Examples paper 1 notes

Nicholas Wong

1 Brief summary of the four subspaces

A *vector subspace* is first of all a *vector space*. Let's recall what that is. A vector space is defined in your notes, and obeys the following:

- $\mathbf{0}$ is always in the space.
- If $\mathbf{a}$ and $\mathbf{b}$ are in the space, then the *linear combination* $\alpha\mathbf{a} + \beta\mathbf{b}$ is also in the space with α, β being scalars.

For example, one vector space is the set of all vectors with three real number components [1]. We can call that $\mathbb{R}^3$.

Now, every vector space can be represented by a *basis*. This is a small set of vectors that belong to the vector space that should represent the vector space. That is, we can construct all other vectors in the vector space as (unique) linear combinations of these vectors. One basis for $\mathbb{R}^3$ is

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \tag{1}$$

because every vector in $\mathbb{R}^3$ can be represented uniquely as a linear combination of these three linearly independent vectors (please revise the definition and intuition behind linear independence if you are unfamiliar; how can we test for linear independence?). Another basis can be

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}. \tag{2}$$

This also contains three linearly independent vectors. Can we have more? No, because another 3-vector cannot possibly be linearly independent from the three already present. Can we have fewer? No, because then some vectors in $\mathbb{R}^3$ cannot be represented. In fact, the number of vectors in the basis is fixed for a vector space; this is called the *dimension* of the vector space. As a quick exercise, think about the following space and explain why it has dimension 2 (the dimension has nothing to do with the number of elements of the vectors contained in the space):

$$\left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \text{ for real numbers } a, b \right\}. \tag{3}$$

Usually, we want to work with *orthogonal* bases, meaning that the vectors are not only linearly independent but also orthogonal to each other (pairwise inner products are all zero).

[1] Let's just work with real numbers and vectors in this section. Generalisation to complex numbers does not require much more machinery.

Now, a vector subspace is a subset of a vector space, but itself also a vector space. For example, a subspace of $\mathbb{R}^3$ is the vector space spanned by

$$\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}. \tag{4}$$

Another subspace is the vector space spanned by

$$\begin{bmatrix}0\\0\\1\end{bmatrix}. \tag{5}$$

In fact, these two subspace can be "added"[2] up to form $\mathbb{R}^3$ (think of it as a set union). Thus, one can say that one subspace is in fact the *complement* of the other. Furthermore, any vector from the first subspace is going to be orthogonal to the other (Why?). Because of this, we call these subspaces *orthogonal complements* of each other.

Matrices are associated with four fundamental subspaces: the column space, the left null space, the row space and the null space. First, consider the matrix

$$\boldsymbol{A} = \begin{bmatrix}1 & 2\\10 & 6\\1 & 3\\3 & 12\end{bmatrix}. \tag{6}$$

The column space is the vector space spanned by the columns. That is,

$$\text{colspan}(\boldsymbol{A}) = \text{span}\left\{\begin{bmatrix}1\\10\\1\\3\end{bmatrix}, \begin{bmatrix}2\\6\\3\\12\end{bmatrix}\right\}. \tag{7}$$

These two vectors are linearly independent and 4-dimensional, so this is a 2-dimensional subspace of $\mathbb{R}^4$. Can you find an orthogonal basis for this space? (Hint: Gram-Schmidt process or QR factorisation)

The left null space is the space of vectors

$$\text{leftnull}(\boldsymbol{A}) = \{\mathbf{x} \text{ such that } \mathbf{x}^T\boldsymbol{A} = 0\}. \tag{8}$$

Is this a vector space? Check that it satisfies the properties!

This space is in fact the orthogonal complement of the column space. Why? Consider $\mathbf{z}$ from the column space and $\mathbf{y}$ from the left null space. Compute the dot product,

$$\mathbf{y}^T\mathbf{z} = \mathbf{y}^T\boldsymbol{A}\mathbf{x} = 0. \tag{9}$$

This is because $\mathbf{z}$ is in the column space, so it can be expressed as $\boldsymbol{A}\mathbf{x}$ for some $\mathbf{x}$. Furthermore, $\mathbf{y}$ satisfies $\mathbf{y}^T\boldsymbol{A} = 0$ because it is from the left null space. This implies that any vector from the column space is orthogonal to any vector in the left null space, and vice versa. Some more food for thought: Why does this mean that $\mathbf{y} \neq \mathbf{z}$ except when both are zero (no non-zero vector can exist in both the column space and left null space!)?

The row space is the space spanned by the rows. For $\boldsymbol{A}$, we have four rows of length 2. They are also all pairwise linearly independent. We only need two linearly independent 2-vectors to span the $\mathbb{R}^2$ and this is more than enough, so the row space is the entirety of $\mathbb{R}^2$, dimension 2.

The null space is

$$\text{null}(\boldsymbol{A}) = \{\mathbf{x} \text{ such that } \boldsymbol{A}\mathbf{x} = 0\}. \tag{10}$$

Exercise for the reader: Go through the same logic as above and conclude that the null space is the orthogonal complement of the row space for any matrix. For $\boldsymbol{A}$, what is the dimension of the null space? List all of its elements.

[2]This is called the *direct sum*.

2 Determinants of transpose and conjugates

Satisfy yourself that

$$\begin{aligned} \det(\boldsymbol{A}) &= \det(\boldsymbol{A}^T) \\ \det(\overline{\boldsymbol{A}}) &= \overline{\det(\boldsymbol{A})}. \end{aligned} \tag{11}$$

3 On question 9

For both of these parts, remember the goal is to prove that the maximum of a certain function over some vector $\mathbf{x}$ is some value, i.e show that

$$\max_{\mathbf{x}\neq 0} f(\boldsymbol{A}, \mathbf{x}) = C(\boldsymbol{A}). \tag{12}$$

The function is also dependent on $\boldsymbol{A}$, and the maximum will also depend on $\boldsymbol{A}$, but we suppress this dependence in the following. In part a), $f(\mathbf{x}) = \frac{\|\boldsymbol{A}\mathbf{x}\|_1}{\|\mathbf{x}\|_1}$; in part b), $f(\mathbf{x}) = \frac{\|\boldsymbol{A}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty}$. The approach we take here is to prove two things:

- Find a value C for which $f(\mathbf{x}) \leq C$ for all $\mathbf{x}$ (aka an upper bound of f).
- Find one $\mathbf{x}^*$ for which $f(\mathbf{x}^*) = C$ exactly (i.e. the upper bound is achievable).

For example, consider $f(x) = -x^2$ defined over all $x \in \mathbb{R}$. We could choose $C = 1$ in the first step since $-x^2 \leq 1$ always, but no value of x achieves this upper bound. Instead, if we choose $C = 0$, then $-x^2 \leq 0$ always, and $f(0)$ exactly equals C, so this is the maximum of $f(x)$.

To begin, we can establish the fact that

$$\max_{\mathbf{x}\neq 0} \frac{\|\boldsymbol{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|\boldsymbol{A}\mathbf{x}\|. \tag{13}$$

That is, instead of searching over all $\mathbf{x}$ (except 0), search over all vectors of norm 1. Why? Since the function is about the ratio between the output norm $\|\boldsymbol{A}\mathbf{x}\|$ and input norm $\|\mathbf{x}\|$, we can "factor" out the input norm by simply considering only unit norm vectors. In other words, for a certain $\mathbf{x}$, $f(c\mathbf{x}) = f(\mathbf{x})$, so the scaling of the vector does not matter.

Now, let's go through the rest of part a), and relegate thinking about part b) as an exercise (which I strongly recommend you do!) Let $g(\mathbf{x}) = \|\boldsymbol{A}\mathbf{x}\|_1$. This string of inequalities completes the first part of the task by finding a valid upper bound C:

$$g(\mathbf{x}) = \|\boldsymbol{A}\mathbf{x}\|_1 =_{(1)} \left\| \sum_{j=1}^{n} \mathbf{a}_j x_j \right\| \leq_{(2)} \sum_{j=1}^{n} \|\mathbf{a}_j\| \, |x_j| \leq_{(3)} \max_{j=1}^{n} \|\mathbf{a}_j\| = C. \tag{14}$$

The explanations for the inequalities are:

1. Column sum representation of matrix-vector multiplication (refer to Part IB notes)
2. Triangle inequality for vectors
3. We stipulated our search space to be $\mathbf{x}$ such that $\|\mathbf{x}\|_1 = 1$. This means that $\sum_{j=1}^{n} |x_j| = 1$, so $|x_j| \leq 1$ for all j. To see this inequality, perhaps it's best to put some concrete numbers in. Let's say

$$\boldsymbol{A} = \begin{bmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{bmatrix}. \tag{15}$$

Then, $\{\|\mathbf{a}_1\|, \|\mathbf{a}_2\|, \|\mathbf{a}_3\|\} = \{12, 15, 18\}$. Given that $\sum_{j=1}^{n} |x_j| = 1$, can you see that

$$\sum_{j=1}^{3} \|\mathbf{a}_j\| \, |x_j| = 18 = \max_{j=1}^{3} \|\mathbf{a}_j\|? \tag{16}$$

So we have found an appropriate upper bound $C = \max_{j=1}^{n} \|\mathbf{a}_j\|$. Can we find one $\mathbf{x}^*$ (where $\|\mathbf{x}^*\|_1 = 1$ such that $g(\mathbf{x}^*) = C$?

Returning to $\boldsymbol{A}$ as in (15), what is the $\mathbf{x}^*$ such that $\|\boldsymbol{A}\mathbf{x}\|_1 = 18$? The answer is

$$\mathbf{x}^* = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{17}$$

This choice gives us equality in both (2) and (3) above. Indeed this is true for any matrix. Why? Thus, We are done.

Remember to repeat this exercise for part (b).

4 On question 12

- Remember the geometric picture for least squares.
- Remember that $(\boldsymbol{A}^H\boldsymbol{A})^{-1} \neq \boldsymbol{A}^{-1}\boldsymbol{A}^{-H}$ here. This is because $\boldsymbol{A}$ is not square, thus non-invertible! Exercise: Why can we guarantee that $\boldsymbol{A}^H\boldsymbol{A}$ is invertible when $\boldsymbol{A}$ is full rank? (Similar to question 13)

5 On question 13

The first thing to note is that $\boldsymbol{A}$ here cannot be tall (can only be either square or wide). Why? Hint: Say $m = 10$ and $n = 3$. Is it possible for the rows to be linearly independent? Consider how many rows there are, and how many elements each row vector has.

The rows are linearly independent, and there are m of them. This implies that the rank of $\boldsymbol{A}$ is m. This also implies that there are exactly m linearly independent columns. Each column has m elements (they are vectors in $\mathbb{C}^m$). Thus, $\mathbb{C}^m$ is completely spanned by the columns. Referring back to the beginning of these notes, we know that the left null space, the orthogonal complement of the column space, can only have the zero vector.

Now, we need to prove that $\boldsymbol{A}\boldsymbol{A}^H$ is invertible. First, a sanity check: $\boldsymbol{A}\boldsymbol{A}^H$ is square regardless of the shape of $\boldsymbol{A}$, so we at least have a chance. In fact, $\boldsymbol{A}\boldsymbol{A}^H$ being invertible implies that the nullspace of $\boldsymbol{A}\boldsymbol{A}^H$ is trivial. Why? If the null space of $\boldsymbol{A}\boldsymbol{A}^H$ is not trivial, let $\mathbf{x} \neq 0$ be in the null space, i.e.

$$\boldsymbol{A}\boldsymbol{A}^H\mathbf{x} = 0. \tag{18}$$

However, this means that there is a linear combination of the columns of $\boldsymbol{A}\boldsymbol{A}^H$ (that is not just multiplying every column of $\boldsymbol{A}\boldsymbol{A}^H$ by zero) that gives us the zero vector. By the definition of linear independence, we know that $\boldsymbol{A}\boldsymbol{A}^H$ does not have linearly independent columns, so it is not full rank, and it is not invertible. (This is called proving the contrapositive—search this on the web if curious!)

Returning to the name of the game: proving the nullspace of $\boldsymbol{A}\boldsymbol{A}^H$ is trivial. Consider then the line in the cribs, where we assume that $\mathbf{x}$ is in the nullspace of $\boldsymbol{A}\boldsymbol{A}^H$. That implies

$$\boldsymbol{A}\boldsymbol{A}^H\mathbf{x} = 0. \tag{19}$$

What does this further imply? It also implies

$$\mathbf{x}^H\boldsymbol{A}\boldsymbol{A}^H\mathbf{x} = 0 \implies (\boldsymbol{A}^H\mathbf{x})^H(\boldsymbol{A}^H\mathbf{x}) = 0 \implies \left\|\boldsymbol{A}^H\mathbf{x}\right\|_2^2 = 0 \implies \boldsymbol{A}^H\mathbf{x} = 0 \tag{20}$$

We have used the fact that *if the norm of the vector* $\boldsymbol{A}^H\mathbf{x}$ *is zero, the vector itself must exactly be the zero vector.*

What do we have then? We have that if $\mathbf{x}$ is in the nullspace of $\boldsymbol{A}\boldsymbol{A}^H$ then it also is in the *left null space of* $\boldsymbol{A}$—it satisfies $\boldsymbol{A}^H\mathbf{x} = 0$. That is,

$$\text{nullspace}(\boldsymbol{A}\boldsymbol{A}^H) \subset \text{leftnull}(\boldsymbol{A}) = \{0\}. \tag{21}$$

Thus, the nullspace of $\boldsymbol{A}\boldsymbol{A}^H$ is indeed trivial, and we are done.

6 More on pseudoinverses

In the notes, we have defined $\boldsymbol{A}^+$ as follows:

$$\boldsymbol{A}^+ = \begin{cases} (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H, & \text{if } \boldsymbol{A} \text{ is tall} \\ \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{A}^H)^{-1}, & \text{if } \boldsymbol{A} \text{ is fat} \end{cases} \tag{22}$$

Is this well-defined (i.e. do the matrix products make sense in terms of sizes? Are the matrices invertible whenever we indicate them?) Consider $(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$ for a fat matrix. Can you explain why this would not work?