Lecture 4-7

Continuing from last time, suppose we are given a linear system $MX = B$, where M is an $m \times n$ matrix (of real numbers, or elements of some field), X is a column vector of unknowns, and B is a column vector of constants. We assume that the system makes sense, so that X has length n while B has length m (we have m linear equations in n unknowns). We learned last time how to bring the augmented matrix of this system (obtained from M by adding the column B as a new column on the right) to echelon form by row operations, so that the pivot(=leftmost nonzero entry) of every nonzero row has all zeroes below it in its column, the pivots move strictly to the right as you go down the rows, and if any row fails to have a pivot (i.e. is zero) then all lower rows are also zero. Note that pivots by definition must lie in the coefficient matrix M, not the rightmost column of its augmented matrix.

We also saw that our original system has a solution if and only if all nonzero rows of the echelon form of the augmented matrix have pivots; that is, no equation of the system in the echelon form reads $0x_1 + \ldots + 0x_n = k$ for $k \neq 0$. The pivots all lie in different columns, by definition of echelon form, so correspond to different variables, called the *pivot* or *bound* variables. The remaining variables, if any, are called *free* and can take any value (if the system is consistent). The values of the pivot variables are completely determined by the values of the free variables, so that the system has a unique solution if and only if it is consistent and there are no free variables. Otherwise it always has infinitely many solutions, with any solution uniquely specified by the values of its free variables.

As an example, the system $MX = B$ where $M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, B = \begin{pmatrix} 6 \\ 15 \\ 24 \end{pmatrix}$ is consistent and has x_3 as its unique free variable. Solving the second equation $-3x_2 - 6x_3 = -9$ of the system in echelon form for x_2 in terms of x_3, we get $x_2 = (-1/3)(-9 + 6x_3) = 3 - 2x_3$; solving the first equation $x_1 + 2x_3 + x_3 = 6$ for x_1 in terms of x_3 we get $x_1 = 6 - 3x_3 - 2x_2 = 6 - 3x_3 - 2(3 - 2x_3) = x_3$. We express the solution set as $\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ 3 - 2x_3 \\ x_3 \end{pmatrix} \right\}$; in this way we indicate that x_3 is the free variable via the last vacuous equation $x_3 = x_3$, while the other variables are explicitly given in terms of x_3.

In particular, if the original system $MX = B$ has fewer equations than unknowns (that is, $m < n$), and if it is consistent, the echelon form must have at least one (and in fact at least $n - m$) free variables, since there cannot be more pivots than columns. If the system $MX = \vec{0}$ is homogeneous (in an obvious sense), then it always has the zero solution, and so must have infinitely many solutions (more precisely parametrized by at least $n - m$ arbitrary real numbers). If however the original system has more equations than unknowns then it is quite likely (but not certain) that it has no solution.

We now take a little break from systems, returning to the abstract setting with which the Treil notes begin. We have already seen that a *real vector space* is an abelian group V such that for every $r \in \mathbb{R}, v \in V$ there is $rv \in V$ such that $r(v_1 + v_2) = rv_1 + rv_2, (r+s)v = rv + sv, r(sv) = (rs)v, 1v = v$, for all $r, s \in \mathbb{R}, v_1, v_2 \in V$. More generally, if these axioms are satisfied with $\mathbb{R}$ replaced by some other field K (e.g. $K = \mathbb{C}$), then we call V a vector space *over* K. Besides the obvious example of $\mathbb{R}^n$ itself we have its *subspaces* S, which are just subgroups of $\mathbb{R}^n$ as an abelian group such that $rs \in S$ whenever $s \in S, r \in \mathbb{R}$. Note that the planes and lines in $\mathbb{R}^3$ studied in Chapter 13 of Salas-Hille qualify as subspaces of $\mathbb{R}^3$ if and only if they contain the origin $\vec{0}$; similarly, we sometimes allow ourselves to call certain subsets S of $\mathbb{R}^n$ subspaces even if they are not actually vector subspaces. The main example that we have seen so far is tangent hyperplanes to level sets in $\mathbb{R}^n$; more generally, if V is a vector subspace of $\mathbb{R}^n$ and $\vec{w} \in \mathbb{R}^n$ then the set $V + \vec{w} = \{\vec{w} + \vec{v} : \vec{v} \in V\}$ is also called a subspace of $\mathbb{R}^n$. As an example of a real vector space that is not a subspace of $\mathbb{R}^n$ we cite the solution set to a linear homogeneous differential equation; this is why we usually refer to the solution space rather than the solution set of such an equation. The solution space to a linear nonhomogeneous differential equation would not then be a vector space, but would be obtained from the solution space to the corresponding homogeneous equation by adding a fixed function to all of its elements, as above. As another example of a vector space different from $\mathbb{R}^n$ we cite the set $M_{m,n}$ of $m \times n$ real matrices: any two such matrices are (by definition) added by adding corresponding entries (this is *not* like matrix multiplication!) while multiplying a matrix by a scalar by definition amounts to multiplying each of its entries by that scalar. The vector space $M_{m,n}$ is really just $\mathbb{R}^{mn}$ in disguise; instead of listing the elements of an mn-tuple in a row, we arrange them in the form of an $m \times n$ matrix.

Now it turns out that every subspace of $\mathbb{R}^n$ looks just like $\mathbb{R}^m$ for some $m \leq n$. To make this last statement more precise we need the notion of linear combination of vectors. Given a set $S = \{v_1, \ldots, v_m\}$ of vectors v_i in a vector space V, the set of all sums $\sum_{i=1}^m a_i v_i$ for $a_i \in \mathbb{R}$ is easily seen to be a subspace W of V; in fact it is the unique smallest subspace containing the v_i. We call an element of W a *linear combination* of the v_i and we say that W is the subspace *generated* (or *spanned* by the v_i. More generally, we could let S be any subset of V; then the subspace of V generated by S consists by definition of all finite linear combinations of elements of S. It is also called the span of S. We never try to define infinite linear combinations of vectors in a vector space.