

217 Transpose and Least Squares
Professor Karen Smith

(c)2015 UM Math Dept
 licensed under a Creative Commons
 By-NC-SA 4.0 International License.

A. Proposition: *If A and B are matrices s.t. the product AB is defined, then $(AB)^T = B^T A^T$.*

1. Verify this proposition for each pair of matrices A and B below.

(a) $A = [a_1 \ a_2 \ a_3], B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$.

Solution note: Both products $(AB)^T$ and $B^T A^T$ yield $a_1 b_1 + a_2 b_2 + a_3 b_3$.

(b) $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

Solution note: Both products yield $\begin{bmatrix} -2 & -4 \\ 1 & 3 \end{bmatrix}$.

2. Think about the dimensions of the matrices A and B in the proposition. Does it make sense that the product $B^T A^T$ is defined and has the same dimensions as $(AB)^T$.
3. Use the Proposition to prove that $(ABC)^T = C^T B^T A^T$.

Solution note: $[(AB)C]^T = C^T(AB)^T = C^T B^T A^T$.

4. Prove the Proposition in the case where A is $1 \times n$ and B is $n \times 1$.

Solution note: Say $A = [a_1 \ \dots \ a_n]$ and $B = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$. Then $AB = \sum_{i=1}^n a_i b_i$, which is a 1×1 matrix so equal to its transpose $(AB)^T$. Also $B^T A^T = [b_1 \ \dots \ b_n] \begin{bmatrix} a_1 \\ \dots \\ a_n \end{bmatrix}$, which gives the same product.

5. Prove the Proposition by writing A as a column of rows $\begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix}$ and B as a row of columns $[C_1 \ C_2 \ \dots \ C_n]$ and computing both products AB and $B^T A^T$.

Solution note: $AB = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} [C_1 \ C_2 \ \cdots \ C_n] = \begin{bmatrix} R_1C_1 & R_1C_2 & \cdots & R_1C_n \\ R_2C_1 & R_2C_2 & \cdots & R_2C_n \\ \vdots & \vdots & \ddots & \vdots \\ R_nC_1 & R_nC_2 & \cdots & R_nC_n \end{bmatrix}$. So

transposing, we get $(AB)^T = \begin{bmatrix} R_1C_1 & R_2C_1 & \cdots & R_nC_1 \\ R_1C_2 & R_2C_2 & \cdots & R_nC_2 \\ \vdots & \vdots & \ddots & \vdots \\ R_1C_n & R_2C_n & \cdots & R_nC_n \end{bmatrix}$. Each $R_iC_j = (R_iC_j)^T = C_j^T R_i^T$, since it is a scalar. So this product is also the same as

$$B^T A^T = \begin{bmatrix} C_1^T \\ C_2^T \\ \vdots \\ C_n^T \end{bmatrix} [R_1^T \ R_2^T \ \cdots \ R_n^T].$$

B. Theorem: Let A be an $m \times n$ matrix. Then for all $\vec{x} \in \mathbb{R}^m$ and $\vec{y} \in \mathbb{R}^n$, we have an equality

$$A\vec{x} \cdot \vec{y} = \vec{x} \cdot A^T\vec{y}.$$

1. Verify this theorem for the matrix $A = I_n$ and any \vec{x} and \vec{y} in \mathbb{R}^n
2. Verify this theorem for the matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$ and any \vec{x} and \vec{y} in \mathbb{R}^2 .
3. Prove the Theorem by interpreting the dot product of two vectors \vec{w} and \vec{v} in \mathbb{R}^n as a matrix product $\vec{w}^T \vec{v}$.

Solution note: $\vec{A}\vec{x} \cdot \vec{y} = (A\vec{x})^T \vec{y} = (\vec{x}^T) A^T \vec{y} = \vec{x}^T (A^T \vec{y}) = \vec{x}^T \cdot A^T \vec{y}$.

C. Theorem: If A is an $m \times n$ matrix, then $\ker A^T = (\text{im } A)^\perp$. Also, $(\ker A^T)^\perp = (\text{im } A)$.

1. Use this theorem to show that A and A^T have the same rank.
2. Prove the theorem. [Hint: Scaffold the proof first, breaking it down into a series of standard steps. Then the theorem from Problem B.]

Solution note: Take $x \in \ker A^T$. We need $y \cdot x = 0$ for all $y \in \text{im } A$. We need $Az \cdot x = 0$ for all z . By the Theorem in B, this is the same as $z \cdot A^T x = 0$ for all z . Of course, this true because $A^T x = 0$ (def of kernel). For the reverse inclusion: say $y \in (\text{im } A)^\perp$. This means $Ax \cdot y = 0$ for all x . This is the same as $x \cdot A^T y = 0$ for all x . This means $A^T y = 0$, so $y \in \ker A^T$. The second statement holds by "perping" both sides.

LEAST SQUARES APPROXIMATION.

D. Let A be a $m \times n$ matrix and let \vec{b} be a vector in \mathbb{R}^m .

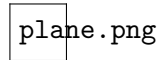
1. **Prove or Disprove:** $A\vec{x} = \vec{b}$ is consistent if and only if \vec{b} is in the image of A .

Solution note: TRUE! The image of A is spanned by the columns of A , so \vec{b} is in the image of A if and only if \vec{b} is a linear combination of the columns of A . Suppose A has columns $\vec{v}_1, \dots, \vec{v}_n$. Then the product

$$A\vec{x} = [\vec{v}_1 \ \vec{v}_2 \ \dots \ \vec{v}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{v}_1 + x_2\vec{v}_2 + \dots + x_n\vec{v}_n$$

which says that $A\vec{x} = \vec{b}$ has a solution if and only if \vec{b} is a linear combination of $\vec{v}_1, \dots, \vec{v}_n$.

2. Suppose you came up with the linear system $A\vec{x} = \vec{b}$ based on real-life measurements in the field (say, 16 equations in 20 variables) whose solution is the answer to all your problems. The figure shows your \vec{b} and the image of your matrix A (meaning the span of its columns or the linear map given by multiplication by A). What do you tell your boss? Can you solve all your problems?



Solution note: No, not exactly. The vector \vec{b} is not in the space spanned by the columns of A , which means the system $A\vec{x} = \vec{b}$ has no solution!

3. Can you come up with a different linear system, say $A\vec{x} = \vec{b}^*$, which is pretty close to yours and *consistent*? Closer than any other system of the form $A\vec{x} = \vec{y}$ for different \vec{y} ? Sketch \vec{b}^* in the figure.

Solution note: Yes! Project \vec{b} onto the space $\text{im } A$. Call the projection \vec{b}^* . The system $A\vec{x} = \vec{b}^*$ is consistent, and has a solution, and we can think of this as the "closest" to a solution for $A\vec{x} = \vec{b}$.

4. Discuss a method with your table for finding an *approximate solution* to $A\vec{x} = \vec{b}$ even when this system is consistent.

Solution note: Solve $A\vec{x} = \vec{b}^*$. Alternatively, you can use the Normal Equation in the book, and solve the system $A^T A\vec{x} = A^T \vec{b}$, which will have the same solutions as $A\vec{x} = \vec{b}^*$.

5. For the toy case

$$\begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \\ -5 & 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix},$$

describe which vectors correspond to \vec{b}, \vec{b}^* and what the green plane is. Explain how to find the least squares solution to this system. Find one.

Solution note: The green plane is the span of the columns of this matrix which is the span of $\begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$, since the third column is dependent on the previous two.

Since this two columns are already perpendicular, it is easy to get an orthonormal basis by scaling each by its length. So an orthonormal basis for $\text{im}A$ is $\vec{u}_1 = \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$

and $\vec{u}_2 = \frac{1}{\sqrt{11}} \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$. We can use this to compute the projection $(\vec{b} \cdot \vec{u}_1)\vec{u}_1 + (\vec{b} \cdot \vec{u}_2)\vec{u}_2$:

$$\begin{aligned} & \left(\begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix} \cdot \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} \right) \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} + \left(\begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix} \cdot \frac{1}{\sqrt{11}} \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{11}} \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} \\ &= \frac{1}{30} \left(\begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} + \frac{1}{11} \left(\begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} \right) \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} \\ &= -2 \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} + \frac{90}{11} \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -112/11 \\ 226/11 \\ 200/11 \end{bmatrix}. \end{aligned}$$

So the consistent linear system we need to solve is

$$\begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \\ -5 & 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -112/11 \\ 226/11 \\ 200/11 \end{bmatrix}.$$

Solve this, using the methods of Chapter 1 (row reduction), to see that the least squares solutions are

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \left\{ \begin{bmatrix} z - 2 \\ -(11z - 90)/11 \\ z \end{bmatrix} \mid z \in \mathbb{R} \right\}$$

6. **Theorem:** *The least squares solutions of $A\vec{x} = \vec{b}$ are the exact solutions of the (consistent) system $A^T A\vec{x} = A^T \vec{b}$. Use this theorem to construct a consistent system whose solutions will best approximate solutions to the original system in (5). Solve it. Do you get the same answer?*

Solution note: The normal equation is $A^T A\vec{x} = A^T \vec{b}$, which give us