Symmetric matrices and positive definiteness

Symmetric matrices are good – their eigenvalues are real and each has a complete set of orthonormal eigenvectors. Positive definite matrices are even better.

Symmetric matrices

A *symmetric matrix* is one for which $A = A^T$. If a matrix has some special property (e.g. it's a Markov matrix), its eigenvalues and eigenvectors are likely to have special properties as well. For a symmetric matrix with real number entries, the eigenvalues are real numbers and it's possible to choose a complete set of eigenvectors that are perpendicular (or even orthonormal).

If A has n independent eigenvectors we can write $A = S\Lambda S^{-1}$. If A is symmetric we can write $A = Q\Lambda Q^{-1} = Q\Lambda Q^T$, where Q is an orthogonal matrix. Mathematicians call this the *spectral theorem* and think of the eigenvalues as the "spectrum" of the matrix. In mechanics it's called the *principal axis theorem*.

In addition, any matrix of the form $Q\Lambda Q^T$ will be symmetric.

Real eigenvalues

Why are the eigenvalues of a symmetric matrix real? Suppose A is symmetric and $A\mathbf{x} = \lambda\mathbf{x}$. Then we can conjugate to get $\overline{A}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. If the entries of A are real, this becomes $A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. (This proves that complex eigenvalues of real valued matrices come in conjugate pairs.)

Now transpose to get $\overline{\mathbf{x}}^T A^T = \overline{\mathbf{x}}^T\overline{\lambda}$. Because A is symmetric we now have $\overline{\mathbf{x}}^T A = \overline{\mathbf{x}}^T\overline{\lambda}$. Multiplying both sides of this equation on the right by $\mathbf{x}$ gives:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x}.$$

On the other hand, we can multiply $A\mathbf{x} = \lambda\mathbf{x}$ on the left by $\overline{\mathbf{x}}^T$ to get:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}.$$

Comparing the two equations we see that $\overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}$ and, unless $\overline{\mathbf{x}}^T\mathbf{x}$ is zero, we can conclude $\lambda = \overline{\lambda}$ is real.

How do we know $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$?

$$\overline{\mathbf{x}}^T\mathbf{x} = \begin{bmatrix} \overline{x}_1 & \overline{x}_2 & \cdots & \overline{x}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2.$$

If $\mathbf{x} \neq \mathbf{0}$ then $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$.

With complex vectors, as with complex numbers, multiplying by the conjugate is often helpful.

Symmetric matrices with real entries have $A = A^T$, real eigenvalues, and perpendicular eigenvectors. If A has complex entries, then it will have real eigenvalues and perpendicular eigenvectors if and only if $A = \overline{A}^T$. (The proof of this follows the same pattern.)

Projection onto eigenvectors

If $A = A^T$, we can write:

$$\begin{aligned} A &= Q\Lambda Q^T \\ &= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\ &= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T \end{aligned}$$

The matrix $\mathbf{q}_k \mathbf{q}_k^T$ is the projection matrix onto $\mathbf{q}_k$, so every symmetric matrix is a combination of perpendicular projection matrices.

Information about eigenvalues

If we know that eigenvalues are real, we can ask whether they are positive or negative. (Remember that the signs of the eigenvalues are important in solving systems of differential equations.)

For very large matrices A, it's impractical to compute eigenvalues by solving $|A - \lambda I| = 0$. However, it's not hard to compute the pivots, and the signs of the pivots of a symmetric matrix are the same as the signs of the eigenvalues:

number of positive pivots = number of positive eigenvalues.

Because the eigenvalues of $A + bI$ are just b more than the eigenvalues of A, we can use this fact to find which eigenvalues of a symmetric matrix are greater or less than any real number b. This tells us a lot about the eigenvalues of A even if we can't compute them directly.

Positive definite matrices

A *positive definite matrix* is a symmetric matrix A for which all eigenvalues are positive. A good way to tell if a matrix is positive definite is to check that all its pivots are positive.

Let $A = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$. The pivots of this matrix are 5 and $(\det A)/5 = 11/5$. The matrix is symmetric and its pivots (and therefore eigenvalues) are positive, so A is a positive definite matrix. Its eigenvalues are the solutions to:

$$|A - \lambda I| = \lambda^2 - 8\lambda + 11 = 0,$$

i.e. $4 \pm \sqrt{5}$.

The determinant of a positive definite matrix is always positive but the determinant of $\begin{bmatrix} -1 & 0 \\ 0 & -3 \end{bmatrix}$ is also positive, and that matrix isn't positive definite. If all of the subdeterminants of A are positive (determinants of the k by k matrices in the upper left corner of A, where $1 \leq k \leq n$), then A is positive definite.

The subject of positive definite matrices brings together what we've learned about pivots, determinants and eigenvalues of square matrices. Soon we'll have a chance to bring together what we've learned in this course and apply it to non-square matrices.

Exercises on symmetric matrices and positive definiteness

Problem 25.1: (6.4 #10. *Introduction to Linear Algebra:* Strang) Here is a quick "proof" that the eigenvalues of all real matrices are real:

$$\textbf{False Proof:}\quad A\mathbf{x} = \lambda\mathbf{x} \text{ gives } \mathbf{x}^T A\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} \text{ so } \lambda = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \text{ is real.}$$

There is a hidden assumption in this proof which is not justified. Find the flaw by testing each step on the 90° rotation matrix:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

with $\lambda = i$ and $\mathbf{x} = (i, 1)$.

Problem 25.2: (6.5 #32.) A *group* of nonsingular matrices includes AB and A^{-1} if it includes A and B. "Products and inverses stay in the group." Which of these are groups?

a) Positive definite symmetric matrices A.

b) Orthogonal matrices Q.

c) All exponentials e^{tA} of a fixed matrix A.

d) Matrices D with determinant 1.

Exercises on symmetric matrices and positive definiteness

Problem 25.1: (6.4 #10. *Introduction to Linear Algebra:* Strang) Here is a quick "proof" that the eigenvalues of all real matrices are real:

$$\textbf{False Proof: } A\mathbf{x} = \lambda\mathbf{x} \text{ gives } \mathbf{x}^T A\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} \text{ so } \lambda = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \text{ is real.}$$

There is a hidden assumption in this proof which is not justified. Find the flaw by testing each step on the 90 $^\circ$ rotation matrix:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

with $\lambda = i$ and $\mathbf{x} = (i, 1)$.

Solution: We can esily confirm that $A\mathbf{x} = \lambda\mathbf{x} = \begin{bmatrix} -1 \\ i \end{bmatrix}$. Next, check if $\mathbf{x^T}A\mathbf{x} = \lambda\mathbf{x^T}\mathbf{x}$ is true for the 90 $^\circ$ rotation matrix:

$$\begin{aligned} \mathbf{x}^T A\mathbf{x} &= \begin{bmatrix} i & 1 \end{bmatrix} \begin{bmatrix} -1 \\ i \end{bmatrix} = 0 \\ \lambda\mathbf{x}^T\mathbf{x} &= i \begin{bmatrix} i & 1 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = 0 \\ \mathbf{x}^T A\mathbf{x} &= \lambda\mathbf{x}^T\mathbf{x}.\checkmark \end{aligned}$$

Note that $\mathbf{x^T}\mathbf{x} = 0$. Since the next and last step involves dividing by this term, the hidden assumption must be that $\mathbf{x^T}\mathbf{x} \neq 0$. If $x = (a, b)$ then

$$\mathbf{x^T}\mathbf{x} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2.$$

The "proof" assumes that the squares of the components of the eigenvector cannot sum to zero: $a^2 + b^2 \neq 0$. This may be false if the components are complex.

Problem 25.2: (6.5 #32.) A *group* of nonsingular matrices includes AB and A^{-1} if it includes A and B. "Products and inverses stay in the group." Which of these are groups?

a) Positive definite symmetric matrices A.

b) Orthogonal matrices Q.

c) All exponentials e^{tA} of a fixed matrix A.

d) Matrices D with determinant 1.

Solution:

a) The positive definite symmetric matrices A **do not form a group.** To show this, we provide a counterexample in the form of two positive definite symmetric matrices A and B whose product is not a positive definite symmetric matrix.

If $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$ then $AB = \begin{bmatrix} 2.5 & 2 \\ 1.5 & 1.5 \end{bmatrix}$ is not symmetric.

b) The orthogonal matrices Q **form a group**. If A and B are orthogonal matrices, then:

$$A^TA = I \Rightarrow A^{-1} = A^T \Rightarrow A^{-1} \text{ is orthogonal, and}$$
$$B^TB = I \Rightarrow (AB)^TAB = B^TA^TAB = B^TB = I \Rightarrow AB \text{ is orthogonal.}$$

c) The exponentials e^{tA} of a fixed matrix A **form a group.** For the elements e^{pA} and e^{qA}:

$$(e^{pA})^{-1} = e^{-pA} \text{ is of the form } e^{tA}$$
$$e^{pA}e^{qA} = e^{(p+q)A} \text{ is of the form } e^{tA}$$

d) The matrices D with determinant 1 **form a group.** If $\det A = 1$ then $\det A^{-1} = 1$. If matrices A and B have determinant 1 then their product also has determinant 1:

$$\det(AB) = \det(A)\det(B) = 1.$$

Complex matrices; fast Fourier transform

Matrices with all real entries can have complex eigenvalues! So we can't avoid working with complex numbers. In this lecture we learn to work with complex vectors and matrices.

The most important complex matrix is the Fourier matrix F_n, which is used for Fourier transforms. Normally, multiplication by F_n would require n^2 multiplications. The fast Fourier transform (FFT) reduces this to roughly $n \log_2 n$ multiplications, a revolutionary improvement.

Complex vectors

Length

Given a vector $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$ with complex entries, how do we find its length? Our old definition:

$$\mathbf{z}^T\mathbf{z} = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

is no good; this quantity isn't always positive. For example:

$$\begin{bmatrix} 1 & i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 0.$$

We don't want to define the length of $\begin{bmatrix} 1 \\ i \end{bmatrix}$ to be 0. The correct definition is: $|\mathbf{z}|^2 = \overline{\mathbf{z}}^T\mathbf{z} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2$. Then we have:

$$\left(\text{length} \begin{bmatrix} 1 \\ i \end{bmatrix}\right)^2 = \begin{bmatrix} 1 & -i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 2.$$

To simplify our notation we write $|\mathbf{z}|^2 = \mathbf{z}^H\mathbf{z}$, where $\mathbf{z}^H = \overline{\mathbf{z}}^T$. The H comes from the name Hermite, and $\mathbf{z}^H\mathbf{z}$ is read "$\mathbf{z}$ Hermitian $\mathbf{z}$".

Inner product

Similarly, the inner or dot product of two complex vectors is not just $\mathbf{y}^T\mathbf{x}$. We must also take the complex conjugate of $\mathbf{y}$:

$$\mathbf{y}^H\mathbf{x} = \overline{\mathbf{y}}^T\mathbf{x} = \overline{y}_1 x_1 + \overline{y}_2 x_2 + \cdots + \overline{y}_n x_n.$$

Complex matrices

Hermitian matrices

Symmetric matrices are real valued matrices for which $A^T = A$. If A is complex, a nicer property is $\overline{A}^T = A$; such a matrix is called *Hermitian* and we abbreviate $\overline{A}^T$ as A^H. Note that the diagonal entries of a Hermitian matrix must be real. For example,

$$\overline{A}^T = A = \begin{bmatrix} 2 & 3+i \\ 3-i & 5 \end{bmatrix}.$$

Similar to symmetric matrices, Hermitian matrices have real eigenvalues and perpendicular eigenvectors.

Unitary matrices

What does it mean for complex vectors $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n$ to be perpendicular (or orthonormal)? We must use our new definition of the inner product. For a collection of $\mathbf{q}_j$ in complex space to be orthonormal, we require:

$$\overline{\mathbf{q}}_j \mathbf{q}_k = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

We can again define $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$, and then $Q^H Q = I$. Just as "Hermitian" is the complex equivalent of "symmetric", the term "*unitary*" is analogous to "orthogonal". A *unitary matrix* is a square matrix with perpendicular columns of unit length.

Discrete Fourier transform

A *Fourier series* is a way of writing a periodic function or *signal* as a sum of functions of different frequencies:

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots .$$

When working with finite data sets, the *discrete Fourier transform* is the key to this decomposition.

In electrical engineering and computer science, the rows and columns of a matrix are numbered starting with 0, not 1 (and ending with $n-1$, not n). We'll follow this convention when discussing the Fourier matrix:

$$F_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & & w^{n-1} \\ 1 & w^2 & w^4 & & w^{2(n-1)} \\ \vdots & & & \ddots & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \cdots & w^{(n-1)^2} \end{bmatrix}.$$

Notice that $F_n = F_n^T$ and $(F_n)_{jk} = w^{jk}$, where $j,k = 0,1,...,n-1$ and the complex number w is $w = e^{i \cdot 2\pi/n}$ (so $w^n = 1$). The columns of this matrix are orthogonal.

All the entries of F_n are on the unit circle in the complex plane, and raising each one to the nth power gives 1. We could write $w = \cos(2\pi/n) + i\sin(2\pi/n)$, but that would just make it harder to compute w^{jk}.

Because $w^4 = 1$ and $w = e^{2\pi i/4} = i$, our best example of a Fourier matrix is:

$$F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}.$$

To find the Fourier transform of a vector with four components (four data points) we multiply by F_4.

It's easy to check that the columns of F_4 are orthogonal, as long as we remember to conjugate when computing the inner product. However, F_4 is not quite unitary because each column has length 2. We could divide each entry by 2 to get a matrix whose columns are orthonormal:

$$\frac{1}{4} F_4^H F_4 = I.$$

An example

The signal corresponding to a single impulse at time zero is (roughly) described by $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. To find the Fourier transform of this signal we compute:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

A single impulse has all frequencies in equal amounts.

If we multiply by F_4 again we almost get back to $(1,0,0,0)$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Because $\frac{1}{\sqrt{n}} F_n$ is unitary, multiplying by F_n and dividing by the scalar n inverts the transform.

Fast Fourier transform

Fourier matrices can be broken down into chunks with lots of zero entries; Fourier probably didn't notice this. Gauss did, but didn't realize how significant a discovery this was.

There's a nice relationship between F_n and F_{2n} related to the fact that $w_{2n}^2 = w_n$:

$$F_{2n} = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_n & 0 \\ 0 & F_n \end{bmatrix} P,$$

where D is a diagonal matrix and P is a $2n$ by $2n$ permutation matrix:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

So, a $2n$ sized Fourier transform F times $\mathbf{x}$ which we might think would require $(2n)^2 = 4n^2$ operations can instead be performed using two size n Fourier transforms ($2n^2$ operations) plus two very simple matrix multiplications which require on the order of n multiplications. The matrix P picks out the even components x_0, x_2, x_4, ... of a vector first, and then the odd ones – this calculation can be done very quickly.

Thus we can do a Fourier transform of size 64 on a vector by separating the vector into its odd and even components, performing a size 32 Fourier transform on each half of its components, then recombining the two halves through a process which involves multiplication by the diagonal matrix D.

$$D = \begin{bmatrix} 1 & & & & \\ & w & & & \\ & & w^2 & & \\ & & & \ddots & \\ & & & & w^{n-1} \end{bmatrix}.$$

Of course we can break each of those copies of F_{32} down into two copies of F_{16}, and so on. In the end, instead of using n^2 operations to multiply by F_n we get the same result using about $\frac{1}{2} n \log_2 n$ operations.

A typical case is $n = 1024 = 2^{10}$. Simply multiplying by F_n requires over a million calculations. The fast Fourier transform can be completed with only $\frac{1}{2} n \log_2 n = 5 \cdot 1024$ calculations. This is 200 times faster!

This is only possible because Fourier matrices are special matrices with orthogonal columns. In the next lecture we'll return to dealing exclusively with real numbers and will learn about positive definite matrices, which are the matrices most often seen in applications.

Exercises on complex matrices; fast Fourier transform

Problem 26.1: Compute the matrix F_2.

Problem 26.2: Find the matrices D and P used in the factorization:

$$F_4 = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_2 & \\ & F_2 \end{bmatrix} P$$

Hint: D is created using fourth roots, not square roots, of 1. Check your answer by multiplying.

Exercises on complex matrices; fast Fourier transform

Problem 26.1: Compute the matrix F_2.

Solution: $F_2 = \begin{bmatrix} 1 & 1 \\ 1 & w \end{bmatrix}$, where $w = e^{i2\pi/2} = -1$. Hence

$$F_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Problem 26.2: Find the matrices D and P used in the factorization:

$$F_4 = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_2 & \\ & F_2 \end{bmatrix} P$$

Hint: D is created using fourth roots, not square roots, of 1. Check your answer by multiplying.

Solution: We computed $F_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ in the previous problem.
$D = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$ contains the first two fourth roots of 1 (and $-D$ contains the others, -1 and -i).

P is a permutation matrix that arranges the components of the incoming vector so that its even components come first. For F_4, that means swapping the first and second components:

$$P \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_0 \\ x_2 \\ x_1 \\ x_3 \end{bmatrix}.$$

$$\text{So, } P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Finally, we check our work by multiplying:

$$
\begin{aligned}
\begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_2 & \\ & F_2 \end{bmatrix} P &= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & i \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -i \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & i \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -i \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = F_4. \ \checkmark
\end{aligned}
$$

Positive definite matrices and minima

Studying positive definite matrices brings the whole course together; we use pivots, determinants, eigenvalues and stability. The new quantity here is $\mathbf{x}^T A\mathbf{x}$; watch for it.

This lecture covers how to tell if a matrix is positive definite, what it means for it to be positive definite, and some geometry.

Positive definite matrices

Given a symmetric two by two matrix $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, here are four ways to tell if it's positive definite:

1. Eigenvalue test: $\lambda_1 > 0$, $\lambda_2 > 0$.
2. Determinants test: $a > 0$, $ac - b^2 > 0$.
3. Pivot test: $a > 0$, $\dfrac{ac - b^2}{a} > 0$.
4. $\mathbf{x}^T A\mathbf{x}$ is positive except when $\mathbf{x} = \mathbf{0}$ (this is usually the definition of positive definiteness).

2 by 2

Using the determinants test, we know that $\begin{bmatrix} 2 & 6 \\ 6 & y \end{bmatrix}$ is positive definite when $2y - 36 > 0$ or when $y > 18$.

The matrix $\begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ is on the borderline of positive definiteness and is called a *positive semidefinite* matrix. It's a singular matrix with eigenvalues 0 and 20. Positive semidefinite matrices have eigenvalues greater than or equal to 0. For a singular matrix, the determinant is 0 and it only has one pivot.

$$\begin{aligned}
\mathbf{x}^T A\mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2x_1 + 6x_2 \\ 6x_1 + 18x_2 \end{bmatrix} \\
&= 2x_1^2 + 12x_1x_2 + 18x_2^2 \\
&= ax_1^2 + 2bx_1x_2 + cx_2^2.
\end{aligned}$$

If this *quadratic form* is positive for every (real) x_1 and x_2 then the matrix is positive definite. In this positive semi-definite example, $2x_1^2 + 12x_1x_2 + 18x_2^2 = 2(x_1 + 3x_2)^2 = 0$ when $x_1 = 3$ and $x_2 = -1$.

Tests for minimum

If we apply the fourth test to the matrix $\begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}$ which is not positive definite, we get the quadratic form $f(x,y) = 2x^2 + 12xy + 7y^2$. The graph of this function has a saddle point at the origin; see Figure 1.

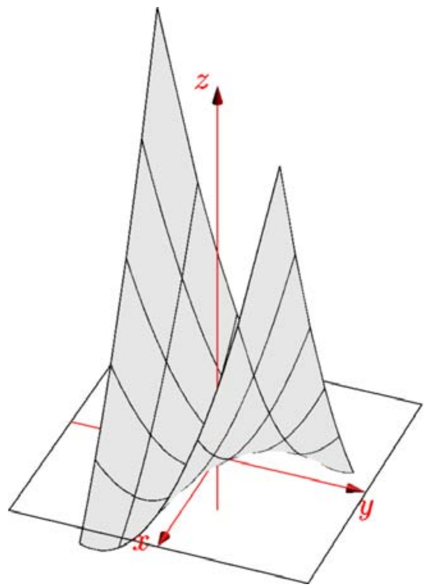

Figure 1: The graph of $f(x,y) = 2x^2 + 12xy + 7y^2$.

The matrix $\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$ is positive definite – its determinant is 4 and its trace is 22 so its eigenvalues are positive. The quadratic form associated with this matrix is $f(x,y) = 2x^2 + 12xy + 20y^2$, which is positive except when $x = y = 0$. The level curves $f(x,y) = k$ of this graph are ellipses; its graph appears in Figure 2. If $a > 0$ and $c > 0$, the quadratic form $ax^2 + 2bxy + cy^2$ is only negative when the value of $2bxy$ is negative and overwhelms the (positive) value of $ax^2 + cy^2$.

The first derivatives f_x and f_y of this function are zero, so its graph is tangent to the xy-plane at $(0,0,0)$; but this was also true of $2x^2 + 12xy + 7y^2$. As in single variable calculus, we need to look at the second derivatives of f to tell whether there is a minimum at the critical point.

We can prove that $2x^2 + 12xy + 20y^2$ is always positive by writing it as a sum of squares. We do this by completing the square:

$$2x^2 + 12xy + 20y^2 = 2(x+3y)^2 + 2y^2.$$

Note that $2(x+3y)^2 = 2x^2 + 12xy + 18y^2$, and 18 was the "borderline" between passing and failing the tests for positive definiteness.

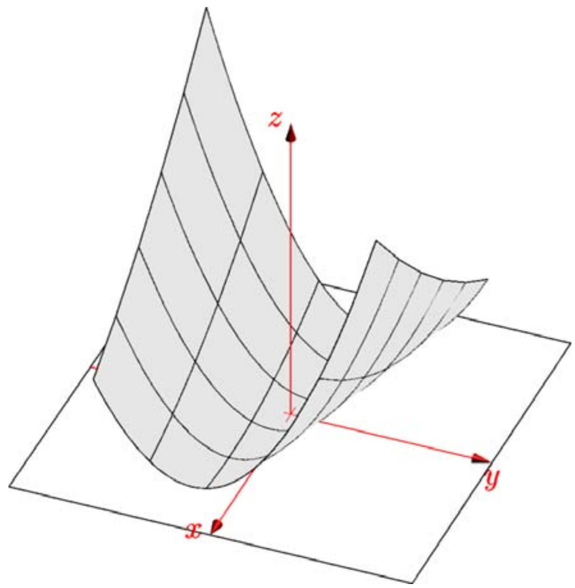

Figure 2: The graph of $f(x, y) = 2x^2 + 12xy + 20y^2$.

When we complete the square for $2x^2 + 12xy + 7y^2$ we get:

$$2x^2 + 12xy + 7y^2 = 2(x + 3y)^2 - 11y^2$$

which may be negative; e.g. when $x = -3$ and $y = 1$.

The coefficients that appear when completing the square are exactly the entries that appear when performing elimination on the original matrix. The two pivots are multiplied by the squares, and the coefficient c in the term $(x - cy)^2$ is the multiple of the first row that's subtracted from the second row.

$$\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix} \xrightarrow{\text{subtract } \mathbf{3} \text{ times row 1}} \begin{bmatrix} \mathbf{2} & 6 \\ 0 & \mathbf{2} \end{bmatrix}.$$

We can see the terms that appear when completing the square in:

$$U = \begin{bmatrix} \mathbf{2} & 6 \\ 0 & \mathbf{2} \end{bmatrix}, \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 \\ \mathbf{3} & 1 \end{bmatrix}.$$

When we complete the square, the numbers multiplied by the squares are the pivots; if the pivots are all positive then the sum of squares will always be positive.

Hessian matrix

The matrix of second derivatives of $f(x, y)$ is:

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}.$$

This matrix is symmetric because $f_{xy} = f_{yx}$. Its determinant is positive when the matrix is positive definite, which matches the $f_{xx}f_{yy} > f_{xy}^2$ test for a minimum that we learned in calculus.

n **by** *n*

A function of several variables $f(x_1, x_2, ..., x_n)$ has a minimum when its matrix of second derivatives is positive definite, and identifying minima of functions is often important. The tests we've just learned for 2 by 2 matrices also apply to n by n matrices.

A 3 by 3 example:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Is this matrix positive definite? Our tests will say *yes*. What's the function $\mathbf{x}^T A\mathbf{x}$ associated with this matrix? Does that function have a minimum at $\mathbf{x} = \mathbf{0}$? What does the graph of its quadratic form look like?

Looking at determinants we see:

$$\det[\,2\,] = 2, \qquad \det\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 5, \qquad \det\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = 4.$$

These are all positive, so A is positive definite.

The pivots of A are 2, 3/2 and 4/3 (all positive) because the products of the pivots equal the determinants.

The eigenvalues of A are positive and their product is 4. It's not difficult to check that they are $2 - \sqrt{2}$, 2 and $2 + \sqrt{2}$ (all positive).

Ellipsoids in $\mathbb{R}^n$

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3.$$

Because A is positive definite, we expect $f(\mathbf{x})$ to be positive except when $\mathbf{x} = \mathbf{0}$. Its graph is a sort of four dimensional bowl or *paraboloid*. If we wrote $f(\mathbf{x})$ as a sum of three squares, those squares would be multiplied by the (positive) pivots of A. Earlier, we said that a horizontal slice of our three dimensional bowl shape would be an ellipse. Here, a horizontal slice of the four dimensional bowl is an ellipsoid – a little bit like a rugby ball. For example, if we cut the graph at height 1 we get a surface whose equation is: $2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 = 1$.

Just as an ellipse has a major and minor axis, an ellipsoid has three axes. If we write $A = Q\Lambda Q^T$, as the principal axis theorem tells us we can, the eigenvectors of A tell us the directions of the principal axes of the ellipsoid. The eigenvalues tell us the lengths of those axes.

Exercises on positive definite matrices and minima

Problem 27.1: (6.5 #33. *Introduction to Linear Algebra:* Strang) When A and B are symmetric positive definite, AB might not even be symmetric, but its eigenvalues are still positive. Start from $AB\mathbf{x} = \lambda\mathbf{x}$ and take dot products with $B\mathbf{x}$. Then prove $\lambda > 0$.

Problem 27.2: Find the quadratic form associated with the matrix $\begin{bmatrix} 1 & 5 \\ 7 & 9 \end{bmatrix}$. Is this function $f(x, y)$ always positive, always negative, or sometimes positive and sometimes negative?

Exercises on positive definite matrices and minima

Problem 27.1: (6.5 #33. *Introduction to Linear Algebra:* Strang) When A and B are symmetric positive definite, AB might not even be symmetric, but its eigenvalues are still positive. Start from $AB\mathbf{x} = \lambda\mathbf{x}$ and take dot products with $B\mathbf{x}$. Then prove $\lambda > 0$.

Solution:

$$\begin{aligned} AB\mathbf{x} &= \lambda\mathbf{x} \\ (AB\mathbf{x})^T B\mathbf{x} &= (\lambda\mathbf{x})^T B\mathbf{x} \\ (B\mathbf{x})^T A^T B\mathbf{x} &= \lambda\mathbf{x}^T B\mathbf{x} \\ (B\mathbf{x})^T A(B\mathbf{x}) &= \lambda(\mathbf{x}^T B\mathbf{x}). \end{aligned}$$

where $A^T = A$ because A is symmetric. Since A is positive definite we know $(B\mathbf{x})^T A(B\mathbf{x}) > 0$, and since B is positive definite $\mathbf{x}^T B\mathbf{x} > 0$. Hence, λ must be positive as well.

Problem 27.2: Find the quadratic form associated with the matrix $\begin{bmatrix} 1 & 5 \\ 7 & 9 \end{bmatrix}$. Is this function $f(x,y)$ always positive, always negative, or sometimes positive and sometimes negative?

Solution: To find the quadratic form, compute $\mathbf{x^T}A\mathbf{x}$:

$$\begin{aligned} f(x,y) &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 7 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= x(x+5y) + y(7x+9y) \\ &= \mathbf{x^2 + 12xy + 9y^2}. \end{aligned}$$

This expression can be positive, e.g. when $y = 0$ and $x \neq 0$.

The expression will sometimes be negative because A is not positive definite. For instance, $f(2,-2) = -8$. Thus the quadratic form associated with the matrix A is **sometimes positive and sometimes negative**. Another way to reach this conclusion is to note that $\det A = -26$ is negative and so A is not positive definite.

Similar matrices and Jordan form

We've nearly covered the entire heart of linear algebra – once we've finished singular value decompositions we'll have seen all the most central topics.

$A^T A$ is positive definite

A matrix is *positive definite* if $\mathbf{x}^T A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$. This is a very important class of matrices; positive definite matrices appear in the form of $A^T A$ when computing least squares solutions. In many situations, a rectangular matrix is multiplied by its transpose to get a square matrix.

Given a symmetric positive definite matrix A, is its inverse also symmetric and positive definite? Yes, because if the (positive) eigenvalues of A are $\lambda_1, \lambda_2, \cdots \lambda_d$ then the eigenvalues $1/\lambda_1, 1/\lambda_2, \cdots 1/\lambda_d$ of A^{-1} are also positive.

If A and B are positive definite, is $A + B$ positive definite? We don't know much about the eigenvalues of $A + B$, but we can use the property $\mathbf{x}^T A\mathbf{x} > 0$ and $\mathbf{x}^T B\mathbf{x} > 0$ to show that $\mathbf{x}^T (A + B)\mathbf{x} > 0$ for $\mathbf{x} \neq 0$ and so $A + B$ is also positive definite.

Now suppose A is a rectangular (m by n) matrix. A is almost certainly not symmetric, but $A^T A$ is square and symmetric. Is $A^T A$ positive definite? We'd rather not try to find the eigenvalues or the pivots of this matrix, so we ask when $\mathbf{x}^T A^T A\mathbf{x}$ is positive.

Simplifying $\mathbf{x}^T A^T A\mathbf{x}$ is just a matter of moving parentheses:

$$\mathbf{x}^T (A^T A)\mathbf{x} = (A\mathbf{x})^T (A\mathbf{x}) = |A\mathbf{x}|^2 \geq 0.$$

The only remaining question is whether $A\mathbf{x} = \mathbf{0}$. If A has rank n (independent columns), then $\mathbf{x}^T (A^T A)\mathbf{x} = \mathbf{0}$ only if $\mathbf{x} = \mathbf{0}$ and A is positive definite.

Another nice feature of positive definite matrices is that you never have to do row exchanges when row reducing – there are never 0's or unsuitably small numbers in their pivot positions.

Similar matrices A and $B = M^{-1} AM$

Two square matrices A and B are *similar* if $B = M^{-1} AM$ for some matrix M. This allows us to put matrices into families in which all the matrices in a family are similar to each other. Then each family can be represented by a diagonal (or nearly diagonal) matrix.

Distinct eigenvalues

If A has a full set of eigenvectors we can create its eigenvector matrix S and write $S^{-1} AS = \Lambda$. So A is similar to Λ (choosing M to be this S).

If $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ then $\Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ and so A is similar to $\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$. But A is also similar to:

$$\begin{aligned} \overset{M^{-1}}{\begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}} \quad \overset{A}{\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}} \quad \overset{M}{\begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}} &= \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 9 \\ 1 & 6 \end{bmatrix} \\ &= \overset{B}{\begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}}. \end{aligned}$$

In addition, B is similar to Λ. All these similar matrices have the same eigenvalues, 3 and 1; we can check this by computing the trace and determinant of A and B.

Similar matrices have the same eigenvalues!

In fact, the matrices similar to A are all the 2 by 2 matrices with eigenvalues 3 and 1. Some other members of this family are $\begin{bmatrix} 3 & 7 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 7 \\ 0 & 3 \end{bmatrix}$. To prove that similar matrices have the same eigenvalues, suppose $A\mathbf{x} = \lambda\mathbf{x}$. We modify this equation to include $B = M^{-1}AM$:

$$\begin{aligned} AMM^{-1}\mathbf{x} &= \lambda\mathbf{x} \\ M^{-1}AMM^{-1}\mathbf{x} &= \lambda M^{-1}\mathbf{x} \\ BM^{-1}\mathbf{x} &= \lambda M^{-1}\mathbf{x}. \end{aligned}$$

The matrix B has the same λ as an eigenvalue. $M^{-1}\mathbf{x}$ is the eigenvector.

If two matrices are similar, they have the same eigenvalues and the same number of independent eigenvectors (but probably not the same eigenvectors).

When we diagonalize A, we're finding a diagonal matrix Λ that is similar to A. If two matrices have the same n distinct eigenvalues, they'll be similar to the same diagonal matrix.

Repeated eigenvalues

If two eigenvalues of A are the same, it may not be possible to diagonalize A. Suppose $\lambda_1 = \lambda_2 = 4$. One family of matrices with eigenvalues 4 and 4 contains only the matrix $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$. The matrix $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ is not in this family.

There are two families of similar matrices with eigenvalues 4 and 4. The larger family includes $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$. Each of the members of this family has only one eigenvector.

The matrix $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ is the only member of the other family, because:

$$M^{-1} \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} M = 4M^{-1}M = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

for any invertible matrix M.

Jordan form

Camille Jordan found a way to choose a "most diagonal" representative from each family of similar matrices; this representative is said to be in *Jordan normal form*. For example, both $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ and $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ are in Jordan form. This form used to be the climax of linear algebra, but not any more. Numerical applications rarely need it.

We can find more members of the family represented by $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ by choosing diagonal entries to get a trace of 4, then choosing off-diagonal entries to get a determinant of 16:

$$\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 17 & 4 \end{bmatrix}, \begin{bmatrix} a & b \\ (8a - a^2 - 16)/b & 8 - a \end{bmatrix}.$$

(None of these are diagonalizable, because if they were they would be similar to $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$. That matrix is only similar to itself.) What about this one?

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Its eigenvalues are four zeros. Its rank is 2 so the dimension of its nullspace is $4 - 2 = 2$. It will have two independent eigenvectors and two "missing" eigenvectors. When we look instead at

$$\begin{bmatrix} 0 & 1 & 7 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

its rank and the dimension of its nullspace are still 2, but it's not as nice as A. B *is* similar to A, which is the Jordan normal form representative of this family. A has a 1 above the diagonal for every missing eigenvector and the rest of its entries are 0.

Now consider:

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Again it has rank 2 and its nullspace has dimension 2. Its four eigenvalues are 0. Surprisingly, it is not similar to A. We can see this by breaking the matrices into their *Jordan blocks*:

$$A = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right], \qquad C = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

A Jordan block J_i has a repeated eigenvalue λ_i on the diagonal, zeros below the diagonal and in the upper right hand corner, and ones above the diagonal:

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \lambda_i & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}.$$

Two matrices may have the same eigenvalues and the same number of eigenvectors, but if their Jordan blocks are different sizes those matrices can not be similar.

Jordan's theorem says that every square matrix A is similar to a Jordan matrix J, with Jordan blocks on the diagonal:

$$J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & J_d \end{bmatrix}.$$

In a Jordan matrix, the eigenvalues are on the diagonal and there may be ones above the diagonal; the rest of the entries are zero. The number of blocks is the number of eigenvectors – there is one eigenvector per block.

To summarize:

- If A has n distinct eigenvalues, it is diagonalizable and its Jordan matrix is the diagonal matrix $J = \Lambda$.
- If A has repeated eigenvalues and "missing" eigenvectors, then its Jordan matrix will have $n - d$ ones above the diagonal.

We have not learned to compute the Jordan matrix of a matrix which is missing eigenvectors, but we do know how to diagonalize a matrix which has n distinct eigenvalues.

Exercises on similar matrices and Jordan form

Problem 28.1: (6.6 #12. *Introduction to Linear Algebra:* Strang) These Jordan matrices have eigenvalues 0, 0, 0, 0. They have two eigenvectors; one from each block. However, their block sizes don't match and they are *not similar*:

$$J = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \text{ and } K = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right].$$

For a generic matrix M, show that if $JM = MK$ then M is not invertible and so J is not similar to K.

Problem 28.2: (6.6 #20.) Why are these statements all true?

a) If A is similar to B then A^2 is similar to B^2.

b) A^2 and B^2 can be similar when A and B are not similar (try $\lambda = 0, 0$.)

c) $\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$ is similar to $\begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix}$.

d) $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ is not similar to $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$.

e) Given a matrix A, let B be the matrix obtained by exchanging rows 1 and 2 of A and then exchanging columns 1 and 2 of A. Show that A is similar to B.

Exercises on similar matrices and Jordan form

Problem 28.1: (6.6 #12. *Introduction to Linear Algebra:* Strang) These Jordan matrices have eigenvalues 0, 0, 0, 0. They have two eigenvectors; one from each block. However, their block sizes don't match and they are *not similar*:

$$J = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \text{ and } K = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right].$$

For a generic matrix M, show that if $JM = MK$ then M is not invertible and so J is not similar to K.

Solution: Let $M = (m_{ij})$. Then:

$$JM = \begin{bmatrix} m_{21} & m_{22} & m_{23} & m_{24} \\ 0 & 0 & 0 & 0 \\ m_{41} & m_{42} & m_{43} & m_{44} \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } MK = \begin{bmatrix} 0 & m_{11} & m_{12} & 0 \\ 0 & m_{21} & m_{22} & 0 \\ 0 & m_{31} & m_{32} & 0 \\ 0 & m_{41} & m_{42} & 0 \end{bmatrix}$$

If $JM = MK$, then $m_{11} = m_{22} = 0$, $m_{21} = 0$, $m_{31} = m_{42} = 0$, and $m_{41} = 0$. Thus, the first column of M is all zeros and M is not invertible.

If J were similar to K there would be an invertible matrix M that satisfies $K = M^{-1}JM$, and so $MK = JM$. We just showed that there can be no such invertible matrix M. Therefore J is not similar to K.

Problem 28.2: (6.6 #20.) Why are these statements all true?

a) If A is similar to B then A^2 is similar to B^2.

b) A^2 and B^2 can be similar when A and B are not similar (try $\lambda = 0, 0$.)

c) $\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$ is similar to $\begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix}$.

d) $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ is not similar to $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$.

e) Given a matrix A, let B be the matrix obtained by exchanging rows 1 and 2 of A and then exchanging columns 1 and 2 of A. Show that A is similar to B.

Solution:

a) If A is similar to B, then:

$$A = M^{-1}BM \Longrightarrow A^2 = M^{-1}BM(M^{-1}BM) = M^{-1}B^2M.$$

Since $A^2 = M^{-1}B^2M$, by definition A^2 is similar to B^2.

b) Let:

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Then $A^2 = B^2$ and so A^2 is similar to B^2, but A is not similar to B because nothing but the zero matrix is similar to the zero matrix.

c) There are multiple ways to verify that the given matrices are similar. One way is to explicitly find the matrix M that satisfies the similarity condition:

$$\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

d) The given matrices are not similar because for the first matrix every vector in the plane is an eigenvector for $\lambda = 3$, whereas the second matrix only has a line ($y = 0$) of eigenvectors corresponding to $\lambda = 3$. We know that similar matrices have the same number of independent eigenvectors, so the given matrices cannot be similar.

e) To exchange the first two rows of A, we multiply A on the left by:

$$M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

In order to exchange the first two columns of A, we multiply it on the right by the same matrix M. We thus have $B = MAM$. Since $M^{-1} = M$, B is similar to A.

Singular value decomposition

The *singular value decomposition* of a matrix is usually referred to as the *SVD*. This is the final and best factorization of a matrix:

$$A = U\Sigma V^T$$

where U is orthogonal, Σ is diagonal, and V is orthogonal.

In the decomoposition $A = U\Sigma V^T$, A can be *any* matrix. We know that if A is symmetric positive definite its eigenvectors are orthogonal and we can write $A = Q\Lambda Q^T$. This is a special case of a SVD, with $U = V = Q$. For more general A, the SVD requires two different matrices U and V.

We've also learned how to write $A = S\Lambda S^{-1}$, where S is the matrix of n distinct eigenvectors of A. However, S may not be orthogonal; the matrices U and V in the SVD will be.

How it works

We can think of A as a linear transformation taking a vector $\mathbf{v}_1$ in its row space to a vector $\mathbf{u}_1 = A\mathbf{v}_1$ in its column space. The SVD arises from finding an orthogonal basis for the row space that gets transformed into an orthogonal basis for the column space: $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.

It's not hard to find an orthogonal basis for the row space – the Gram-Schmidt process gives us one right away. But in general, there's no reason to expect A to transform that basis to another orthogonal basis.

You may be wondering about the vectors in the nullspaces of A and A^T. These are no problem – zeros on the diagonal of Σ will take care of them.

Matrix language

The heart of the problem is to find an orthonormal basis $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$ for the row space of A for which

$$\begin{aligned} A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} &= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_r\mathbf{u}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}, \end{aligned}$$

with $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_r$ an orthonormal basis for the column space of A. Once we add in the nullspaces, this equation will become $AV = U\Sigma$. (We can complete the orthonormal bases $\mathbf{v}_1, \ldots \mathbf{v}_r$ and $\mathbf{u}_1, \ldots \mathbf{u}_r$ to orthonormal bases for the entire space any way we want. Since $\mathbf{v}_{r+1}, \ldots \mathbf{v}_n$ will be in the nullspace of A, the diagonal entries $\sigma_{r+1}, \ldots \sigma_n$ will be 0.)

The columns of U and V are bases for the row and column spaces, respectively. Usually $U \neq V$, but if A is positive definite we can use the *same* basis for its row and column space!

Calculation

Suppose A is the invertible matrix $\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We want to find vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in the row space $\mathbb{R}^2$, $\mathbf{u}_1$ and $\mathbf{u}_2$ in the column space $\mathbb{R}^2$, and positive numbers σ_1 and σ_2 so that the $\mathbf{v}_i$ are orthonormal, the $\mathbf{u}_i$ are orthonormal, and the σ_i are the scaling factors for which $A\mathbf{v}_i = \sigma_i u_i$.

This is a big step toward finding orthonormal matrices V and U and a diagonal matrix Σ for which:

$$AV = U\Sigma.$$

Since V is orthogonal, we can multiply both sides by $V^{-1} = V^T$ to get:

$$A = U\Sigma V^T.$$

Rather than solving for U, V and Σ simultaneously, we multiply both sides by $A^T = V\Sigma^T U^T$ to get:

$$\begin{aligned} A^T A &= V\Sigma U^{-1} U\Sigma V^T \\ &= V\Sigma^2 V^T \\ &= V \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} V^T. \end{aligned}$$

This is in the form $Q\Lambda Q^T$; we can now find V by diagonalizing the symmetric positive definite (or semidefinite) matrix $A^T A$. The columns of V are eigenvectors of $A^T A$ and the eigenvalues of $A^T A$ are the values σ_i^2. (We choose σ_i to be the positive square root of λ_i.)

To find U, we do the same thing with AA^T.

SVD example

We return to our matrix $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We start by computing

$$\begin{aligned} A^T A &= \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}. \end{aligned}$$

The eigenvectors of this matrix will give us the vectors $\mathbf{v}_i$, and the eigenvalues will gives us the numbers σ_i.

Two orthogonal eigenvectors of A^TA are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. To get an orthonormal basis, let $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$. These have eigenvalues $\sigma_1^2 = 32$ and $\sigma_2^2 = 18$. We now have:

$$\underset{A}{\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}} = \underset{U}{\begin{bmatrix} & \\ & \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix}} \underset{V^T}{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}.$$

We could solve this for U, but for practice we'll find U by finding orthonormal eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ for $AA^T = U\Sigma^2U^T$.

$$\begin{aligned} AA^T &= \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}. \end{aligned}$$

Luckily, AA^T happens to be diagonal. It's tempting to let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, as Professor Strang did in the lecture, but because $A\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3\sqrt{2} \end{bmatrix}$ we instead have $\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Note that this also gives us a chance to double check our calculation of σ_1 and σ_2.

Thus, the SVD of A is:

$$\underset{A}{\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}} = \underset{U}{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix}} \underset{V^T}{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}.$$

Example with a nullspace

Now let $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$. This has a one dimensional nullspace and one dimensional row and column spaces.

The row space of A consists of the multiples of $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$. The column space of A is made up of multiples of $\begin{bmatrix} 4 \\ 8 \end{bmatrix}$. The nullspace and left nullspace are perpendicular to the row and column spaces, respectively.

Unit basis vectors of the row and column spaces are $\mathbf{v}_1 = \begin{bmatrix} .8 \\ .6 \end{bmatrix}$ and $\mathbf{u}_1 =$

$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$. To compute σ_1 we find the nonzero eigenvalue of $A^T A$.

$$\begin{aligned} A^T A &= \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}. \end{aligned}$$

Because this is a rank 1 matrix, one eigenvalue must be 0. The other must equal the trace, so $\sigma_1^2 = 125$. After finding unit vectors perpendicular to $\mathbf{u}_1$ and $\mathbf{v}_1$ (basis vectors for the left nullspace and nullspace, respectively) we see that the SVD of A is:

$$\underset{A}{\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}} = \underset{U}{\frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix}}.$$

The singular value decomposition combines topics in linear algebra ranging from positive definite matrices to the four fundamental subspaces.

$\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$	is an orthonormal basis for the row space.
$\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_r$	is an orthonormal basis for the column space.
$\mathbf{v}_{r+1}, \ldots \mathbf{v}_n$	is an orthonormal basis for the nullspace.
$\mathbf{u}_{r+1}, \ldots \mathbf{u}_m$	is an orthonormal basis for the left nullspace.

These are the "right" bases to use, because $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.

Exercises on singular value decomposition

Problem 29.1: (Based on 6.7 #4. *Introduction to Linear Algebra:* Strang) Verify that if we compute the singular value decomposition $A = U\Sigma V^T$ of the Fibonacci matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$,

$$\Sigma = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{\sqrt{5}-1}{2} \end{bmatrix}.$$

Problem 29.2: (6.7 #11.) Suppose A has orthogonal columns $\mathbf{w}_1$, $\mathbf{w}_2$, ..., $\mathbf{w}_n$ of lengths σ_1, σ_2, ..., σ_n. Calculate A^TA. What are U, Σ, and V in the SVD?

Exercises on singular value decomposition

Problem 29.1: (Based on 6.7 #4. *Introduction to Linear Algebra:* Strang) Verify that if we compute the singular value decomposition $A = U\Sigma V^T$ of the Fibonacci matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$,

$$\Sigma = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{\sqrt{5}-1}{2} \end{bmatrix}.$$

Solution:

$$A^T A = A A^T = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

The eigenvalues of this matrix are the roots of $x^2 - 3x + 1$, which are $\dfrac{3 \pm \sqrt{5}}{2}$. Thus we have:

$$\sigma_1^2 = \frac{3+\sqrt{5}}{2} \quad \text{and} \quad \sigma_2^2 = \frac{3-\sqrt{5}}{2}.$$

To check that $\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$, we will square the entries of the matrix Σ given above.

$$\left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{1+2\sqrt{5}+5}{4} = \frac{3+\sqrt{5}}{2}. \checkmark$$

$$\left(\frac{\sqrt{5}-1}{2}\right)^2 = \frac{5-2\sqrt{5}+1}{4} = \frac{3-\sqrt{5}}{2}. \checkmark$$

Problem 29.2: (6.7 #11.) Suppose A has orthogonal columns $\mathbf{w}_1$, $\mathbf{w}_2$, ..., $\mathbf{w}_n$ of lengths σ_1, σ_2, ..., σ_n. Calculate $A^T A$. What are U, Σ, and V in the SVD?

Solution: Since the columns of A are orthogonal, $A^T A$ is a diagonal matrix with entries ${\sigma_1}^2$, ..., ${\sigma_n}^2$. Since $A^T A = V\Sigma^2 V^T$, we find that Σ^2 is the matrix with diagonal entries ${\sigma_1}^2$, ...,${\sigma_n}^2$ and thus that Σ is the matrix with diagonal entries σ_1, ..., σ_n.

Referring again to the equation $A^T A = V\Sigma^2 V^T$, we conclude also that $V = I$.

The equation $A = U\Sigma V^T$ then tells us that U must be the matrix whose columns are $\dfrac{1}{\sigma_i}\mathbf{w}_i$.

Linear transformations and their matrices

In older linear algebra courses, linear transformations were introduced before matrices. This geometric approach to linear algebra initially avoids the need for coordinates. But eventually there must be coordinates and matrices when the need for computation arises.

Without coordinates (no matrix)

Example 1: Projection

We can describe a projection as a *linear transformation* T which takes every vector in $\mathbb{R}^2$ into another vector in $\mathbb{R}^2$. In other words,

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2.$$

The rule for this *mapping* is that every vector $\mathbf{v}$ is projected onto a vector $T(\mathbf{v})$ on the line of the projection. Projection is a linear transformation.

Definition of *linear*

A transformation T is *linear* if:

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$$

and

$$T(c\mathbf{v}) = cT(\mathbf{v})$$

for all vectors $\mathbf{v}$ and $\mathbf{w}$ and for all scalars c. Equivalently,

$$T(c\mathbf{v} + d\mathbf{w}) = cT(\mathbf{v}) + dT(\mathbf{w})$$

for all vectors $\mathbf{v}$ and $\mathbf{w}$ and scalars c and d. It's worth noticing that $T(\mathbf{0}) = \mathbf{0}$, because if not it couldn't be true that $T(c\mathbf{0}) = cT(\mathbf{0})$.

Non-example 1: Shift the whole plane

Consider the transformation $T(\mathbf{v}) = \mathbf{v} + \mathbf{v}_0$ that shifts every vector in the plane by adding some fixed vector $\mathbf{v}_0$ to it. This is *not* a linear transformation because $T(2\mathbf{v}) = 2\mathbf{v} + \mathbf{v}_0 \neq 2T(\mathbf{v})$.

Non-example 2: $T(\mathbf{v}) = ||\mathbf{v}||$

The transformation $T(\mathbf{v}) = ||\mathbf{v}||$ that takes any vector to its length is not a linear transformation because $T(c\mathbf{v}) \neq cT(\mathbf{v})$ if $c < 0$.

We're not going to study transformations that aren't linear. From here on, we'll only use T to stand for linear transformations.

Example 2: Rotation by $45°$

This transformation $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ takes an input vector $\mathbf{v}$ and outputs the vector $T(\mathbf{v})$ that comes from rotating $\mathbf{v}$ counterclockwise by $45°$ about the origin. Note that we can describe this and see that it's linear without using any coordinates.

The big picture

One advantage of describing transformations geometrically is that it helps us to see the big picture, as opposed to focusing on the transformation's effect on a single point. We can quickly see how rotation by $45°$ will transform a picture of a house in the plane. If the transformation was described in terms of a matrix rather than as a rotation, it would be harder to guess what the house would be mapped to.

Frequently, the best way to understand a linear transformation is to find the matrix that lies behind the transformation. To do this, we have to choose a basis and bring in coordinates.

With coordinates (matrix!)

All of the linear transformations we've discussed above can be described in terms of matrices. In a sense, linear transformations are an abstract description of multiplication by a matrix, as in the following example.

Example 3: $T(\mathbf{v}) = A\mathbf{v}$

Given a matrix A, define $T(\mathbf{v}) = A\mathbf{v}$. This is a linear transformation:

$$A(\mathbf{v} + \mathbf{w}) = A(\mathbf{v}) + A(\mathbf{w})$$

and

$$A(c\mathbf{v}) = cA(\mathbf{v}).$$

Example 4

Suppose $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. How would we describe the transformation $T(\mathbf{v}) = A\mathbf{v}$ geometrically?

When we multiply A by a vector $\mathbf{v}$ in $\mathbb{R}^2$, the x component of the vector is unchanged and the sign of the y component of the vector is reversed. The transformation $\mathbf{v} \mapsto A\mathbf{v}$ reflects the xy-plane across the x axis.

Example 5

How could we find a linear transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ that takes three dimensional space to two dimensional space? Choose any 2 by 3 matrix A and define $T(\mathbf{v}) = A\mathbf{v}$.

Describing $T(\mathbf{v})$

How much information do we need about T to to determine $T(\mathbf{v})$ for all $\mathbf{v}$? If we know how T transforms a single vector $\mathbf{v}_1$, we can use the fact that T is a linear transformation to calculate $T(c\mathbf{v}_1)$ for any scalar c. If we know $T(\mathbf{v}_1)$ and $T(\mathbf{v}_2)$ for two independent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, we can predict how T will transform any vector $c\mathbf{v}_1 + d\mathbf{v}_2$ in the plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. If we wish to know $T(\mathbf{v})$ for all vectors $\mathbf{v}$ in $\mathbb{R}^n$, we just need to know $T(\mathbf{v}_1), T(\mathbf{v}_2), ..., T(\mathbf{v}_n)$ for any basis $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ of the input space. This is because any $\mathbf{v}$ in the input space can be written as a linear combination of basis vectors, and we know that T is linear:

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \\ T(\mathbf{v}) &= c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n). \end{aligned}$$

This is how we get from a (coordinate-free) linear transformation to a (coordinate based) matrix; the c_i are our coordinates. Once we've chosen a basis, every vector $\mathbf{v}$ in the space can be written as a combination of basis vectors in exactly one way. The coefficients of those vectors are the *coordinates* of $\mathbf{v}$ in that basis.

Coordinates come from a basis; changing the basis changes the coordinates of vectors in the space. We may not use the standard basis all the time – we sometimes want to use a basis of eigenvectors or some other basis.

The matrix of a linear transformation

Given a linear transformation T, how do we construct a matrix A that represents it?

First, we have to choose two bases, say $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ of $\mathbb{R}^n$ to give coordinates to the input vectors and $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m$ of $\mathbb{R}^m$ to give coordinates to the output vectors. We want to find a matrix A so that $T(\mathbf{v}) = A\mathbf{v}$, where $\mathbf{v}$ and $A\mathbf{v}$ get their coordinates from these bases.

The first column of A consists of the coefficients $a_{11}, a_{21}, ..., a_{1m}$ of $T(\mathbf{v}_1) = a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + \cdots + a_{1m}\mathbf{w}_m$. The entries of column i of the matrix A are determined by $T(\mathbf{v}_i) = a_{1i}\mathbf{w}_1 + a_{2i}\mathbf{w}_2 + \cdots + a_{1i}\mathbf{w}_m$. Because we've guaranteed that $T(\mathbf{v}_i) = A\mathbf{v}_i$ for each basis vector $\mathbf{v}_i$ and because T is linear, we know that $T(\mathbf{v}) = A\mathbf{v}$ for all vectors $\mathbf{v}$ in the input space.

In the example of the projection matrix, $n = m = 2$. The transformation T projects every vector in the plane onto a line. In this example, it makes sense to use the same basis for the input and the output. To make our calculations as simple as possible, we'll choose $\mathbf{v}_1$ to be a unit vector on the line of projection and $\mathbf{v}_2$ to be a unit vector perpendicular to $\mathbf{v}_1$. Then

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{v}_1 + \mathbf{0}$$

and the matrix of the projection transformation is just $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

$$A\mathbf{v} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix}.$$

This is a nice matrix! If our chosen basis consists of eigenvectors then the matrix of the transformation will be the diagonal matrix Λ with eigenvalues on the diagonal.

To see how important the choice of basis is, let's use the standard basis for the linear transformation that projects the plane onto a line at a 45° angle. If we choose $\mathbf{v}_1 = \mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we get the projection matrix $P = \dfrac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$. We can check by graphing that this is the correct matrix, but calculating P directly is more difficult for this basis than it was with a basis of eigenvectors.

Example 6: $T = \dfrac{d}{dx}$

Let T be a transformation that takes the derivative:

$$T(c_1 + c_2x + c_3x^2) = c_2 + 2c_3x. \tag{1}$$

The input space is the three dimensional space of quadratic polynomials $c_1 + c_2x + c_3x^2$ with basis $\mathbf{v}_1 = 1$, $\mathbf{v}_2 = x$ and $\mathbf{v}_3 = x^2$. The output space is a two dimensional subspace of the input space with basis $\mathbf{w}_1 = \mathbf{v}_1 = 1$ and $\mathbf{w}_2 = \mathbf{v}_2 = x$.

This is a linear transformation! So we can find $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ and write the transformation (1) as a matrix multiplication (2):

$$T\left(\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}\right) = A \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} c_2 \\ 2c_3 \end{bmatrix}. \tag{2}$$

Conclusion

For any linear transformation T we can find a matrix A so that $T(\mathbf{v}) = A\mathbf{v}$. If the transformation is invertible, the inverse transformation has the matrix A^{-1}. The product of two transformations $T_1 : \mathbf{v} \mapsto A_1\mathbf{v}$ and $T_2 : \mathbf{w} \mapsto A_2\mathbf{w}$ corresponds to the product A_2A_1 of their matrices. This is where matrix multiplication came from!

Exercises on linear transformations and their matrices

Problem 30.1: Consider the transformation T that doubles the distance between each point and the origin without changing the direction from the origin to the points. In polar coordinates this is described by

$$T(r, \theta) = (2r, \theta).$$

a) Yes or no: is T a linear transformation?

b) Describe T using Cartesian (xy) coordinates. Check your work by confirming that the transformation doubles the lengths of vectors.

c) If your answer to (a) was "yes", find the matrix of T. If your answer to (a) was "no", explain why the T isn't linear.

Problem 30.2: Describe a transformation which leaves the zero vector fixed but which is not a linear transformation.

Exercises on linear transformations and their matrices

Problem 30.1: Consider the transformation T that doubles the distance between each point and the origin without changing the direction from the origin to the points. In polar coordinates this is described by

$$T(r,\theta) = (2r,\theta).$$

a) Yes or no: is T a linear transformation?

b) Describe T using Cartesian (xy) coordinates. Check your work by confirming that the transformation doubles the lengths of vectors.

c) If your answer to (a) was "yes", find the matrix of T. If your answer to (a) was "no", explain why the T isn't linear.

Solution:

a) Yes. In terms of vectors, $T(\mathbf{v}) = 2\mathbf{v}$, so $T(\mathbf{v}_1 + \mathbf{v}_2) = 2\mathbf{v}_1 + 2\mathbf{v}_2 = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ and $T(c\mathbf{v}) = 2c\mathbf{v} = cT(\mathbf{v})$.

b) $T(x,y) = (2x, 2y)$. We know $\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \sqrt{x^2+y^2}$ and can calculate

$$\left| T\left(\begin{bmatrix} x \\ y \end{bmatrix} \right) \right| = \left\| \begin{bmatrix} 2x \\ 2y \end{bmatrix} \right\| = \sqrt{4(x^2+y^2)} = 2\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|.$$

This confirms that T doubles the lengths of vectors.

c) We can use our answer to (b) to find that the matrix of T is $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

Problem 30.2: Describe a transformation which leaves the zero vector fixed but which is not a linear transformation.

Solution: If we limit ourselves to "simple" transformations, this is not an easy task!

One way to solve this is to find a transformation that acts differently on different parts of the plane. If $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ |y| \end{bmatrix}$ then

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

is not equal to

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) + T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

Another approach is to use a nonlinear function to define the transformation: if $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y^2 \end{bmatrix}$ then $T\left(c\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} cx \\ c^2y^2 \end{bmatrix}$ is not equal to $cT\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} cx \\ cy^2 \end{bmatrix}$.

Change of basis; image compression

We've learned that computations can be made easier by an appropriate choice of basis. One application of this principle is to image compression. Lecture videos, music, and other data sources contain a lot of information; that information can be efficiently stored and transmitted only after we change the basis used to record it.

Compression of images

Suppose one frame of our lecture video is 512 by 512 pixels and that the video is recorded in black and white. The camera records a brightness level for each of the $(512)^2$ pixels; in this sense, each frame of video is a vector in a $(512)^2$ dimensional vector space.

The standard basis for this space has a vector for each pixel. Transmitting the values of all $(512)^2$ components of each frame using the standard basis would require far too much bandwidth, but if we change our basis according to the JPEG image compression standard we can transmit a fairly good copy the video very efficiently.

For example, if we're reporting light levels pixel by pixel, there's no efficient way to transmit the information "the entire frame is black". However, if one of our basis vectors corresponds to all pixels having the same light level (say 1), we can very efficiently transmit a recording of a blank blackboard.

Along with a vector of all 1's, we might choose a basis vector that alternates 1's and -1's, or one that's half 1's and half -1's corresponding to an image that's bright on the left and dark on the right. Our choice of basis will directly affect how much data we need to download to watch a video, and the best choice of basis for algebra lectures might differ from the best choice for action movies!

Fourier basis vectors

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \\ \omega^6 \\ \omega^7 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega^2 \\ \omega^4 \\ \omega^6 \\ \omega^8 \\ \omega^{10} \\ \omega^{12} \\ \omega^{14} \end{bmatrix}, \ldots, \begin{bmatrix} 1 \\ \omega^7 \\ \omega^{14} \\ \omega^{21} \\ \omega^{28} \\ \omega^{35} \\ \omega^{42} \\ \omega^{49} \end{bmatrix}$$

The best known basis is the Fourier basis, which is closely related to the Fourier matrices we studied earlier. The basis used by JPEG is made up of cosines – the real parts of ω^{jk}.

This method breaks the 512 by 512 rectangle of pixels into blocks that are 8 pixels on a side, each block containing 64 pixels total. The brightness information for those pixels is then compressed, possibly by eliminating all coefficients below some threshold chosen so that we can hardly see the difference once they're gone.

$$\text{signal } \mathbf{x} \overset{\text{lossless}}{\longrightarrow} 64 \text{ coefficients } c \overset{\text{lossy compression}}{\longrightarrow} \hat{c} \text{ (many zeros)} \longrightarrow \hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$$

In video, not only should we consider compressing each frame, we can also consider compressing sequences of frames. There's very little difference between one frame and the next. If we do it right, we only need to encode and compress the differences between frames, not every frame in its entirety.

The Haar wavelet basis

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

The closest competitor to the JPEG encoding method uses a wavelet basis. (JPEG2000 improves on the Haar wavelets above.) In Haar's wavelet basis for $\mathbb{R}^8$, the non-zero entries are half 1's and half -1's (except for the vector of all 1's). However, half or even three quarters of a basis vector's entries may be 0. These vectors are chosen to be orthogonal and can be adjusted to be orthonormal.

Compression and matrices

Linear algebra is used to find the coefficients c_i in the change of basis from the standard basis (light levels for each pixel) to the Fourier or wavelet basis. For example, we might want to write:

$$\mathbf{x} = c_1\mathbf{w}_1 + \cdots + c_8\mathbf{w}_8.$$

But this is just a linear combination of the wavelet basis vectors. If W is the matrix whose columns are the wavelet vectors, then our task is simply to solve for $\mathbf{c}$:

$$\mathbf{x} = W \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_8 \end{bmatrix}.$$

So $\mathbf{c} = W^{-1}\mathbf{x}$.

Our calculations will be faster and easier if we don't have to spend a lot of time inverting a matrix (e.g. if $W^{-1} = W^T$) or multiplying by the inverse. So in the field of image compression, the criteria for a good basis are:

- Multiplication by the basis matrix and its inverse is fast (as in the FFT or in the wavelet basis).
- Good compression – the image can be approximated using only a few basis vectors. Most components c_i are small – safely set to zero.

Change of basis

Vectors

Let the columns of matrix W be the basis vectors of the new basis. Then if $\mathbf{x}$ is a vector in the old basis, we can convert it to a vector $\mathbf{c}$ in the new basis using the relation:

$$\mathbf{x} = W\mathbf{c}.$$

Transformation matrices

Suppose we have a linear transformation T. If T has the matrix A when working with the basis $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_8$ and T has the matrix B when working with the basis $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_8$, it turns out that A and B must be similar matrices. In other words, $B = M^{-1}AM$ for some change of basis matrix M.

Reminder: If we have a basis $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_8$ and we know $T(\mathbf{v}_i)$ for each i, then we can use the fact that T is a linear transformation to find $T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + c_2 T(\mathbf{v}_2) + \cdots + c_8 T(\mathbf{v}_8)$ for any vector $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_8\mathbf{v}_8$ in the space. The entries of column i of the matrix A are the coefficients of the output vector $T(\mathbf{v}_i)$.

If our basis consists of eigenvectors of our transformation, i.e. if $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$, then $A = \Lambda$, the (diagonal) matrix of eigenvalues. It would be wonderful to use a basis of eigenvectors for image processing, but finding such a basis requires far more computation than simply using a Fourier or wavelet basis.

Summary

When we change bases, the coefficients of our vectors change according to the rule $\mathbf{x} = W\mathbf{c}$. Matrix entries change according to a rule $B = M^{-1}AM$.

Exercises on change of basis; image compression

Problem 31.1: Verify that the vectors of the Haar wavelet basis, given in lecture, are orthogonal. Adjust their lengths so that the resulting basis vectors are orthonormal.

Problem 31.2: We can think of the set of all two by two matrices with real valued entries as a vector space. Describe two different bases for this space. Is one of your bases better than the other for describing diagonal matrices? What about triangular matrices? Symmetric matrices?

Exercises on change of basis; image compression

Problem 31.1: Verify that the vectors of the Haar wavelet basis, given in lecture, are orthogonal. Adjust their lengths so that the resulting basis vectors are orthonormal.

Solution: The vectors given for the Haar wavelet basis were:

$$\begin{bmatrix} 1\\1\\1\\1\\1\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\1\\-1\\-1\\-1\\-1 \end{bmatrix}, \begin{bmatrix} 1\\1\\-1\\-1\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\1\\1\\-1\\-1 \end{bmatrix}, \begin{bmatrix} 1\\-1\\0\\0\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\-1\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\1\\-1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\0\\0\\1\\-1 \end{bmatrix}.$$

It's easy to see that the second through eighth vectors are orthogonal to the first; their inner products are sums of equal numbers of ones and negative ones. The same thing happens when taking the dot products of the third through eighth vectors with the second. The remaining inner products are either sums of equal numbers of ones and negative ones or simply sums of zeros. Since all the pairwise inner products are zero, the vectors are orthogonal.

To make the basis orthonormal, divide by the lengths of the vectors:

$$\begin{bmatrix} 1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8}\\1/\sqrt{8}\\-1/\sqrt{8}\\-1/\sqrt{8}\\-1/\sqrt{8}\\-1/\sqrt{8} \end{bmatrix}, \begin{bmatrix} 1/2\\1/2\\-1/2\\-1/2\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\1/2\\1/2\\-1/2\\-1/2 \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2}\\-1/\sqrt{2}\\0\\0\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1/\sqrt{2}\\-1/\sqrt{2}\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\1/\sqrt{2}\\-1/\sqrt{2}\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\0\\0\\1/\sqrt{2}\\-1/\sqrt{2} \end{bmatrix}.$$

Problem 31.2: We can think of the set of all two by two matrices with real valued entries as a vector space. Describe two different bases for this space. Is one of your bases better than the other for describing diagonal matrices? What about triangular matrices? Symmetric matrices?

Solution: There are many different ways to answer this question correctly.
The most obvious choice of basis is:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hopefully, the second basis you chose differs from this one significantly. Here are some possibilities:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

or

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

If you have trouble checking that the matrices in your basis are independent, think of them as vectors in $\mathbb{R}^4$ and check that those vectors are independent. The examples above are easy to check because, when considered as vectors in $\mathbb{R}^4$, they are orthogonal.

Of these bases, the first may be best for describing diagonal matrices; the third is also reasonably good for that task:

$$\begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} = 3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The first basis also looks best for describing triangular matrices – any triangular matrix is a combination of only three of the four vectors. The third is best for describing symmetric matrices because symmetric matrices will always have a zero coefficient for the $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ component.

Left and right inverses; pseudoinverse

Although pseudoinverses will not appear on the exam, this lecture will help us to prepare.

Two sided inverse

A *2-sided inverse* of a matrix A is a matrix A^{-1} for which $AA^{-1} = I = A^{-1}A$. This is what we've called the *inverse* of A. Here $r = n = m$; the matrix A has full rank.

Left inverse

Recall that A has full column rank if its columns are independent; i.e. if $r = n$. In this case the nullspace of A contains just the zero vector. The equation $A\mathbf{x} = \mathbf{b}$ either has exactly one solution $\mathbf{x}$ or is not solvable.

The matrix A^TA is an invertible n by n symmetric matrix, so $(A^TA)^{-1}A^TA = I$. We say $A_{\text{left}}^{-1} = (A^TA)^{-1}A^T$ is a *left inverse* of A. (There may be other left inverses as well, but this is our favorite.) The fact that A^TA is invertible when A has full column rank was central to our discussion of least squares.

Note that AA_{left}^{-1} is an m by m matrix which only equals the identity if $m = n$. A rectangular matrix can't have a two sided inverse because either that matrix or its transpose has a nonzero nullspace.

Right inverse

If A has full row rank, then $r = m$. The nullspace of A^T contains only the zero vector; the rows of A are independent. The equation $A\mathbf{x} = \mathbf{b}$ always has at least one solution; the nullspace of A has dimension $n - m$, so there will be $n - m$ free variables and (if $n > m$) infinitely many solutions!

Matrices with full row rank have right inverses A_{right}^{-1} with $AA_{\text{right}}^{-1} = I$. The nicest one of these is $A^T(AA^T)^{-1}$. Check: A times $A^T(AAT)^{-1}$ is I.

Pseudoinverse

An invertible matrix ($r = m = n$) has only the zero vector in its nullspace and left nullspace. A matrix with full column rank $r = n$ has only the zero vector in its nullspace. A matrix with full row rank $r = m$ has only the zero vector in its left nullspace. The remaining case to consider is a matrix A for which $r < n$ and $r < m$.

If A has full column rank and $A_{\text{left}}^{-1} = (A^TA)^{-1}A^T$, then

$$AA_{\text{left}}^{-1} = A(A^TA)^{-1}A^T = P$$

is the matrix which projects $\mathbb{R}^m$ onto the column space of A. This is as close as we can get to the product $AM = I$.

Similarly, if A has full row rank then $A^{-1}_{\text{right}}A = A^T(AA^T)^{-1}A$ is the matrix which projects $\mathbb{R}^n$ onto the row space of A.

It's nontrivial nullspaces that cause trouble when we try to invert matrices. If $A\mathbf{x} = \mathbf{0}$ for some nonzero $\mathbf{x}$, then there's no hope of finding a matrix A^{-1} that will reverse this process to give $A^{-1}\mathbf{0} = \mathbf{x}$.

The vector $A\mathbf{x}$ is always in the column space of A. In fact, the correspondence between vectors $\mathbf{x}$ in the (r dimensional) row space and vectors $A\mathbf{x}$ in the (r dimensional) column space is one-to-one. In other words, if $\mathbf{x} \neq \mathbf{y}$ are vectors in the row space of A then $A\mathbf{x} \neq A\mathbf{y}$ in the column space of A. (The proof of this would make a good exam question.)

Proof that if $\mathbf{x} \neq \mathbf{y}$ then $A\mathbf{x} \neq A\mathbf{y}$

Suppose the statement is false. Then we can find $\mathbf{x} \neq \mathbf{y}$ in the row space of A for which $A\mathbf{x} = A\mathbf{y}$. But then $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$, so $\mathbf{x} - \mathbf{y}$ is in the nullspace of A. But the row space of A is closed under linear combinations (like subtraction), so $\mathbf{x} - \mathbf{y}$ is also in the row space. The only vector in both the nullspace and the row space is the zero vector, so $\mathbf{x} - \mathbf{y} = \mathbf{0}$. This contradicts our assumption that $\mathbf{x}$ and $\mathbf{y}$ are not equal to each other.

We conclude that the mapping $\mathbf{x} \mapsto A\mathbf{x}$ from row space to column space is invertible. The inverse of this operation is called the *pseudoinverse* and is very useful to statisticians in their work with linear regression – they might not be able to guarantee that their matrices have full column rank $r = n$.

Finding the pseudoinverse A^+

The *pseudoinverse* A^+ of A is the matrix for which $\mathbf{x} = A^+A\mathbf{x}$ for all $\mathbf{x}$ in the row space of A. The nullspace of A^+ is the nullspace of A^T.

We start from the singular value decomposition $A = U\Sigma V^T$. Recall that Σ is a m by n matrix whose entries are zero except for the singular values $\sigma_1, \sigma_2, ..., \sigma_r$ which appear on the diagonal of its first r rows. The matrices U and V are orthonormal and therefore easy to invert. We only need to find a pseudoinverse for Σ.

The closest we can get to an inverse for Σ is an n by m matrix Σ^+ whose first r rows have $1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_r$ on the diagonal. If $r = n = m$ then $\Sigma^+ = \Sigma^{-1}$. Always, the product of Σ and Σ^+ is a square matrix whose first r diagonal entries are 1 and whose other entries are 0.

If $A = U\Sigma V^T$ then its pseudoinverse is $A^+ = V\Sigma^+\mathbf{U}^T$. (Recall that $Q^T = Q^{-1}$ for orthogonal matrices U, V or Q.)

We would get a similar result if we included non-zero entries in the lower right corner of Σ^+, but we prefer not to have extra non-zero entries.

Conclusion

Although pseudoinverses will not appear on the exam, many of the topics we covered while discussing them (the four subspaces, the SVD, orthogonal matrices) are likely to appear.

Exercises on left and right inverses; pseudoinverse

Problem 32.1: Find a right inverse for $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Problem 32.2: Does the matrix $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$ have a left inverse? A right inverse? A pseudoinverse? If the answer to any of these questions is "yes", find the appropriate inverse.

Exercises on left and right inverses; pseudoinverse

Problem 32.1: Find a right inverse for $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Solution: We apply the formula $A^{-1}_{\text{right}} = A^T(AA^T)^{-1}$:

$$\begin{aligned} A^T &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \\ AA^T &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \\ (AA^T)^{-1} &= \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \\ A^T(AA^T)^{-1} &= \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \\ 1/2 & 0 \end{bmatrix}. \end{aligned}$$

Thus, $A^{-1}_{\text{right}} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \\ 1/2 & 0 \end{bmatrix}$ is one right inverse of A. We can quickly check that $AA^{-1}_{\text{right}} = I$.

Problem 32.2: Does the matrix $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$ have a left inverse? A right inverse? A pseudoinverse? If the answer to any of these questions is "yes", find the appropriate inverse.

Solution: The second row of A is a multiple of the first row, so A has rank 1 and $\det A = 0$. Because A is a square matrix its determinant is defined, and we can use the fact that $\det C \cdot \det D = \det(CD)$ to prove that A can't have a left or right inverse. (If $AB = I$, then $\det A \det B = \det I$ implies $0 = 1$.)

We *can* find a pseudoinverse $A^+ = V\Sigma^+U^T$ for A. We start by finding the singular value decomposition $U\Sigma V^T$ of A.

The SVD of A was calculated in the lecture on singular value decomposition, so we know that

$$\underbrace{\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}}_{A} = \underbrace{\tfrac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix}}_{V^T}.$$

Hence, $\Sigma^+ = \begin{bmatrix} 1/\sqrt{125} & 0 \\ 0 & 0 \end{bmatrix}$ and

$$\begin{aligned} A^+ &= V\Sigma^+U^T \\ &= \begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix} \begin{bmatrix} 1/\sqrt{125} & 0 \\ 0 & 0 \end{bmatrix} \left(\frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \right) \\ &= \begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix} \begin{bmatrix} 1/25 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \\ &= \begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix} \begin{bmatrix} 1/25 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \\ &= \frac{1}{125} \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix}. \end{aligned}$$

To check our work, we confirm that A^+ reverses the operation of A on its row space using the bases we found while computing its SVD. Recall that

$$A\mathbf{v}_j = \begin{cases} \sigma_j \mathbf{u}_j & \text{for } j \le r \\ \mathbf{0} & \text{for } j > r. \end{cases}$$

Here $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$ and $A^+\mathbf{u}_1 = \frac{1}{\sqrt{125}} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \frac{1}{\sigma_1}\mathbf{v}_1$. We can also check that $A^+\mathbf{u}_2 = \frac{1}{125} \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \mathbf{0}$.

Exam 3 review

The exam will cover the material through the singular value decomposition. Linear transformations and change of basis will be covered on the final.

The main topics on this exam are:

- Eigenvalues and eigenvectors
- Differential equations $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$ and exponentials e^{At}
- Symmetric matrices $A = A^T$: These always have real eigenvalues, and they always have "enough" eigenvectors. The eigenvector matrix Q can be an orthogonal matrix, with $A = Q\Lambda Q^T$.
- Positive definite matrices
- Similar matrices $B = M^{-1}AM$. Matrices A and B have the same eigenvalues; powers of A will "look like" powers of B.
- Singular value decomposition

Sample problems

1. This is a question about a differential equation with a skew symmetric matrix.

   Suppose

   $$\frac{d\mathbf{u}}{dt} = A\mathbf{u} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{u}.$$

   The general solution to this equation will look like

   $$\mathbf{u}(t) = c_1 e^{\lambda_1 t}\mathbf{x}_1 + c_2 e^{\lambda_2 t}\mathbf{x}_2 + c_e e^{\lambda_3 t}\mathbf{x}_3.$$

   a) What are the eigenvalues of A?

   The matrix A is singular; the first and third rows are dependent, so one eigenvalue is $\lambda_1 = 0$. We might also notice that A is antisymmetric ($A^T = -A$) and realize that its eigenvalues will be imaginary.

   To find the other two eigenvalues, we'll solve the equation $|A - \lambda I| = 0$.

   $$\begin{vmatrix} -\lambda & -1 & 0 \\ 1 & -\lambda & -1 \\ 0 & 1 & -\lambda \end{vmatrix} = -\lambda^3 - 2\lambda = 0.$$

   We conclude $\lambda_2 = \sqrt{2}\,i$ and $\lambda_3 = -\sqrt{2}\,i$.

At this point we know that our solution will look like:

$$\mathbf{u}(t) = c_1\mathbf{x}_1 + c_2 e^{\sqrt{2}\,it}\mathbf{x}_2 + c_e e^{-\sqrt{2}\,it}\mathbf{x}_3.$$

We can now see that the solution doesn't increase without bound or decay to zero. The size of $e^{i\theta}$ is the same for any θ; the exponentials here correspond to points on the unit circle.

b) The solution is periodic. When does it return to its original value? (What is its period?)

This is not likely to be on the exam, but we can quickly remark that $e^{\sqrt{2}\,it} = e^0$ when $\sqrt{2}t = 2\pi$, or when $t = \pi\sqrt{2}$.

c) Show that two eigenvectors of A are orthogonal.

The eigenvectors of a symmetric matrix or a skew symmetric matrix are always orthogonal. One choice of eigenvectors of A is:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ \mathbf{x}_2 = \begin{bmatrix} -1 \\ \sqrt{2}\,i \\ 1 \end{bmatrix},\ \mathbf{x}_3 = \begin{bmatrix} 1 \\ \sqrt{2}i \\ -1 \end{bmatrix}.$$

Don't forget to conjugate the first vector when computing the inner product of vectors with complex number entries.

d) The solution to this differential equation is $\mathbf{u}(t) = e^{At}\mathbf{u}(0)$. How would we compute e^{At}?

If $A = S\Lambda S^{-1}$ then $e^{At} = Se^{\Lambda t}S^{-1}$ where

$$e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix}.$$

So e^{At} comes from the eigenvalues in Λ and the eigenvectors in S.

Fact: A matrix has orthogonal eigenvectors exactly when $AA^T = A^TA$; i.e. when A commutes with its transpose. This is true of symmetric, skew symmetric and orthogonal matrices.

2. We're told that a three by three matrix A has eigenvalues $\lambda_1 = 0$, $\lambda_2 = c$ and $\lambda_3 = 2$ and eigenvectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

a) For which c is the matrix diagonalizable?

The matrix is diagonalizable if it has 3 independent eigenvectors. Not only are $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ independent, they're orthogonal. So the matrix is diagonalizable for all values of c.

b) For which values of c is the matrix symmetric?

If $A = Q\Lambda Q^T$ is symmetric its eigenvalues (the entries of Λ) are real. On the other hand, if c is real, then $A^T = Q^T \Lambda^T Q = A$ is symmetric. The matrix is symmetric for all real numbers c.

c) For which values of c is the matrix positive definite?

All positive definite matrices are symmetric, so c must be real. The eigenvalues of a positive definite matrix must be positive. The eigenvalue 0 is not positive, so this matrix is not positive definite for any values of c. (If $c \geq 0$ then the matrix is positive semidefinite.)

d) Is it a Markov matrix?

In a Markov matrix, one eigenvalue is 1 and the other eigenvalues are smaller than 1. Because $\lambda_3 = 2$, this cannot be a Markov matrix for any value of c.

e) Could $P = \frac{1}{2}A$ be a projection matrix?

Projection matrices are real and symmetric so their eigenvalues are real. In addition, we know that their eigenvalues are 1 and 0 because $P^2 = P$ implies $\lambda^2 = \lambda$. So $\frac{1}{2}A$ could be a projection matrix if $c = 0$ or $c = 2$.

Note that it was the fact that the eigenvectors were orthogonal that made it possible to answer many of these questions.

Singular value decomposition (SVD) is a factorization

$$A = \text{(orthogonal)(diagonal)(orthogonal)} = U\Sigma V^T.$$

We can do this for any matrix A. The key is to look at the symmetric matrix $A^T A = V\Sigma^T \Sigma V^T$; here V is the eigenvector matrix for $A^T A$ and $\Sigma^T\Sigma$ is the matrix of eigenvalues σ_i^2 of $A^T A$. Similarly, $AA^T = U\Sigma\Sigma^T U^T$ and U is the eigenvector matrix for AA^T. (Note that we can introduce a sign error if we're unlucky in choosing eigenvectors for the columns of U. To avoid this, use the formula $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ to calculate U from V.)

On the exam, you might be asked to find the SVD of a matrix A or you might be given information on U, Σ and V and asked about A.

3. Suppose $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ and U and V each have two columns.

a) What can we say about A?

We know A is a two by two matrix, and because U, V and Σ are all invertible we know A is nonsingular.

b) What if $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}$?

This is not a valid possibility for Σ. The singular values – the diagonal entries of Σ – are never negative in a singular value decomposition.

c) What if $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$?

Then A is a singular matrix of rank 1 and its nullspace has dimension 1. The four fundamental subspaces associated with A are spanned by orthonormal bases made up of selected columns of U and V. In this example, the second column of V is a basis for the nullspace of A.

4. We're told that A is symmetric and orthogonal.

a) What can we say about its eigenvalues?

The eigenvalues of symmetric matrices are real. The eigenvalues of orthogonal matrices Q have $|\lambda| = 1$; multiplication by an orthogonal matrix doesn't change the length of a vector. So the eigenvalues of A can only be 1 or -1.

b) True or false: A is sure to be positive definite.

False – it could have an eigenvalue of -1, as in $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

c) True or false: A has no repeated eigenvalues.

False – if A is a three by three matrix or larger, it's guaranteed to have repeated eigenvalues because every λ is 1 or -1.

d) Is A diagonalizable?

Yes, because all symmetric and all orthogonal matrices can be diagonalized. In fact, we can choose the eigenvectors of A to be orthogonal.

e) Is A nonsingular?

Yes; orthogonal matrices are all nonsingular.

f) Show $P = \frac{1}{2}(A + I)$ is a projection matrix.

We could check that P is symmetric and that $P^2 = P$:

$$P^2 = \left(\frac{1}{2}(A+I)\right)^2 = \frac{1}{4}(A^2 + 2A + I).$$

Because A is orthogonal and symmetric, $A^2 = A^T A = I$, so

$$P^2 = \frac{1}{4}(A^2 + 2A + I) = \frac{1}{2}(A + I) = P.$$

Or we could note that since the eigenvalues of A are 1 and -1 then the eigenvalues of $\frac{1}{2}(A + I)$ must be 1 and 0.

These questions all dealt with eigenvalues and special matrices; that's what the exam is about.

18.06SC Unit 3 Exam

1 (34 pts.) (a) If a square matrix A has all n of its *singular values* equal to 1 in the SVD, what basic classes of matrices does A belong to? (Singular, symmetric, orthogonal, positive definite or semidefinite, diagonal)

(b) Suppose the (orthonormal) columns of H are eigenvectors of B:

$$H = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \qquad H^{-1} = H^{\mathrm{T}}$$

The eigenvalues of B are $\lambda = 0, 1, 2, 3$. Write B as the product of 3 specific matrices. Write $C = (B + I)^{-1}$ as the product of 3 matrices.

(c) Using the list in question (a), which basic classes of matrices do B and C belong to? (Separate question for B and C)

2 (33 pts.) (a) Find three eigenvalues of A, and an eigenvector matrix S:

$$A = \begin{bmatrix} -1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 0 & 1 \end{bmatrix}$$

(b) Explain why $A^{1001} = A$. Is $A^{1000} = I$? Find the three diagonal entries of e^{At}.

(c) The matrix $A^{\mathrm{T}}A$ (for the same A) is

$$A^{\mathrm{T}}A = \begin{bmatrix} 1 & -2 & -4 \\ -2 & 4 & 8 \\ -4 & 8 & 42 \end{bmatrix}.$$

How many eigenvalues of $A^{\mathrm{T}}A$ are positive? zero? negative? (Don't compute them but explain your answer.) Does $A^{\mathrm{T}}A$ have the same eigenvectors as A?

3 (33 pts.) Suppose the n by n matrix A has n orthonormal eigenvectors $q_1, \ldots, q_n$ and n positive eigenvalues $\lambda_1, \ldots, \lambda_n$. Thus $Aq_j = \lambda_j q_j$.

(a) What are the eigenvalues and eigenvectors of A^{-1}? *Prove that your answer is correct.*

(b) Any vector b is a combination of the eigenvectors:

$$b = c_1 q_1 + c_2 q_2 + \cdots + c_n q_n \,.$$

What is a quick formula for c_1 using orthogonality of the q's?

(c) The solution to $Ax = b$ is also a combination of the eigenvectors:

$$A^{-1} b = d_1 q_1 + d_2 q_2 + \cdots + d_n q_n \,.$$

What is a quick formula for d_1? You can use the c's even if you didn't answer part (b).

18.06SC Unit 3 Exam Solutions

1 (34 pts.) (a) If a square matrix A has all n of its *singular values* equal to 1 in the SVD, what basic classes of matrices does A belong to? (Singular, symmetric, orthogonal, positive definite or semidefinite, diagonal)

(b) Suppose the (orthonormal) columns of H are eigenvectors of B:

$$H = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \qquad H^{-1} = H^{\mathrm{T}}$$

The eigenvalues of B are $\lambda = 0, 1, 2, 3$. Write B as the product of 3 specific matrices. Write $C = (B + I)^{-1}$ as the product of 3 matrices.

(c) Using the list in question (a), which basic classes of matrices do B and C belong to? (Separate question for B and C)

Solution.

(a) If $\sigma = I$ then $A = UV^{\mathrm{T}}$ = product of orthogonal matrices = orthogonal matrix.

2nd proof: All $\sigma_i = 1$ implies $A^{\mathrm{T}}A = I$. So A is orthogonal.

(A is *never* singular, and it won't always be symmetric — take $U = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $V = I$, for example. This also shows it can't be diagonal, or positive definite or semidefinite.)

(b) $B = H\Lambda H^{-1}$ with $\Lambda = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 2 & \\ & & & 3 \end{bmatrix}$

$(B+I)^{-1} = H(\Lambda+I)^{-1}H^{-1}$ with (same eigenvectors) $(\Lambda+I)^{-1} = \begin{bmatrix} 1 & & & \\ & 1/2 & & \\ & & 1/3 & \\ & & & 1/4 \end{bmatrix}$

(c) B is singular, symmetric, positive semidefinite.

C is symmetric positive definite.

2 (33 pts.) (a) Find three eigenvalues of A, and an eigenvector matrix S:

$$A = \begin{bmatrix} -1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 0 & 1 \end{bmatrix}$$

(b) Explain why $A^{1001} = A$. Is $A^{1000} = I$? Find the three diagonal entries of e^{At}.

(c) The matrix $A^{\mathrm{T}}A$ (for the same A) is

$$A^{\mathrm{T}}A = \begin{bmatrix} 1 & -2 & -4 \\ -2 & 4 & 8 \\ -4 & 8 & 42 \end{bmatrix}.$$

How many eigenvalues of $A^{\mathrm{T}}A$ are positive? zero? negative? (Don't compute them but explain your answer.) Does $A^{\mathrm{T}}A$ have the same eigenvectors as A?

Solution.

(a) The eigenvalues are $-1, 0, 1$ since A is triangular.

$$\lambda = -1 \text{ has } x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \lambda = 0 \text{ has } x = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \qquad \lambda = 1 \text{ has } x = \begin{bmatrix} 7 \\ 5 \\ 1 \end{bmatrix}.$$

Those vectors x are the columns of S (upper triangular!).

(b) $A = \bullet \Lambda S^{-1}$ and $A^{1001} = S\Lambda^{1001}S^{-1}$. Notice $\Lambda^{1001} = \Lambda$, $A^{1000} \neq I$ (A is singular) $(0^{1000} = 0 \neq 1)$.

e^{At} has e^{-1t}, $e^{0t} = 1$, e^{t} on its diagonal. *Proof using series:*

$\sum_0^\infty (At)^n/n!$ has triangular matrices so the diagonal has $\sum(-t)^n/n! = e^{-t}$, $\sum 0^n/n! = 1$, $\sum t^n/n! = e^t$.

Proof using $S\Lambda S^{-1}$:

$$e^{At} = Se^{\Lambda t}S^{-1} = \begin{bmatrix} 1 & \times & \times \\ 0 & 1 & \times \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^{-t} & & \\ & 1 & \\ & & e^{t} \end{bmatrix} \begin{bmatrix} 1 & \times & \times \\ 0 & 1 & \times \\ 0 & 0 & 1 \end{bmatrix}.$$

(c) $A^{\mathrm{T}}A$ has 2 positive eigenvalues (it has rank 2, its eigenvalues can never be negative).

One eigenvalue is zero because $A^{\mathrm{T}}A$ is singular. And $3 - 2 = 1$.

(Or: $A^{\mathrm{T}}A$ is symmetric, so the eigenvalues have the same signs as the pivots.

Do elimination: the pivots are 1, 0, and $42 - 16 = 26$.)

3 (33 pts.) Suppose the n by n matrix A has n orthonormal eigenvectors $q_1, \ldots, q_n$ and n positive eigenvalues $\lambda_1, \ldots, \lambda_n$. Thus $Aq_j = \lambda_j q_j$.

(a) What are the eigenvalues and eigenvectors of A^{-1}? *Prove that your answer is correct.*

(b) Any vector b is a combination of the eigenvectors:

$$b = c_1 q_1 + c_2 q_2 + \cdots + c_n q_n \,.$$

What is a quick formula for c_1 using orthogonality of the q's?

(c) The solution to $Ax = b$ is also a combination of the eigenvectors:

$$A^{-1}b = d_1 q_1 + d_2 q_2 + \cdots + d_n q_n \,.$$

What is a quick formula for d_1? You can use the c's even if you didn't answer part (b).

Solution.

(a) A^{-1} has eigenvalues $\dfrac{1}{\lambda_j}$ with the same eigenvectors

$$Aq_j = \lambda_j q_j \longrightarrow q_j = \lambda_j A^{-1} q_j \longrightarrow A^{-1} q_j = \frac{1}{\lambda_j}\, q_j\,.$$

(b) Multiply $b = c_1 q_1 + \cdots + c_n q_n$ by q_1^{T}.

Orthogonality gives $q_1^{\mathrm{T}} b = c_1 q_1^{\mathrm{T}} q_1$ so $c_1 = \dfrac{q_1^{\mathrm{T}} b}{q_1^{\mathrm{T}} q_1} = q_1^{\mathrm{T}} b$.

(c) Multiplying b by A^{-1} will multiply each q_i by $\dfrac{1}{\lambda_i}$ (part (a)). So c_i becomes

$$d_1 = \frac{c_1}{\lambda_1} \quad \left(= \frac{q_1^{\mathrm{T}} b}{\lambda_1 q_1^{\mathrm{T}} q_1} \text{ or } \frac{q_1^{\mathrm{T}} b}{\lambda_1} \right).$$