

Common Mistakes on Midterm 2

- On this exam, there were a small, but still alarming, number of mistakes against the basic rules of matrix multiplication.

It is impossible to multiply matrices unless the number of **columns** of the matrix on the left is equal to the number of **rows** of the matrix on the right, in which case they multiply like this:

$$\underbrace{\begin{matrix} & n \\ \left[\begin{array}{c} \\ \\ \end{array} \right] & \\ m & \end{matrix}}_{m \times n \text{ matrix}} \underbrace{\begin{matrix} & p \\ \left[\begin{array}{c} \\ \\ \end{array} \right] & \\ n & \end{matrix}}_{n \times p \text{ matrix}} = \underbrace{\begin{matrix} & p \\ \left[\begin{array}{c} \\ \\ \end{array} \right] & \\ m & \end{matrix}}_{m \times p \text{ matrix}}$$

If you have matrices of sizes that do not match this pattern, and you think that you're supposed to multiply them anyway, then put bluntly, you are wrong. It is not an invitation to start throwing away rows or columns until you've gotten something you can multiply (even if the rows or columns are all zeros), or to invent an ad hoc, meaningless way of moving the entries of a matrix around.

- If $\vec{v}$ is a vector in $\mathbb{R}^n$, and $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis of $\mathbb{R}^n$, then to compute $[\vec{v}]_{\mathcal{B}}$ in the basis $\mathcal{B}$, the formula is

$$[\vec{v}]_{\mathcal{B}} = S^{-1}\vec{v} \quad \text{where } S = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$

If A is an $n \times n$ matrix, and $\mathcal{B}$ is a basis of $\mathbb{R}^n$, then “the matrix A expressed in the basis $\mathcal{B}$ ” (which arguably deserves the notation $[A]_{\mathcal{B}}$, but is not given this notation in the book) is computed by

$$[A]_{\mathcal{B}} = S^{-1}AS$$

So it is not merely incorrect to write

$$[\vec{v}]_{\mathcal{B}} = S^{-1}\vec{v}S$$

it is nonsensical in light of the rules for matrix multiplication I reiterated above:

$$\underbrace{\begin{matrix} S^{-1} \\ n \times n \end{matrix}}_{n \times n} \underbrace{\begin{matrix} \vec{v} \\ n \times 1 \end{matrix}}_{n \times 1} \underbrace{\begin{matrix} S \\ n \times n \end{matrix}}_{n \times n}$$

not compatible

Why does the formula for $[A]_{\mathcal{B}}$ have both S and S^{-1} ? Because we want the matrix $[A]_{\mathcal{B}}$ to work like this: for any vector $\vec{v} \in \mathbb{R}^n$,

$$[A]_{\mathcal{B}}[\vec{v}]_{\mathcal{B}} = [A\vec{v}]_{\mathcal{B}}$$

which is true because

$$[A]_{\mathcal{B}}[\vec{v}]_{\mathcal{B}} = S^{-1}AS[\vec{v}]_{\mathcal{B}} = S^{-1}A\vec{v} = [A\vec{v}]_{\mathcal{B}}$$

- Gram-Schmidt in problem 3: when using $\vec{v}_3$ to make a new vector that is perpendicular to the first two new vectors, the formula is

$$\vec{v}_3^\perp = \vec{v}_3 - \underbrace{(\vec{v}_3 \cdot \vec{u}_1)\vec{u}_1}_{\text{Don't forget } \vec{u}_1!} - (\vec{v}_3 \cdot \vec{u}_2)\vec{u}_2 \quad \left(\text{and then of course } \vec{u}_3 = \frac{1}{\|\vec{v}_3^\perp\|} \vec{v}_3^\perp\right)$$

Lots of people forgot the underlined part, which resulted in their getting a vector which was not perpendicular to u_1 .

- Given a subspace V of $\mathbb{R}^n$, there is a **single** matrix that is referred to as “orthogonal projection onto V ”, and it has the notation proj_V . It is an $n \times n$ matrix, reflecting the fact that orthogonal projection onto V is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$,

$$\text{proj}_V : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

and you can compute it with either the formula

$$\text{proj}_V = QQ^T \quad \text{where } Q = \underbrace{n \begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix}}_{n \times k \text{ matrix}} \quad \text{where } v_1, \dots, v_k \text{ is any orthonormal basis of } V$$

or alternatively,

$$\text{proj}_V = A(A^T A)^{-1} A^T \quad \text{where } A = \underbrace{n \begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix}}_{n \times k \text{ matrix}} \quad \text{where } v_1, \dots, v_k \text{ is any basis of } V \text{ at all}$$

(note that k is the dimension of V)

While it's true and important that

$$\text{im}(\text{proj}_V) = V$$

not every matrix whose image is V is orthogonal projection onto V . Many people seemed to make this mistake, and just wrote a matrix like A or Q as their answer.

- For any matrix A whatsoever, the matrix $A^T A$ is always a **symmetric matrix**, i.e. equal to its own transpose $(A^T A)^T = A^T A$, i.e. symmetric around this diagonal:

$$\begin{bmatrix} \cdot & & \\ & \ddots & \\ & & \cdot \end{bmatrix}$$

For example if $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 0 & 4 \end{bmatrix}$, then

$$A^T A = \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 10 & 2 & 15 \\ 2 & 4 & 6 \\ 15 & 6 & 25 \end{bmatrix}$$

which is symmetric. So when doing least-squares calculations and you compute $A^T A$, make sure your answer is symmetric – a lot of people made arithmetic mistakes in the last problem, which caused them to mistakenly conclude that there did not exist any least-squares solutions since they got an inconsistent system of linear equations.

- Which brings me to: **there always exists at least one least-squares solution!**

Sometimes a system of linear equations $A\vec{x} = \vec{b}$ has one exact solution, or many exact solutions, and sometimes it has no exact solutions. That's fine.

But the least-squares solutions are the choices of $\vec{x}$ that make $A\vec{x}$ approximate $\vec{b}$ as well as possible, and no matter if the approximation is terrible, there's always **at least one** choice of $\vec{x}$ that is the **least bad**. Sometimes there are multiple choices of $\vec{x}$ like this, but it makes no sense to say that there exist no least-squares solutions.

- Lastly, to find the least-squares solutions of $A\vec{x} = \vec{b}$, you should solve the system of linear equations

$$(A^T A)\vec{x} = (A^T \vec{b})$$

If, and only if, the kernel of A is equal to $\{\vec{0}\}$, you will have that the matrix $A^T A$ is invertible, and then formula

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

makes sense. When this happens, there is a **single** least-squares solution, namely the result of the formula.

However, in the last problem on the exam, the kernel of A was not just $\{\vec{0}\}$, it was bigger – that makes the matrix $A^T A$ not invertible. But many people tried to use the formula

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

and, finding that $A^T A$ had no inverse, concluded that there did not exist any least squares solutions, which is impossible. The general approach, which always works, is to solve the system of linear equations

$$(A^T A)\vec{x} = (A^T \vec{b})$$

which, remember, is **guaranteed** to have at least one solution – if you find that it is inconsistent, then you made an arithmetic mistake.