STRANG
FIX

An Analysis of the Finite Element Method

describes the finite element technique for approximate solution of engineering problems, and establishes its mathematical foundations

AN ANALYSIS OF THE FINITE ELEMENT METHOD

GILBERT STRANG

Massachusetts Institute of Technology

GEORGE J. FIX

University of Maryland

PRENTICE-HALL, INC.

ENGLEWOOD CLIFFS, N.J.

Library of Congress Cataloging in Publication Data

STRANG, GILBERT,
An analysis of the finite element method.

(Prentice-Hall series in automatic computation)
Bibliography: p.
1. Finite element method. I. Fix, George J.,
joint author. II. Title.
TA335.S77 515'.624 72-12642
ISBN 0-13-032946-0

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America.

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

PREFACE

The finite element method has been an astonishing success. It was created to solve the complicated equations of elasticity and structural mechanics, and for those problems it has essentially superseded the method of finite differences. Now other applications are rapidly developing. Whenever flexibility in the geometry is important—and the power of the computer is needed not only to *solve* a system of equations, but also to *formulate* and *assemble* the discrete approximation in the first place—the finite element method has something to contribute.

From a mathematical point of view, the method is an extension of the Rayleigh–Ritz–Galerkin technique. It therefore applies to a wide class of partial differential equations. The Ritz technique does not, however, operate directly with the differential equation; instead, the continuous problem is put into an equivalent variational form, and the approximate solution is assumed to be a combination $\sum q_j \varphi_j$ of given trial functions $\varphi_j(x)$. This is the method of weighted residuals, and the weights q_j are computed from the underlying variational principle. It is this discrete problem which the computer actually solves.

So far the idea is an old one. What is new is the choice of trial functions: in the finite element method they are *piecewise polynomials*. That choice is responsible for the method's success. Each function φ_j is zero over most of the domain, and enters the computation only in the neighborhood of a particular node. In that neighborhood φ_j is pieced together from polynomials of low degree, and the computations are as simple as possible. It is remarkable that simultaneously, and quite independently, piecewise polynomials have become preeminent in the mathematical theory of approximation of functions. Apparently it was the right idea at the right time.

Because the mathematical foundations are sound, it is possible to understand why the method works. This is the real reason for our book. Its purpose

is to explain the effect of each of the approximations that are essential for the finite element technique to be computationally efficient. We list here some of these approximations:

(1) interpolation of the original physical data
(2) choice of a finite number of polynomial trial functions
(3) simplification of the geometry of the domain
(4) modification of the boundary conditions
(5) numerical integration of the underlying functional in the variational principle
(6) roundoff error in the solution of the discrete system.

These questions are fundamentally mathematical, and so are the authors. Nevertheless this book is *absolutely not* intended for the exclusive use of specialists in numerical analysis. On the contrary, we hope it may help establish closer communication between the mathematical engineer and the mathematical analyst. It seems to us that the finite element method provides a special opportunity for this communication: the theory is attractive, the applications are growing, and best of all, the method is so new that the gap between theory and application ought not yet to be insurmountable.

Of course we recognize that there are obstacles which cannot be made to disappear. One of them is the language itself; we have kept the mathematical notations to a minimum, and indexed them (with definitions) at the end of the book. We also know that, even after a norm has been interpreted as a natural measure of strain energy, and a Hilbert space identified with the class of admissible functions in a physically derived variational principle, there still remains the hardest problem: to become comfortable with these ideas, and to make them one's own. This requires genuine patience and tolerance on both sides, as well as effort. Perhaps this book at least exhibits the kind of problems which a mathematician is trained to solve, and those for which he is useless.

In the last few years a great many numerical analysts have turned to finite elements, and we are very much in their debt. This is acknowledged explicitly throughout the book, and implicitly in the bibliography, even though we have by no means attempted a formal history. Here, before the book begins, we want to thank two others—engineers rather than mathematicians—for help that was the most important of all. One is Isaac Fried, whose influence led us to abandon an earlier (and completed) "Fourier Analysis of the Finite Element Method," and to study instead the real thing. The other is Bruce Irons, whose remarkable intuitions are described (and proved correct, as often as we can) in the book itself.

Chapter 1 is very much longer than the others, and was used by the first author as the text in an introductory course at M.I.T. The only homework

was to go out and program some finite elements. Where such programs are already available, students could be asked to combine computational experiments with a theoretical seminar based on the book.

Chapters 2 to 5 were also written by the first author. The last three chapters were drafted by the second author, and then revised and "homogenized" by the first. And the whole was typed by Mrs. Ingrid Naaman, who has gracefully allowed us to believe that she enjoyed it; thank you.

GILBERT STRANG

GEORGE J. FIX

Cambridge, Massachusetts

CONTENTS

1 AN INTRODUCTION TO THE THEORY

1.1. THE BASIC IDEAS

The finite element method can be described in a few words. Suppose that the problem to be solved is in variational form—it may be required to find the function u which minimizes a given expression of potential energy. This minimizing property leads to a differential equation for u (the Euler equation), but normally an exact solution is impossible and some approximation is necessary. The Rayleigh–Ritz–Galerkin idea is to choose a finite number of trial functions $\varphi_1, \ldots, \varphi_N$, and among all their linear combinations $\sum q_j \varphi_j$ to find the one which is minimizing. This is the Ritz approximation. The unknown weights q_j are determined, not by a differential equation, but by a system of N discrete algebraic equations which the computer can handle. The theoretical justification for this method is simple, and compelling: *The minimizing process automatically seeks out the combination which is closest to* u. Therefore, the goal is to choose trial functions φ_j which are convenient enough for the potential energy to be computed and minimized, and at the same time general enough to approximate closely the unknown solution u.

The real difficulty is the first one, to achieve convenience and computability. In theory there always exists a set of trial functions which is complete—their linear combinations fill the space of all possible solutions as $N \longrightarrow \infty$, and therefore the Ritz approximations converge—but to be able to compute with them is another matter. This is what finite elements have accomplished.

The underlying idea is simple. It starts by a subdivision of the structure, or the region of physical interest, into smaller pieces. These pieces must be easy for the computer to record and identify; they may be triangles or rectangles. Then within each piece the trial functions are given an extremely

simple form—normally they are polynomials, of at most the third or fifth degree. Boundary conditions are infinitely easier to impose locally, along the edge of a triangle or rectangle, than globally along a more complicated boundary. The accuracy of the approximation can be increased, if that is necessary, but not by the classical Ritz method of including more and more complex trial functions. Instead, the same polynomials are retained, and the subdivision is refined. The computer follows a nearly identical set of instructions, and just takes longer to finish. In fact, a large-scale finite element system can use the power of the computer, for the *formulation* of approximate equations as well as their solution, to a degree never before achieved in complicated physical problems.

Unhappily none of the credit for this idea goes to numerical analysts. The method was created by structural engineers, and it was not recognized at the start as an instance of the Rayleigh–Ritz principle. The subdivision into simpler pieces, and the equations of equilibrium and compatibility between the pieces, were initially constructed on the basis of physical reasoning. The later development of more accurate elements happened in a similar way; it was recognized that increasing the degree of the polynomials would greatly improve the accuracy, *but the unknowns q_j computed in the discrete approximation have always retained a physical significance.* In this respect the computer output is much easier to interpret than the weights produced by the classical method.

The whole procedure became mathematically respectable at the moment when the unknowns were identified as the coefficients in a Ritz approximation $u \approx \sum q_j \varphi_j$, and the discrete equations were seen to be exactly the conditions for minimizing the potential energy. Surely Argyris in Germany and England, and Martin and Clough in America, were among those responsible; we dare not guess who was first. The effect was instantly to provide a sound theoretical basis for the method. As the techniques of constructing more refined elements have matured, the underlying theory has also begun to take shape.

The fundamental problem is to discover how closely piecewise polynomials can approximate an unknown solution u. In other words, we must determine how well finite elements—which were developed on the basis of computational simplicity—satisfy the second requirement of good trial functions, to be effective in approximation. Intuitively, any reasonable function u can be approached to arbitrary accuracy by piecewise linear functions. The mathematical task is to estimate the error as closely as possible and to determine how rapidly the error decreases as the number of pieces (or the degree of the polynomial within each piece) is increased. Of course, the finite element method can proceed without the support of precise mathematical theorems; it got on pretty well for more than 10 years. But we believe it will be useful, especially in the future development of the method, to understand and consolidate what has already been done.

We have attempted a fairly complete analysis of *linear problems* and the *displacement method.* A comparable theory for fully nonlinear equations does not yet exist, although it would certainly be possible to treat semilinear equations—in which the difficulties are confined to lower-order terms. We make a few preliminary comments on nonlinear equations, but this remains an outstanding problem for the future. In our choice of the displacement method over the alternative variational formulations described in Chapter 2, we have opted to side with the majority. This is the most commonly used version of the finite element method. Of course, the approximation theory would be the same for all formulations, and the duality which is so rampant throughout the whole subject makes the conversion between displacement methods and force methods nearly automatic.

Our goal in this chapter is to illustrate the basic steps in the finite element method:

1. The variational formulation of the problem.
2. The construction of piecewise polynomial trial functions.
3. The computation of the stiffness matrix and solution of the discrete system.
4. The estimation of accuracy in the final Ritz approximation.

We take the opportunity, when stating the problem variationally, to insert some of the key mathematical ideas needed for a precise theory—the Hilbert spaces $\mathcal{H}^s$ and their norms, the estimates for the solution in terms of the data, and the energy inner product which is naturally associated with the specific problem. With these tools, the convergence of finite elements can be proved even for a very complicated geometry. In fact, the simplicity of variational arguments permits an analysis which already goes beyond what has been achieved for finite differences.

1.2. A TWO-POINT BOUNDARY-VALUE PROBLEM

Our plan is to introduce the finite element method, and the mathematics which lies behind it, in terms of a specific and familiar example. It makes sense to choose a one-dimensional problem, in order that the construction of elements shall be simple and natural, and also in order that the mathematical manipulations shall be straightforward—requiring integration by parts rather than some general Green's formula. Therefore, our choice falls on the equation

$$-\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + q(x)u = f(x). \tag{1}$$

With suitable boundary conditions at the endpoints $x = 0$ and $x = \pi$, this is a classical Sturm–Liouville problem. It represents a number of different

physical processes—the distribution of temperature along a rod, for example, or the displacement of a rotating string. Mathematically, the first point to emphasize is that the equation and boundary conditions arise from a *steady-state problem*, and not one which unfolds in time from initial conditions of displacement and velocity. It would correspond in more space dimensions to an *elliptic boundary-value problem*, governed for example by Laplace's equation.

In order to illustrate the treatment of different types of boundary conditions, especially in the variational statement of the problem, we fix the left-hand end of the string and let the other be free. Thus at the end $x = 0$ there is an *essential* (or *kinematic*, or *restrained*, or *geometric*) boundary condition, in other words, one of *Dirichlet* type:

$$u(0) = 0.$$

At the right-hand end $x = \pi$, the string is not constrained, and it assumes a *natural* (or *dynamic*, or *stress*) boundary condition, in other words, one of *Neumann* type:

$$u'(\pi) = 0.$$

We propose to consider this model problem from four different points of view, in the following order:

1. Pure mathematics.
2. Applied mathematics.
3. Numerical approximation by finite differences.
4. Numerical approximation by finite elements.

It is essential to recognize the common features of these four approaches to the same problem; the tools which are useful to the pure mathematician in proving the existence and uniqueness of the solution, and to the applied mathematician in understanding its qualitative behavior, ought to be applied also to the study of the numerical algorithms.

We begin with the pure mathematician, who combines the differential equation and boundary conditions into a single equation,

$$Lu = f.$$

L is a linear operator, acting on a certain class of functions—those which in some sense satisfy the boundary conditions and can be differentiated twice. Mathematically, the fundamental question is precisely this: *to match such a space of functions u with a class of inhomogeneous terms f*, in such a way that *to each f there corresponds one and only one solution u*. Once this correspondence between f and u has been established, the problem $Lu = f$ is in an abstract

sense "solved." Of course there is still a little way to go in actually discovering which solution u corresponds to a given f. That step is the real subject of this book. But we believe it is worthwhile, and not just useless fussiness, to try first to get these function spaces right. In fact, it is of special importance for the variational principles and their approximation to know exactly which space of functions is admissible. (The references to "spaces" of functions carry the implication that if u_1 and u_2 are admissible, then so is $c_1u_1 + c_2u_2$; this superposition is a natural property in linear problems.)

We want to consider one specific choice, the one which is perhaps most important to the theory, for the space of inhomogeneous data: Those f are admitted which have *finite energy*. This means that

$$\int_0^\pi (f(x))^2\,dx < \infty. \tag{2}$$

Any piecewise smooth function f is thereby included, but the Dirac δ-function is not; we shall return later to this case of a "point load." The space of functions satisfying (2) is often denoted by L_2; we prefer the notation $\mathcal{H}^0$, indicating by the superscript how many derivatives of f are required to have finite energy (in this case it is only f itself).

For the simplest Sturm–Liouville equation $-u'' = f$, it is not hard to guess the corresponding space of solutions. This solution space is denoted by $\mathcal{H}_B^2$—the subscript B refers to the boundary conditions $u(0) = u'(\pi) = 0$, and the superscript 2 requires that the second derivative of u has finite energy.† The role of the pure mathematician is then to show, under the assumptions $p(x) \geq p_{\min} > 0$ and $q(x) \geq 0$, that $\mathcal{H}_B^2$ is still the solution space for the more general equation $-(pu')' + qu = f$. In fact, his final theorem can be stated in the following way:

The operator L is a one-to-one transformation from $\mathcal{H}_B^2$ onto $\mathcal{H}^0$, so that for each f in $\mathcal{H}^0$ the differential equation (1) *has a unique solution u in $\mathcal{H}_B^2$. Furthermore, the solution depends continuously on the data: If f is small, then so is u.*

The last sentence requires further explanation; we need *norms* in which to measure the size of f and u. The two norms will be different, since the data space and solution space are different. Fortunately, there is a natural choice for the norms in terms of the energy, or rather its square root:

$$\|f\|_0 = \left[\int (f(x))^2\,dx\right]^{1/2},$$

$$\|u\|_2 = \left[\int \left((u''(x))^2 + (u'(x))^2 + (u(x))^2\right) dx\right]^{1/2}.$$

†These spaces are defined again in the index of notations at the end of the book.

With these definitions, the continuous dependence of the solution on the data can be expressed in a quantitative form: There exists a constant C such that

$$\|u\|_2 \leq C\|f\|_0. \tag{3}$$

The uniqueness of the solution follows immediately from this estimate: If $f = 0$, then necessarily $u = 0$. In fact, it is such estimates which lie at the very center of the modern theory of partial differential equations. A general technique for proving (3), which applies also to boundary-value problems in several space dimensions, has been created only in the last generation. In this book, we shall accept such estimates as proved: For elliptic equations of order $2m$, this means that

$$\|u\|_{2m} \leq C\|f\|_0. \tag{4}$$

We move now to a more applied question, the actual construction of the solution. If the coefficients p and q are constant, then this can be carried out in terms of an infinite series. The key lies in knowing the eigenvalues and eigenfunctions of L:

$$u_n(x) = \sqrt{\frac{\pi}{2}}\sin(n - \tfrac{1}{2})x, \qquad \lambda_n = p(n - \tfrac{1}{2})^2 + q. \tag{5}$$

It is immediate to check that $Lu_n = -pu_n'' + qu_n = \lambda_n u_n$, that the functions u_n satisfy the boundary conditions and therefore lie in $\mathfrak{C}_B^2$, and that they are orthonormal:

$$\int_0^\pi u_n(x)u_m(x)\,dx = \delta_{nm}.$$

Suppose the inhomogeneous term is expanded in a series of eigenfunctions:

$$f(x) = \sum_{n=1}^{\infty} a_n\sqrt{\frac{\pi}{2}}\sin(n - \tfrac{1}{2})x. \tag{6}$$

Then integrating formally, the orthogonality of the u_n gives

$$\|f\|_0^2 = \int_0^\pi f^2\,dx = \sum_{n=1}^{\infty} a_n^2.$$

The functions f in $\mathfrak{C}^0$ are exactly those which admit a harmonic expansion of the form (6), with coefficients satisfying $\sum a_n^2 < \infty$. Actually, this ought to seem a little paradoxical, since apparently every f of the form (6) will satisfy $f(0) = 0$, $f'(\pi) = 0$, whereas no boundary conditions were meant to be imposed on f: The elements of $\mathfrak{C}^0$ are required only to have finite

energy, $\int f^2 < \infty$. The paradox is resolved by the completeness of the eigenfunctions u_n in $\mathcal{C}^0$. Whether f satisfies these spurious boundary conditions or not, its expansion is valid in the mean-square sense,

$$\int_0^\pi \left(f(x) - \sum_1^N a_n \sqrt{\frac{\pi}{2}} \sin(n - \tfrac{1}{2})x \right)^2 dx \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty.$$

The boundary conditions on f are thus unstable and disappear in the limit as $N \longrightarrow \infty$. Figure 1.1 shows how a sequence of functions f_N, all lying in $\mathcal{C}_B^2$, could still converge to a function f outside that space.

The Sturm-Liouville differential equation $Lu = f$ is now ready to be solved: if $f = \sum a_n u_n$, then u has the expansion

$$u = \sum \frac{a_n}{\lambda_n} u_n = \sqrt{\frac{\pi}{2}} \sum_1^\infty \frac{a_n \sin(n - \frac{1}{2})x}{p(n - \frac{1}{2})^2 + q}. \tag{7}$$

With this explicit construction, the estimate $\| u \|_2 \leq C \| f \|_0$ and the matching of data space $\mathcal{C}^0$ with solution space $\mathcal{C}_B^2$ can be verified directly.

The question of the boundary conditions is more subtle and deserves further comment. We have seen already that even though f can be expanded in terms of the u_n, which do satisfy the boundary conditions, still f has absolutely nothing to do with these conditions. Therefore the question is: What is different about u? Why does u satisfy the boundary conditions? The answer is that the series expansion for u converges in a much stronger sense than the expansion for f: not only does $\sum a_n u_n / \lambda_n$ converge to u in the mean-square sense, but so do its first and second derivatives. More precisely,

$$\| u - \sum_1^N \frac{a_n}{\lambda_n} u_n \|_2 \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty.$$

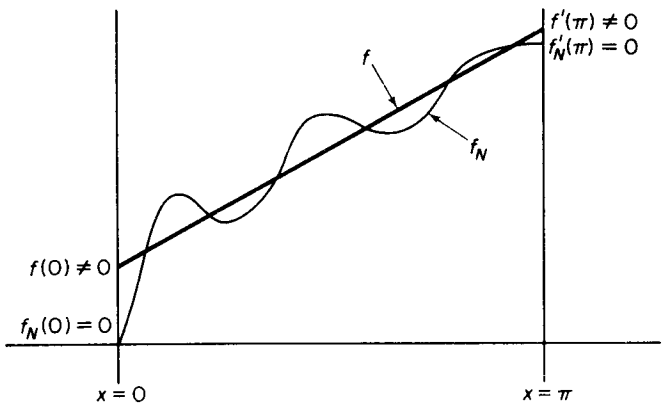

Fig. 1.1 f_N in $\mathcal{C}_B^2$ approximating a general function f.

The point is that when even the second derivatives converge, the boundary conditions are *stable*; the limit function u is compelled to satisfy the boundary conditions. (Note that in Figure 1.1, the second derivatives of f_N did not converge to those of f; therefore the limit f did not have to satisfy the boundary conditions, and was outside the space $\mathcal{H}_B^2$. This is what will not happen for u.)

The general rule is this: *boundary conditions which involve only derivatives below order s will make sense in the* $\mathcal{H}^s$ *norm*; those involving derivatives of order s or higher will be unstable and will not apply to functions in the space $\mathcal{H}^s$. We shall see that this is the rule which distinguishes between essential boundary conditions, which stay, and natural boundary conditions, which go. The distinction becomes crucial in the variational problem, whose statement is in terms of *first derivatives*, that is, the $\mathcal{H}^1$ norm. The finite element approximations will be required to satisfy all boundary conditions below order 1—that means the condition $u(0) = 0$—but they will not be required to satisfy the condition on the first derivative. This leniency at $x = \pi$ will not prevent the finite element approximations from converging in the $\mathcal{H}^1$ norm to a solution u which does satisfy $u'(\pi) = 0$. This is the key to the following section, which extends the "pure mathematics" standpoint to the equivalent variational problem.

1.3. THE VARIATIONAL FORM OF THE PROBLEM

The linear equation $Lu = f$ is related to the quadratic functional

$$I(v) = (Lv, v) - 2(f, v)$$

in the following way: $I(v)$ is minimized at $v = u$ only if its derivative (or first variation) vanishes there, and the condition for the vanishing of this derivative is exactly the *Euler equation* $Lu = f$. The problems of inverting L and minimizing I are equivalent; they produce the same solution u. Therefore, such problems can be investigated either in an *operational form*, in terms of the linear operator L, or in *variational form*, in terms of the quadratic I. The goal in this section is to find the exact variational equivalent of our two-point boundary-value problem.

This equivalence of differential equations with variational problems is basic also to the choice of a computational scheme. The differential equation may be approximated by a discrete system, using finite differences, or the variational integral can be minimized over a discrete class of functions, as in the finite element method. In many applications—particularly in steady-state rather than transient problems—the variational statement is the primary physical principle, and the differential equation only a secondary consequence. Therefore, it is not surprising to find in such applications a