# LETTERS, OPINIONS, AND COMMENTS

## The teaching of engineering mathematics

This note is about a subject that is easy to neglect. It is the basic course in applied mathematics, which comes after the student has first met calculus and differential equations and matrices. It may be an undergraduate course or a first-year graduate course, intended to provide the mathematics that is going to be needed and useful. Often it is taught by the mathematics department, which may not recognize what modern engineering has become—and therefore, with the best intentions, it is not exciting. It becomes a "service course," but it misses the chance to be of real service. Inertia wins, but the discipline loses.

I believe it is possible to do better. That course can bring together the classical theories of differential equations and Fourier analysis and the modern methods of solving those equations and carrying out that analysis. The crucial point is that the new methods have not rejected the old ones. The exact opposite is the case—they are totally dependent on the central ideas of applied mathematics. The fast Fourier transform relies, as completely as any infinite series expansion, on orthogonality. Finite elements can go nowhere without the equation of virtual work, or the variational principles of mechanics. Iterative methods converge or diverge at speeds that are controlled entirely by eigenvalues. And for special geometries, the old idea of separation of variables is implemented by the spectral method. Good algorithms run parallel to good theory, and by making that theory concrete and explicit they bring it to life.

The last paragraph suggests one change that I hope to see. Along with continuum problems should come discrete problems—differential equations should appear together with matrix equations. However, I am absolutely not proposing that this should become a course in numerical methods. There is something more important to teach, and it must come early and clear. It is the *basic framework of applied mathematics*, and it is the source of the equations in the first place.

This framework is shared by one application after another—that is the reason why a mathematics course provides such an opportunity to be useful—and it starts with problems of *equilibrium*. In describing it I am only highlighting what is familiar; the model is not new or revolutionary. My goal is to make it familiar to our students too.

The common feature of equilibrium equations is the appearance of a triple product $A^TCA$. In differential equations it will govern an elastic rod, or a beam, or a membrane (or potential flow or electrostatics or thermal equilibrium):

$$-\frac{d}{dx}\left(c\,\frac{du}{dx}\right)=f$$

or

$$\frac{d^2}{dx^2}\left(c\,\frac{d^2u}{dx^2}\right)=f$$

or

$$-\operatorname{div}(c \operatorname{grad} u)=f.$$

In each case there are three factors on the left side. The problems are *symmetric*, because $A^T$ appears together with $A$. In the discrete case, that is a rectangular matrix and its transpose, with a square symmetric $C$ in between. The product $A^TCA$ is the stiffness matrix! In the continuous case we meet derivatives $A = d/dx$ or $A = d^2/dx^2$ or $A =$ gradient, and part of the mathematics is to identify the corresponding $A^T = -d/dx$ or $(-d/dx)^2$ or $-$ divergence. Those come from integration by parts, or Green's theorem in higher dimensions, or the equation of virtual work. They put calculus to use, and I apologize for rushing past all the fun. (The boundary conditions come too.) What has to be emphasized is the importance of this framework.

You recognize it in the deeper problems of continuum mechanics:

$$\text{displacement} \xrightarrow{A} \text{strain}$$

$$\xrightarrow{C} \text{stress} \xrightarrow{A^T} \text{force}.$$

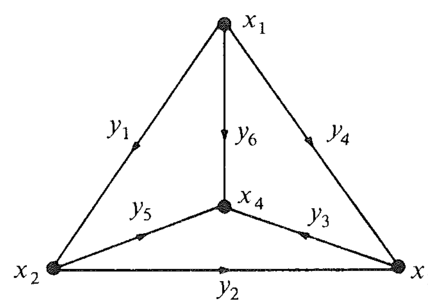
The constitutive law is expressed by $C$. It could be nonlinear, but in this course Hooke's law is enough. The point is that the mathematics is not divorced from the engineering and left to solve equations in a vacuum. If the connections are not made now, most of them will never be made.

Another connection is to *flow through a network*. That may not be mechanics—$C$ now represents Ohm's law—but it is absolutely basic:

$$\text{potential} \xrightarrow{A} \text{potential difference}$$

$$\xrightarrow{C} \text{current} \xrightarrow{A^T} \text{current source}.$$

The last step is Kirchhoff's current law: The net flow into each node is zero. It is the analog of balance of forces, and the outstanding example of a conservation law. In the continuous case it illustrates the divergence theorem (in use). The net flow out of a region is zero when the divergence—represented by $A^T$— is zero inside.




In the matrix case, $A$ takes differences instead of derivatives. Each row contains a $+1$ and a $-1$, to give the potential difference across an edge. It is an *incidence matrix* or a *connectivity matrix*, with a row for every edge and a column for every node. In the figure below it is $6 \times 4$:

$$A=\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

Notice one important point. The four columns are not independent. Their sum is a column of zeros. The constant vector $c = (1,1,1,1)$ satisfies $Ac = 0$, and the potentials cannot be determined from the potential differences. All potentials can be increased by a constant (homogeneous solution added to particular solution) unless we *ground a node*. That is the "boundary condition" in the discrete case, like fixing the end of a rod or the boundary of a membrane. It removes the last column of $A$.

The matrix $A^TCA$ is now symmetric and *positive definite*. That property brings together the key facts of matrix theory, in a basic application of linear algebra. The eigenvalues are positive, controlling the dynamics. The energy $\frac{1}{2}x^TA^TCAx$ is positive, controlling the statics. The pivots are positive, controlling the stability of the numerical algorithm (elimination). For a symmetric matrix those conditions are equivalent, and it is important to see the matrices themselves:

$$A^TA = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix},$$

$$A^TCA =$$

$$\begin{bmatrix} c_1 + c_4 + c_6 & -c_1 & -c_4 \\ -c_1 & c_1 + c_2 + c_5 & -c_2 \\ -c_4 & -c_2 & c_2 + c_3 + c_4 \end{bmatrix}.$$

Edges 3, 5, 6 go to ground; thus the conductances $c_3$, $c_5$, $c_6$ appear only on the diagonal and make the matrix positive definite.

The same framework governs least-squares estimation, and the mechanics of a truss. For a plane truss the flexible wires become solid bars, and there are two displacements (instead of one potential) at every joint. The original $A$ becomes $6 \times 8$. If there are supports, they act to prevent rigid motion and to remove columns corresponding to fixed displacements (just as grounding a node removed the unknown constant from the potential). The final matrix exposes the stability or instability of the truss:

Stable truss ($A$ has independent columns)

(1) Statically determinate:
  $A$ is square and invertible
(2) Statically indeterminate:
  more rows than columns

Unstable truss ($A$ has dependent columns)

(3) Rigid motion: too few supports
(4) Mechanism: a displacement without stress, from $Ax = 0$

An unstable truss could move rigidly and also deform, if it is really unsafe. Of course, robotics looks at the same matrix $A^TCA$.

In a continuum, the direct analog of network flow is potential flow. The current law becomes "divergence = zero" and the voltage law is "curl = zero." Again vector calculus enters, to show that the curl is zero when the flow comes from a potential. The identity curl grad = 0 brings in the gradient—just as potential differences appeared, to make the circulation zero around every loop. I find that these analogies, when done properly and carefully, are welcomed by the students. They see the framework, they see the purpose of vector calculus, and they also see the purpose of the course.

In fact, there are several purposes. One is to formulate the equation, in this case $A^TAu = 0$ (Laplace's equation div grad $u = 0$). Another is to describe analytical methods for its solution—Fourier series or Bessel functions or complex variables. A third purpose is to propose numerical methods, in this case finite differences or finite elements. Note how the computer has an important place, but the ideas come first.

I will try to comment briefly on some individual topics. Then I will refer briefly to my new textbook (Strang, 1986) and its goal, to combine these topics into a course on modern applied mathematics:

*1. Partial differential equations*: That is the central thread of the course, to introduce those equations and solve them. For equilibrium equations (*boundary-value problems*) the framework is $A^TCAu = f$. For dynamic equations (*initial-value problems*) there is an inertial term $Mu_{tt}$. The positive definiteness of $K = A^TCA$ produces oscillating solutions in the wave equation and decaying solutions in the heat equation. The underlying eigenvalue problem is $Ku = \lambda Mu$, and $\lambda$ had better not be negative.

These are ordinary differential equations when $K$ is a matrix. They are partial differential equations when $A$ is $d/dx$; I do not believe that step to be an overwhelming obstacle. The key is to solve the equations for forces that are interesting —especially a uniform load $f = 1$ or a point load $f = \delta$ or an exponential $f = e^{ikx}$. The response to $f = \delta$ is the Green's function. The responses to exponentials are more exponentials, and we come naturally to transforms.

*2. Fourier analysis*: This remains vital to every problem that is linear and "stationary." In differential equations the coefficients are constant; in integral equations there is a convolution; in matrix equations there are constants down each diagonal. The eigenfunctions are exponentials. The eigenvalues are the transforms. The problem changes from analysis to algebra, when the harmonics are uncoupled.

The underlying solution process has three steps. The data is expanded into exponentials, each frequency is followed separately, and the pieces are recombined into the solution. I believe that the first step may have had too much emphasis—computing Fourier coefficients has come to dominate the homework. That is basically an exercise in integration, not so often done in practice. The real key is in the second step, where the system converts input to output. It is the properties of this transfer function that matter, and the effect it produces on the Fourier coefficients. Their behavior at high frequencies decides whether jumps are formed or maintained or destroyed.

*3. Complex variables:* They are essential to transform methods, and there is a special relation between analytic functions and Laplace's equation. But I do not think residue methods or conformal mapping should dominate the course—a balance is necessary, and other topics have moved forward.

*4. Vector calculus:* A necessary tool, but not to be confused with the real content of applied mathematics.

*5. Linear algebra:* This is indispensable. That was not always recognized when the engineering curriculum was set; matrices were important but differential equations were absolutely preeminent. The balance has been adjusted by the needs of scientific computing. The continuous problem is now made discrete, the nonlinear problem is linearized at each iteration, and the dynamic problem is chopped into finite time steps. The digital computer defeated the analog computer, and even Fourier analysis is changed; signals are discrete and so are their transforms. The result is a flood of linear equations that have to be solved efficiently.

The key to a linear system is almost always a *factorization* of the underlying matrix. In elimination it is $A = LU$. The equation is broken into two triangular systems, and for large problems we try to make them sparse. Similarly, Gram–Schmidt leads to $A = QR$—with orthonormal columns in $Q$ and a triangular $R$. The principal axis theorem diagonalizes a symmetric matrix by using its eigenvectors, $A = Q\Lambda Q^T$. In the singular value decomposition $A = Q_1\Sigma Q_2^T$, the diagonal $\Sigma$ now contains the eigenvalues of $A^TA$. The polar decomposition separates stretching from rotation, and the fast Fourier transform is really a factorization of the Fourier matrix (whose entries are roots of unity $w^{jk}$) into $\log n$ factors each with $O(n)$ nonzeros.

Applied linear algebra has become too important to be picked up on the streets. It can be presented early (Strang, 1980) or later (Golub and Van Loan, 1983), but a purely abstract course (which some mathematicians have been known to teach, the author included) is generally an ineffective preparation. It is also less fun, when there are no applications.

*6. Optimization and the calculus of variations:* I am seduced by the elegance of a minimum principle. The potential energy $P = \frac{1}{2}u^TA^TCAu - u^Tf$ is minimized at equilibrium. So is the complementary energy $Q=\frac{1}{2}\sigma^T C^{-1}\sigma$, subject to the constraint $A^T\sigma=f$—and this brings in Lagrange multipliers, which turn out to equal $u$. Additional constraints are possible; they enter $P$ and $Q$ in a dual way. Nonlinear problems fit perfectly as long as they are conservative—and the subject is unified.

*7. Numerical methods:* I believe that algorithms belong with the problems they solve! Finite differences and finite elements and the fast Fourier transform and iterative methods should not wait for some uncertain future course. It is *this* course that should recognize what the computer can do, without being dominated by it. I finally tried a computing project instead of a takehome exam, and the response was tremendous—we could see what Newton's method was doing, and whether it worked. If a specific calculation is suggested but not required, then students can choose to be safe or original. In the long run, it is what *they* do that is learned and remembered.

In my own class I became convinced that the textbook was crucial. A course needs structure, and a pattern into which the applications will fit. There must be examples and exercises to reinforce the theory. The goal is to teach more of what is really done in modern applied mathematics, and perhaps my experience was typical—to become responsible for a course that needed renewal. It led to a new text (Strang, 1986) for MIT's engineering mathematics course, covering the seven areas outlined above. Special topics are there for reference—the Kalman filter, chaos and strange attractors, entropy and jump conditions at shocks, networks and graphs, and Karmarkar's new method for linear programming. No course should cover them all, but this subject is alive and growing! It is vital to teach it that way.

## REFERENCES

Golub, G, and Van Loan, C (1983). *Matrix computations*, Johns Hopkins Press, Baltimore MD.
Strang, G (1986). *Introduction to applied mathematics*, Wellesley-Cambridge Press, Box 157, Wellesley MA.
Strang, G (1980). *Linear algebra and its applications*, Academic, New York.

*Gilbert Strang*
*Massachusetts Institute of Technology*
*Cambridge MA 02139*