + c$ for every i. This is a linear equation ("$A\mathbf{x} = \mathbf{b}$") in a, b, c:

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ & \cdots & \\ x_n^2 & x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

In general this is lots of equations in three variables, and there are no solutions (the right side isn't in the column space of the matrix). That means there aren't any parabolas through all the points. But you can solve $A^TA\mathbf{x} = A^T\mathbf{b}$ instead, and that will be the best fit.

9 Gram-Schmidt and $A = QR$

Say we're handed a list of independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The Gram-Schmidt process will start with these and then produce a new set vectors with the same span, but which are orthonormal. This is convenient, since dealing with projection matrices and many other things is easier when you have an orthonormal basis. Here's how it works: the first vector $\mathbf{w}_1$ is just the original vector you started with. To get $\mathbf{w}_2$, take $\mathbf{v}_2$, and subtract its component in the $\mathbf{w}_1$ direction: $\mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1$. Similarly, $\mathbf{w}_3 = \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2}\mathbf{w}_2$. This produces a set of orthogonal vectors. At the end, you want to normalize by dividing each of your vectors by its length.

If you imagine putting your vectors as the columns of a matrix, and then carrying out Gram-Schmidt by doing basic column operations, each step corresponds to multiplying A on the right (since we're doing column instead of row operations). Combining all the elimination steps, in the end you get $AS = Q$, where S is upper triangular and Q is orthogonal. Multiply both sides by S^{-1} on the right, and we get $A = QS^{-1} = QR$.

Keeping track of all the elimination matrices can get annoying, so here's a shortcut. First just compute Q via Gram-Schmidt as usual. We know that $Q^TA = Q^TQR = R$, since Q is orthogonal. If you don't want to keep track of elimination matrices at every step, you can just compute Q first by Gram-Schmidt, and then get R using $R = Q^TA$.

10 Determinants

A determinant takes a square matrix and outputs a number. Be sure you know the following useful properties of this operation:

- Determinant is linear in each row and column.
- Determinant is unchanged when adding one row/column to another row/column.
- Determinant changes sign when swapping two rows/columns.
- $\det A = 0$ if and only if A has a nonzero nullspace (i.e. it's not invertible).
- The determinant of A is equal to the volume of the parallelpiped defined by the columns of A; this shows up as the Jacobian in change of coordinates for 2- or 3-dimensional integrals.

We have three main methods for computing determinants. I'm not going to explain how each goes, but make sure you know them all. Here they are, with some suggestions one when each one is useful.

- Cofactor expansion. Especially suitable when the matrix has a row or column with only a couple nonzero entries. Also good in a lot of these induction-type problems where you want to know the determinants of a sequence of $n \times n$ matrices of similar shape.
- The big formula. Good if there are only a couple nonzero terms, or if the matrix is 2×2 or 3×3 (in which case the big formula isn't so big after all, having only 2 or 6 terms).
- Product of pivots. If a matrix doesn't fit any of the above (i.e. it's at least 4×4 and doesn't have many zeroes), this is probably the way to go.

Perhaps the following should qualify as a basic methods:

- Triangular matrices: just multiply the diagonal entries.
- Reduce to simpler matrix by row/column operations. These don't change the determinant, so if you can do something sneaky like reduce to an triangular matrix or a matrix with lots of 0's using just a couple row/column operations (including row swaps), you're in business.

11 Cramer's rule

Cramer's rule is a way to solve linear systems $A\mathbf{x} = \mathbf{b}$, where A is square. Here's what you do. Let B_i be the matrix obtained by replacing the i^{th} column of A with the vector $\mathbf{b}$. The $x_i = \det B_i / \det A$. As a consequence, we get a formula for A^{-1}: we have $A^{-1} = C^T / \det A$, where C is the matrix of cofactors of A. These are not efficient ways to solve $A\mathbf{x} = \mathbf{b}$ in general, since you have to take a lot of determinant. Nevertheless it can sometimes be very helpful to have an explicit formula like this.

12 Eigenvalues and eigenvectors

A vector $\mathbf{x}$ is called an eigenvector of the matrix A if $A\mathbf{x} = \lambda\mathbf{x}$, where λ is a number (called the eigenvalue): this means that $A\mathbf{x}$ points in the same direction as $\mathbf{x}$ (or the opposite direction, in case $\lambda < 0$).

This is equivalent to saying that $(A - \lambda I)\mathbf{x} = \vec{0}$, which is to say that $\mathbf{x}$ is in the nullspace of $A - \lambda I$. This is the observation that lets us find the eigenvectors. For most values of λ, the matrix $A - \lambda I$ won't have a nullspace at all. The only times it does is when $\det(A - \lambda I) = 0$, and so the eigenvalues are precisely the solutions of $\det(A - \lambda I) = 0$. Once you've find an eigenvalue, the way to find the corresponding eigenvector is to write down the matrix $A - \lambda I$ for that value of λ, and then find something in the nullspace using the usual procedure (elimination, special solutions).

There are a couple cases where this has particular geometric significance: $\mathbf{x}$ is an eigenvector with $\lambda = 1$ means that $\mathbf{x}$ doesn't change when you apply A; with $\lambda = 0$ means that $A\mathbf{x} = \vec{0}$, i.e. $\mathbf{x}$ is in the nullspace of A; with $\lambda = -1$ means that the direction of $\mathbf{x}$ is reversed when we apply A to $\mathbf{x}$.

The sum of the eigenvectors is equal to the trace, which is defined to be the sum of the diagonal entries. The product of the eigenvectors is equal to the determinant. This can be useful either for finding the determinant/trace if you somehow know all the eigenvalues, or finding one eigenvalue if you know the determinant/trace and all of the other eigenvalues.

13 Eigenvalues and eigenvectors

A vector $\mathbf{x}$ is called an eigenvector of the matrix A if $A\mathbf{x} = \lambda\mathbf{x}$, where λ is a number (called the eigenvalue): this means that $A\mathbf{x}$ points in the same direction as $\mathbf{x}$ (or the opposite direction, in case $\lambda < 0$).

This is equivalent to saying that $(A - \lambda I)\mathbf{x} = \vec{0}$, which is to say that $\mathbf{x}$ is in the nullspace of $A - \lambda I$. This is the observation that lets us find the eigenvectors. For most values of λ, the matrix $A - \lambda I$ won't have a nullspace at all. The only times it does is when $\det(A - \lambda I) = 0$, and so the eigenvalues are precisely the solutions of $\det(A - \lambda I) = 0$. Once you've found an eigenvalue, the way to find the corresponding eigenvector is to write down the matrix $A - \lambda I$ for that value of λ, and then find something in the nullspace using the usual procedure (elimination+special solutions, or guessing).

There are a couple cases where this has particular geometric significance: $\mathbf{x}$ is an eigenvector with $\lambda = 1$ means that $\mathbf{x}$ doesn't change when you apply A; with $\lambda = 0$ means that $A\mathbf{x} = \vec{0}$, i.e. $\mathbf{x}$ is in the nullspace of A; with $\lambda = -1$ means that the direction of $\mathbf{x}$ is reversed when we apply A to $\mathbf{x}$.

The sum of the eigenvalues is equal to the trace, which is defined to be the sum of the diagonal entries. The product of the eigenvalues is equal to the determinant. This can be useful either for finding the determinant/trace if you somehow know all the eigenvalues, or finding one eigenvalue if you know the determinant/trace and all of the other eigenvalues.

14 Diagonalization

If we have a bunch of eigenvectors for a matrix A, we can put all of them as the columns of a matrix S. This will satisfy $AS = S\Lambda$, where Λ is a matrix with the corresponding eigenvalues on the main diagonal: AS is the matrix whose columns are A applied to the columns of S, and since those columns are eigenvectors, we just need to multiply each column by the corresponding eigenvalue. That's exactly what $S\Lambda$ is. Note that using ΛS on the right would instead multiply the rows by the eigenvalues. We don't want that.

If we actually had n linearly independent eigenvectors, the matrix S would be square and in fact invertible. Then we'd be able to write $A = S\Lambda S^{-1}$, i.e. diagonalize A.

Part of doing diagonalization is knowing how to invert the matrix S, something we covered earlier in the class, so be ready for it. I'd guess that on the test it's unlikely you'll have to diagonalize anything bigger than 2×2, so make sure you remember the quick way to invert a 2×2 matrix. This will let you find S^{-1} without having to think too hard.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Not every matrix is diagonalizable. If A has n *distinct* eigenvalues, then for each eigenvalue we can find an eigenvector. Eigenvectors with different eigenvalues are automatically independent, so that gives us n independent eigenvectors. Put those into a matrix S as above, and tada, it's diagonalized. Let me stress: if the eigenvalues are all distinct, diagonalization is automatic. If there's a repeated eigenvalue, things can go either way: maybe it's diagonalizable, maybe it isn't. You have to check.

Problems can arise when A has a repeated eigenvalue. It's only guaranteed that we can find a single eigenvector for that eigenvalue, which isn't enough to make a square matrix S. It's still possible that A can be diagonalized, but you actually need to check for eigenvectors. The typical example of this is something like $\left(\begin{smallmatrix}\lambda & 1\\ 0 & \lambda\end{smallmatrix}\right)$, which has only a single eigenvector. Here's some terminology: the algebraic multiplicity is the number of times λ shows up as an eigenvalue. The geometric multiplicity is the number of linearly independent λ-eigenvectors you can find. AM is always greater than or equal to GM; it's the case when they're equal that you can diagonalize.

15 Markov matrices etc.

In the happy case that a matrix can be diagonalized, we can quickly compute its powers (actually we can do that in other cases too, using Jordan form). If $A = S\Lambda S^{-1}$, then $A^k = S\Lambda^k S^{-1}$. The matrix Λ^k just has the powers of the λ_i along its diagonal.

Computing these powers is actually a useful thing to do in lots of cases. Maybe we have a sequence of vectors $\mathbf{u}_k$, and at each step the next vector is determined by applying a fixed matrix A: that is, $\mathbf{u}_{k+1} = A\mathbf{u}_k$. Then $\mathbf{u}_k = A^k\mathbf{u}_0 = S\Lambda^k S^{-1}\mathbf{u}_0$. If you write $\mathbf{u}_0 = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$ as a combination of the eigenvectors, then

$$\mathbf{u}_k = c_1(\lambda_1)^k\mathbf{x}_1 + \cdots + c_n(\lambda_n)^k\mathbf{x}_n.$$

This gives a formula for $\mathbf{u}_k$! An important instance of this is when A is a Markov matrix: this means that the columns of A add up to 1, and all the entries are nonnegative. Another way to write the first part (helpful if you have to prove something about Markov matrices) is that $A^T(1,\ldots,1) = (1,\ldots,1)$. Make sure you understand what a Markov matrix represents; it's things like the "rental cars" example on page 432.

One cool thing about Markov matrices is that $\lambda = 1$ is an eigenvector, and (at least if all the entries are nonzero) the other eigenvectors are less than 1. In the above equation, when k gets really big, all the terms will go to 0 except for the $c_1(\lambda_1)^k\mathbf{x}_1$. So no matter what vector $\mathbf{u}_0$ you start with, it eventually gets close to a multiple of $\mathbf{x}_1$. This is called the steady state; to find it just compute the eigenvector for $\lambda = 1$.

16 Diff eq stuff

For a single differential equation in a single variable the solution to $\frac{du}{dt} = \lambda u(t)$ is $u(t) = Ce^{\lambda t}$. If we have n functions $u_1,\ldots,u_n$, and they satisfy a bunch differential equations $\frac{du_1}{dt} = a_{11}u_1(t) + \cdots + a_{1n}u_n(t)$ etc., we can pack this all in a single matrix equation $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$.

The first thing one checks is that one way to get solutions is by using $\mathbf{u}(t) = e^{\lambda_i t}\mathbf{x}_i$, where $\mathbf{x}_i$ is any eigenvector of A. Any combination of solutions is a solution too, so in fact the general solution to $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$ is of the form

$$u(t) = c_1e^{\lambda_1 t}\mathbf{x}_1 + \cdots + c_ne^{\lambda_n t}\mathbf{x}_n.$$

If you have initial conditions (i.e. values for $u_i(0)$), you can use these to solve for the constants c_i. Indeed, we need $\mathbf{u}(0) = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$. Sticking all the eigenvectors $\mathbf{x}_i$ as columns of a matrix S, this just says that $S\mathbf{c} = \mathbf{u}(0)$, and inverting we get the initial conditions as $\mathbf{c} = S^{-1}\mathbf{u}(0)$.

Note that there is a useful trick for converting a second-order linear differential equation into a system of first-order ones, which can then be solved by this method. If you want to solve something like $y''+3y'+2y = 0$, just introduce a new function $z = y'$. Then our equation is converted into two first-order equations: $y' = z$, and $z' = y'' = -3y' - 2y = -2y - 3z$. In matrix form, this is

$$\begin{pmatrix} y' \\ z' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}$$

This is something we know how to solve!

It's important to think about what these solutions actually looks like, in terms of what the λ_i are. I'll simply note here that if all of them have real part less than 0, $u(t)$ is going to go to 0 as $t \to \infty$. For more, I refer you to the book (there's a bit of this in my review sheet for 18.03 last year: `http://math.mit.edu/~johnl/1803/review.pdf`).

There's another way to go about this stuff, via the matrix exponential. If A is a square matrix, we define

$$e^{At} = I + At + \frac{1}{2}(At)^2 + \frac{1}{6}(At)^3 + \cdots .$$

If $A = S\Lambda S^{-1}$ is diagonalizable, this is very easy to compute: use the rule $e^{At} = Se^{\Lambda t}S^{-1}$. The reason this works is that

$$\begin{aligned} e^{At} &= I + At + \frac{1}{2}(At)^2 + \frac{1}{6}(At)^3 + \cdots \\ &= I + (S\Lambda S^{-1})t + \frac{1}{2}((S\Lambda S^{-1})t)^2 + \frac{1}{6}((S\Lambda S^{-1})t)^3 + \cdots \\ &= I + S(\Lambda t)S^{-1} + \frac{1}{2}(S(\Lambda^2 t^2)S^{-1}) + \frac{1}{6}(S\Lambda^3 t^3 S^{-1}) + \cdots \\ &= S(I + \Lambda t + \frac{1}{2}(\Lambda t)^2 + \frac{1}{6}(\Lambda t)^3 + \cdots)S^{-1} \\ &= Se^{\Lambda t}S^{-1} \end{aligned}$$

Now $e^{\Lambda t}$ is just a diagonal matrix with entries $e^{\lambda_1 t}$, …, $e^{\lambda_2 t}$ (this can be checked from the definition).

If A isn't diagonalizable, you can't use this formula. Not all hope is lost, though: if the powers of A are periodic, or some power is 0, you can use the defining formula directly. Using the Jordan form of A is another useful technique, though we didn't really talk about it.

Matrix exponential gives another way to express the solution above: the solution to $\frac{d\mathbf{u}}{dt}$ is $\mathbf{u}(t) = e^{At}\mathbf{u}(0)$. If you unravel this by writing $e^{At} = Se^{\Lambda t}S^{-1}$, and noting $e^{\Lambda t}$ just has diagonal entries $e^{\lambda_1 t}, \ldots, e^{\lambda_n t}$, a bit of unraveling shows this is really just another way of writing the same formula from above. More precisely,

$$e^{At}\mathbf{u}(0) = S\begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} S^{-1}\mathbf{u}(0) = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = c_1 e^{\lambda_1 t}\mathbf{x}_1 + c_2 e^{\lambda_2 t}\mathbf{x}_2.$$

Other facts: the inverse of e^{At} is e^{-At}, the eigenvalues of e^{At} are $e^{\lambda t}$. When A is skew-symmetric, e^{At} is orthogonal. This chapter has a lot of stuff going on that's packed into innocuous-looking notation – make sure to do some examples to see what's really going on.

17 Symmetric matrices and positive-definite matrices

A matrix is symmetric if $A^T = A$. It turns out that in this situation the eigenvalues and eigenvectors have very special properties:

1. The eigenvalues are all real numbers.
2. The eigenvectors are orthonormal.

This means that when we write $A = S\Lambda S^{-1}$, the matrix S is actually an orthogonal matrix: $A = Q\Lambda Q^T$. Another special property is that A is always diagonalizable – remember that this is not the case for every matrix!

We say that a symmetric matrix is positive definite if it has the following properties (all of these things are equivalent). After each, I've listed a setting in which it might be useful.

1. All n pivots are positive. (Fast to compute if you don't know anything else about a matrix).
2. All n upper-left determinants are positive. (Easy to check, and helpful if you have a matrix that depends on a parameter, and you want to know what values of the parameter will make it definite).
3. All n eigenvalues are positive. (Helpful in some proofs, or if you already know the eigenvalues of the matrix somehow).
4. $\mathbf{x}^T A\mathbf{x} > 0$ for any nonzero vector $\mathbf{x}$. (Hard to check directly, but greatly simplifies some proofs).
5. We can write $A = R^T R$ for some matrix R whose columns are independent. (Useful in some numerical applications, also in some proofs.)

Most of these are pretty self-explanatory. #4 is worth interpreting a bit further: if $\mathbf{x} = (x_1, \ldots, x_n)$, then $\mathbf{x}^T A\mathbf{x}$ is some polynomial in the variables $x_1, \ldots, x_n$, with every term quadratic. The matrix being positive definite just means that whatever you plug in for these variables, the polynomial takes a positive value. Also bear in mind that pivots and eigenvalues are not the same thing, but the number of positive pivots equals the number of positive eigenvalues.

A good trick is that you can use this to figure out the picture of an ellipse, even when its equation has "xy" terms, which means it isn't parallel to the axes. The book spells out a step-by-step procedure for this on page 346, and I'm not feeling ambitious enough to recreate the diagrams here. Summary: the ellipse $\mathbf{x}^T A\mathbf{x} = 1$ has axes along the eigenvectors of A, with axes of length $1/\sqrt{\lambda_i}$.

18 Similar matrices

A matrix A is said to be similar to B if there exists an invertible matrix M such that $B = M^{-1}AM$. If A is similar to B, they have the same eigenvalues. If $\mathbf{v}$ is an eigenvector for A, then $M^{-1}v$ is an eigenvector for A.

If two matrices are similar, they have the same eigenvalues. If they have the same eigenvalues, it's not always true that they are similar: the basic problem arises when you have diagonalizable and non-diagonalizable matrices whose eigenvalues are the same. However, if A and B have the same eigenvalues, are those eigenvalues are all distinct, then A and B are

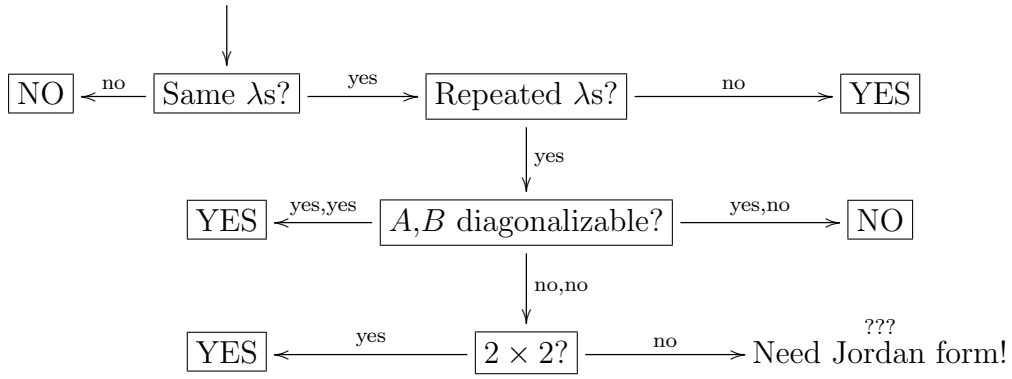
automatically similar. The reason is that both are diagonalizable, and similar to the same diagonal matrix!

While not every matrix is similar to a diagonal one, every matrix is similar to a Jordan matrix. This is a block-diagonal matrix with

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & & \\ & \ddots & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}.$$

The main fact here is that two matrices are similar if and only if they have the same Jordan decomposition. A lot of the tricks we managed for diagonalizable matrices can also be done for others, using Jordan decomposition instead of diagonalization.

Here's a flowchart for checking if two matrices A and B are similar.



The reason any two non-diagonalizable 2×2 matrices with repeat eigenvalues are similar is that there's only one possible non-diagonal Jordan form for 2×2. In the "???" case, the thing to do is compute the Jordan form of the matrices, which we didn't really cover how to do. I doubt it'll be on the exam.

19 Singular value decomposition

Given any matrix A, you can write it as $A = U\Sigma V^T$. The matrix is sort of "diagonalized" in the SVD: we have $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ (in matrix form: $AV = U\Sigma$). Here A is $m \times n$ (we don't assume anything about it), and

1. U is an orthogonal $m \times m$ matrix.
2. V is an orthogonal $n \times n$ matrix.
3. The nonzero diagonal entries of Σ ("singular values") are the square roots of the eigenvalues of A^TA (which are the same as those of AA^T, though these have different dimensions).

Here's how to compute it:

1. The first r columns of V are the eigenvectors of A^TA with nonzero eigenvalue. NB: because we want the entries of Σ to be in decreasing order, you should start with the largest eigenvalue for $\mathbf{v}_1$ and work your way down.

2. The other $n - r$ columns are an orthonormal basis for the nullspace of A (which is the same as the nullspace of $A^T A$!)
3. Don't forget to transpose: $A = U\Sigma V^{\mathbf{T}}$)
4. The diagonal entries on Σ are square roots of eigenvalues of $A^T A$ (note that $A^T A$ is a symmetric, positive-definite matrix – so said eigenvalues are positive real numbers)
5. To get the first r columns of $\mathbf{u}$, use the rule $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ (you know $\mathbf{v}_i$ and σ_i)
6. The other $m - r$ columns are an orthonormal basis for the nullspace of A^T (same as nullspace of AA^T).

Observe that if $\mathbf{v}_i$ is one of the first r columns of V, then $A^T A v_i = \sigma_i^2 v_i$, which tells us that $\mathbf{v}_i$ is in the column space of A^T, hence in the row space of A. The other $\mathbf{v}_i$ are in the null space of A. Similarly, the first r $\mathbf{u}_i$'s are in the column space of A^T, and the rest are in the null space of A^T. So all four of the fundamental subspaces of A can be found in the SVD!

1. Row space of A = first r columns of V.
2. Nullspace of A = last $n - r$ columns of V.
3. Column space of A = first r columns of U.
4. Left nullspace of A = last $m - r$ columns of U.