Princeton University MAT 202 Spring 2008

Willie W. Wong

Mar 24 - 28

Mar 24, 2008

- Motivation: remember a side remark I made the week before midterms: dot-products, in general, should only be evaluated in the standard basis[1]. Today we look at other bases in which the dot-product can be evaluated with no problem.
- Definition: a linear transformation $T : \mathbb{R}^M \to \mathbb{R}^M$ is said to be *orthogonal* if it preserves dot products. In mathematical notation:

  $$T(\boldsymbol{v}) \cdot T(\boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{w}$$

  for all vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^M$.
- (Equivalently, as given by your textbook, a linear transformation is orthogonal if it preserves lengths of vectors:

  $$|T(\boldsymbol{v})| = |\boldsymbol{v}|$$

  for all vectors $\boldsymbol{v}$. To see that my definition implies that of the textbook, it only suffices to remember $|\boldsymbol{v}| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$; to see that the textbook definition implies mine, the proof is slightly more tricky, it involves the use of the Pythagorean theorem, basically we can follow the textbook's derivation that orthogonal transformations preserve orthogonality [Fact 5.3.2], and expand the dot-product using the Pythagorean theorem.)
- (My definition is in many ways superior: it is inherently equivalent to the textbook definition and many facts requires less "proving".)

[1] By evaluating the dot-product, I mean that given two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, and a basis $\mathcal{B}$, in which $[\boldsymbol{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_j \end{bmatrix}$ and $[\boldsymbol{w}]_{\mathcal{B}} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_j \end{bmatrix}$, forming the number $c_1 d_1 + c_2 d_2 + \cdots c_j d_j$. In standard basis, this number is the dot product. In other basis, often this number is not quite as meaningful.

- A square matrix A is orthogonal if its associated transformation T is an orthogonal transformation.

- Example: let T be the rotation in $\mathbb{R}^2$ by an angle θ. Imaging you have two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, if you rotate both of them by the same angle, their *relative* angle is not changed, and their lengths are not changed: therefore their dot-product remains the same. So a rotation transformation is orthogonal.

- Example: Let V be a 2-dimensional subspace of $\mathbb{R}^3$ and T is the reflection about V. Then T is an orthogonal transformation. Consider two vectors $\boldsymbol{v}, \boldsymbol{w}$ in V: since T keep them fixed, their inner product is preserved. For two arbitrary vectors $\boldsymbol{x}, \boldsymbol{y}$, we decompose them as the parallel part and the orthogonal parts, so by definition

  $$\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^{\parallel} \cdot \boldsymbol{y}^{\parallel} + \boldsymbol{x}^{\perp} \cdot \boldsymbol{y}^{\perp}$$

  Remember that

  $$T(\boldsymbol{x}) = \boldsymbol{x}^{\parallel} - \boldsymbol{x}^{\perp}$$

  we write

  $$T(\boldsymbol{x}) \cdot T(\boldsymbol{y}) = \boldsymbol{x}^{\parallel} \cdot \boldsymbol{y}^{\parallel} + (-\boldsymbol{x}^{\perp}) \cdot (-\boldsymbol{y}^{\perp}) = \boldsymbol{x} \cdot \boldsymbol{y}$$

- In the above example, we relied on/demonstrated a crucial aspect of orthogonal transformations: they preserve orthogonality. In other words: if $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal, then $T(\boldsymbol{v})$ and $T(\boldsymbol{w})$ are orthogonal. This is rather obvious as T by definition preserves dot products, and the dot products of orthogonal vectors are 0.

- Using the fact that the orthogonal transformation $T : \mathbb{R}^M \to \mathbb{R}^M$ preserves dot products, we see that, given an orthonormal basis $\boldsymbol{u}_1, \ldots \boldsymbol{u}_M$ (say the standard basis for example), then $T(\boldsymbol{u}_1), \ldots, T(\boldsymbol{u}_M)$ is also an orthonormal basis: we have by assumption

  $$\boldsymbol{u}_i \cdot \boldsymbol{u}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

  So since T is orthogonal

  $$T(\boldsymbol{u}_i) \cdot T(\boldsymbol{u}_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

- A consequence of this is that: a square matrix A is orthogonal only when its column vectors form an orthonormal basis.

- Since our definition of orthogonal matrices are based on orthonormal transformations, we also have

  - The product of two orthogonal matrices is orthogonal. This fact follows because the product of two matrices is the matrix for the composition of two transformations. And the composition will preserve dot-products when the two intermediate transformations both do.

- Similarly, the inverse of an orthogonal matrix is orthogonal. All orthogonal matrices are invertible: since T preserves dot-products, T cannot have a non-trivial kernel, and so its nullity is zero. So the rank of an orthogonal $T : \mathbb{R}^M \to \mathbb{R}^M$ must by M, and T is invertible.

- Before continuing, we'll finally introduce the useful concept known as the transpose. Let A be an $M \times N$ matrix. We say that B is the *transpose* of A (written $B = A^T$) when

  - B is an $N \times M$ matrix
  - The entries b_{ij} (the ith row and jth column of B) is equal to a_{ji} (the jth row and ith column of A).

- Basically, what is a row in A is now a column in A^T and vice versa.

- We call a matrix *symmetric* if $A = A^T$. We say it is *skew-symmetric* if $A = -A^T$. Notice that in both this cases it is implicit that A is square. Give general forms.

- Some properties of transposes

  - $(AB)^T = B^T A^T$. The entry in the ith row and jth column of $(AB)^T$ is the entry in the jth row and the ith column of AB, which is the dot product of the jth row vector of A against the ith column vector of B. Similarly, the entry in the ith row and jth column of $B^T A^T$ is the inner product of the ith row of B^T against the jth column of A^T which is the same as the inner product of the ith column of B against jth row of A.
  - If A is invertible, then so is A^T, and

    $$(A^T)^{-1} = (A^{-1})^T$$

    This comes from the previous property, and the fact that $I_M^T = I_M$.
  - The rank of A is the same as rank of A^T. The proof is slightly involved. Let A be $N \times M$. Let V be the subspace spanned by the column vectors of A, so

    $$\dim(V) = rank(A)$$

    Let $V^\perp$ be the orthogonal complement to V. Notice that

    $$\dim(V^\perp) + \dim(V) = N$$

    Then for all $\boldsymbol{u} \in V^\perp$ and all $\boldsymbol{x} \in \mathbb{R}^M$, we have

    $$0 = \boldsymbol{u} \cdot A\boldsymbol{x} = \boldsymbol{u}^T A\boldsymbol{x}$$

    (see below... drat, I wrote this in the wrong order). We can take the transpose of that equation and arrive at (since the transpose of a scalar is still the same scalar)

    $$0 = \boldsymbol{x}^T A^T \boldsymbol{u}$$

for all $\boldsymbol{x}$ in $\mathbb{R}^M$. This is in particular true if we substitute for $\boldsymbol{x}$ the standard basis vectors. This means that $A^T\boldsymbol{u} = 0$. So $\boldsymbol{u}$ is in the kernel of A^T. Similarly, you can show that as long as $\boldsymbol{v} \in V$, there exist some $\boldsymbol{x}$ (in particular, an $\boldsymbol{x}$ that solves $A\boldsymbol{x} = \boldsymbol{v}$) such that

$$0 \neq \boldsymbol{x}^T A^T \boldsymbol{v}$$

and so any vector in V is not in the kernel of A^T. This implies that the kernel of A^T is exactly $V^\perp$. And so the nullity of A^T is $N - \dim(V)$. Lastly, applying the Rank-Nullity theorem we have that

$$rank(A^T) + (N - \dim(V)) = N$$

so

$$rank(A^T) = rank(A)$$

- The notion of transpose allows us to write dot-products as follows:

$$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^T \boldsymbol{w}$$

where on the right hand side we use matrix multiplication:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_M \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix}$$

- From here we have another characterization of the orthogonal matrices: by definition

$$T(\boldsymbol{v}) \cdot T(\boldsymbol{w}) = (A\boldsymbol{v})^T (A\boldsymbol{w}) = \boldsymbol{v}^T A^T A \boldsymbol{w}$$

So if $A^T A = I_M$, then

$$T(\boldsymbol{v}) \cdot T(\boldsymbol{w}) = \boldsymbol{v}^T I_M \boldsymbol{w} = \boldsymbol{v} \cdot \boldsymbol{w}$$

In fact, using the property that the column vectors of A need to form an orthonormal basis, we can also show that $A^T A = I_M$ whenever A is orthogonal. (Check that this is true for rotation, reflection in 2D.)

- A summary: The following statements are equivalent for $M \times M$ matrix A
  - A is orthogonal
  - The linear transformation associated to A is length-preserving
  - The linear transformation associated to A is dot-product preserving
  - The columns of A form an orthonormal basis
  - $A^T A = I_M$; in other words, $A^T = A^{-1}$

- An application of the transpose matrix is to give us a good way of writing the matrix corresponding to an projection: let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P$ be an orthonormal basis of a subspace V in $\mathbb{R}^M$. Then the projection onto V is

$$\underset{V}{proj}(\boldsymbol{x}) = \boldsymbol{u}_1(\boldsymbol{u}_1 \cdot \boldsymbol{x}) + \cdots + \boldsymbol{u}_P(\boldsymbol{u}_P \cdot \boldsymbol{x}) = U\boldsymbol{y}$$

where U is the matrix

$$\begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_P \end{bmatrix}$$

and $\boldsymbol{y}$ is the vector

$$\begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{x} \\ \boldsymbol{u}_2 \cdot \boldsymbol{x} \\ \vdots \\ \boldsymbol{u}_P \cdot \boldsymbol{x} \end{bmatrix}$$

Using the definition of matrix multiplication, we see that $\boldsymbol{y}$ is the result of the product $B\boldsymbol{x}$, where the row vectors of B are the vectors $\boldsymbol{u}$s. *In other word, the matrix B is just like the matrix U except the rows and columns are exchanged.* So $B = U^T$. Putting it all together

$$\underset{V}{proj}(\boldsymbol{x}) = UU^T\boldsymbol{x}$$

where U is the $M \times P$ matrix with column vectors given by the basis vectors $\boldsymbol{u}$s. (This is a generalization of the form of the matrix for projection onto a line given in Week 2.)

MAR 26, 2008

- On Monday, when showing that rank(A) =rank(A^T), we've effectively shown the following fact: *The orthogonal complement of the image of A is the kernel of the transpose of A.* In other words:

$$(im(A))^\perp = \ker(A^T)$$

With a little of work, we have the following useful facts:

  – $\ker(A) = \ker(A^TA)$: Suppose A is $N \times M$, then A^TA is a transformation from $\mathbb{R}^M \to \mathbb{R}^M$ that factors through $\mathbb{R}^N$. There are two ways that a vector $\boldsymbol{v}$ can live in the kernel of A^TA: either $\boldsymbol{v}$ is in $\ker(A)$, or $A\boldsymbol{v}$ is a non-zero vector that lives in $\ker(A^T)$. But using the orthogonality of $\ker(A^T)$ and $im(A)$, we conclude that the latter is impossible, so that $\boldsymbol{v}$ must live in $\ker(A)$.

  – If $\ker(A)$ is trivial, then the square matrix A^TA is invertible. (This fact follows directly from above.)

- A geometric interpretation of the projection: let V be a subspace and $\boldsymbol{x}$ a vector, then the vector $\boldsymbol{x}^* = proj_V(\boldsymbol{x})$ is *the vector in V that is closest to x*; in other words, for any other vector $\boldsymbol{y}$ in V,

$$|\boldsymbol{x} - \boldsymbol{x}^*| < |\boldsymbol{x} - \boldsymbol{y}|$$

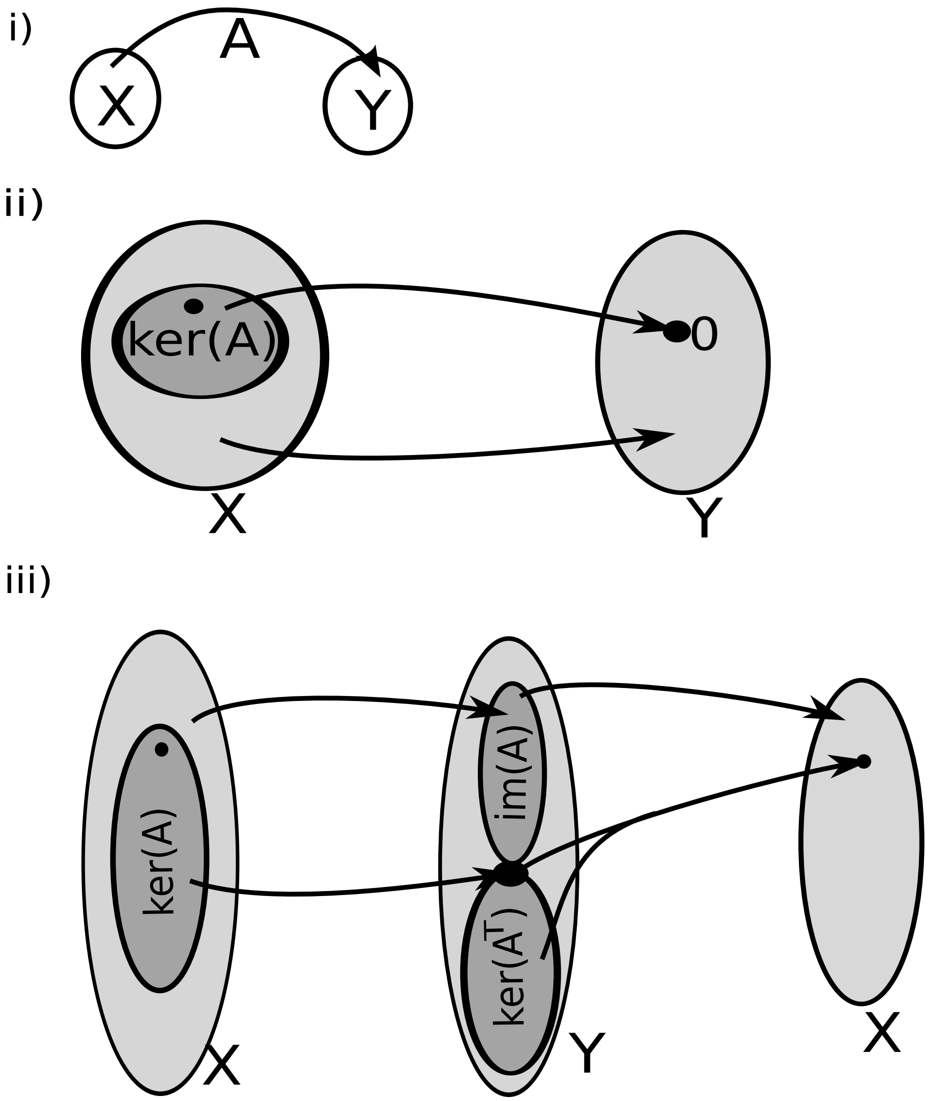

Figure 1: i) Let A be a linear map from $X \rightarrow Y$ ii) We can split X into two parts: $\ker(A)$ and everything that is not in the kernel. As shown in the picture, A takes $\ker(A)$ (including the 0 vector in X, represented by the dot) to the 0 vector in Y, and takes everything else to something that is non-zero iii) Now we look at the case for A^TA. $A : X \rightarrow Y$ and $A^T : Y \rightarrow X$, for each we split the domain up into the kernel and everything else. Since we've established that the image of A is the orthogonal complement of the $\ker(A^T)$, we show them as two different subsets intersecting only at the 0 vector. Backtrekking from the 0 on the far right, we see that only vectors in $\ker(A)$ can be in $\ker(A^TA)$.

This is because the endpoints of the vectors $\boldsymbol{x}, \boldsymbol{x}^*$ and $\boldsymbol{y}$ form a triangle with sides $\boldsymbol{x} - \boldsymbol{x}^*$, $\boldsymbol{x} - \boldsymbol{y}$, and $\boldsymbol{y} - \boldsymbol{x}^*$. Since $\boldsymbol{y}$ and $\boldsymbol{x}^*$ are both in a subspace V, their difference is in V, and so $\boldsymbol{y} - \boldsymbol{x}^* \perp \boldsymbol{x} - \boldsymbol{x}^*$, meaning that the triangle formed is a right triangle, with $\boldsymbol{x} - \boldsymbol{y}$ as the hypotenuse. And we all know that for a right triangle the hypotenuse is longer than either legs.

- This geometric interpretation allows us to introduce one of the most important applications of linear algebra: *The Method of Least Squares*. (The name "least squares" comes from the fact that when you measure the distance in Cartesian/standard coördinates, you sum the square of the coördinate values.)

  – Let's motivate our discussion with an example first: it is well known that from theoretical considerations, the motions of planets/comets/asteroids/other bodies orbiting the sun follow a trajectory described by a conic section (a circle, an ellipse, a parabola, or a hyperbola) with the sun being one of the foci. When a solar-system object is first observed, we try to determine its trajectory (if the trajectory is a circle or an ellipse, the object is bound to the solar system; if the trajectory is a parabola or a hyperbola, the object is transient: it came from outer-space and it will again fly out to out-space after this fly-by).

  – The general equation for a conic section (with one focal point at the origin) in a plane can be written in polar coördinates (r, θ), where r is the distance to the origin and θ is the angle measured counter-clockwise from the x-axis. The equation is

  $$r = \frac{l}{1 + c\cos\theta - d\sin\theta}$$

  with three parameters l, c, d. l is known as the semi-latus rectum. The number $e = \sqrt{c^2 + d^2}$ is the eccentricity ($e = 0$ for circles, $0 < e < 1$ for ellipses, $e = 1$ for a parabola, and $e > 1$ for hyperbolas), and $\phi = \sin^{-1}(c/e)$ is the angle of deviation between the major axis of the conic section and the x-axis of the plane.

  – The equation of the conic section can be re-written as

  $$r = l - cr\cos\theta + dr\sin\theta$$

  To determine the trajectory of a celestial object, it suffices that we determine the three parameters l, c, d based on measurements of r and θ. Since we take our coördinate system such that the origin is a focal point, we can take the origin to be the sun (a little bit of heliocentrism here; see figure below).

  – To determine l, c, d, we measure the position of the celestial object at some different times. Since the object is known to follow a trajectory that satisfies an equation of the given form, the measurements give us a system of linear equations for the coefficients

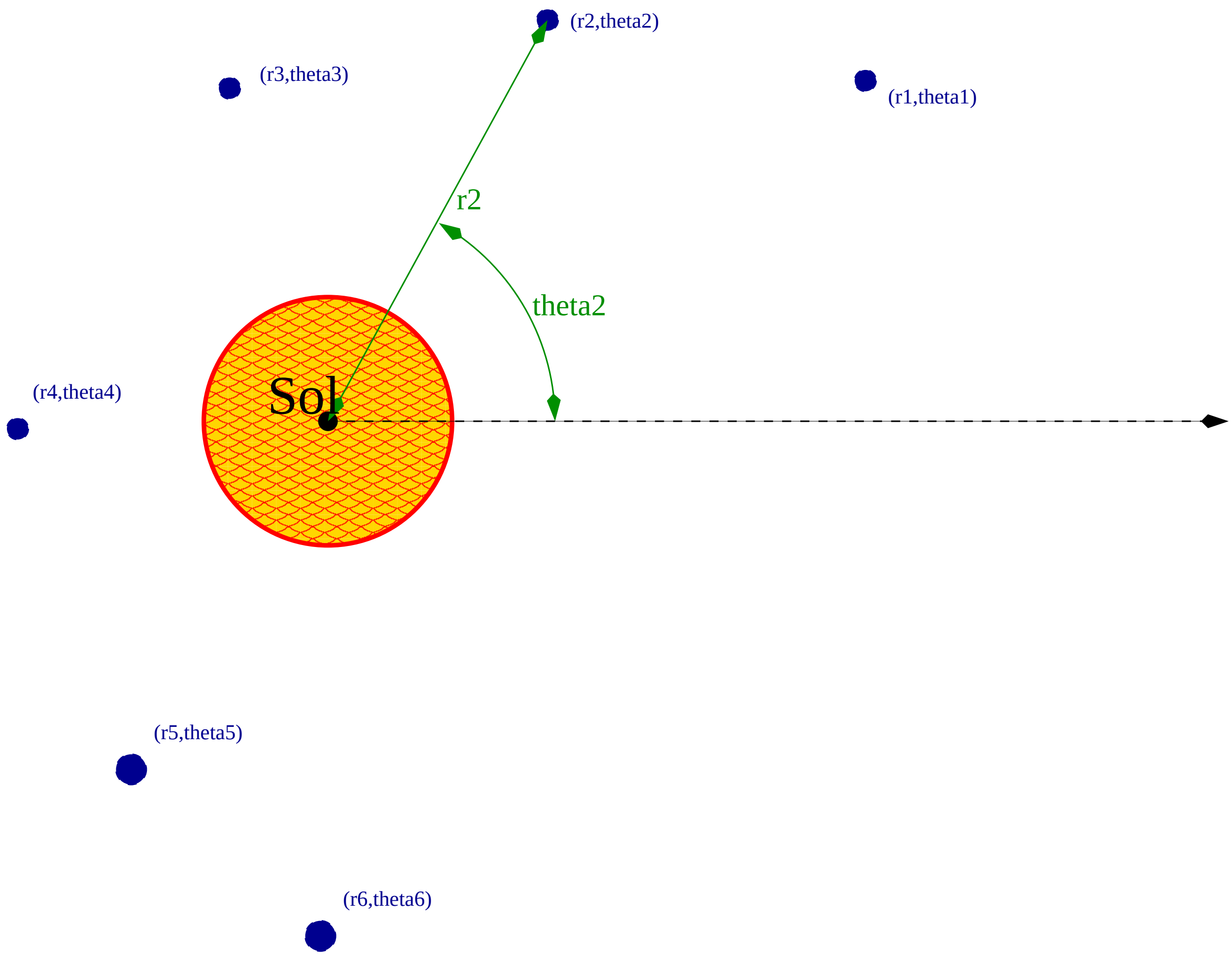

Figure 2: The sun (Sol) is the center of our coördinate system here. The black arrow toward the right denote the x-axis. The blue blob represent some asteroid/comet/something that is flying around. At six different times an astronomer made the measurements (in real life it is not so simple to measure the distance of an object to the sun, but we'll ignore that bit of technicality here) of r, the distance to the sun, and θ, the angle relative to the chosen axis. We can use this to calculate whether the object is in orbit or is just flying-by.

l, c, d:

$$\begin{cases} l - cr_1 \cos\theta_1 + dr_1 \sin\theta_1 & = r_1 \\ l - cr_2 \cos\theta_2 + dr_2 \sin\theta_2 & = r_2 \\ \vdots & \\ l - cr_6 \cos\theta_6 + dr_6 \sin\theta_6 & = r_6 \end{cases}$$

which we can write in matrix form

$$A\boldsymbol{v} = \begin{bmatrix} 1 & -r_1 \cos\theta_1 & r_1 \sin\theta_1 \\ 1 & -r_2 \cos\theta_2 & r_2 \sin\theta_2 \\ & \vdots & \\ 1 & -r_6 \cos\theta_6 & r_6 \sin\theta_6 \end{bmatrix} \begin{bmatrix} l \\ c \\ d \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_6 \end{bmatrix} = \boldsymbol{b}$$

where A is a $P \times 3$ matrix, P being the number of measurements taken.

- A first thing to notice: if rank of A is equal to P, then the system will have at least 1 solution; and if rank is equal to 3, then the solution will be unique. For this system, a simple bit of arithmetic shows that if $P \leq 3$, the system cannot be inconsistent: if $P \leq 3$, and the reduced form of A has a row of zeroes, the corresponding entry in $\boldsymbol{b}$ must also be zero. Thus a property common to this type of "curve-fitting" problems[2]: if you take fewer measurements than there are parameters, you can always find a curve that fits. In other words, to be physically meaningful, one needs more measurements than there are parameters.
- Now, if the system $A\boldsymbol{v} = \boldsymbol{b}$ is consistent. Then we can solve it and be done with the problem.
- But what happens if we make, say, 6 measurements, and the system turns out to be inconsistent? This means that our theory (that the trajectory follows a conic section) and/or our data (the physical measurements) may be incorrect. It is possible that the theory is only an approximation (in our case, the conic section trajectory is based on Newtonian gravity, which has been superseded by general relativity in the past century), and that the experimenter is not exactly precise and/or accurate in reporting the data (to err is human). But when we plot the 6 points on a graph, they "look like" they belong in some conic section orbit: here is the power of the method of least squares.
- Instead of trying to find an exact solution to the problem, we look for parameter values for l, c, d such that the curve approximates the data as best as possible. Here we go back to the geometric interpretation of the projection. In our system $A\boldsymbol{v} = \boldsymbol{b}$, A is a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^P$ (in our example with 6 measurement, $P = 6$). So the image of A is a subspace W of $\mathbb{R}^P$. In order to find the best approximation $\boldsymbol{v}$, we first find the point $\boldsymbol{w} \in W$ such that $\boldsymbol{w}$ is *as close to* $\boldsymbol{b}$ *as possible*. Since $\boldsymbol{w}$ is in the image of A, we can solve for some $\boldsymbol{v}$ with $A\boldsymbol{v} = \boldsymbol{w}$.

[2] Namely, given a known parametrized form of a curve, and given several points on the curve, we are asked to find the parameters that realizes the curve.

– What is $\boldsymbol{w}$ then? By the geometric interpretation, $\boldsymbol{w} = proj_W(\boldsymbol{b})$. This is the method of least squares: *Given a system of equations $A\boldsymbol{v} = \boldsymbol{b}$ where A is an $N \times M$ matrix. Let W denote the image set of A in $\mathbb{R}^N$. Then the* least squares solution *to the system is the vector $\boldsymbol{v}$ satisfying*

$$A\boldsymbol{v} = \underset{W}{proj}(\boldsymbol{b})$$

in particular, if the system $A\boldsymbol{v} = \boldsymbol{b}$ is consistent, then $proj_W(\boldsymbol{b}) = \boldsymbol{b}$, and $\boldsymbol{v}$ is an exact solution. Furthermore, the least squares solution $\boldsymbol{v}$ satisfies the property

$$|\boldsymbol{b} - A\boldsymbol{v}| \leq |\boldsymbol{b} - A\boldsymbol{y}|$$

for any other vector $\boldsymbol{y} \in \mathbb{R}^M$.

- The problem is, finding $proj_W(\boldsymbol{b})$ is hard: it involves taking the linear span of the column vectors of A, using Grahm-Schmidt process to extract a set of orthonormal basis, and from there calculating the projection matrix and then the projected vector. We want something that's easier for calculation.

- For that, we go back to our useful fact given in the beginning of class. Let $\boldsymbol{v}$ be the least squares solution to $A\boldsymbol{v} = \boldsymbol{b}$. Then by definition,

$$\boldsymbol{b} - A\boldsymbol{v} = \boldsymbol{b} - \boldsymbol{b}^{\parallel} = \boldsymbol{b}^{\perp}$$

  So $\boldsymbol{b} - A\boldsymbol{v}$ belongs in $W^{\perp} = (im(A))^{\perp}$ which, by our useful fact, $W^{\perp} = \ker(A^T)$. This means that the least squares solution $\boldsymbol{v}$ also satisfies

$$0 = A^T(\boldsymbol{b} - A\boldsymbol{v}) \Rightarrow A^T\boldsymbol{b} = A^TA\boldsymbol{v}$$

  and so we have

- *Revised method of least squares:* given the system $A\boldsymbol{v} = \boldsymbol{b}$, the approximation given by the least squares solution $\boldsymbol{v}$ can be found by solving

$$A^TA\boldsymbol{v} = A^T\boldsymbol{b}$$

- The system $A^TA\boldsymbol{v} = A^T\boldsymbol{b}$ is, by construction, consistent, and we call it the *normal equation* of $A\boldsymbol{v} = \boldsymbol{b}$.

- (This is a big improvement: taking the transpose of a matrix involves almost zero computation, and multiplying matrices/vectors are much less computationally intensive than the Grahm-Schmidt process.)

- Note: since $\ker(A^TA) = \ker(A)$, you won't get additional phantom solutions to

$$A^TA\boldsymbol{v} = A^T\boldsymbol{b}$$

  when compared to

$$A\boldsymbol{v} = \boldsymbol{b}$$

- In the case $\ker(A) = \{0\}$, we can even do "better": by another useful fact, $\ker(A) = \{0\}$ implies $A^T A$ is invertible. So

$$\boldsymbol{v} = (A^T A)^{-1} A^T \boldsymbol{b}$$

is the least squares solution in that case. (Notice we cannot say

$$(A^T A)^{-1} A^T \boldsymbol{b} = A^{-1} (A^T)^{-1} A^T \boldsymbol{b} = A^{-1} \boldsymbol{b}$$

Why not?) Computationally, however, taking the inverse of $A^T A$ is usually more difficult than solving $A^T A \boldsymbol{v} = A^T \boldsymbol{b}$ using Gauss-Jordan elimination.

- The above fact can be used to construct the matrix of an orthogonal projection without finding the orthonormal basis. (Remember that on Monday we took U to be the matrix of the orthonormal basis for a subspace, then UU^T is the matrix for the orthogonal projection onto the subspace.) Letting W be a subspace and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_P$ be a basis, and A be the matrix whose column vectors are the basis vectors. We claim that the matrix for the orthogonal projection is

$$A(A^T A)^{-1} A^T$$

This follows from the fact that when $\boldsymbol{v}$ is a least squares solution to $A\boldsymbol{v} = \boldsymbol{b}$, $A\boldsymbol{v} = proj_W \, \boldsymbol{b}$. But since the column vectors of A are linearly independent, $\ker(A) = \{0\}$, and from above we know that $\boldsymbol{v} = (A^T A)^{-1} A^T \boldsymbol{b}$.

- Finish with an example: find the trajectory of the celestial object if the 6 data points are

	r	θ	$\cos\theta$	$\sin\theta$
1	2.7	$\pi/6$	$\sqrt{3}/2 \sim 0.866$	$1/2$
2	4.4	$\pi/3$	$1/2$	$\sqrt{3}/2$
3	10	$5\pi/9$	-0.174	0.985
4	5.8	π	-1	0
5	2.3	$4\pi/3$	$-1/2$	$-\sqrt{3}/2$
6	1.8	$3\pi/2$	0	-1

From the data we can write

$$\boldsymbol{b} = \begin{bmatrix} 2.7 \\ 4.4 \\ 10 \\ 5.8 \\ 2.3 \\ 1.8 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & -2.3 & 1.4 \\ 1 & -2.2 & 3.8 \\ 1 & 1.7 & 9.8 \\ 1 & 5.8 & 0 \\ 1 & 1.2 & -2.0 \\ 1 & 0 & -1.8 \end{bmatrix}$$

(I am going to cheat a bit and use the calculator in this case...) which we can evaluate for (rounding to 2 significant figures)

$$A^T A = \begin{bmatrix} 6 & 4.1 & 11 \\ 4.1 & 48 & 3.3 \\ 11 & 3.3 & 120 \end{bmatrix}, \quad A^T \boldsymbol{b} = \begin{bmatrix} 27 \\ 38 \\ 110 \end{bmatrix}$$

From here we can solve

$$A^T A\boldsymbol{v} = A^T\boldsymbol{b}$$

and get

$$\boldsymbol{v} = \begin{bmatrix} 3.0 \\ 0.48 \\ 0.63 \end{bmatrix}$$

(Not bad considering that I "generated" the "data" from $l = 3, c = 12/25, d = 16/25$ and rounded off after two digits. The original equation $A\boldsymbol{v} = \boldsymbol{b}$ is inconsistent if one checks its rref.)

- The calculated eccentricity is then 0.79, showing that the trajectory is a very elongated ellipse.

MAR 28, 2008

- (Seems my dates here are slightly misaligned with Real-Life)
- Today we start on determinants.
- Remember way-back-when we gave a quick way of checking whether a 2×2 matrix is invertible:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then A is invertible if and only if the quantity

$$\det(A) = ad - bc \neq 0$$

and

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Now we ask whether there is a way of defining the *determinant* for arbitrary square matrices, a quantity that determines whether the matrix is invertible?

- The answer (to the needlessly rhetorical question) is yes. First let's look at the meaning of the 2×2 determinant.
  - The first thing to remember is that when a 2×2 matrix is invertible, it means that its column vectors form a basis. In particular, they must be linearly independent.
  - How do we write down the condition that two vectors are linearly independent? Linearly independent means that the two vectors are not multiples of each other, which means that there *does not* exist any constant k such that

$$\boldsymbol{v}_1 = k\boldsymbol{v}_2$$

In $\mathbb{R}^2$, we write for the vectors the column vectors of A:

$$\boldsymbol{v}_1 = \begin{bmatrix} a \\ c \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} b \\ d \end{bmatrix}$$

So we want the non-existence of the constant k in the set of equations

$$a = kb\,, \quad c = kd$$

Should k exist, then we can solve for k and have

$$\frac{a}{b} = k = \frac{c}{d}$$

and so cross multiplying:

$$ad = bc$$

- So we have that $ad = bc$ if and only if we can find a k that shows $\boldsymbol{v}_1$ is a multiple of $\boldsymbol{v}_2$.
- This just goes to show that the determinant of a 2×2 matrix, $ad - bc = \det(A)$, just encodes whether the two column vectors are parallel to each other.

- So the natural idea is to generalize this concept: the $\det(A)$ of an $N \times N$ matrix A should be a quantity that measures whether the column vectors are linearly independent.

- Let's start with the case of $N = 3$. In this case, there's a easy geometric interpretation of the condition. Let

$$A = \begin{bmatrix} \boldsymbol{u} & \boldsymbol{v} & \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}$$

Something you should've seen in multivariable calculus is the notation

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}$$

for the cross product in $\mathbb{R}^3$. Remember that the cross product of two vectors in $\mathbb{R}^3$ has the following nice properties:

  - If $\boldsymbol{u}$ and $\boldsymbol{v}$ are parallel (in other words, linearly dependent), then $\boldsymbol{u} \times \boldsymbol{v} = 0$.
  - If $\boldsymbol{u}$ and $\boldsymbol{v}$ span a 2 dimensional plane in $\mathbb{R}^3$ (in other words, they are linearly independent), then $\boldsymbol{u} \times \boldsymbol{v}$ is a vector perpendicular to the plane that they span.

- So $\boldsymbol{u} \times \boldsymbol{v}$ can already be used to rule out the case when the first two vectors are linearly dependent. How do we deal with the third vector $\boldsymbol{w}$? Now suppose $\boldsymbol{w}$ is a linear combination of $\boldsymbol{u}$ and $\boldsymbol{v}$ (meaning that $\boldsymbol{w}$ is redundant), then $\boldsymbol{w}$ sits inside the span of $\boldsymbol{u}$ and $\boldsymbol{v}$; in other words, $\boldsymbol{w}$ is in the plane defined by $\boldsymbol{u}$ and $\boldsymbol{v}$. This in particular means that $\boldsymbol{w}$ must be perpendicular to the vector $\boldsymbol{u} \times \boldsymbol{v}$!

- This motivates us to define the determinant of a 3×3 matrix thus: given A as written above,

$$\det(A) = (\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$$

  - When $\boldsymbol{u}$ and $\boldsymbol{v}$ are multiples of one another, then $\boldsymbol{u} \times \boldsymbol{v} = 0$ and so $\det(A) = 0$.
  - If $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent, but $\boldsymbol{w}$ is redundant, then since $\boldsymbol{w} \perp (\boldsymbol{u} \times \boldsymbol{v})$, $\det(A) = 0$.
  - Only when $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are linearly independent (and since there are three of them, they must form a basis) can $\det(A) \neq 0$

- A computational aid: Sarrus's rule. By expanding algebraically the expression $\det(A) = (\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$, we have the following way of computing the determinant: first copy down the first two columns at the end of the matrix (in order!):

$$\begin{bmatrix} a_{11} & a_{12} & a_{12} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

  Calling the NW-SE direction "positive" and the NE-SW direction "negative", we sum over the products of all 6 diagonal lines with the associated signs:

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$$

- Example:

$$\det(\begin{bmatrix} 1 & 3 & 5 \\ 2 & 2 & 3 \\ 7 & 1 & 2 \end{bmatrix}) = 4 + 63 + 10 - 70 - 3 - 12 = -8$$

HOMEWORK FOR THIS WEEK

5.3: 6, 8, 28, 29, 32, 38, 40, 42

5.4: 2, 3, 4, 8, 17, 24, 32, 36