

Pure Mathematics? Positively Applied!

Introduction. Even amongst mathematically literate scientists there is considerable ignorance about what mathematicians in general, and pure mathematicians in particular, are and what they do. Paradoxically, this ignorance seems more widespread amongst those who use more mathematics in their own work.

The division between mathematicians and non-mathematicians is probably the easiest to define. Mathematicians *create* new mathematics, non-mathematicians don't! The creation of new mathematics may take many forms and be performed for a variety of reasons, but that is surely what distinguishes (say) a mathematician doing physics from a physicist using some mathematics in his/her research.

[And yes, I know there are hybrids, but outside mathematics no definition is perfect!]

The distinction between various kinds of mathematicians is harder to make. Certainly, so-called “applied mathematicians” create their mathematics with particular applications in mind, but so do many “pure mathematicians”. If there is a dividing line, which is not at all clear to me, one possible definition would be that applied mathematicians create their mathematics to solve one particular problem, whilst pure mathematicians tend to generalise and try to produce tools that are applicable to a wide variety of problems.

There is a wide spectrum of different kinds of pure mathematician. At one extreme there are those who are actively inspired by a particular problem from some other subject area (at one time I would have said “science”, but now many other subjects are becoming mathematical and these new areas are providing some of the more interesting of the current mathematical problems). At the other extreme are those who disavow all connections with reality and pursue mathematics for its own internal beauty. An extreme example of such a person was the famous English Analyst G.H. Hardy who wrote a book entitled “A Mathematician’s Apology” to justify his life’s work. In it he cited a particularly beautiful, but useless, example of his work on number theory. That work now has important applications in cryptography! The majority of pure mathematicians lie somewhere in the middle, vaguely aware that their work will or may have utility somewhere but happy enough to pursue it for its own sake.

As Hardy's example shows, it is not possible to say in advance what mathematics will turn out to be useful and in what fields. Sometimes applications drive the development of new mathematics—more often mathematicians have the tools waiting for the new applications. The success of mathematical descriptions of the world has often been termed “unreasonable”—pure mathematicians' success at anticipating what mathematics will be needed in decades to come seems to me equally unreasonable! Yes, 95% of pure mathematics is never used for any application but I suspect that this is neither a better nor a worse record than any other academic discipline.

In my talk today, I want to illustrate the general nature of the mathematics created by pure mathematicians by giving a brief introduction to the area of mathematics that I research into myself and giving two applications of one of the central results in the subject. One of these applications prompted the simultaneous development of this subject in the 1940's and early 1950's in both the USA and the USSR. The other is a very recent application in a very different subject area.

Positive Matrices. Almost anyone who has used mathematics in any non-trivial manner will be familiar with the notion of an $n \times n$ *matrix* $\mathbf{A}$. This is simply n^2 numbers written out in a square array (conventionally surrounded by some kind of bracket) thus:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Apart from any other considerations, these are useful for denoting the coefficients of a system of n linear expressions in n unknowns. If $\mathbf{x}$ denotes a column vector of n unknowns

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

then $\mathbf{A}.\mathbf{x}$ is used as a shorthand to denote the column vector

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{pmatrix}$$

Although this is a very simple notational convention, such conventions have played a very important part in mathematics. For example, if $\mathbf{A}$ and $\mathbf{B}$ are two $n \times n$ matrices and $\mathbf{x}$ a column vector of length n , then we can form $\mathbf{A}(\mathbf{B}.\mathbf{x})$ and the notation suggests rewriting this as $(\mathbf{A}.\mathbf{B}).\mathbf{x}$. It is not too difficult to show that this is a valid definition of the matrix product $\mathbf{A}.\mathbf{B}$ and this is indeed a much easier and more useful way of thinking of $\mathbf{A}.\mathbf{B}$ than the normal manner of writing out horrible summation formulae for its entries. Once we accept this, it is but a small step to writing $\mathbf{A}^k$ for $\mathbf{A}.\mathbf{A} \dots \mathbf{A}$ with k factors.

Most people first meet matrix notation in connection with the solution of sets of simultaneous linear equations, where we are given the matrix $\mathbf{A}$ and a vector $\mathbf{b}$ thus:

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

and we want to solve the equation $\mathbf{A}.\mathbf{x} = \mathbf{b}$, which is just a shorthand for the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

In many physical situations the vector $\mathbf{b}$ is also unknown, but it might be desired (for example) that $\mathbf{A}.\mathbf{x} = \mathbf{x}$. This set of equations cannot be solved in general (apart from the trivial solution where every entry in $\mathbf{x}$ is zero) but there is often enough leeway in the physical problem behind the mathematics that it suffices to be able to find $\mathbf{x}$ (with at least one of its entries non-zero) and a number λ such that $\mathbf{A}.\mathbf{x} = \lambda\mathbf{x}$. Such an $\mathbf{x}$ is called an *eigenvector* for $\mathbf{A}$ and λ is known as the (corresponding) *eigenvalue*.

Amongst all $n \times n$ matrices, that one with all the diagonal entries equal to 1 and others being zero is important enough to warrant a special name. It is usually denoted by $\mathbf{I}_n$, or simply $\mathbf{I}$ if the size is not important or is clear. Here is $\mathbf{I}_4$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

This is called the *identity matrix* of order n , since if $\mathbf{x}$ is an n -dimensional column vector then $\mathbf{I}_n \cdot \mathbf{x} = \mathbf{x}$.

The *inverse* of an $n \times n$ matrix $\mathbf{A}$ is that matrix $\mathbf{B}$ with $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$, if such a matrix exists. Not all $n \times n$ matrices have inverses. If $\mathbf{A}$ has an inverse it is denoted by $\mathbf{A}^{-1}$. There is an important connection between inverses and eigenvalues:

The number λ is an eigenvalue of the $n \times n$ matrix $\mathbf{A}$ if and only if $\mathbf{A} - \lambda \mathbf{I}_n$ does NOT have an inverse.

Every square matrix has eigenvalues, as long as one is prepared to accept eigenvalues and eigenvectors involving complex numbers rather than real ones. Without allowing that there may be no eigenvalues. For example even in the simple 2×2 case where $\mathbf{A}$ is

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

if $\mathbf{A} \cdot \mathbf{x} = \lambda \mathbf{x}$ then we require x_1 and x_2 to satisfy the two equations

$$\begin{aligned} x_2 &= \lambda x_1 \\ -x_1 &= \lambda x_2. \end{aligned}$$

It follows that $x_2 = \lambda x_1 = \lambda(-\lambda x_2) = -\lambda^2 x_2$. If $x_2 = 0$ then $x_1 = -\lambda x_2 = 0$ also and we do not have an eigenvector! But if $x_2 \neq 0$ then we can divide by x_2 in the equality $x_2 = -\lambda^2 x_2$ to find that $\lambda^2 = -1$. No real number has its square equal to -1 , so that λ is not real.

In problems that correspond in a direct manner to the physical world, it will only be real eigenvectors and real eigenvalues that have any physical meaning. Although in any specific problem it may be possible to find all the (possibly complex) eigenvalues and eigenvectors and then determine whether or not any of them are real, it would be vastly preferable to be able to assert that there must be real eigenvectors and eigenvalues.

We have seen that in general this is not possible, but there are special cases where it is. A vector $\mathbf{x}$ is *positive* (and we write $\mathbf{x} \geq 0$) if each $x_i \geq 0$ and is *fully positive* if each $x_i > 0$, in which case we write $\mathbf{x} > 0$. Similarly a matrix $\mathbf{A}$ is called *positive* if each $a_{ij} \geq 0$ and is called *fully positive* if each $a_{ij} > 0$. We again write $\mathbf{A} \geq 0$ or $\mathbf{A} > 0$ as the case may be. In many physical models the vectors and matrices that arise *must* be positive. For example the vector could describe proportions or populations for several alternative which physically have to be positive vectors if they are to correspond to physically meaningful solutions. An important alternative way of describing a positive matrix $\mathbf{A}$ is by saying that if $\mathbf{x} \geq 0$ then $\mathbf{A}\mathbf{x} \geq 0$. I.e. $\mathbf{A}$ takes physically meaningful vectors into physically meaningful ones.

Now we must give a proper mathematical definition.

The *spectral radius* of a matrix is the largest absolute value of any eigenvalue (real or complex) of that matrix.

There are always eigenvalues (as long as we allow complex ones), so this always make sense. The matrix

$$\begin{pmatrix} -2 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

has eigenvalues $-2, i$ and $-i$ which have absolute values $2, 1, 1$ respectively, so this has a spectral radius of 2 . Note that 2 is *not* an eigenvalue. [**Picture 1**].

And now our first major result.

If $\mathbf{A}$ is a positive matrix then its spectral radius is an eigenvalue of $\mathbf{A}$ and there is at least one corresponding eigenvector $\mathbf{x} \geq 0$.

The matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

has eigenvalues $-1, 1$ and 2 , so has spectral radius equal to 2 , which is an eigenvalue. [**Picture 2**] All the other eigenvalues have smaller absolute value. On the other hand the matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}$$

has eigenvalues of -2 and 2 , so again the spectral radius is an eigenvalue, but so is -2 which has the same absolute value. **[Picture 3]**

Although regarded as an elementary result in this subject area, the proof of it is nowhere near as easy as its statement, so I will omit it (you will be glad to hear that I am going to leave out all of the hard bits!) except to say that it is definitely analysis rather than linear algebra. For anyone interested, the proof involves bounds for the Neumann series of a resolvent. There is no guarantee that any other eigenvalue is real nor that the corresponding eigenvectors are reals.

If we have non-zero eigenvalues then a little bit of scaling (in the physical set-up) can often change things so that $\mathbf{A}\mathbf{x} = \mathbf{x}$. This is not possible if the eigenvalue is zero, so estimates of the size of the eigenvalue are important. There is a simple estimate for the size of the spectral radius for positive matrices that does not work in general.

Add up the entries in each row of the matrix to obtain a column vector of non-negative numbers. The spectral radius lies somewhere between the largest and the smallest of these numbers.

In particular it can only be zero if one row of the matrix consists entirely of zeros. The matrix

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

has rowsums of

$$\begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix}$$

so that the spectral radius lies between 2 and 3. In fact the eigenvalues are 0 and 2, so the spectral radius is 2. In some special cases this estimate allows one to find the spectral radius exactly.

Although rather harder even to state, a stronger result due to Perron (1907) and Frobenius (1909 & 1912) is much more useful. Again this requires a definition.

The $n \times n$ square matrix $\mathbf{A}$ is called *reducible* if there is a proper subset of numbers $\{k_1, k_2, \dots, k_m\} \subset \{1, 2, \dots, n\}$ such that if $\mathbf{y} = \mathbf{A}\mathbf{x}$ and $x_k = 0$ for $k = k_1, k_2, \dots, k_m$ then $y_k = 0$ for $k = k_1, k_2, \dots, k_m$. A matrix that is not reducible is called *irreducible*. These are the important ones.

Fully positive matrices are irreducible, but so are some others. Even more is actually true for fully positive matrices—they are primitive.

A square matrix $\mathbf{A}$ is *primitive* if $\mathbf{A}^k$ is irreducible for all positive integers k .

There are many wonderful properties of primitive matrices, for example:

An irreducible positive square matrix $\mathbf{A}$ is primitive if and only if the only eigenvalue, λ , of $\mathbf{A}$ with $|\lambda|$ equal to the spectral radius of $\mathbf{A}$ is the spectral radius itself.

The matrix

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

is irreducible, but its eigenvalues are the three complex cube roots of 1. **[Picture 4]** These all have absolute value equal to 1, so the matrix is not primitive. In fact the cube of this matrix is I_3 which is reducible. On the other hand

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

is also irreducible and has eigenvalues -1 and 2 **[Picture 5]** so is primitive.

The Perron-Frobenius theorem asserts many important properties of irreducible positive matrices. I will mention only one:

If $\mathbf{A}$ is an irreducible positive matrix then its spectral radius is an eigenvalue of $\mathbf{A}$ and there is at least one corresponding eigenvector $\mathbf{x} > 0$.

A simple and early application of this theory was to *Markov chains*.

Suppose we have a physical system capable of n distinct states. It is observed at discrete times $k = 1, 2, \dots$. The sequence of observations $(X_k)_{k=1}^{\infty}$ giving the state of this system at time k is a Markov process if the probability of X_k taking a particular value depends only on the value of X_{k-1} and *not* any previous observation, so the process that creates this sequence of states has a very short memory. Such a process is described entirely by the n^2 numbers p_{ij} which are the probability that $X_k = j$ given that $X_{k-1} = i$. These give us an $n \times n$ positive matrix $\mathbf{P}$ (positive because probabilities lie between 0 and 1). No matter what X_{k-1} may be, X_k must take precisely

one of the n possible states so that

$$\sum_{j=1}^n p_{ij} = 1$$

for $i = 1, 2, \dots, n$. We saw earlier that we could estimate the spectral radius of a positive matrix by looking at the sums of the entries in each row. In this case these sums are all equal to 1, so the spectral radius of $\mathbf{P}$ is precisely 1.

If $\mathbf{x}$ is a vector describing the probability of the system being in each of the n possible states at a given point in time, then $\mathbf{P}\mathbf{x}$ gives the probabilities of it being in each state at the next moment in time. As a first consequence of the theory of positive matrices we find:

For any Markov process there is stationary distribution of the system.

I.e. a distribution $\mathbf{x}$ such that $\mathbf{P}\mathbf{x} = \mathbf{x}$. This is a direct consequence of our first result about positive matrices, using the fact that the spectral radius is 1 to conclude that 1 is an eigenvalue and that there corresponds to it a *positive* eigenvector. Now it only remains to divide each entry in this eigenvector by the sum of all its entries to get an eigenvector that is a probability distribution.

Much more can be deduced if the matrix $\mathbf{P}$ is primitive—in particular if every $p_{ij} > 0$.

If $\mathbf{P}$ is primitive then there is a fully positive probability distribution $\mathbf{q}$ such that for any starting distribution $\mathbf{x}$, $\mathbf{P}^k\mathbf{x}$ converges to $\mathbf{q}$.

If $\mathbf{P}$ is merely required to be irreducible then slightly more complicated things happen that I will not go into here. It suffices to point out that the general theory of positive matrices allows a complete description of what happens.

Banach lattices and positive operators. Although positive matrices are very well-behaved amongst matrices in general and many applications of matrix theory actually involve positive matrices, unfortunately matrices are not sufficient to model very many physical situations. Even in the last, rather simple, example we can see that not many physical systems have only a finite number of states. Often a description of the state of a system

involves one or more variables which can take infinitely many different values. These can sometimes be discrete (but with no *a priori* bound to it), like the Gross National Product in pounds but more often continuous like the energy of an atomic particle. In both these cases there are arguments for treating the measurement as both discrete and continuous, so there is a good case for having a general theory that can handle all such cases at once.

The notion of a *Banach lattice* is a generalisation of that of an n -tuple of numbers to an infinite-dimensional setting. The precise definition does not matter for our purposes—it suffices to think of elements of a Banach lattice as generalised vectors that we can add, multiply by scalars, calculate the length of and recognise which ones are positive.

Instead of a definition, I will give some examples.

- (i) ℓ_∞ denotes the collection of all sequences of real numbers (x_n) such that there is a number K with $|x_n| \leq K$ for all positive integers n . Sums and multiples of such vectors are easy to compute:

$$\begin{aligned}(x_n) + (y_n) &= (x_n + y_n) \\ \lambda(x_n) &= (\lambda x_n)\end{aligned}$$

The length of (x_n) is just the smallest K for which every $|x_n| \leq K$ and (x_n) is positive if $x_n \geq 0$ for all n . This is a straightforward generalisation of n -tuples of scalars to an infinite sequence, except for the boundedness condition.

- (ii) The space of continuous real-valued functions on the unit interval, $C([0, 1])$, is also a Banach lattice. This time we have

$$\begin{aligned}(f + g)(t) &= f(t) + g(t) \\ (\lambda f)(t) &= \lambda f(t),\end{aligned}$$

the length of f is the smallest K such that $|f(t)| \leq K$ for all $t \in [0, 1]$ and f is positive if $f(t) \geq 0$ for all $t \in [0, 1]$.

- (iii) The space of (Lebesgue) integrable real-valued functions on $[0, 1]$ is again a Banach lattice. Sums and multiples are calculated as in the last example. This time the length of f is

$$\int_0^1 |f(t)| dt$$

and f is positive if $f(t) \geq 0$ *almost everywhere*. This last condition simply says that $f(t)$ can be negative, but only on a small set of points (for example we can forget about countably many points where it is negative.)

The important connection between matrices and n -tuples of numbers is that the equations represented by $\mathbf{y} = \mathbf{A}.\mathbf{x}$ are linear. Thus for any two vectors, $\mathbf{x}_1$ and $\mathbf{x}_2$, and any scalar λ we have

$$\begin{aligned}\mathbf{A} . (\mathbf{x}_1 + \mathbf{x}_2) &= \mathbf{A} . \mathbf{x}_1 + \mathbf{A} . \mathbf{x}_2 \\ \mathbf{A} . (\lambda \mathbf{x}_1) &= \lambda \mathbf{A} . \mathbf{x}_1\end{aligned}$$

Both of these identities make perfectly good sense (although they may or may not hold) for a mapping $\mathbf{A}$ from a Banach lattice into itself and are the definition of a *linear operator* on a Banach lattice. Linear operators are not, in general, easily described. There is certainly no simple description in terms of matrices, so we can't use the entries to define positivity of a linear operator. The other characterisation that we had for positivity of a matrix works though:

A linear operator $\mathbf{A}$ on a Banach lattice is *positive* if $\mathbf{A} . \mathbf{x}$ is positive whenever $\mathbf{x}$ is positive.

Many of the definitions from the elementary theory of matrices may be re-used in this setting. The *identity* linear operator $\mathbf{I}$ is that operator with $\mathbf{I}\mathbf{x} = \mathbf{x}$ for every vector $\mathbf{x}$ in the Banach lattice, the product $\mathbf{A}.\mathbf{B}$ is that operator which takes $\mathbf{x}$ to $\mathbf{A} . (\mathbf{B} . \mathbf{x})$ and the *inverse* of $\mathbf{A}$ is that operator $\mathbf{B}$ (if it exists) with $\mathbf{A} . \mathbf{B} = \mathbf{B} . \mathbf{A} = \mathbf{I}$.

Although positive operators are a considerable generalisation of the notion of a positive matrix, many (but not all) of the special properties of positive matrices extend to positive operators. That is not the case for properties of general matrices to anywhere near as large an extent.

In general, even when they are positive, linear operators need not have any eigenvectors. However the *spectrum*, the set of those λ such that $\mathbf{A} - \lambda\mathbf{I}$ is not invertible, must be non-empty. We saw earlier that for matrices this was the same as the set of eigenvalues so this is a reasonable substitute in the infinite dimensional setting. The *spectral radius* is now defined to be the largest value of $|\lambda|$ for λ in the spectrum. With this change to the definition we have:

For any positive operator $\mathbf{A}$ on a Banach lattice, the spectral radius is in the spectrum.

What is missing from this theorem is any notion of an eigenvector (although there are approximate eigenvectors around) and any indication of when the spectral radius is non-zero. There is a very special (but very important as far as applications are concerned) class of operators for which the spectrum consists of eigenvalues. These are the *compact* operators.

An operator $\mathbf{A}$ on a Banach lattice is *compact* if whenever $(\mathbf{x}_n)$ is a sequence of vectors of length 1, then we can choose $n_1 < n_2 < n_3 \dots$ such that $\mathbf{A}\mathbf{x}_{n_1}, \mathbf{A}\mathbf{x}_{n_2}, \mathbf{A}\mathbf{x}_{n_3} \dots$ converges to something in the Banach lattice.

For any compact positive operator $\mathbf{A}$ on a Banach lattice, the spectral radius is an eigenvalue of $\mathbf{A}$ and there is a corresponding eigenvector which is positive.

The hypotheses of this theorem can be varied in various ways whilst still maintaining the same conclusion. For example the assumption of compactness can be weakened if the assumption of positivity is strengthened or if extra assumptions are made about the Banach lattice. Such results tend to be proved on an *ad-hoc* basis when needed for particular applications.

When is the spectral radius positive? We don't have a matrix representation, so we can't start looking at row sums to get an estimate. One extra condition that guarantees positivity of the spectral radius is that $\mathbf{A}$ also be irreducible. The definition of this is slightly technical, but is similar to that for matrices except that instead of involving $\mathbf{x}_k$ being zero for certain values of k , it will involve (say) functions which are zero on a certain set.

Any irreducible compact positive operator on a Banach lattice has strictly positive spectral radius.

This is the most recent (and most difficult) of all the theorems that I have mentioned today. It was proved in the mid-1980's by a young Dutch mathematician, Ben de Pagter.

Two applications. I am going to give very brief descriptions of two applications of the theory that I outlined earlier.

Consider the distribution of neutrons in a nuclear reactor. How can we describe this distribution? We cannot specify each neutron, instead we must talk about the number of neutrons in a particular region. Things are rather more complicated as we do not only want to specify where neutrons are, but also how fast they are moving and in what direction. The *phase space* of a nuclear reactor is (essentially) a geometric model of the reactor but with each point replaced by a three-dimensional vector space of possible neutron velocities. The *neutron distribution* of the reactor is a measure on the Borel subsets of the phase space (roughly speaking, a count of how many neutrons with a specified range of velocities are in any given region of the reactor). The collection of all neutron distributions on a given nuclear reactor is a Banach lattice where positivity of a neutron distribution has its intuitively

obvious definition that no region contains a negative (or complex) number of neutrons. Clearly only positive elements of this Banach lattice have a physical meaning.

In a simplified description of such a reactor, two things can happen to a given neutron in a specified (small) time interval. It can bounce off a nucleus and be *scattered* or it can cause *fission* of a nucleus and generate (on average) ν neutrons. There is a positive linear operator $\mathbf{S}$ (with a rather nasty definition) on this Banach lattice such that given a neutron distribution $\mathbf{m}$ now, the scattered neutrons will have distribution $\mathbf{S}\mathbf{m}$ in the next generation. There is similarly a positive linear operator $\mathbf{F}$ such that $\mathbf{F}\mathbf{m}$ is the distribution of fission events caused by the neutron distribution $\mathbf{m}$. To work out the distribution of fission neutrons emitted in successive generations, one must take into account neutrons that have been scattered an arbitrary number of times and it turns out that this is described by an operator

$$\mathbf{P} = \nu(\mathbf{I} + \mathbf{S} + \mathbf{S}^2 + \mathbf{S}^3 + \cdots)\mathbf{F} = \nu(\mathbf{I} - \mathbf{S})^{-1}\mathbf{F},$$

which is also positive. If ν is chosen such that $\mathbf{P}$ has spectral radius equal to 1 (by altering the geometry of the reactor, for example) then 1 is in the spectrum of $\mathbf{P}$. Various assumptions about the reactor design suffice to establish that 1 will actually be an eigenvalue of $\mathbf{P}$ and that there will be an eigenvector corresponding to it that is positive. More importantly, refinements of the results I have presented show that in these circumstances, no matter what the starting neutron distribution, the neutron distribution converges to this eigenvector in the long term, so that the reactor is stable.

My second example concerns the role of the balance between mutation and selection processes in population dynamics. Consider a large (effectively infinite) population that reproduces asexually and suppose that reproduction occurs in discrete generations. Consider some property of an individual that can be measured numerically. This can include any property that allows only finitely many possibilities by assigning a number to each of those possibilities. There may be finitely many distinct types (e.g. presence or absence of some feature), there may be (theoretically) infinitely many types indexed by the integers or the types may be indexed by a real number in a certain range (e.g. size). The possible set of types is denoted by $\mathcal{M}$. When we write $\int_{\mathcal{M}} f(x) dx$, we mean normal integration when $\mathcal{M}$ is an interval in the reals and a sum in the discrete case, but bearing in mind that this might be an infinite sum and that convergence might be a problem.

In all of these cases the distribution of the population at any given time will be represented by a positive (we can't have a negative number of individuals of a given type) vector in a certain Banach lattice. In the case of n different types it will be an n -tuple, but in general will be represented by a function on the set of all possible types. Rather than actual numbers of individuals, it is conventional to use this vector to denote the proportion of the population in each type.

Two factors affect the population distribution. There will be mutation from one type to another and some types may be better fitted for survival than others. There is a function $u(x, y)$ describing the conditional probability that mutation from type y to type x occurs, given that mutation occurs at all, a number μ giving the overall mutation rate, and a function $w(x)$ giving the probability of survival of a type x individual from one generation to the next. The evolution of the population from one generation to the next is described by the positive linear operator $\mathbf{A}$ defined by

$$\mathbf{A}\mathbf{p}(x) = (1 - \mu)\mathbf{p}_s(x) + \mu \int_{\mathcal{M}} u(x, y)\mathbf{p}_s(y) dy$$

where

$$\mathbf{p}_s(x) = \frac{w(x)\mathbf{p}(x)}{\int_{\mathcal{M}} w(y)\mathbf{p}(y) dy}.$$

is the proportion of type x out of those surviving for one generation.

There are many results in the literature for this model assuming various extra conditions (which cannot be omitted) on u and w . In all cases it is necessary to assume that $w(x) > 0$ everywhere but that $\lim_{x \rightarrow \infty} w(x) = 0$ —i.e. extreme cases can occur but are not well-suited to survival. In the case of finite numbers of types one need only assume further that all probabilities are non-zero (this case is just a Markov chain problem), in other cases it is necessary to assume that the mutation rate $u(x, y)$ depends either on x only or on $x - y$. Both of these mathematical assumptions correspond to biologically reasonable hypotheses. In all these cases it can be shown that the operator $\mathbf{A}$ has a fully positive eigenvector and hence:

There is a steady-state fully positive population distribution and the population converges to that no matter what the starting distribution.

Summary. My two applications used very similar mathematical tools, but were taken from two entirely different subject areas. I could also have used examples from many other fields. Bertrand Russell once described mathematicians as people who didn't know whether what they were saying was true or false and didn't care. He was referring to the axiomatic nature of modern mathematics, but since has frequently been misquoted in terms of abuse. I, for one, am not ashamed to admit that I don't know what research I am doing! As a mathematician, I don't know whether the work that I am doing today is going to provide the mathematical foundation of some branch of theoretical physics, or population dynamics, or finance, or statistical decision theory, or economics (all of which currently use some aspects of vector lattice theory), some discipline that has not yet discovered the delights of positive operators or even some discipline which hasn't been invented yet!

Should I know? Should I care? I leave it to you to decide.