A couple caveats:

- This is only a brief summary of the topics on the test – make sure to study the notes as well!
- 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.
- 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 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.

2 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.

3 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 $\lambda = 1$.

4 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

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

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.

5 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 $1/\sqrt{\lambda_i}$.

6 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.

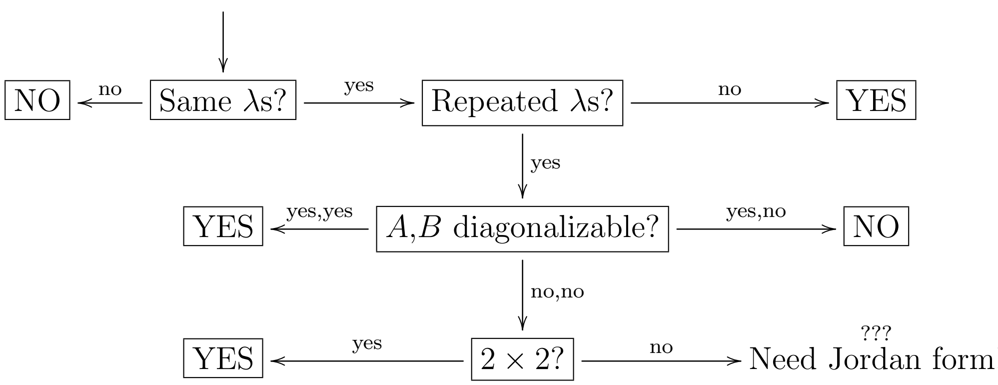
If two matrices are similar, they have the same eigenvalues. If they have the same eigenvalues, it's not always true that they are similar: the basic problem arises when you have diagonalizable and non-diagonalizable matrices whose eigenvalues are the same. However, if A and B have the same eigenvalues,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

7 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.