Math 52: Course Summary

Rich Schwartz

September 2, 2009

General Information: Math 52 is a first course in linear algebra. It is a transition between the lower level calculus courses and the upper level math courses. I would say that Math 52 has four general aims:

- To clarify and unify all the stuff you learned in high school about linear equations and matrices.
- To introduce you to higher dimensional Euclidean spaces and the easiest kinds of mappings between these spaces – the linear maps.
- To introduce you to the abstract formalism of vector spaces and linear transformations. The upper level math classes are filled with abstract structures like this, and the ones in M52 are a good warm-up.
- To introduce you to the idea of proving theorems in mathematics. Proofs abound in the upper level courses, and M52 gives an easy context in which to get started.

M52 is traditionally taken after finishing the calculus series, but there is some argument for taking M52 in advance of several variable calculus – e.g. M18, M20, M35. Several variable calculus uses linear algebra implicitly, and the constructions are more awkward without the linear algebra formalisms.

Math 54 is a somewhat more advanced version of Math 52. In math 54, less attention is paid to concrete calculations with matrices and such, and more attention is paid to the theoretical side of the subject.

High School Algebra Revisited: Suppose you want to solve the linear system

$$3x + 4y = 2$$
$$7x - 3y = 5$$

It doesn't hurt to multiply both sides of the first equation by 3 and both sides of the second equation by 4. This gives

$$\begin{aligned} 9x + 12y &= 6 \\ 28x - 12y &= 20 \end{aligned}$$

It doesn't hurt to add the equations together. This gives

$$37x = 26 \qquad \Longrightarrow \qquad x = 26/37.$$

A similar trick will get y. Continuing with the example above, the only thing really important is the 3×2 array of numbers

$$\begin{array}{cc|c} 3 & 4 & 2 \\ 7 & -3 & 5 \end{array}.$$

You can simplify this array by adding multiples of the first column to the second column or vice versa. After some effort, you get down to an array that looks like

$$\begin{array}{cc|c} 1 & 0 & \frac{26}{37} \\ 0 & 1 & \frac{-1}{37} \end{array}$$

This final array tells us that $x = 26/37$ and $y = -1/37$.

The same idea works for systems of equations with more variables. You might have an equation like

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_4 + a_{14}x_4 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_4 + a_{24}x_4 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_4 + a_{34}x_4 &= b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_4 + a_{44}x_4 &= b_4 \end{aligned},$$

where a_{11} etc. and b_1 etc are numbers and the variables x_1, x_2, x_3, x_4 are the unknowns. The important thing in solving this system is the 5×4 array

$$\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & b_3 \\ a_{41} & a_{42} & a_{43} & a_{44} & b_4 \end{array}$$

You can solve this system (or see that it has no solutions) by adding multiples of one row to another until you have simplified the thing as much as possible. The first main thing you learn in M52 is an algorithmic way to do this, called

Gaussian elimination. This is really nothing more than a systematization of what you already knew how to do.

Higher dimensional space: Here's another view of our first example above. Each of the two equations specifies a line in the plane, and then we are finding the point where these lines intersect. What about the second example? If there were just 3 variables, then each of the equations would specify a plane in space, and we would be looking for the point where the 3 planes intersect. Incidentally, one can see geometrically why such systems might not have solutions. For instance, the lines and planes might be parallel. In the actual second example, we could say that each of the equations in 4 variables specifies a copy of 3-dimensional space sitting inside 4-dimensional space, and we are looking for the quadruple intersection point.

In M52 you learn that n-dimensional Euclidean space is nothing more than the set of all n-tuples $(x_1, ..., x_n)$ where these variables can take on any real number. Some people find this to be quite exhilarating. Most people are mystified by *four dimensional* space, and here you can think about a billion dimensional space without really sweating. Otjer people find this to be a letdown, because the concrete description we've just given completely demystifies higher dimensional space.

Once you have the underlying space, you can talk about distances between points. The distance between $(x_1, ..., x_n)$ and $(y_1, ..., y_n)$ is

$$\sqrt{(x_1 - y_1)^2 + ... + (x_n - y_n)^2}.$$

This is really just a generalization of the Pythagorean theorem. Once you have distance, you can talk about rotations of higher dimensional space, mirror reflections, translations, and so on. This is all done in M52. So, this part of M52 gives you a foundation for working with higher dimensional Euclidean geometry.

Linear Mappings of Euclidean Space: A function of one variable takes a number as input and spits out another number. The simplest kinds of functions have the form $f(x) = ax$. These functions, called *linear functions*, are characterized by the following two properties.

- $f(x + y) = f(x) + f(y)$
- $f(bx) = bf(x)$. (On the right side, I mean b times $f(x)$.)

They are known as *linear functions*.

n-dimensional Euclidean space is denoted by $\boldsymbol{R}^n$. To generalize the notion of a function, we could say that a *map* from $\boldsymbol{R}^n$ to $\boldsymbol{R}^m$ is a gadget that takes a point in $\boldsymbol{R}^n$ as input and spits out a point in $\boldsymbol{R}^m$. For instance

$$f(x_1, x_2, x_3) = (x_2, 5x_1 + 7x_2 - x_3)$$

is a map from $\boldsymbol{R}^3$ to $\boldsymbol{R}^2$. In general, we write $f : \boldsymbol{R}^n \to \boldsymbol{R}^m$. Since it makes sense to add and scale points in $\boldsymbol{R}^n$, we call f *linear* if it satisfies the two properties mentioned above, namely $f(x + y) = f(x) + f(y)$ and $f(bx) = bf(x)$.

Note that we have suddenly made a jump in abstraction. In our last equations, x and y represent points in $\boldsymbol{R}^n$. That is, $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ and $x + y = (x_1 + y_1, ..., x_n + y_n)$ and $bx = (bx_1, ..., bx_n)$. Just as the single number a characterizes the linear function $f(x) = ax$, the linear maps are characterized by $m \times n$ matrices. These are just like the arrays of numbers we discussed above.

To facilitate the study of linear maps, one defines a way for matrices to *act* on vectors, as follows:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 \end{bmatrix}.$$

We took $m = 3$ and $n = 4$ just to make the notation easier, but the general case is essentially the same. Inspecting this last equation, we see that our 4×3 matrix sets up a map from $\boldsymbol{R}^4$ to $\boldsymbol{R}^3$. The input is (x_1, x_2, x_3, x_4) and the output is the right hand side of the equation. This map turns out to be linear, and all linear maps have this form.

In calculus, you learned about composing functions, and perhaps about composing maps. (Remember the chain rule.) When linear maps are represented as matrices, the composition of linear maps corresponds to an operation called *matrix multiplication*. The example we gave above illustrates how the products of 3×4 matrix and a 4×1 matrix is a 3×1 matrix. In general, the product of an $m \times n$ matrix and an $n \times p$ matrix is an $m \times p$ matrix. The complete correspondence between linear maps and matrices is worked out in M54.

Vector Spaces: A real vector space is a set V of elements called *vectors*. This set has two operations on it, addition and scaling by real numbers. The addition is denoted by $(+)$ and the scaling is written without an intermediate symbol: rx means scale x by the real number r.. These rules obey the following axioms.

- $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$ for all $v_1, v_2, v_3 \in V$.
- $v_1 + v_2 = v_2 + v_1$ for all $v_1, v_2 \in V$.
- There exists an element, called 0, such that $0 + v = v + 0$ for all $v \in V$.
- For any $v \in V$ there exists $w \in V$ such that $v + w = 0$.
- $r(sx) = (rs)x$.
- $r(v_1 + v_2) = rv_1 + rv_2$.

Here are some examples of vector spaces.

- The space $\boldsymbol{R}^n$ is a vector space. When considered as a vector space, the *points* in $\boldsymbol{R}^n$ are called vectors. The addition and scaling of vectors are the obvious ones.
- Let P_n denote the space of real polynomials of degree n. Polynomials in P_n are added and scaled in the obvious way, e.g.

$$3(1 + 2x + 7x^2) = 3 + 6x + 21x^2.$$

- The space of functions $f : \boldsymbol{R} \to \boldsymbol{R}$. Given f and g, the function $h = f + g$ has the property that $h(x) + f(x) + g(x)$.

One point of introducing vector spaces is that a lot of basic properties of $\boldsymbol{R}^n$ carry over to other examples. Sometimes the added abstraction clarifies what is going on. For instance, Fourier series, a subject from analysis, can be nicely formulated in terms of vector spaces.

Bases, Dimension, Isomorphism: Once vector spaces are defined in M52, their basic properties are studied in a systematic way. I'll give you a feel for the sort of results you'll see. A subset of vectors $v_1, ..., v_k \in V$ is called *dependent* if there are constants $a_1, ..., a_k$, not all zero, such that

$$a_1 v_1 + ... + a_k v_k = 0.$$

Otherwise, the set is called *independent*. A set of independent vectors is called a *basis* if it is maximal: The inclusion of any new vector creates a dependent set. One foundational result is this: If a vector space has a finite basis, then all of the bases have the same number of elements. This common number is called the *dimension* of the vector space.

Linear maps between vector spaces are defined just as for $\boldsymbol{R}^n$. A linear map $T : V \to W$ between vector spaces is called an *isomorphism* if T is one-to-one and onto. In this case, V and W are said to be *isomorphic*. Intuitively, this means that V and W are really *the same vector space*, up to a different convention for naming the elements. One standard result in M52 is that an n-dimensional vector space is isomorphic to $\boldsymbol{R}^n$. For instance, a basis for the vector space of quadratic polynomials is given by $\{1, x, x^2\}$, and so P_2 is isomorphic to $\boldsymbol{R}^3$. An isomorphism is given by $T(ax^2 + bc + c) = (a, b, c)$. Our third example of a vector space is infinite dimensional. It is not isomorphic to $\boldsymbol{R}^n$ for any n.

Quadratic Forms and Inner Products: A *quadratic form* on a real vector space V is an operation Q which maps pairs of vectors to the reals. Q satisfies the following axioms.

- $Q(v_1, v_2) = Q(v_2, v_1)$ for all $v_1, v_2 \in V$.
- $Q(v_1 + v_2, v_3) = Q(v_1, v_3) + Q(v_2, v_3)$ for all $v_1, v_2, v_3 \in V$.
- $Q(rv) = rQ(v)$ for all $r \in \boldsymbol{R}$ and all $v \in V$.

The dot product on $\boldsymbol{R}^n$ is a classic example of a quadratic form. The formula is

$$Q(x, y) = x_1 y_1 + ... + x_n y_n.$$

The dot product is a special case of an *inner product*. The quadratic form Q is an inner product if it is such that $Q(v, v) > 0$ for all nonzero $v \in V$.

Given a pair (V, Q), where V is a vector space and Q is a quadratic form, one can consider $O(V, Q)$, the set of linear isomorphisms from V to itself such that $Q(T(v), T(w)) = Q(v, w)$ for all $v, w \in V$. Then $V = \boldsymbol{R}^n$ and Q is the dot product, we are just talking about the set of rotations of $\boldsymbol{R}^n$ that fix the origin. Other examples of Q and V (when Q is not necessarily an inner product) lead to very interesting geometrical objects, like hyperbolic geometry and Minkowski spacetime. In particular, the linear algebra in M52 is a nice framework for special relativity.

Eigenvalues and Eigenvectors: Let $T : V \to V$ be a linear transformation from V to itself. An *eigenvector* of T is a vector v such that

$$T(v) = \lambda v,$$

for some real λ. The number λ is called the *eigenvalue* associated to the eigenvector. In M52 you learn the significance of eigenvalues and eigenvectors, as well as methods for computing.

An example will illustrate some of the significance of these concepts. Suppose that you have two kinds of cells, Type A and Type B. Each day, a Type A cell mutates to a Type B cell. At the same time, each Type B cell splits into a Type A cell and a Type B cell. Starting with, say, a single cell of Type A, one might wonder roughly how many cells one has after N days. One can think of the growth of these cells as a linear transformation on $\boldsymbol{R}^2$. The point (x, y) represents a population of x cells of Type A and y cells of Type B. Then

$$T(x, y) = (y, x + y),$$

represents the population the next day.

Let

$$\phi = \frac{1 + \sqrt{5}}{2} = 1.618...$$

be the famous golden ratio. It turns out that T has a basis of eigenvectors, given by

$$v_1 = (\phi, 1); \qquad v_2 = (-\phi, 1),$$

with corresponding eigenvalues

$$\lambda_1 = \phi; \qquad \lambda_2 = -1/\phi.$$

Since $\{v_1, v_2\}$ is a basis, and other vector will be a combination of these. For such a combination, of the form $a_1 v_1 + a_2 v_2$, we compute

$$T(a_1 v_1 + a_2 v_2) = a_1 T(v_1) + a_2 T(v_2) = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2.$$

In general, we have

$$T^n(a_1 v_2 + a_2 v + 2) = a_1 \lambda_1^n v_1 + a_2 \lambda_2^n v_2.$$

Note that $|\lambda_1| > 1$ and $|\lambda_2| < 1$. So, for large n, the second term is negligible. That is

$$T^n(a_1v_1 + a_2v_2) \approx \phi^n \times (a_1v_1).$$

This tells us that we can expect an exponential growth of cells, with the growth rate being ϕ. That is, after many days have passed, the population of cells just about grows by a factor of ϕ each day.

In general, studying the eigenvalues and eigenvectors of a linear transformation gives you a lot of information about what happens when this linear transformation is applied over and over again. Our example is a prototypical application of linear algebra to fields such as tumor growth or population genetics.