Matrices of Rank One

- When $r=1$, every row is a multiple of the same row, and likewise with columns
- The row space is a line in $\mathbb{R}^n$ and the column space is a line in $\mathbb{R}^m$
- Every rank one matrix is expressible as an outer product uv^T
- The nullspace is the plane perpendicular to v
  $Ax=0 \Rightarrow (uv^T)x=0 \Rightarrow u(v^Tx)=0 \Rightarrow v^Tx=0 \Rightarrow v$ and x are perpendicular
  - It is this perpendicularity of the subspaces that will be part 2 of the Fundamental Theorem

Problem Set 3.6

1.

a) $m=7$
$n=9$
$\dim(C(A)) = \dim(C(A^T)) = r = 5$
$\dim(N(A)) = n-r = 9-5 = 4$
$\dim(N(A^T)) = m-r = 9-7 = 2$

b) $m=3$, $n=4$
$\dim(C(A)) = \dim(C(A^T)) = r = 3$
$C(A) = \mathbb{R}^3$
$\dim(N(A^T)) = m-r = 3-3 = 0 = Z$

2.

$A=\begin{bmatrix}1&2&4\\2&4&8\end{bmatrix}$

row basis: $\{(1,2,4)\}$
column basis: $\{(1,2)\}$
$\dim(C(A)) = \text{rank}(A) = 1$
$\dim(C(A^T)) = \text{rank}(A) = 1$
$\dim(N(A)) = n-r = 2$
$\dim(N(A^T)) = m-r = 1$

$B=\begin{bmatrix}1&2&4\\2&5&8\end{bmatrix}$

row basis: $\{(1,2,4),(2,5,8)\}$
column basis: $\{(1,2),(2,5)\}$

4.

a) $\begin{bmatrix}1&0\\1&0\\0&1\end{bmatrix}$ b) not possible

11.

$r<m$ and $r=n$
$r<m$ and $r<n$
A^Ty is short and wide, which always has solutions

12.

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

$\dim(C(A^T)) = r = 1$,
then $\dim(N(A)) = n-r = 3-1 = 2$

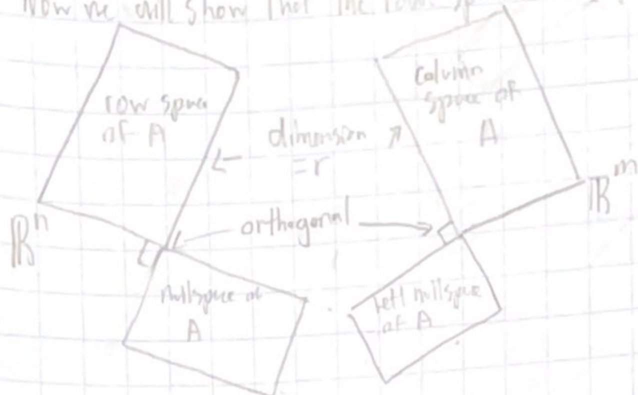
14. row space, left nullspace

Chapter 4: Orthogonality

Chapter 4.1: Orthogonality of the Four Subspaces

- Two vectors are orthogonal when their dot products are zero: $v^Tw = v\cdot w = 0$
- Orthogonal vectors also satisfy $\|v\|^2 + \|w\|^2 = \|v+w\|^2$
  $(v+w)^T(v+w) = v^Tv + w^Tw + 2v^Tw = \|v\|^2 + \|w\|^2 + 2v^Tw$ must be zero
- The row space is perpendicular to the nullspace. Every row of A is perpendicular to every solution of $Ax=0$
- The column space is perpendicular to the left nullspace. When b is outside the column space, then this nullspace of A^T comes into its own. It contains the error $e = b - Ax$ in the "least-squares" solution, a key application of linear algebra.

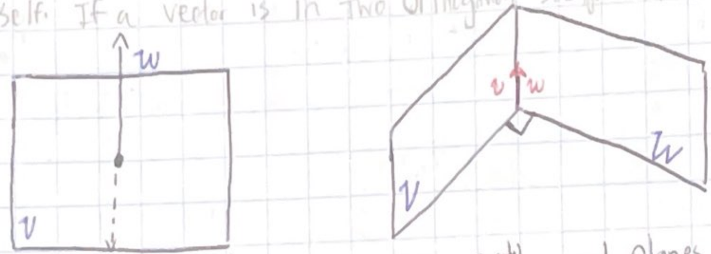
- Pat one of the Fundamental Theorem gives the dimension of the subspaces.
- Now we will show that the row space and nullspace are orthogonal subspaces inside $\mathbb{R}^n$



- Two subspaces V and W are orthogonal if every vector v in V is perpendicular to every w in W. That is:

$$v^T w = v \cdot w = 0 \;\; \forall v \in V, w \in W$$

- Examples (see fig 4.1.1)
  - The floor of your room (extended to infinity) is a subspace V. The line where two walls meet is a subspace W. The subspaces are orthogonal
  - The two walls may look like it, but they are not orthogonal subspaces. The intersection line is in both V and W, and a line is not perpendicular to itself. If a vector is in two orthogonal subspaces it must be the zero vector



orthogonal line and plane — non-orthogonal planes

Fig 4.1.1

- Orthogonality is impossible when dim V + dim W > dimension of the whole space
- Zero is the only point where the nullspace meets the row space. More than that, the nullspace and row space of A meet at 90°. This fact comes directly from $Ax=0$
  - Every vector x in the nullspace is perpendicular to every row of A because $Ax=0$

$$Ax=0 \Rightarrow \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} x = \begin{bmatrix} a_1 \cdot x \\ a_2 \cdot x \\ \vdots \\ a_n \cdot x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \forall x \in N(A)$$

Then, x is also perpendicular to all linear combinations of the rows

$$x \cdot (c_1 a_1 + c_2 a_2 + \dots c_n a_n) = c_1 \overbrace{(x \cdot a_1)}^{0} + c_2 \overbrace{(x \cdot a_2)}^{0} + \dots + c_n \overbrace{(x \cdot a_n)}^{0} = 0$$

  - The nullspace $N(A)$ and the row space are orthogonal subspaces of $\mathbb{R}^n$

- Example
  - The rows of A are perpendicular to $v = (1,1,-1)$

$$Ax = \begin{bmatrix} 1 & 3 & 4 \\ 5 & 2 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Likewise, the columns of A and the left nullspace are orthogonal.
  - Every vector y in the nullspace of A^T is perpendicular to every column of A
  - The left nullspace $N(A^T)$ and the column space $C(A)$ are orthogonal in $\mathbb{R}^m$
  - It is proved similarly:

$$A^T y = 0 \Rightarrow \begin{bmatrix} (\text{column } 1)^T \\ \vdots \\ (\text{column } n)^T \end{bmatrix} y = \begin{bmatrix} (\text{column } 1)^T \cdot y \\ \vdots \\ (\text{column } 2)^T \cdot y \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad \forall y \in N(A^T)$$

Then y is perpendicular to all linear combination of the columns

$$[C_1(\text{column } 1) + \dots + C_n(\text{column } n)] \cdot y = C_1(y \cdot \text{column } 1) + \dots + C_n(y \cdot \text{column } n) = 0$$

Orthogonal Complements

- The fundamental subspaces are more than just orthogonal (in pairs). Their dimensions are also right.
- Two lines could be perpendicular in $\mathbb{R}^3$ but those lines could not be the row space and nullspace of a 3 by 3 matrix. The lines have dimensions 1 and 1, summing to 2. The correct dimensions r and n-r must sum to n.
  - The fundamental subspaces have dimensions 2 and 1 or 3 and 0. These subspaces are not only orthogonal, they are orthogonal complements.
- The orthogonal complement of a subspace V contains every vector perpendicular to V. It is noted as $V^\perp$ (pronounced "V perp")
- By this definition, the nullspace is the orthogonal complement of the row space since every x perpendicular to the rows satisfies $Ax = 0$
- The reverse is also true. If V is orthogonal to the nullspace it must be in the row space. Otherwise we could add this V as an extra row without changing the nullspace. The row space would grow, which breaks the law $r + (n-r) = n$
  - Thus $N(A)^\perp = C(A^T)$ and $C(A^T)^\perp = N(A)$
- Fundamental Theorem of Linear Algebra, Part Two
  - $N(A)$ is the orthogonal complement of the row space $C(A^T)$ (in $\mathbb{R}^n$)
  - $N(A^T)$ is the orthogonal complement of the column space $C(A)$ (in $\mathbb{R}^m$)
- The point of complements is that every x can be split into a row space component x_r and a nullspace component x_n. When A multiplies x

$$Ax = A(x_r + x_n) = Ax_r + Ax_n = Ax_r + 0 = Ax_r = Ax$$

- Every vector goes to the column space.
  - The nullspace component goes to zero: $Ax_n = 0$
  - The row space component goes to the column space: $Ax_r = Ax$

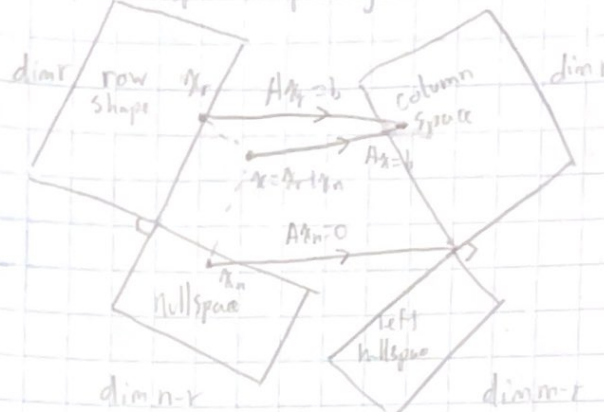


Fig 4.1.2

- Fig 4.1.2 shows the true action of A on $x = x_r + x_n$. A transforms (or "maps") x_r to the column space and the nullspace vector x_n to 0.
  - More than that, every vector b in the column space comes from one and only one vector in the row space. If $Ax_r = Ax_r'$, the difference $x_r - x_r'$ is in the nullspace. It is also in the row space since x_r and x_r' also did, and $x_r - x_r'$ is a linear combination. This difference $x_r' - x_r$ must be zero since the nullspace and row space are perpendicular. Therefore $x_r' = x_r$.
- There is an r by r invertible matrix hiding inside A if we throw away the two nullspaces. From the row space to the column space, A is invertible. The "pseudoinverse" will invert it in Section 7.3
- Example

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$ contains the submatrix $\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$

  - The other 11 zeroes are responsible for the nullspaces

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 1 & 2 & 4 & 5 & 6 \end{bmatrix}$$ contains $\begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}$ in the pivot rows and columns

Combining Bases from Subspaces

- Any n independent vectors in $\mathbb{R}^n$ must span $\mathbb{R}^n$. So they are a basis.
- Any n vectors that span $\mathbb{R}^n$ must be independent. So they are a basis
- Starting with the correct number of vectors, one property of a basis produces the other
- When the vectors go into the columns, the above two facts are:
  - If the n columns of A are independent, they span $\mathbb{R}^n$ and $Ax = b$ is solvable
  - If the n columns span $\mathbb{R}^n$ they are independent. $Ax = b$ has only 1 solution
- Example

$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ split $x = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ into $x_r + x_n = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

Problem Set 4.1

1.

$$\begin{bmatrix}1&2&3\\2&4&6\end{bmatrix} \rightarrow \begin{bmatrix}1&2&3\\0&0&0\end{bmatrix}$$

basis for row space: $\{(1,2,3)\}$

basis for null space: $\{(-2,1,0),(-3,0,1)\}$

3.

c) $\begin{bmatrix}1&2&-3\\2&-3&1\\-3&5&-2\end{bmatrix}$

b) not possible
row space vectors should all be orthogonal to $(1,1,1)$
$(2,-3,5)\cdot(1,1,1) \neq 0$

8.

Since $x = x_p + x_n$

$Ax = A(x_p + x_n) = Ax_r + Ax_n$
$= Ax_r$

13.

If $V^T W = 0$, then the corresponding columns of V and W are perpendicular, which imply the spans of the columns of V and W are orthogonal

15.

$p+q \leq n$

19.

Plane, line, L

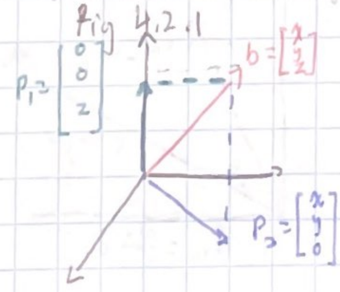
24.

row 2 \ to n

Chapter 4.2 Projections

- Two introductory questions:
  1. What are projections of $b=(2,3,4)$ onto the z axis and the xy plane?
  2. What matrices produce these projections onto a line and plane?
- When b is projected onto a line, its projection p is the part of b along that line. If b is projected onto a plane, p is the part in that plane. The projection p is Pb. The projection matrix P multiplies b to give p.
- The projection onto the z-axis we'll call P_1 and the projection onto the xy-plane we'll call P_2. The image in your mind should be that of fig 4.2.1.
  $P_1 = (0,0,4)$, $P_2 = (2,3,0)$. These are the parts of b along the z-axis and xy-plane

fig 4.2.1



- The projection matrices P_1 and P_2 are 3 by 3. They multiply b with 3 components to produce p with 3 components.
  - Projection onto a line comes from a rank one matrix. Projection onto a plane comes from a rank two matrix.

- For the second question:

$$P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- In this case the projections p_1 and p_2 are perpendicular. The xy plane and z-axis are orthogonal subspaces. More than that, they are orthogonal complements. Their dimensions add to $1+2=3$. Every vector b in the whole space can be expressed as a sum of its parts in the two subspaces.
  - The projection p_1 and p_2 are exactly those parts.

    The vectors give $p_1 + p_2 = b$. The matrices give $P_1 + P_2 = I$

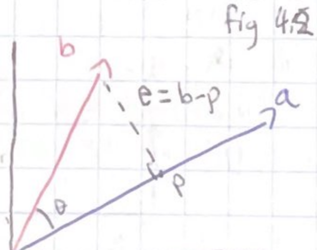
- We can use this to express x as $x_r + x_n$. The goal is to find the part p in each subspace and the corresponding projection matrix that produces $p = Pb$
  - Every subspace of $\mathbb{R}^m$ has its own m by m projection matrix
- To compute P, we need a good description of the subspace it projects onto. The best such description is a basis!
  - We put the basis vectors into the columns of A. Now we are projecting onto the column space of A.

    For z-axis: $A_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ For xy plane: $A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 0 \end{bmatrix}$ Both are bases. Both work.
  - We want to be able to project any b onto the column space of any m by n matrix.

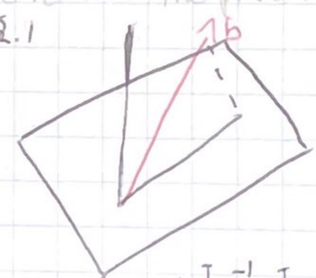
Projection onto a Line

- A line goes through the origin in the direction of $a = (a_1, a_2, a_3, \ldots, a_n)$. Along that line, we want the point p closest to $b = (b_1, b_2, b_3, \ldots, b_n)$. The key to projection is orthogonality. The line from b to p is perpendicular to the vector a.
- The projection p is some scalar multiple of a. Call it $p = \hat{x}a$. Computing the scalar $\hat{x}$ will give the vector p. Then from the formula for p, we read off the projection matrix P.
- The line from b to p is denoted by a dotted line in fig 4.2.1.
  - It is $e = b - p = b - \hat{x}a$, which is perpendicular to a.

fig 4.2.1



$p = \hat{x}a = \frac{a^Tb}{a^Ta}a$



$p = A\hat{x} = A(A^TA)^{-1}A^Tb = Pb$

$$a \cdot (b - \hat{x}a) = 0 \Rightarrow a \cdot b - \hat{x}a \cdot a = 0$$

$$\Rightarrow \hat{x}a \cdot a = a \cdot b$$

$$\hat{x} = \frac{a \cdot b}{a \cdot a} = \frac{a^Tb}{a^Ta}$$

- Then $p = \hat{x}a = \frac{a^Tb}{a^Ta}a$ is the projection of b onto the line through a.

- Special Cases
  - If $b = a$, then $\hat{x} = 1$. The projection a onto a is itself. $Pa = a$
  - If $b \perp a$, then $a^T b = 0$. The projection is $p = 0$
- Example
  - project $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ to find p

$$p = \frac{a^T b}{a^T a} a = \frac{5}{9} a = \left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9}\right) \text{ and } e = b - p = \left(\frac{4}{9}, -\frac{1}{9}, -\frac{1}{9}\right)$$

- Now for the projection matrix:
  - What matrix is multiplying b in $p = Pb$?

$$p = a\hat{x} = a\frac{a^T b}{a^T a} = Pb \Rightarrow P = \frac{a a^T}{a^T a}$$

  - P is a column times a row (outer product). The column is a, the row is a^T, and then we divide by $a^T a$. The projection matrix is m by m, but its rank is one. We are projecting onto a one-dimensional subspace
- Example:
  - Find the projection matrix $P = \frac{aa^T}{a^T a}$ onto the line through $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

$$P = \frac{aa^T}{a^T a} = \frac{1}{9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}$$

Check:

$$p = Pb = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{9} \\ \frac{10}{9} \\ \frac{10}{9} \end{bmatrix}$$

- If the vector a is doubled, the matrix P stays the same. It still projects onto the same line (the line through 2a). If the matrix is squared, $P^2 = P$. Projecting twice doesn't change anything, so $P^2 = P$.
- The matrix $I - P$ is another projection matrix. It produces the part of b perpendicular to a (the "e" part). Note that $(I - P)b = Ib - Pb = b - p$. When P projects onto one subspace, $I - P$ projects onto the perpendicular subspace. Here $I - P$ projects onto the plane perpendicular to a.
- Projection onto an n-dimensional subspace of $\mathbb{R}^m$ takes more effort.

Projection Onto a Subspace

- Start with n vectors $a_1, \dots, a_n$ in $\mathbb{R}^m$. Assume these a's are linearly independent
- We want to find the combination $p = \hat{x}_1 a_1 + \dots + \hat{x}_n a_n$ closest to a given vector b. We are projecting each b in $\mathbb{R}^m$ onto the subspace spanned by the a's to get p
- With $n = 1$, this is projection onto a line. The line is the column space of $A = [a_1]$ with only 1 column
- The combinations for p in $\mathbb{R}^m$ are the vectors Ax in the column space. We are looking to the particular combination $p = A\hat{x}$ closest to b.

- We compute projections onto n-dimensional subspaces in three steps as before: Find the vector $\hat{x}$, find the projection $p = A\hat{x}$, then find the matrix P.
- The dotted line in Fig 4.2.1 goes from b to the nearest point $A\hat{x}$ in the subspace. This error vector $b - A\hat{x}$ is perpendicular to the subspace.
- The error $b - A\hat{x}$ makes a right angle with all the vectors $a_1, \dots, a_n$. That gives us n equation

$$\begin{matrix} a_1^T(b - Ax) = 0 \\ \vdots \\ a_n^T(b - Ax) = 0 \end{matrix} \quad \text{or} \quad \begin{bmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_n^T & - \end{bmatrix} \begin{bmatrix} b - Ax \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

  - The matrix with those rows is A^T. The n equations are exactly $A^T(b - A\hat{x}) = 0$. Rewrite to $A^Tb - A^TAx = 0 \Rightarrow A^TA\hat{x} = A^Tb$. This in the equation for $\hat{x}$ and the coefficient matrix is A^TA. Now we can find $\hat{x}$, p, and P.
- The combination $p = \hat{x}_1 a_1 + \dots + \hat{x}_n a_n = A\hat{x}$ that is closest to b comes from:

$$A^T(b - A\hat{x}) = 0 \quad \text{or} \quad A^TA\hat{x} = A^Tb$$

  - This symmetric matrix A^TA is n by n. It is invertible if the a's are independent. The solution is $\hat{x} = (A^TA)^{-1}A^Tb$. Then the projection is:

$$p = A\hat{x} = A(A^TA)^{-1}A^Tb$$

  - The projection matrix P is then

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

- Looking at the respective equations for $n=1$:

$$\hat{x} = \frac{a^Tb}{a^Ta} \qquad p = a\frac{a^Tb}{a^Ta} \qquad P = \frac{aa^T}{a^Ta}$$

they are very similar. Instead of a^Ta, we have A^TA. Instead of dividing by the scalar, we invert the matrix: $(A^TA)^{-1}$ instead of $\frac{1}{a^Ta}$. Since $a_1, \dots, a_n$ are linearly independent, this inverse matrix is guaranteed to exist.
- The key step is $A^T(b - A\hat{x}) = 0$:
  1. Our subspace is the column space of A
  2. The error vector $b - A\hat{x}$ is perpendicular to that column space
  3. Therefore $b - A\hat{x}$ is in the nullspace of A^T (by the Fundamental Theorem, part 2). Then $A^T(b - A\hat{x}) = 0$
- The left nullspace is important in projection. It contains the error vector $e = b - A\hat{x}$. The vector b is split into the error $e = b - p$ and the projection p

- Example
  - If $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$, find $\hat{x}$, p, P

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \text{ and } A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

Now solve

$$A^TA\hat{x} = A^Tb \Rightarrow \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \text{ gives } \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

$$p = A\hat{x} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}. \text{ The error is } e = b - p = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

To find P, compute $P = A(A^TA)^{-1}A^T$

$$(A^TA)^{-1} = \frac{1}{\det(A^TA)} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \text{ and } P = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

- Warning: $P = A(A^TA)^{-1}A^T$ is deceptive. $(A^TA)^{-1} \neq A^{-1}A^{T^{-1}}$. If we do that, then $P = AA^{-1}(A^T)^{-1}A^T = I$. This is not true.
  A is a rectangular and it has no inverse matrix
- A^TA is invertible if and only if A has linearly independent columns.
  - A^TA is a square matrix. For every matrix A, we will show that A^TA has the same nullspace as A. When the columns of A are linearly independent, its nullspace contains only the zero vector. Then A^TA with this same nullspace is invertible.
  - Let A be any matrix. If x is in its nullspace, the $Ax = 0$. Multiplying by A^T gives $A^TAx = 0$, so x is also in the nullspace of A^TA
  - Now we need to prove the opposite, that from $A^TAx = 0$ we get $Ax = 0$
    We can't multiply by $(A^T)^{-1}$, which generally doesn't exist. Just multiply by x^T
    $(x^T)A^TAx \Rightarrow (Ax)^T(Ax) = 0 \Rightarrow \|Ax\|^2 = 0$
  - This says if $A^TAx = 0$ then Ax has length zero. Therefore $Ax = 0$
    Every vector x in one nullspace is in the other.
- When A has independent columns, A^TA is square symmetric and invertible
  - A^TA is (n by m) times (m by n). Then A^TA is square (n by n)
  - It is symmetric because $(A^TA)^T = A^T(A^T)^T = A^TA$

$$\overset{A^T}{\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \end{bmatrix}} \overset{A}{\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 0 \end{bmatrix}} = \overset{A^TA}{\begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}} \qquad \overset{A^T}{\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix}} \overset{A}{\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}} = \overset{A^TA}{\begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix}}$$

dep. singular $\qquad$ indep. invertible

Problem Set 4.2

1.

a)

$$p = \frac{a^T b}{a^T a} a = \frac{5}{3}(1,1,1) = \left(\frac{5}{3}, \frac{5}{3}, \frac{5}{3}\right)$$

$$e = b - p = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

b)

$$p = \frac{a^T b}{a^T a} a = \frac{-11}{11}(-1,-3,-1) = (1,3,1)$$

$$e = b - p = (0,0,0)$$

2.

a)

$$p = \frac{a^T b}{a^T a} a = \frac{\cos\theta}{\cos^2\theta + \sin^2\theta}(1,0) = \cos\theta(1,0) = (\cos\theta, 0)$$

$$e = b - p = (0, \sin\theta)$$

3.

a)

$$p = \frac{a a^T}{a^T a} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

b)

$$p = \frac{a a^T}{a^T a} = \frac{1}{11}\begin{bmatrix} 1 & 3 & 1 \\ 3 & 9 & 3 \\ 1 & 3 & 1 \end{bmatrix}$$

6.

$$P_1 = \frac{a_1 a_1^T}{a_1^T a_1} = \frac{1}{9}\begin{bmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{bmatrix} \qquad P_2 = \frac{a_2 a_2^T}{a_2^T a_2} = \frac{1}{9}\begin{bmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{bmatrix}$$

$$P_1 P_2 = \frac{1}{81}\begin{bmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{bmatrix}\begin{bmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{bmatrix} = \frac{1}{81}\begin{bmatrix} -8 & -8 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

9.

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

$$A^T A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{4}\begin{bmatrix} 5 & -1 \\ -1 & 1 \end{bmatrix}$$

$$P = A(A^T A)^{-1} A^T = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 5 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

11

a)

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad b = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\hat{x} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

$$p = A\hat{x} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$

11

b)

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 14 \end{bmatrix}$$

$$A^T A\hat{x} = A^T b$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 14 \end{bmatrix}$$

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$$

$$p = A\hat{x} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$$

$$e = b - p = (0, 0, 0)$$

12.

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

$$(A^T A)^{-1} = \frac{1}{2}\begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}$$

$$P = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

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

$$Pb = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 8 \\ 8 \\ 12 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$$

13.

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

$p=(1,2,3,0)$

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

14.

$$A=\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix} \quad b=\begin{bmatrix}0\\2\\4\end{bmatrix}$$

$$A^TA=\begin{bmatrix}0&1&2\\1&2&0\end{bmatrix}\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix}=\begin{bmatrix}5&2\\2&5\end{bmatrix}$$

$$A^Tb=\begin{bmatrix}0&1&2\\1&2&0\end{bmatrix}\begin{bmatrix}0\\2\\4\end{bmatrix}=\begin{bmatrix}10\\4\end{bmatrix}$$

$$A^TA\hat{x}=A^Tb$$

$$\begin{bmatrix}5&2\\2&5\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}10\\4\end{bmatrix} \rightarrow \begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}2\\0\end{bmatrix}$$

$$A\hat{x}=p$$

$$p=\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix}\begin{bmatrix}2\\0\end{bmatrix}=\begin{bmatrix}0\\2\\4\end{bmatrix}$$

$$(A^TA)^{-1}=\frac{1}{21}\begin{bmatrix}5&-2\\-2&5\end{bmatrix}$$

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

$$=\frac{1}{21}\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix}\begin{bmatrix}5&-2\\-2&5\end{bmatrix}\begin{bmatrix}0&1&2\\1&2&0\end{bmatrix}$$

$$=\frac{1}{21}\begin{bmatrix}0&1\\1&2\\2&0\end{bmatrix}\begin{bmatrix}-2&1&10\\5&5&-4\end{bmatrix}$$

$$=\frac{1}{21}\begin{bmatrix}5&8&-4\\8&17&-8\\-4&2&20\end{bmatrix}$$

17.

$$(I-P)^2=I^2-IP-PI+P^2$$
$$=I-P-P+P$$
$$=I-P$$

left nullspace

21. itself

$$P^2=[A(A^TA)^{-1}A^T][A(A^TA)^{-1}A^T]$$
$$=A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T$$
$$=A(A^TA)^{-1}A^T$$
$$=P$$

27. A^T, A

23.

$P=I_n$

A covers the entire $\mathbb{R}^n$ space, so a projection of any vector onto $\mathbb{R}^n$ is itself. $E=0$

Chapter 4.3: Least Squares Approximations

- It often happens that $Ax=b$ has no solution, and usually it's because there are too many equations. The matrix has more rows than columns. There are more equation than unknowns ($m>n$). The n columns spans a small part of $\mathbb{R}^m$ and b happens to be outside the column space.
- We can't always get the error $e=b-Ax$ down to zero, but it is practical to minimize it. When the length of e is minimized, $\hat{x}$ is a least squares solution.
- When $Ax=b$ has no solution, multiply by A^T and solve $A^TA\hat{x}=A^Tb$
- Example 1: Fitting a straight line
  - Given $(0,6)$, $(1,0)$, $(2,0)$; no straight line $b=C+Dt$ goes through the three points. We are asking for two numbers C and D that satisfies 3 equations

$t=0$ The first point is on the line $b=C+Dt$ if $C+D\cdot 0=6$

$t=1$ The second point is on the line $b=C+Dt$ if $C+D\cdot 1=0$

$t=2$ The third point is on the line $b=C+Dt$ if $C+D\cdot 2=0$

- We have three equations

$$\begin{cases} C + 0D = 6 \\ C + 1D = 0 \\ C + 2D = 0 \end{cases} \quad \text{Rewrite} \quad \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \quad Ax = b \text{ is not solvable}$$

Find $\hat{x}$ via $A^TA\hat{x} = A^Tb$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \quad ; \quad A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb \Rightarrow \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} C \\ D \end{bmatrix}$$

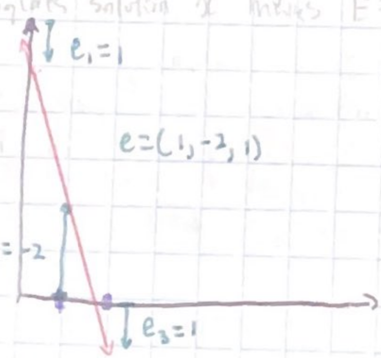
$\therefore$ $5 - 3t$ is the line of best fit for those three lines

Minimizing the Error

- How do we minimize $e = b - Ax$, we can find $\hat{x}$ (the best choice) by geometry, linear algebra, or calculus
- By geometry
  - Every Ax lies within the column space of A. In that space, we look for the point closest to b, which is the projection p.
  - The best choice for $A\hat{x}$ is p. The smallest possible error is $e = b - p$.
  - The three points of height (p_1, p_2, p_3) do lie on a line because p is in the column space. In fitting a straight line, $\hat{x}$ gives the best choice for (C, D)
- By algebra
  - Every vector b splits into two parts. The part in the column space is p. The perpendicular part in the nullspace of A^T is e.
  - We cannot solve $Ax = b$, but we can solve $A\hat{x} = p$ since p is in the column space

    $Ax = b = p + e$ is impossible ; $A\hat{x} = p$ is solvable

  - The solution to $A\hat{x} = p$ leaves the least possible error (which is e)
    Squared length for any x: $\|Ax - b\|^2 = \|Ax - p\|^2 + \|e\|^2$
    This is from the pythagorean theorem. The vector $Ax - p$ in the column space is perpendicular to e in the left nullspace.
  - We reduce $Ax - p$ to zero by choosing x to be $\hat{x}$. This leaves the smallest possible error $e = (e_1, e_2, e_3)$
- We minimize the squared length of $Ax - b$, hence the name
- The least squares solution $\hat{x}$ makes $E = \|Ax - b\|^2$ as small as possible



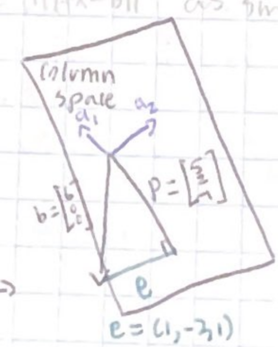


fig 4.3.1
A line of best fit vs. its linear algebra counterpart

- fig 4.3.1 shows the closest line, which minimizes the vertical distance between corresponding points on the line.
  It also shows the linear algebra approach
- By Calculus
  - We want to minimized the error function $E = e_1^2 + e_2^2 + e_3^2$

$$E = \|Ax - b\|^2 = (C + 0D - 6)^2 + (C + D - 0)^2 + (C + 2D - 0)^2$$

$$= (C-6)^2 + (C+D)^2 + (C+2D)^2$$

  - Since we have 2 unknowns (C and D), we have two derivatives. These are partial derivatives with respect to each unknown which give us two equations when set to zero

$$\frac{\partial E}{\partial C} = 2(C-6) + 2(C+D) + 2(C+2D) \quad \text{set } \frac{\partial E}{\partial C} = 0 \Rightarrow 3C + 3D = 6$$

$$\frac{\partial E}{\partial D} = 2(C+D) + 4(C+2D) \quad \text{set } \frac{\partial E}{\partial D} = 0 \Rightarrow 3C + 5D = 0$$

Write equations as matrix

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \text{ is our } A^TA$$

$$A^TA \quad \hat{x} \quad = A^Tb$$

  - These equations are the same as the ones we got through linear algebra
    The partial derivatives of $\|Ax - b\|^2$ are zero when $A^TA\hat{x} = A^Tb$

The Big Picture

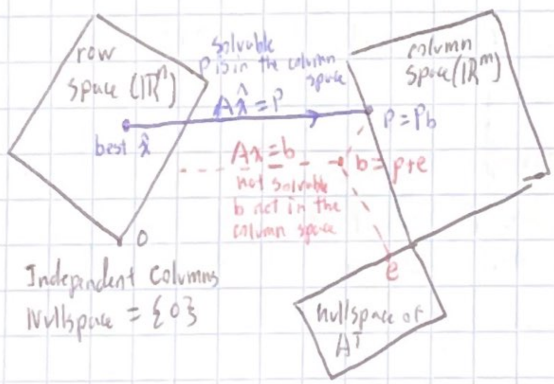


Fig 4.3.2
The projection $p = A\hat{x}$ is closest to b, so $\hat{x}$ minimizes $E = \|b - Ax\|^2$

- In previous chapters, we split x into $x_r + x_n$. There were many solutions to $Ax = b$. Now, we do the opposite. There are no solutions to $Ax = b$. Instead of splitting x, we are splitting b into p (in the column space) and an orthogonal e (in the left nullspace). Now we solve $A\hat{x} = p$; the error $e = b - p$ is unavoidable.
- Notice how the nullspace $N(A)$ is very small; just Z. This is because the columns are independent. Then A^TA is invertible. The equation $A^TA\hat{x} = A^Tb$ describes the best vector $\hat{x}$. The error has $A^Te = 0$

Fitting a Straight Line

- Fitting a clear line is the clearest application of least squares. It starts with $m > 2$ points. At times $t_1, \dots, t_m$ those m points are at heights $b_1, \dots, b_m$. The best line $C + Dt$ misses the points by vertical distances $e_1, \dots, e_m$. Least squares minimizes $e_1^2 + \dots + e_m^2$.
- A line goes through the m points when we exactly solve $Ax = b$. Generally we can't do it.
- To fit the m points, we try to solve m equations with two unknowns each.

$$Ax = b \text{ is } \begin{cases} C + Dt_1 = b_1 \\ \vdots \\ C + Dt_m = b_m \end{cases} \quad \text{with } A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$$

- The column space is so thin that b is almost certainly outside of it. When b happens to be in the column space, the points lie on a line and $b = p$. Then $Ax = b$ is solvable and the errors are $e = (0, \dots, 0)$.
- The closest line $C + Dt$ has heights $p_1, \dots, p_m$ with errors $e_1, \dots, e_m$. Solve $A^T A\hat{x} = A^T b$ for $\hat{x} = (C, D)$. The errors are $e_i = b_i - C - Dt_i$
  - The two columns of A are independent unless all times t_i are the same.

Dot-product Matrix $A^T A = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum_1 t_i \\ \sum_1 t_i & \sum_1 t_i^2 \end{bmatrix}$

On the right side:

$A^T b \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum_1 b_i \\ \sum_1 t_i b_i \end{bmatrix}$

Making the substitutions

The line $C + Dt$ minimizes $e_1^2 + \dots + e_m^2 = \|Ax - b\|^2$ when $A^T A\hat{x} = A^T b$

$$\begin{bmatrix} m & \sum_1 t_i \\ \sum_1 t_i & \sum_1 t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum_1 b_i \\ \sum_1 t_i b_i \end{bmatrix}$$

- The vertical errors at the m points of the line are the components of $e = b - p$. This error vector (the residual) $b - A\hat{x}$ is perpendicular to the columns of A.
- The best $\hat{x} = (C, D)$ minimizes the sum of the errors squared

$$E = \|Ax - b\|^2 = (C + Dt_1 - b_1)^2 + \dots + (C + Dt_m - b_m)^2$$

  - When the derivatives $\frac{\partial E}{\partial D}$ and $\frac{\partial E}{\partial C}$ are set to zero, it produces $A^T A\hat{x} = A^T b$
- Other least squares can have more parameters. Fitting a parabola takes 3 coefficients. In general we are fitting m data points by n parameters $x_1, \dots, x_n$. The matrix has n columns and $n < m$.
- The derivative of a square is linear, which is why we can approach this with calculus and linear algebra

- Example 2
  - A has orthogonal columns when the measurement times t_i add to zero. Suppose $b = 1, 2, 4$ at times $t = -2, 0, 2$. Those times add to zero. The columns of A here a zero dot product.

$$\begin{cases} C + D(-2) = 1 \\ C + D(0) = 2 \\ C + D(2) = 4 \end{cases} \quad \text{or} \quad Ax = b \Rightarrow \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb \text{ is } \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$$

  - Notice how A^TA is now diagonal. We can solve for $C = \frac{7}{3}$ and $D = \frac{3}{4}$ separately. The diagonal matrix A^TA with entries $m = 3$ and $\sum_1^3 t_i^2 = t_1^2 + t_2^2 + t_3^2 = 8$ is virtually as good as the identity matrix.
  - Orthogonal columns are so useful it is worth moving the time origin to produce them. To do that, we subtract the average time $\hat{t} = (t_1 + t_2 + \ldots + t_m)/m$ from each time. The shifted times $T_i = t_i - \hat{t}$ sum to $\sum_1^m T_i = m\hat{t} - m\hat{t} = 0$, which makes A^TA diagonal. A^TA's entries are now m and $\sum_1^m T_i^2$, giving the best C and D direct formulas

$$T_i = t_i - \hat{t} \qquad C = \frac{b_1 + \ldots + b_m}{m} \quad \text{and} \quad D = \frac{b_1T_1 + \ldots + b_mT_m}{T_1^2 + \ldots + T_m^2}$$

and the best line is $C + DT$ or $C + D(t - \hat{t})$.

  - This time shift which makes A^TA diagonal is an example of the Gram-Schmidt process: orthogonalize the columns in advance.

Fitting a Parabola

- A parabola $b = C + Dt + t^2$, even though it is quadratic, it coefficients are still linear.
- Problem: Fit heights $b_1, \ldots, b_m$ at times $t_1, \ldots, t_m$ by a parabola $C + Dt + Et^2$. Solution: with $m > 3$ points, the m equations for an exact fit are generally unsolvable

$$\begin{cases} C + Dt_1 + Et_1^2 = b_1 \\ \vdots \\ C + Dt_m + Et_m^2 = b_m \end{cases} \quad \text{or} \quad Ax = b \Rightarrow \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Least Squares: The closest parabola $C + Dt + Et^2$ chooses $\hat{x} = (C, D, E)$ to satisfy the three normal equations $A^TA\hat{x} = A^Tb$

- The column space of A has dimension 3. We project b onto the column space of A. Our error is $e_i = b_i - C - Dt_i - Et_i^2$. Total error $= \sum_1^m e_i^2 = \sum_1^m (b_i - C - Dt_i - Et_i^2)^2$
- By calculus, we take partial derivatives with respect to C, D, and E and solve the 3 by 3 system generated

- Example 3
  - Given the points $(0,6), (1,0), (2,0)$

$$\begin{cases} C + D(0) + (0)^2 = 6 \\ C + D(1) + E(1)^2 = 0 \\ C + D(2) + E(2)^2 = 0 \end{cases} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \Rightarrow x = \begin{bmatrix} 6 \\ -9 \\ 3 \end{bmatrix}$$

we can solve this directly since $m=3$.

Problem Set 4.3

1.

$$\begin{cases} C + (0)D = 0 \\ C + (1)D = 8 \\ C + (3)D = 8 \\ C + (4)D = 20 \end{cases} \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb \Rightarrow \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \end{bmatrix}$$

$$\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \Rightarrow b = 1 + 4t$$

2.

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 13 \\ 17 \end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

4.

$$E = \|Ax - b\|^2 = (C + (0)D - 0)^2 + (C + (1)D - 8)^2 + (C + (3)D - 8)^2 + (C + (4)D - 20)^2$$

$$\frac{\partial E}{\partial C} = 2C + 2(C + D - 8) + 2(C + 3D - 8) + 2(C + 4D - 20) = 0$$

$$= 8C + 16D - 72 = 0$$

$$4C + 8D = 36$$

$$\frac{\partial E}{\partial D} = 2(C + D - 8) + 6(C + 3D - 8) + 8(C + 4D - 20)$$

$$= 16C + 52D - 224 = 0$$

$$8C + 26D = 112$$

5.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [C] = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 4 \end{bmatrix}$$

$$A^TA = [1\ 1\ 1\ 1] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = [4]$$

$$A^Tb = [1\ 1\ 1\ 1] \begin{bmatrix} 0 \\ 8 \\ 8 \\ 4 \end{bmatrix} = 20$$

$4C = 20$

$C = 5$

6.

$$\hat{x} = \frac{a^Tb}{a^Ta} = \frac{36}{4} = 9$$

$p = \hat{x}a = (9,9,9,9)$

$e = b - p = (0,8,8,20) - (9,9,9,9)$

$= (-9,-1,-1,11)$

9.

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{bmatrix} \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \\ 400 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} = \begin{bmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \\ 400 \end{bmatrix} \Rightarrow \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{4}{3} \\ \frac{2}{3} \end{bmatrix}$$

$$y = \frac{2}{3}x^2 + \frac{4}{3}x + 2$$

12. a)

$\hat{x} = \frac{a^T b}{a^T a}$

$= \frac{b_1 + b_2 + \dots + b_m}{m}$

b)

$P = a\hat{x}$

$= \begin{bmatrix} \frac{b_1 + b_2 + \dots + b_m}{m} \\ \vdots \\ \frac{b_1 + b_2 + \dots + b_m}{m} \end{bmatrix}$

$e = b - p$

$= \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} - \begin{bmatrix} \frac{b_1 + \dots + b_m}{m} \\ \vdots \\ \frac{b_1 + \dots + b_m}{m} \end{bmatrix}$

$= \begin{bmatrix} b_1 - \frac{b_1 + \dots + b_m}{m} \\ \vdots \\ b_m - \frac{b_1 + \dots + b_m}{m} \end{bmatrix}$

$\|e\|^2 = \sum_1^m (b_i - \bar{b})^2$

17.

$\vec{PQ} = (y - x, 3y - x, -1 - x)$

$\vec{PQ} \cdot (1,1,1) = 0$

$y - x + 3y - x - 1 - x = 0$

$-3x + 4y = 1$

$\vec{PQ} \cdot (1,3,0) = 0$

$y - x + 9y - 3x = 0$

$-4x + 10y = 0$

$x = \frac{10}{4}y$

$-3(\frac{10}{4}y) + 4y = 1$

$y = -\frac{2}{7}$

$x = -\frac{5}{7}$

13.

$(A^TA)^{-1}A^Te = (A^TA)^{-1}A^Tb - (A^TA)^{-1}A^TAx$

$= \hat{x} - x$

17.

$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix}$ no solution

$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

$A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix} = \begin{bmatrix} 35 \\ 42 \end{bmatrix}$

$A^TAx = A^Tb \Rightarrow \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 35 \\ 42 \end{bmatrix}$

$\Rightarrow \begin{bmatrix} 3 & 2 \\ 0 & \frac{14}{3} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 35 \\ \frac{56}{3} \end{bmatrix}$

$\hat{x} = \begin{bmatrix} 9 \\ 4 \end{bmatrix}$ $\quad y = 9 + 4x$

18.

$p = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 9 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 13 \\ 17 \end{bmatrix}$

$e = b - p = \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix} - \begin{bmatrix} 5 \\ 13 \\ 7 \end{bmatrix}$

$= \begin{bmatrix} 2 \\ -6 \\ 14 \end{bmatrix}$

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

$(A^TA)^{-1} = \frac{1}{14} \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}$

$P = \frac{1}{14} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix}$

$= \frac{1}{14} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 8 & 4 & 2 \\ -5 & 1 & 4 \end{bmatrix}$

$= \frac{1}{14} \begin{bmatrix} 13 & 3 & -2 \\ 3 & 5 & 6 \\ -2 & 6 & 10 \end{bmatrix}$

$Pe = \frac{1}{14} \begin{bmatrix} 13 & 3 & -2 \\ 3 & 5 & 6 \\ -2 & 6 & 10 \end{bmatrix} \begin{bmatrix} 2 \\ -6 \\ 14 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} \ \end{bmatrix}$

Chapter 4.4: Orthogonal Bases and Gram-Schmidt

- This section has two goals: to illustrate the usefulness of orthogonal bases in finding $\hat{x}$, p, and P. The dot products are zero, so A^TA becomes diagonal. Secondly, is to construct orthogonal vectors. We will pick combinations of the original vectors to produce right angle. The original vectors are columns of A, probably not orthogonal. Orthogonal vectors will be columns of a new matrix Q.
- The vectors $q_1, q_2, \ldots, q_n$ are orthogonal when their dot products $q_i \cdot q_j$ are zero. More specifically $q_i^T q_j = 0$ when $i \neq j$. If we divide all vectors by its length to get unit vectors, we get orthogonal unit vectors. Then this basis is called orthonormal.
- The vectors $q_1, \ldots, q_n$ are orthonormal if:

$$q_i^T q_j = \begin{cases} 0 & \text{when } i \neq j \text{ (orthogonal vectors)} \\ 1 & \text{when } i = j \text{ (unit vectors: } \|q_i\| = 1) \end{cases}$$

A matrix with orthonormal columns is assigned the letter Q

- The matrix Q is easy to work with because $Q^TQ = I$. Q does not have to be square.

$$Q^TQ = \begin{bmatrix} - & q_1^T & - \\ - & q_2^T & - \\ & \vdots & \\ - & q_n^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- When Q is square, $Q^TQ = I$ means $Q^T = Q^{-1}$: transpose = inverse.
- If the columns are only orthogonal, dot products still give a diagonal matrix.
- $Q^TQ = I$ even when Q is rectangular, but it is only a left-inverse.

For square matrices, we have $Q^TQ = QQ^T = I$. The rows of a square Q is also orthonormal columns. We call a square Q an orthogonal/orthonormal matrix.

- Here are three examples of orthogonal matrices:

1. Rotation
   - Q rotates every vector in the plane clockwise by the angle θ

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{and} \quad Q^T = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

   - These columns give an orthonormal basis for $\mathbb{R}^2$
   - Q^T rotates by $-\theta$

2. Permutation
   - Q changes the order of the entries of a vector

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

   - The inverse of any permutation matrix is its transpose and vice versa

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

3. Reflection

- If u is any unit vector, set $Q = I - 2uu^T$. Then $Q^T = Q^{-1} = Q$

$$Q^T = I - 2uu^T = Q \quad \text{and} \quad Q^TQ = I - 4uu^T + 4uu^Tuu^T = I$$

- Reflection matrices $I - 2uu^T$ are symmetric and also orthogonal. Squaring them gets the identity matrix $Q^2 = Q^TQ = I$. Reflecting twice gets back the original vector
- Example:

Given $u = (1,0)$

$$I - 2uu^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = Q$$

→ Q reflects $(x,0)$ across the y axis to $(-x,0)$

- Multiplication by any orthogonal matrix leaves lengths and angles unchanged.
- If Q has orthonormal columns ($Q^TQ = I$), it leaves lengths unchanged.
  - Same length: $\|Qx\| = \|x\| \ \forall x \in V$
  - Dot products preserved: $(Qx)^T(Qy) = x^TQ^TQy = x^Ty$
- Orthogonal matrices are very useful in computation because it keeps numbers relatively small.

Projections using Orthogonal bases: Q replaces A

- In all our formulas for $\hat{x}$, p, and P onto subspaces, we see A^TA. Imagine if they were orthonormal: Then Q^TQ simplifies to I. We have:

$$Q^TQ\hat{x} = Q^Tb \Rightarrow \hat{x} = Q^Tb, \quad p = Q\hat{x} \quad P = Q(Q^TQ)^{-1}Q^T = QQ^T$$

  - The least squares solution of $Qx = b$ is $\hat{x} = Q^Tb$. The project matrix is $P = QQ^T$

$$P = QQ^Tb = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix}\begin{bmatrix} q_1^Tb \\ \vdots \\ q_n^Tb \end{bmatrix} = q_1(q_1^Tb) + \dots + q_n(q_n^Tb)$$

- When Q is square, the subspace is the whole space. Then $Q^T = Q^{-1}$ and $\hat{x} = Q^Tb = Q^{-1}b$ and we have an exact solution. In this case $P = QQ^T$ because b is being projected onto a subspace containing it.
  - When $p = b$, our formula constructs b out of one dimensional projections

$$q_i\left(\frac{q_i^Tb}{q_i^Tq_i}\right) = p \quad q_i^Tq_i = 1 \text{ because } \|q_i\| = 1. \text{ Therefore, we get } q_i(q_i^Tb) = p$$

  - This is the foundation of famous transforms, like the Fourier transform, which break vectors / functions into perpendicular pieces. The inverse transform then builds these pieces together

- Example 4:

$$Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$$

The separate projections of $b = (0,0,1)$ onto q_1 and q_2 and q_3 are p_1 and p_2 and p_3.

$$p_1 = q_1(q_1^T b) = \frac{2}{3}q_1 \qquad p_2 = q_2(q_2^T b) = \frac{2}{3}q_2 \qquad p_3 = q_3(q_3^T b) = -\frac{1}{3}q_3$$

The sum of the first two is the projection of b onto the plane of q_1 and q_2.
The sum of all three is the projection onto the whole space, which is b itself

$$b = p_1 + p_2 + p_3 \qquad \frac{2}{3}q_1 + \frac{2}{3}q_2 - \frac{1}{3}q_3 = \frac{1}{9}\begin{bmatrix} -2+4-2 \\ 4-2-2 \\ 4+4+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = b$$

The Gram-Schmitt Process

- Orthonormal vectors "uncouple" the projections. A projection can be thought of as the "part" of a vector in a subspace. If our vectors are orthogonal to each other, no vector "bleeds into" another.
- How do we create orthonormal vectors?
  - We start with three independent vectors a, b, c. We intend to construct three orthogonal vectors A, B, and C. Then we divide by the lengths to get unit vectors.
  - Gram-Schmit: Begin by choosing $A = a$. The next direction B must be perpendicular to A. We start with b and subtract its projection along A. This leaves us the perpendicular part, the vector B (the error vector e)
  - A and B are then orthogonal. $A^T B = A^T b - A^T b = 0$.
  - The third direction starts with C. This is not a combination of A and B (because c is not a combination of a and b). Most likely c is not perpendicular to A and B, so we subtract off its components in those 2 directions to get C

$$A = a$$

$$B = b - \text{proj}_A(b) = b - \frac{A^T b}{A^T A}A$$

$$C = c - \text{proj}_A(c) - \text{proj}_B(c) = c - \frac{A^T c}{A^T A}A - \frac{B^T c}{B^T B}B$$

$$u_k = v_k - \sum_{j=1}^{k-1} \text{proj}_{u_j}(v_k) = v_k - \sum_{j=1}^{k-1} \frac{u_j^T v_k}{u_j^T u_j}$$

- Example

$$a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \quad c = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}$$

$A = a \qquad q_1 = \frac{A}{\|A\|} = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0\right)$

$B = b - \frac{A^T b}{A^T A} A = b - \frac{2}{2} A = (1, 1, -2) \qquad q_2 = \frac{B}{\|B\|} = \left(\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, -\frac{\sqrt{6}}{3}\right)$

$C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B = c - \frac{6}{2} A + \frac{6}{6} B = (1, 1, 1) \qquad q_3 = \frac{C}{\|C\|} = \left(\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}\right)$

The Factorization $A = QR$

- We started with a matrix A with columns a, b, c. We ended with a matrix Q with columns q_1, q_2, q_3. Since the vectors a, b, c are combinations of q_1, q_2, q_3 and vice versa, there must be a third matrix connecting A to Q. This matrix is a triangular R in $A = QR$.
- The first step is $q_1 = \frac{a}{\|a\|}$ (other vectors not involved). Then we iterate, subtract from each new vector its projections onto previous vectors. This non-involvement of later vectors is the key point of Gram-Schmidt.
- The vectors a, A, q_1 are along one line
- The vectors a, b, A, B, q_1, q_2 are in the same plane
- The vectors $a, b, c, A, B, C, q_1, q_2, q_3$ are in one subspace ($\mathbb{R}^3$)
- At every step $a_1, \ldots, a_k$ are combinations of $q_1, \ldots, q_k$. Later q's are not involved. Therefore the connecting matrix R is triangular and we have $A = QR$

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^T a & q_1^T b & q_1^T c \\ & q_2^T b & q_2^T c \\ & & q_3^T c \end{bmatrix} \quad \text{or} \quad A = QR$$

- From independent vectors $a_1, \ldots, a_n$, Gram-Schmidt constructs orthonormal vectors $q_1, \ldots, q_n$. The matrices with these columns satisfy $A = QR$. Then $R = Q^T A$ is upper triangular because the later q's are orthogonal to earlier a's.
- From the example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & -\frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ & \sqrt{6} & -\sqrt{6} \\ & & \sqrt{3} \end{bmatrix}$$

- Any m by n matrix A with independent columns can be factored into QR. The m by n matrix Q has orthonormal columns and the matrix R is upper triangular with a positive diagonal.
- For least squares, $A^T A = R^T Q^T Q R = R^T R$ and the least squares equation simplifies $A^T A \hat{x} = A^T b \Rightarrow R^T R \hat{x} = R^T Q^T b \Rightarrow R\hat{x} = Q^T b$ or $\hat{x} = R^{-1} Q^T b$.
- We solve $R\hat{x} = Q^T b$ by back substitution, which is really fast. The real cost is mn^2 multiplications for the Gram-Schmidt process.

Problem Set 4.4

11

a)

$A = a = (1,3,4,5,7)$

$B = b - \frac{A^Tb}{A^TA}A$

$= b - \frac{100}{100}A$

$= (-7,3,4,-5,1)$

$q_1 = \frac{1}{100}A = (\frac{1}{100}, \frac{3}{100}, \frac{1}{25}, \frac{1}{20}, \frac{7}{100})$

$q_2 = \frac{1}{100}B = (-\frac{7}{100}, \frac{3}{100}, \frac{1}{25}, -\frac{1}{20}, \frac{1}{100})$

b)

$Q^TQ\hat{x} = Q^Tb$

$\hat{x} = Q^Tb$

$\hat{x} = \begin{bmatrix} \frac{1}{100} & \frac{3}{100} & \frac{1}{25} & \frac{1}{20} & \frac{7}{100} \\ -\frac{7}{100} & \frac{3}{100} & \frac{1}{25} & \frac{1}{20} & \frac{1}{100} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$= \begin{bmatrix} \frac{1}{100} \\ -\frac{7}{100} \end{bmatrix}$

$p = A\hat{x} = \begin{bmatrix} \frac{1}{100} & -\frac{7}{100} \\ \frac{3}{100} & \frac{3}{100} \\ \frac{1}{25} & \frac{1}{25} \\ \frac{1}{20} & \frac{1}{20} \\ \frac{7}{100} & \frac{1}{100} \end{bmatrix} \begin{bmatrix} \frac{1}{100} \\ -\frac{7}{100} \end{bmatrix} = \begin{bmatrix} \frac{1}{200} \\ 0 \\ -\frac{1}{500} \\ \\ \\ \end{bmatrix}$

15.

$a = (1,2,-2)$

$b = (1,-1,4)$

$A = a = (1,2,-2)$

$B = b - \frac{A^Tb}{A^TA}A$

$= b - \frac{-9}{9}A$

$= b + A$

$= (2,1,2)$

$q_1 = (\frac{1}{3}, \frac{2}{3}, -\frac{2}{3})$

$q_2 = (\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$

$q_3 = (\frac{2}{3}, -\frac{2}{3}, -\frac{1}{3}) = q_1 \times q_2$

$q_3 \in N(A^T)$

$Q^Tb = \frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} -9 \\ 18 \end{bmatrix}$ $R = Q^{-1}A$

$\cancel{Q^TA = QR}$

$Ax = b$

$Rx = Q^Tb$

$R = Q^TA$

$= \frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{bmatrix}$

$= \frac{1}{3}\begin{bmatrix} 9 & -9 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} 3 & -3 \\ 0 & 3 \end{bmatrix}$

$\begin{bmatrix} 3 & -3 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -9 \\ 18 \end{bmatrix}$

$\hat{x} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$

17.

$\text{proj}_a(b) = \frac{a^Tb}{a^Ta}a = \frac{9}{3}a = 3a = (3,3,3)$

$e = b - p = (1,3,5) - (3,3,3) = (-2,0,2)$

$q_1 = \frac{1}{\sqrt{3}}(1,1,1)$

$q_2 = \frac{1}{2\sqrt{2}}(-2,0,2)$

18.

$A = a = (1,-1,0,0)$

$B = b - \frac{A^Tb}{A^TA}A$

$= b + \frac{1}{2}A$

$= (\frac{1}{2}, \frac{1}{2}, -1, 0)$

$C = c - \frac{A^Tc}{A^TA}A - \frac{B^Tc}{B^TB}B$

$= c - 0A + \frac{2}{3}B$

$= (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -1)$

19.

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

$a_1 = (-1,2,2), a_2 = (1,1,4)$

$v_1 = a_1 = (-1,2,2)$

$v_2 = a_2 - \frac{a_1^Ta_2}{a_1^Ta_1}a_1$

$= a_2 - \frac{9}{9}a_1$

$= a_2 - a_1$

$= (2,-1,2)$

$q_1 = \frac{1}{\|v_1\|}v_1 = \frac{1}{\sqrt{9}}v_1$

$= (-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$

$q_2 = \frac{1}{\|v_2\|}v_2 = \frac{1}{\sqrt{9}}v_2$

$= (\frac{2}{3}, -\frac{1}{3}, \frac{2}{3})$

$Q = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$

$A = QR$

$A^TA = (QR)^T(QR)$

$= R^TQ^TQR$

$= R^TR$

$A^TA = \begin{bmatrix} -1 & 2 & 2 \\ 1 & 1 & 4 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 2 & 1 \\ 2 & 4 \end{bmatrix}$

$= \begin{bmatrix} 9 & 9 \\ 9 & 18 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 9 & 9 \\ 0 & 9 \end{bmatrix}$

$= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 9 & \\ & 9 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}$

21.

$$A = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{bmatrix}$$

$a_1 = (1, 1, 1, 1)$

$a_2 = (-2, 0, 1, 3)$

$v_1 = a_1 = (1, 1, 1, 1)$

$v_2 = a_2 - \frac{v_1^T a_2}{v_1^T v_1} v_1$

$= a_2 - \frac{2}{4} v_1$

$= a_2 - \frac{1}{2} v_1$

$= (-2, 0, 1, 3) - \frac{1}{2}(1, 1, 1, 1)$

$= (-\frac{5}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{5}{2})$

$q_1 = \frac{1}{\|v_1\|} v_1 = \frac{1}{2}(1, 1, 1, 1)$

$= (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

$q_2 = \frac{1}{\|v_2\|} v_2 = \frac{1}{\sqrt{13}}(-\frac{5}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{5}{2})$

$= (-\frac{5\sqrt{13}}{26}, -\frac{\sqrt{13}}{26}, \frac{\sqrt{13}}{26}, \frac{5\sqrt{13}}{26})$

$$Q = \begin{bmatrix} \frac{1}{2} & -\frac{5\sqrt{13}}{26} \\ \frac{1}{2} & -\frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{5\sqrt{13}}{26} \end{bmatrix}$$

$Q^T Q \hat{x} = Q^T b$

$\hat{x} = Q^T b$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{5\sqrt{13}}{26} & -\frac{\sqrt{13}}{26} & \frac{\sqrt{13}}{26} & \frac{5\sqrt{13}}{26} \end{bmatrix} \begin{bmatrix} -4 \\ -3 \\ 3 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -2 \\ \sqrt{13} \end{bmatrix}$$

$p = Q\hat{x}$

$$= \begin{bmatrix} \frac{1}{2} & -\frac{5\sqrt{13}}{26} \\ \frac{1}{2} & -\frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{\sqrt{13}}{26} \\ \frac{1}{2} & \frac{5\sqrt{13}}{26} \end{bmatrix} \begin{bmatrix} -2 \\ \sqrt{13} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{7}{2} \\ -\frac{3}{2} \\ -\frac{1}{2} \\ \frac{3}{2} \end{bmatrix}$$

23.

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$$

$a_1 = (1, 0, 0)$

$a_2 = (2, 0, 3)$

$a_3 = (4, 5, 6)$

$v_1 = a_1 = (1, 0, 0)$

$v_2 = a_2 - \frac{v_1^T a_2}{v_1^T v_1} v_1$

$= a_2 - \frac{2}{1} v_1$

$= a_2 - 2v_1$

$= (0, 0, 3)$

$v_3 = a_3 - \frac{v_1^T a_3}{v_1^T v_1} v_1 - \frac{v_2^T a_3}{v_2^T v_2} v_2$

$= a_3 - \frac{4}{1} v_1 - \frac{18}{9} v_2$

$= a_3 - 4v_1 - 2v_2$

$= (0, 5, 0)$

$q_1 = (1, 0, 0)$

$q_2 = (0, 0, 1)$

$q_3 = (0, 1, 0)$

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

$A = QR$

$R = Q^{-1}A = Q^T A$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

30.

$$W = \frac{1}{4} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{\sqrt{2}}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix} = \frac{1}{4} Q$$

$W^{-1} = 4Q^T$

$$= \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & -2 & -2 \\ 2\sqrt{2} & -2\sqrt{2} & 0 & 0 \\ 0 & 0 & 2\sqrt{2} & 2\sqrt{2} \end{bmatrix}$$