# MIT 18.06 Exam 2 Solutions, Fall 2017
# Johnson

## Problem 1 (40 points):

The complete solution to $Ax = b$ is $x = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \alpha_1 \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$ for all possible scalars $\alpha_1$ and $\alpha_2$.

(a) $A$ is an $m \times n$ matrix of rank $r$. Describe all possible values of $m$, $n$, and $r$.

(b) If $b = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, give a possible matrix $A$. (Look carefuly at $x$: can you identify likely free and pivot columns of $A$ from how we usually construct the particular and special solutions?)

(c) Look carefully at $x$, and write down the matrix $P$ that performs orthogonal projection onto $N(A)$. (Not much calculation should be needed!)

## Solution:

(a) For the matrix multiplication $Ax$ to make sense, it must be that $A$ has 4 columns, i.e. that $\boxed{n = 4}$. The vectors sitting after $\alpha_1$ and $\alpha_2$ must span the nullspace; as they are nonzero and not scalar multiples of one another, they are linearly independent, so form a basis for the nullspace. The nullspace therefore has dimension 2. By the rank nullity theorem, we therefore have that $r + 2 = 4$, so $\boxed{r = 2}$. We always have to have $m \geq r$, so we must have $\boxed{m \geq 2}$. To see that this is all that is required of $m$, consider the rowspace of $C(A^T)$ of $A$, and let $b = (b_1, b_2)$ be any nonzero vector of length 2 (we can't have $b = 0$, because otherwise the full solution would include 0, and ours doesn't - so there's also a condition on$b$ lurking around that we didn't ask you about). As the vector (1, 0, 1, 0) is not in $N(A)$, it cannot be orthogonal to all vectors in the rowspace $C(A^T)$, because $N(A)$ is the orthogonal complement of $C(A^T)$ and doesn't contain (1, 0, 1, 0). We can therefore find a basis $\{v, w\}$ for $C(A^T) = N(A)^\perp$ such that $v^T(1,0,1,0) = b_1$ and $w^T(1,0,1,0) = b_2$, because $(b_1, b_2) \neq 0$. Then the $2 \times 4$ matrix $A$ with first row $v$ and second row $w$ does the trick, so $m$ as small as 2 is possible. Appending zero rows/entries to $A$ and $b$ shows that any $m \geq 2$ is also achievable, so $m \geq 2$ is exactly what's required of $m$.

(b) By looking at $x$, we can first see that the first and third columns of $A$ need to add up to (1,2,1). To make sure that our matrix has rank 2, we should take two linearly independent vectors that do that, e.g. (1,1,1) and (0, 1, 0). To make sure that the two nullspace vectors (1, 1, -1, 0) and (1, 0, 1, 1) are in the nullspace (which is all we need to arrange, because the rank is 2 in this case, so the nullspace is 2-dimesional and hence is exactly the span of these two linearly independent vectors like it's supposed to be), we then need to make (1) the second column be the third minus the first, i.e. the second column should be (-1, 0, -1) and (2) the last column should be -first + -third, i.e. (-1, -2, -1). This gives the matrix

$$\boxed{A = \begin{pmatrix} 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & -2 \\ 1 & -1 & 0 & -1 \end{pmatrix}}.$$

It's then easy to check that this matrix works.

Of course, there are many possible solutions here, but the key thing is to check that both the particular solution and the nullspace vectors work.

(c) The vectors sitting after the $\alpha$'s are not only a basis for the nullspace, but they are orthogonal, so there are an orthogonal basis for the nullspace! Call the first one $v_1 = (1, 1, -1, 0)$ and the second on $v_2 = (1, 0, 1, 1)$. Then from class/homework we can write down the projection:

$$P = \frac{v_1 v_1^T}{v_1^T v_1} + \frac{v_2 v_2^T}{v_2^T v_2} = \frac{1}{3}\begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

$$= \boxed{\frac{1}{3}\begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ 0 & -1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}}$$

## Problem 2 (30 points):

(a) Give a possible $4 \times 3$ matrix $A$ with three *different, nonzero* columns such that blindly applying Gram–Schmidt to the columns of $A$ will lead you to **divide by zero** at some point.

(b) The reason Gram–Schmidt didn't work is that your $A$ does not have ____________________________________.

(c) To find an orthonormal basis for $C(A)$, you should instead apply Gram–Schmidt to what matrix (for your $A$)?

## Solution:

(a) As we'll say in the next part, any $A$ that doesn't have full column rank will do the trick. So taking any nonzero vector for our first column and scaling it in different nonzero ways will produce an answer, for example:

$$\boxed{\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}}.$$

This $A$ is rank 1 (the smallest possible rank since $A$ was required to be nonzero). Alternatively, we could give a rank 2 example. Just take any two linearly independent first two columns, and (for example) their sum as the third column:

$$\boxed{\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 1 & 2 \end{pmatrix}}.$$

(b) The reason Gram–Schmidt didn't work is that your $A$ does not have $\boxed{\text{full column rank}}$ (i.e. columns are not linearly independent).

(c) We just need to throw out enough columns that we're left with a basis for the original column space. In our first example (the rank 1 example) the column space is visibly one-dimensional (the simplest way you can solve this problem), and so we can keep just the first column:

$$\boxed{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}.$$

In the second example (the rank 2 example), we should keep the first two columns (actually, any two of the columns will work!), giving

$$\boxed{\begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}}.$$

## Problem 3 (30 points):

Given two $m \times n$ matrices $A$ and $B$, and two right-hand sides $b, c \in \mathbb{R}^m$, suppose that we want to minimize

$$f(x) = \|b - Ax\|^2 + \|c - Bx\|^2$$

over all $x \in \mathbb{R}^n$, i.e. we want to minimize the *sum of two least-squares fitting errors.*

(a) $\|b\|^2 + \|c\|^2$ can be written as the length squared $\|w\|^2$ of a single vector $w$. What is $w$?

(b) Write down a matrix equation $C\hat{x} = d$ whose solution $\hat{x}$ is the minimum of $f(x)$. (Give explicit formulas for $C$ and $d$ in terms of $A, B, b, c$.) Hint: your answer from the previous part should give you an idea to convert this into a "normal" least-squares problem.

## Solution:

(a) Lengths squared are just the sum of squares of the coordinates, so a natural answer is just to append the vector $c$ onto the vector $b$:

$$\boxed{w = \begin{pmatrix} b \\ c \end{pmatrix}}$$

(block matrix notation, so the above $w$ is in $\mathbb{R}^{2m}$).

Of course, there are other possible solutions here. You could add rows of zeros to $w$, or re-order the rows (or in fact multiply it by any orthogonal matrix), but all of these are unnecessarily complicated.

Some students instead proposed a vector $w \in \mathbb{R}^m$ with entries $w_i = \sqrt{b_i^2 + c_i^2}$. This indeed has the same norm, but it is not at all useful in the second part of the question.

(b) Let's start by rewriting $f(x)$. We have, taking part (1) as motivation,

$$f(x) = \|b - Ax\|^2 + \|c - Bx\|^2 = \left\| \begin{pmatrix} b \\ c \end{pmatrix} - \begin{pmatrix} Ax \\ Bx \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} b \\ c \end{pmatrix} - \begin{pmatrix} A \\ B \end{pmatrix} x \right\|^2.$$

In particular, minimizing $f(x)$ is then just exactly a usual least squares problem, but for the matrix equation

$$\begin{pmatrix} A \\ B \end{pmatrix} x = \begin{pmatrix} b \\ c \end{pmatrix}.$$

So, $\hat{x}$ minimizes $f(x)$ if and only if it satisfies the normal equations for this least squares problem, which reads

$$\begin{pmatrix} A \\ B \end{pmatrix}^T \begin{pmatrix} A \\ B \end{pmatrix} \hat{x} = \begin{pmatrix} A \\ B \end{pmatrix}^T \begin{pmatrix} b \\ c \end{pmatrix}.$$

So, taking

$$\boxed{d = \begin{pmatrix} A \\ B \end{pmatrix}^T \begin{pmatrix} b \\ c \end{pmatrix} = A^T b + B^T c}$$

and

$$\boxed{C = \begin{pmatrix} A \\ B \end{pmatrix}^T \begin{pmatrix} A \\ B \end{pmatrix} = A^T A + B^T B},$$

we have our answer. Note that $\begin{pmatrix} A \\ B \end{pmatrix}^T = \begin{pmatrix} A^T & B^T \end{pmatrix}$, as should be apparent from the definition of transposition.