

MATH 167: APPLIED LINEAR ALGEBRA

Chapter 3

Jesús De Loera, UC Davis

February 18, 2012

Orthogonal Vectors and Subspaces (3.1).

- In real life vector spaces come with additional **METRIC properties!!** We have notions of distance and angles!! You are familiar with the **Euclidean vector space** \mathbb{R}^n :
- Since kindergarden you know that the distance between two vectors $x = (x_1, \dots, x_n)$ $y = (y_1, \dots, y_n)$ is given by

$$\text{dist}(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

- We say vectors x, y are **perpendicular** when they make a 90 degree angle. When that happens the triangle they define is right triangle! (WHY?)
- **Lemma** Two vectors x, y in \mathbb{R}^n are perpendicular if and only if

$$x_1y_1 + \dots + x_ny_n = xy^T = 0$$

When this last equation holds we say x, y are **orthogonal**.

- **Orthogonal Bases:** A basis u_1, \dots, u_n of V is orthogonal if $\langle u_i, u_j \rangle = 0$ for all $i \neq j$.
- **Lemma** If v_1, v_2, \dots, v_k are orthogonal then they are linearly independent.

The Orthogonality of the Subspaces

- **Definition** We say two subspaces V, W of \mathbb{R}^n are **orthogonal** if for $u \in V$ and $w \in W$ we have $uw^T = 0$.
- Can you see a way to detect when two subspaces are orthogonal?? **Through their bases!**
- **Theorem:** The row space and the nullspace are orthogonal. Similarly the column space is orthogonal to the left nullspace.
- **proof:** The dot product between the rows of A^T and the respective entries in the vector y is zero.
- Therefore the rows of A^T are perpendicular to any $y \in N(A^T)$

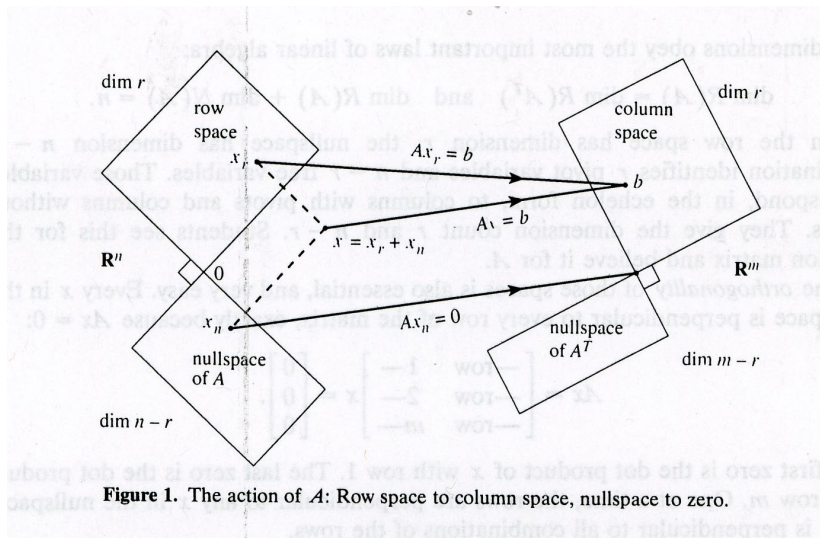
$$A^T y = \begin{bmatrix} \text{Column 1 of } A \\ \vdots \\ \text{Column } n \text{ of } A \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $y \in N(A^T)$.

- There is a stronger relation, for a subspace V of \mathbb{R}^n the set of all vectors orthogonal to V is the **orthogonal complement** of V , denoted V^\perp .
- **Warning** Spaces can be orthogonal without being complements!
- **Exercise** Let W be a subspace, its orthogonal complement is a subspace, and $W \cap W^\perp = 0$.
- **Exercise** If $V \subset W$ subspaces, then $W^\perp \subset V^\perp$.
- **Theorem** (Fundamental theorem part II) $C(A^T)^\perp = N(A)$ and $N(A)^\perp = C(A^T)$. **Why?**
- **proof:** First equation is easy because x is orthogonal to all vectors of row space $\leftrightarrow x$ is orthogonal to each of the rows $\leftrightarrow x \in N(A)$. The other equality follows from exercises.
- **Corollary** Given an $m \times n$ matrix A , the nullspace is the orthogonal complement of the row space in \mathbb{R}^n . Similarly, the left nullspace is the orthogonal complement of the column space inside \mathbb{R}^m
- WHY is this such a big deal?

- **Theorem** Given an $m \times n$ matrix A , every vector x in \mathbb{R}^n can be written in a unique way as $x_n + x_r$ where x_n is in the nullspace and x_r is in the row space of A .
- **proof** Pick x_n to be the orthogonal projection of x into $N(A)$ and x_r to be the orthogonal projection into $C(A^T)$. Clearly x is a sum of both, but why are they unique?
- If $x_n + x_r = x'_n + x'_r$, then $x_n - x'_n = x'_r - x_r$. Thus there must be the zero vector because $N(A)$ is orthogonal to $C(A^T)$.
- This has a beautiful consequence: Every matrix A , when we think of it as a linear map, transforms the row space into its column space!!!

An important picture

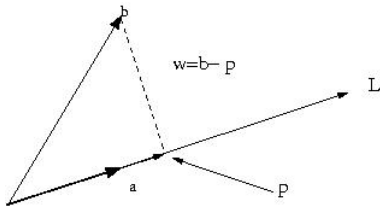


Other USEFUL inner products!

- For image processing it is important to consider the vector space $C^0[a, b]$ consisting of all continuous scalar functions $f(x)$ inside the interval $[a, b]$.
- Inner product is $\langle f, g \rangle = \int_a^b f(x)g(x)dx$
- **Example:** Suppose $a = 0, b = \pi/2$,
 $\langle \sin(x), \cos(x) \rangle = \frac{1}{2} \sin^2(x) \Big|_{x=0}^{\pi/2} = 1/2$
- We can define the **norm** of a vector (remember vector=functions) by $\|f\| = \sqrt{\int_a^b f(x)^2}$. For example $\|\sin(x)\| = \sqrt{\frac{\pi}{4}}$. Here is a weird one the function that is constant $c(x) = 1$ has norm $\sqrt{\pi/2}$.
- This inner product satisfies similar properties as the usual **dot product** xy^T (bilinearity, symmetry, positivity).
- Some well-know properties are preserved, like the idea of triangle inequality!

Cosines and Projections onto Lines (3.2)

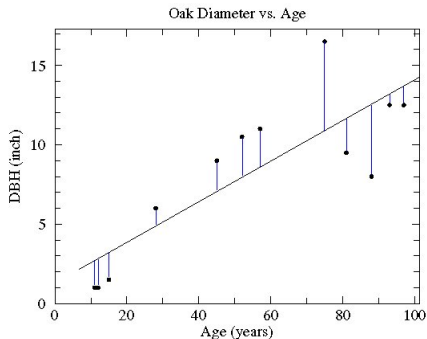
- Fundamental question: If we have a subspace S , is there a formula for the projection p of a vector b into that subspace?
- Imaging b as **data from experiments**, b is not in S , due to error of measurement, its projection p is the best choice to replace b . Key idea of **LEAST SQUARES** for **regression analysis**
- Let us learn how to do this projection for a line! b is projected into the line L given by the vector a . (PICTURE!).



- The projection of vector b onto the line in the direction a is
$$p = \frac{a^T b}{a^T a} a.$$
- **Corollary Cauchy-Schwarz inequality.** For all vectors a, b in \mathbb{R}^n satisfy $|a^T b| \leq \|a\| \|b\|$.

Projections and Least Squares (3.2)

- Suppose we do a series of experiments, collecting data.
Suppose we expect the output b to be a linear function of the input t $b = \alpha + t\beta$, but we need to **determine** α, β .
- At different times t_i we measure a quantity b_i . E.g, Say a police man is interested on clocking the speed of a vehicle by using measurements of its relative distance. Assuming the vehicle is traveling at constant speed, so we know linear formula, but errors exist.
- At $t = t_i$, the error between the measured value b_i and the value predicted by the function is $e_i = b_i - (\alpha + \beta t_i)$.
- We can write it as $e = b - Ax$ where $x = (\alpha, \beta)$. e is the **error vector**, b is the **data vector**. A is an $m \times 2$ matrix.
- We seek the line that minimizes the total squared error or Euclidean norm $\|e\| = \sqrt{\sum_{i=1}^m e_i^2}$.



We are looking for x that minimizes $\|b - Ax\|$.

NOTE: Sometimes data does not follow a linear model, say we measure amount of radioactive material in a rock, data follows an exponential change, but it can be converted into the linear case by taking logarithms.

- Note that $\|b - Ax\|$ is the distance from b to the point Ax which is element of the column space!
- The error vector $e = b - Ax$ must be perpendicular to the column space (picture!).
- Thus for each column a_i we have $a_i^T(b - Ax) = 0$. Thus in matrix notation: $A^T(b - Ax) = 0$, This gives the normal equations $A^T A = A^T b$.
- **Theorem** The best estimate is given by $x = (A^T A)^{-1} A^T b$. and its projection is $p = A((A^T A)^{-1} A^T)b$.
- **Lemma** $A^T A$ is a symmetric matrix. $A^T A$ has the same Nullspace as A .
Why? if $x \in N(A)$, then clearly $A^T Ax = 0$. Conversely, if $A^T Ax = 0$ then $x^T A^T Ax = \|Ax\|^2 = 0$, thus $Ax = 0$.
- **Corollary** If A has independent columns, then $A^T A$ is square, symmetric and invertible.

- **Example** Consider the problem $Ax = b$ with

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & -1 & -2 \\ 2 & 1 & -1 \end{bmatrix} \quad b^T = (1, 0, -1, 2, 2).$$

- We can see that there is no EXACT solution to $Ax = b$, use NORMAL EQUATION!

$$A^T A = \begin{bmatrix} 16 & -2 & -2 \\ -2 & 11 & 2 \\ -2 & 2 & 7 \end{bmatrix} \quad A^T b = \begin{bmatrix} 8 \\ 0 \\ -7 \end{bmatrix}$$

- Solving $A^T Ax = A^T b$ we get the least square solution $x^* \approx (0.4119, 0.2482, -0.9532)^T$ with error $\|b - Ax^*\| \approx 0.1799$.

- **Example** A sample of lead-210 measured the following radioactivity data at the given times (time in days). Can YOU predict how long will it take until one percent of the original amount remains?

time in days	0	4	8	10	14	18
mg	10	8.8	7.8	7.3	6.4	6.4

- A linear model does not work. There is an exponential decay on the material $m(t) = m_0 e^{\beta t}$, where m_0 is the initial radioactive material and β the decay rate. By taking logarithms

$$y(t) = \log(m(t)) = \log(m_0) + \beta t$$

- Thus we can now use **linear** least squares to fit on the logarithms $y_i = \log(m_i)$ of the radioactive mass data. In this case we have

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 4 & 8 & 10 & 14 & 18 \end{bmatrix} \quad b^T = [2.302585093, 2.174751721]$$

- Thus $A^T A = \begin{bmatrix} 6 & 54 \\ 54 & 700 \end{bmatrix}$. Solving the NORMAL form system we get $\log(m_0) = 2.277327661$ and $\beta = -0.0265191683$

Thus the original amount was 10 mg. After 173 days it will be below one percent of the radioactive material.

- There is nothing special about polynomials or exponential functions in the application. We can deal with approximating function is a linear combination of some prescribed functions $h_1(t), h_2(t), \dots, h_n(t)$. Then we receive data y_i at time t_i and the matrix A has entry $A_{ij} = h_i(t_j)$.
- The least squares method can be applied when the measurement of error is not can be applied to situations when not all observations are trusted the same way!
- Now the error is $\sqrt{(b - Ax)^T C (b - Ax)}$. Then the weighted least square error is given by the new equations

$$A^T C A x = A^T C b, \quad \text{and} \quad x = (A^T C A)^{-1} A^T C b.$$

Orthogonal Bases and Gram-Schmidt (3.4)

- A basis u_1, \dots, u_n of a vector space V is **orthonormal** if it is orthogonal and each vector has unit length.
- **Observation** If the vectors u_1, \dots, u_n are orthogonal basis, their normalizations $\frac{u_i}{\|u_i\|}$ form an orthonormal basis.
- **Examples** Of course the standard unit vectors are orthonormal.

Consider the vector space of all quadratic polynomials $p(x) = a + bx + cx^2$, using the L^2 inner product of integration:

$$\langle p, q \rangle = \int_0^1 p(x)q(x)dx$$

The standard monomials $1, x, x^2$ form a basis, but do **not** form an orthogonal basis!

$$\langle 1, x \rangle = 1/2, \quad \langle 1, x^2 \rangle = 1/3, \quad \langle x, x^2 \rangle = 1/4$$

- An orthonormal basis is given by

$$u_1(x) = 1, \quad u_2(x) = \sqrt{3}(2x-1), \quad u_3(x) = \sqrt{5}(6x^2-6x+1).$$

Why do we care about orthonormal bases?

- **Theorem** Let u_1, \dots, u_n be an orthonormal bases for a vector space with inner product V . The one can write any element $v \in V$ as a linear combination $v = c_1 u_1 + \dots + c_n u_n$ where $c_i = \langle v, u_i \rangle$, for $i = 1, \dots, n$. Moreover the norm

$$\|v\| = \sqrt{\sum c_i^2}. \text{ Why?}$$

- **Example** Let us rewrite the vector $v = (1, 1, 1)^T$ in terms of the orthonormal basis

$$u_1 = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)^T, u_2 = \left(0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right), u_3 = \left(\frac{5}{\sqrt{30}}, \frac{-2}{\sqrt{30}}, \frac{1}{\sqrt{30}}\right)$$

Computing the dot products $v^T u_1 = \frac{2}{\sqrt{6}}$, $v^T u_2 = \frac{3}{\sqrt{5}}$, and $v^T u_3 = \frac{4}{\sqrt{30}}$. Thus

$$v = \frac{2}{\sqrt{6}} u_1 + \frac{3}{\sqrt{5}} u_2 + \frac{4}{\sqrt{30}} u_3$$

- **Challenge:** Figure out the same kind of formulas if the vectors are just orthogonal!!!

- A key reason to like matrices that have orthonormal vectors:
The least-squares equations are even nicer!!!
- **Lemma** If Q is a rectangular matrix with orthonormal columns, then the normal equations simplify because $Q^T Q = I$:
 - $Q^T Qx = Q^T b$ simplifies to $x = Q^T b$
 - Projection matrix simplifies $Q(Q^T Q)^{-1}Q^T = QIQ^T = QQ^T$.
 - Thus the projection point is $p = QQ^T b$, thus

$$p = (q_1^T b)q_1 + (q_2^T b)q_2 + \cdots + (q_n^T b)q_n$$

- So how do we compute orthogonal/orthonormal bases for a space?? We use the GRAM-SCHMIDT ALGORITHM.
- **input** Starting with a linear independent vectors a_1, \dots, a_n ,
output: orthonormal vectors q_1, \dots, q_n .
 - At step j it subtracts from a_j its components in the direction q_1, \dots, q_{j-1}

$$A_j = a_j - (q_1^T a_j)q_1 - (q_2^T a_j)q_2 - \cdots - (q_{j-1}^T a_j)q_{j-1}$$

- Then $q_j = A_j / \|A_j\|$.

- In this way, the original basis vectors a_1, \dots, a_n can be written in a “triangular” way!

$$a_1 = r_{11}q_1, \quad (1)$$

$$a_2 = r_{12}q_1 + r_{22}q_2 \quad (2)$$

$$a_3 = r_{13}q_1 + r_{23}q_2 + r_{33}q_3 \quad (3)$$

$$\vdots \quad (4)$$

$$a_n = r_{1n}q_1 + r_{2n}q_2 + \dots + r_{nn}q_n. \quad (5)$$

Where $r_{ij} = a_j^T q_i$.

- Write in matrix form! $A = QR$ where $A = (a_1 \ \dots \ a_n)$ and $Q = (q_1 \ q_2 \ \dots \ q_n)$ and $R = (r_{ij})$.
- **Theorem (QR decomposition)** Every $m \times n$ matrix A with independent columns can be factor as $A = QR$ where the columns of Q are orthonormal and R is upper triangular and invertible.

- Another famous application is **Fourier Analysis**. Useful in sound and image processing. **Key idea**: Decompose a **signal** as a linear combination of sines and cosines.
- The key space is T_n of all **trigonometric polynomials** of the form

$$T(x) = \sum a_{jk} (\sin(x))^j (\cos(x))^k$$

- The individual monomials $(\sin(x))^j (\cos(x))^k$ but they do not form a basis (we saw this in class!). A basis is given by the $2n + 1$ functions

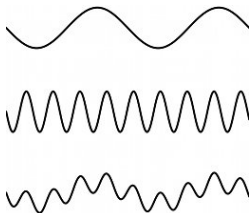
$$1, \cos(x), \sin(x), \cos(2x), \sin(2x), \dots, \cos(nx), \sin(nx)$$

- **Lemma** They are an orthogonal basis for T_n with respect to

$$\langle p, q \rangle = \int_{-\pi}^{\pi} p(x)q(x)dx$$

Thus the vector space T_n is $2n + 1$ dimensional.

- To investigate functions that arise in **Signal processing** of images and sound, one uses **Fourier Analysis**: We decompose a **signal** as a linear combination of sines and cosines. In signal processing applications we take a complicated signal f and rewrite it as a sum of sin, cos functions.



- Suppose $f(x)$ represents a signal defined in the interval $a \leq x \leq b$. The computer stores its measured values at finite number of **sample points** $a \leq x_0 < x_1 < \dots < x_n \leq b$. Normally the sample points are equally spaced $x_j = a + jh$, $j = 0, \dots, n$, where $h = \frac{b-a}{n}$. Say $a = 0$, $b = 2\pi$, thus $x_j = \frac{2j\pi}{n}$.

- The vector of samples is $f = (f(x_0), f(x_1), \dots, f(x_{n-1}))^T$.
where $f(x_j) = f(\frac{2j\pi}{n})$.
- **Note** Sampling cannot distinguished from functions that have the same values at the sample points! For example $f(x) = \cos(nx) + i\sin(nx)$ has value one at all sampled points so we cannot distinguished from the constant one function!
- Similarly $e^{i(k+n)x}$ and e^{ikx} are indistinguishable! But one has high frequency while the other goes “slowly”.
- **Definition:** The **Fourier series** of a function $f(x)$ is its (infinite) expansion into sines and cosines:

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

- How to find the Fourier coefficients a_i, b_i ? Integration! to find a_k multiply by $\cos(kx)$ and then integrate in the desired interval. Similarly for b_i .
- The passage from a signal to its Fourier coefficients is known as the **Finite Fourier Transform (FFT)**.