A SHORT SUMMARY OF VECTOR SPACES AND MATRICES

This is a little summary of some of the essential points of linear algebra we have covered so far. If you have followed the course so far you should have no trouble understanding these notes. I suggest that you flesh out this text with your own examples.

1. Solving linear equations

We are given an $m \times n$ matrix A and a vector $\vec{b}$ and consider the system of linear equations

$$A\vec{x} = \vec{b} \ . \tag{1}$$

The basic questions are:

a) Is there a solution?

b) Is there a solution for every $\vec{b} \in \mathbb{R}^m$?

c) If there is one, is it unique?

d) If it is not unique, how can we describe all of them?

The technical tool that we use to answer these questions is **elimination** which leads to the **row reduced echelon form** R. More precisely, elimination for the augmented matrix $[A|\vec{b}]$ leads to $[R|\vec{c}]$ where R is in row reduced echelon form. Each pivot is one, and above and below a pivot there are only zeros. This form is unique, i.e., it does not depend on how you do the elimination. I will assume that you know how to work the row reduction algorithm.
The set of solutions of (1) *and of*

$$R\vec{x} = \vec{c} \tag{2}$$

is the same.
The number of pivots is called **the rank** of the matrix A and denoted by $r(A)$ or just r.

There are four spaces associated with the matrix A. The **column space** $C(A) \subset \mathbb{R}^m$, the **null space** $N(A) \subset \mathbb{R}^n$, the **row space** which is the same as the column space of the transposed matrix A^T, $C(A^T) \subset \mathbb{R}^n$ and the null space $N(A^T) \subset \mathbb{R}^m$. The column space $C(A)$ consists of all **linear combinations** of the column vectors and the null space consists of all vectors $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{0}$. Note that $C(A), N(A^T)$ are subspace of the *same* space $\mathbb{R}^m$ and $C(A^T), N(A)$ are subspaces of the *same* space $\mathbb{R}^n$. Elimination does not change the row space, and it does not change the null space of A. Hence we have

$$C(A^T) = C(R^T) \ , \ N(A) = N(R) \ .$$

The spaces $C(A)$ and $N(A^T)$ will change, however.

The **dimensions** of these various spaces are related by

$$\dim C(A) = r(A) = \dim C(A^T)$$

and

$$\dim N(A) = n - r(A) \ , \dim N(A^T) = m - r(A) \ .$$

These relations follow from the row reduced echelon form. Here is the argument: For $\vec{b}$ to be in the column space it must be a linear combination of the column vectors of $A = [\vec{a}_1, \dots, \vec{a}_n]$, i.e., of the form

$$\vec{b} = \sum_{j=1}^{n} y_j \vec{a}_j \ ,$$

thus, the vector $\vec{y} = (y_1, \dots, y_n)$ is a solution of the equation (1). After row reduction this equation transforms into

$$\vec{c} = \sum_{j=1}^{n} y_j \vec{r}_j$$

where $\vec{r}_j$ are the column vectors of the matrix R. The y's stay the same because the solutions of (1) are the same as the ones of (2)! Thus, whenever, a set of column vectors of A are **linearly independent** so are the corresponding vectors in the matrix R and conversely. Thus, the dimensions of the column space of A and of R are the same. *Not the spaces, but the dimensions.* The dimension of the column space of R equals the number of pivots, i.e., the rank of A. In fact the columns of R that contain a pivot are the basis vectors for $C(R)$. Therefore the corresponding columns in A, which we call the pivot columns of A, form a **basis** for $C(A)$ (Why?). Again, let me emphasize that $C(R) \neq C(A)$. A little bit easier is the argument for the row space $C(A^T)$, since this space does not change and the rows in R that contain a pivot form a basis for $C(A^T)$. Hence dime$C(A^T) = r(A)$. The rest follows by noting that the number of free variables is the dimension of the null space of A and this number plus R yields the number of columns.

Now we can give some answers to the questions mentioned at the beginning. There **exists** a solution if and only if $\vec{b} \in C(A)$. This is a bit of a triviality but a usefull one, as we shall see later. The equation(1) has a solution for every $\vec{b} \in \mathbb{R}^m$ if an only if $r(A) = m$. This simply expresses that fact the $C(A)$ being a subspace of $\mathbb{R}^m$ and having dimension m must be equals to $\mathbb{R}^m$.

Recall that the $n \times m$ matrix B is a **right inverse** for the matrix A if $AB = I_m$ where I_m is the identity matrix in $\mathbb{R}^m$. Note that a right inverse exists if and only if (1) has a solution for every $\vec{b} \in \mathbb{R}^m$. If the right inverse B exists, then $B\vec{b}$ solves $A(B\vec{b}) = I_m\vec{b} = \vec{b}$. So we have a solution for every $\vec{b} \in \mathbb{R}^m$. To see the converse, we know, by assumption, that $A\vec{x}_i = \vec{e}_i$ has a solution for every $e_i \in \mathbb{R}^m$. The solution might not be unique but that does not matter. From the matrix $B = [\vec{x}_1, \dots, \vec{x}_m]$ and note that $AB = I_m$. Another way of saying this is that A has a right inverse if and only if $r(A) = m$.

To make further progress we observe: *Every solution of* (1) *is of the form* $\vec{x}_p + \vec{z}$ *where* $\vec{z}$ *is a solution of the* **homogeneous** *equation* $A\vec{z} = \vec{0}$ *and* $\vec{x}_p$ *is any particular solution of* (1). Thus, we find that the solution is **unique** if and only if $N(A) = \{\vec{0}\}$ which is the same as dim$N(A) = 0$. In this case $r(A) = n$. If the solution is not unique we can describe all the solutions of the equation (1). Pick any **basis** of $N(A)$, $\{\vec{v}_1, \dots, \vec{v}_{n-r}\}$. The complete solution set of (1) is then given by

$$\vec{x}_p + \sum_{j=1}^{n-r} t_j \vec{v}_j$$

where $t_1, \dots, t_{n-r}$ are real numbers and $\vec{x}_p$ is any particular solution. A basis for $N(A)$ can be found by solving the row reduced system in terms of the free variables.

The matrix A has a **left inverse** if the exists an $n \times m$ matrix C with $CA = I_n$. The matrix A has a left inverse if and only if $N(A) = \{\vec{0}\}$. If A has a left inverse C, then applying C to the equation $A\vec{x} = 0$ yields $\vec{x} = CA\vec{x} = 0$. Hence $N(A) = \{\vec{0}\}$. Conversely, if $N(A) = \{\vec{0}\}$ then $r(A) = n$. This means that the equation $A^T\vec{y} = \vec{c}$ has a solution for every $\vec{c} \in \mathbb{R}^n$. Hence, A^T has a right inverse $A^T D = I_n$ and therefore D^T is a left inverse of A. To summarize, the matrix A has a left inverse if and only if dim$N(A) = 0$.

Recall that the matrix A is **invertible if it has a right and a left inverse** in which case the two matrices B and C are the same. Thus, for A invertible it is necessary and sufficient that both, $n = m$ *and* $r(A) = n$ hold for the matrix A. Equivalently, A is invertible if and only if $n = m$ and dim$N(A) = 0$. Another way of saying this is, that the matrix A is invertible if and only if the equation (1) has a unique solution for any $\vec{b} \in \mathbb{R}^m$. For this to be possible we necessarily need that $m = n$. Still another way of characterizing an invertible matrix is by saying that the column vectors are linearly independent, since n such vectors form for a basis for $\mathbb{R}^n$.

You see that this part of linear algebra is about formulating the same statement in many different ways. This somewhat confusing but useful. Think of it as acquiring a language.

2. Least square approximations

So far we argued only via the dimensions of the various subspaces, but they have special positions. Recall that the **orthogonal complement** of a subspace $V \subset \mathbb{R}^n$ consists of all vectors in $\mathbb{R}^n$ that are perpendicular to every vector in V. The orthogonal complement is denoted by $V^\perp$. V and $V^\perp$ have only the zero vector in common and we have that

$$\dim V + \dim V^\perp = n \ .$$

Moreover, ${V^\perp}^\perp = V$. The orthogonal complement of the row space of A consists of all vectors that are perpendicular to all row vectors of A and hence to all column vectors of A^T. Hence

$$C(A^T)^\perp = N(A) \ , \ N(A)^\perp = C(A^T)$$

and

$$C(A)^\perp = N(A^T) \ , \ N(A^T)^\perp = C(A) \ .$$

With these concepts one can formulate some interesting problems. What is the distance of the tip of a vector $\vec{b} \in \mathbb{R}^n$ to a subspace $V \subset \mathbb{R}^n$, or better what is vector in V that is closest to $\vec{b}$. If $\vec{b}$ is in V then, of course, $\vec{b}$ itself is the answer and the distance is zero. So we assume that $\vec{b} \notin V$. Suppose we have a vector $\vec{v} \in V$ with the property that $\vec{b} - \vec{v}$ is perpendicular to V, or what amounts to the same, that $\vec{b} - \vec{v} \in V^\perp$. I claim that $\|\vec{b} - \vec{v}\|$ is the distance between V and the tip of $\vec{b}$. To see this, pick any other vector $\vec{w} \in V$. Then, we write

$$\|\vec{b} - \vec{w}\|^2 = \|(\vec{b} - \vec{v}) + (\vec{v} - \vec{w})\|^2 = \|\vec{b} - \vec{v}\|^2 + \|\vec{v} - \vec{w})\|^2 + 2(\vec{b} - \vec{v}) \cdot (\vec{v} - \vec{w})$$

where the last term vanishes, since $(\vec{v} - \vec{w}) \in V$ and $(\vec{b} - \vec{v}) \in V^\perp$. Hence,

$$\|\vec{b} - \vec{w}\|^2 = \|\vec{b} - \vec{v}\|^2 + \|\vec{v} - \vec{w})\|^2 \geq \|\vec{b} - \vec{v}\|^2$$

with equality only if $\vec{w} = \vec{v}$. Thus, we have to find $\vec{v}$ with $\vec{b} - \vec{v} \perp V$ in order to solve our distance problem.

This can be readily solved. Pick any basis $\vec{v}_1, \ldots, \vec{v}_k$ in V. The vector $\vec{v}$ in question must be of the form $\vec{v} = \sum_{j=1}^k x_k \vec{v}_k$. We can reformulate this by using the matrix

$$A = [\vec{v}_1, \ldots, \vec{v}_k] \ .$$

so that $\vec{v} = A\vec{x}$. Now $\vec{b} - A\vec{x} \perp V = C(A)$ and hence $\vec{b} - A\vec{x} \in N(A^T)$ so that

$$A^T(\vec{b} - A\vec{x}) = 0$$

or

$$A^T A\vec{x} = A^T\vec{b} \ .$$

These are the **normal equations**. The always have a unique solution, in fact the matrix $A^T A$ is invertible. It suffices to show that the column vectors of $A^T A$ are linearly independent. To check this we solve $A^T A\vec{y} = 0$ which means that $A\vec{y} \in N(A^T)$. Moreover $A\vec{y} \in C(A)$ and hence $A\vec{y} = 0$. Since the column vectors of A are linearly independent, $\vec{y} = 0$. The vector $\vec{v}$ is now given by

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

and we set

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

which is the **orthogonal projection** of $\mathbb{R}^n$ onto V. Note that P_V does not depend on the choice of basis. Why?

We have that

$$P_V^T = P_V$$

and

$$P_V^2 = P_V \ .$$

These observations are the foundations of the least square approximation. Imagine that a data set is given by vectors in $\mathbb{R}^n$ and you would like to see that this data set fits onto some lower dimensional subspace, e.g., a line. Very often this problem can be brought into the form of seeking the **least square approximation** to a system $A\vec{x} - \vec{b}$. This is almost the problem we solved above. The vector $\vec{b}$ is given as well as the matrix A. The column vectors, however, do not in general form a basis for $C(A)$, they may be linearly dependent. Moreover, we are not primarily interested in the vector $\vec{v} \in C(A)$ that is closest to the vector $\vec{b}$. Our interest is in $\vec{x}$. The normal equations are still valid

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

and this system of equation always has a solution, the right side is always in the column space of A^T and hence of the column space of $A^T A$. The solution for $\vec{x}$, however, need not be unique. The error of the least square approximation is given by

$$\|A\vec{x}_\star - \vec{b}\|$$

where $\vec{x}_\star$ is any solution of the normal equations. One might think that this number depends on which solution we choose, but this is not the case (Why?).