1

Overview

TABLE OF CONTENTS

This book is an introduction to the analysis of linear elastic structures by the Finite Element Method (FEM). This Chapter presents an overview of where the book fits, and what finite elements are.

§1.1. Where this Material Fits

The field of Mechanics can be subdivided into three major areas:

$$\text{Mechanics} \begin{cases} \textit{Theoretical} \\ \textit{Applied} \\ \textit{Computational} \end{cases} \tag{1.1}$$

Theoretical mechanics deals with fundamental laws and principles of mechanics studied for their intrinsic scientific value. *Applied mechanics* transfers this theoretical knowledge to scientific and engineering applications, especially as regards the construction of mathematical models of physical phenomena. *Computational mechanics* solves specific problems by simulation through numerical methods implemented on digital computers.

Remark 1.1. Paraphrasing an old joke about mathematicians, one may define a computational mechanician as a person who searches for solutions to given problems, an applied mechanician as a person who searches for problems that fit given solutions, and a theoretical mechanician as a person who can prove the existence of problems and solutions.

§1.1.1. Computational Mechanics

Several branches of computational mechanics can be distinguished according to the *physical scale* of the focus of attention:

$$\text{Computational Mechanics} \begin{cases} \textit{Nanomechanics and micromechanics} \\ \textit{Continuum mechanics} \begin{cases} \text{Solids and Structures} \\ \text{Fluids} \\ \text{Multiphysics} \end{cases} \\ \textit{Systems} \end{cases} \tag{1.2}$$

Nanomechanics deals with phenomena at the molecular and atomic levels of matter. As such it is closely linked to particle physics and chemistry. Micromechanics looks primarily at the crystallographic and granular levels of matter. Its main technological application is the design and fabrication of materials and microdevices.

Continuum mechanics studies bodies at the macroscopic level, using continuum models in which the microstructure is homogenized by phenomenological averages. The two traditional areas of application are solid and fluid mechanics. The former includes *structures* which, for obvious reasons, are fabricated with solids. Computational solid mechanics takes an applied sciences approach, whereas computational structural mechanics emphasizes technological applications to the analysis and design of structures.

Computational fluid mechanics deals with problems that involve the equilibrium and motion of liquid and gases. Well developed subsidiaries are hydrodynamics, aerodynamics, acoustics, atmospheric physics, shock, combustion and propulsion.

Multiphysics is a more recent newcomer. This area is meant to include mechanical systems that transcend the classical boundaries of solid and fluid mechanics, as in interacting fluids and structures. Phase change problems such as ice melting and metal solidification fit into this category, as do the interaction of control, mechanical and electromagnetic systems.

Finally, *system* identifies mechanical objects, whether natural or artificial, that perform a distinguishable function. Examples of man-made systems are airplanes, buildings, bridges, engines, cars, microchips, radio telescopes, robots, roller skates and garden sprinklers. Biological systems, such as a whale, amoeba, inner ear, or pine tree are included if studied from the viewpoint of biomechanics. Ecological, astronomical and cosmological entities also form systems.[1]

In the progression of (1.2) *system* is the most general concept. A system is studied by *decomposition*: its behavior is that of its components plus the interaction between components. Components are broken down into subcomponents and so on. As this hierarchical breakdown process continues, individual components become simple enough to be treated by individual disciplines, but component interactions get more complex. Consequently there is a tradeoff art in deciding where to stop.[2]

§1.1.2. Statics vs. Dynamics

Continuum mechanics problems may be subdivided according to whether inertial effects are taken into account or not:

$$\text{Continuum mechanics}\ \begin{cases} \textit{Statics} \\ \textit{Dynamics} \end{cases} \tag{1.3}$$

In dynamics actual time dependence must be explicitly considered, because the calculation of inertial (and/or damping) forces requires derivatives respect to actual time to be taken.

Problems in statics may also be time dependent but with inertial forces ignored or neglected. Accordingly static problems may be classed into strictly static and quasi-static. For the former time need not be considered explicitly; any historical time-like response-ordering parameter, if one is needed, will do. In quasi-static problems such as foundation settlement, metal creep, rate-dependent plasticity or fatigue cycling, a realistic measure of time is required but inertial forces are still neglected.

§1.1.3. Linear vs. Nonlinear

A classification of static problems that is particularly relevant to this book is

$$\text{Statics}\ \begin{cases} \textit{Linear} \\ \textit{Nonlinear} \end{cases} \tag{1.4}$$

Linear static analysis deals with static problems in which the *response* is linear in the cause-and-effect sense. For example: if the applied forces are doubled, the displacements and internal stresses also double. Problems outside this domain are classified as nonlinear.

---

[1] Except that their function may not be clear to us. "The usual approach of science of constructing a mathematical model cannot answer the questions of why there should be a universe for the model to describe. Why does the universe go to all the bother of existing? Is the unified theory so compelling that it brings about its own existence? Or does it need a creator, and, if so, does he have any other effect on the universe? And who created him?" (Stephen Hawking).

[2] Thus in breaking down a car engine for engineering analysis, say, the decomposition does not usually proceed beyond the components you can buy at a parts shop.

§1.1.4. Discretization methods

A final classification of CSM static analysis is based on the discretization method by which the continuum mathematical model is *discretized* in space, *i.e.*, converted to a discrete model with a finite number of degrees of freedom:

$$\text{Spatial discretization method} \begin{cases} \textit{Finite Element (FEM)} \\ \textit{Boundary Element (BEM)} \\ \textit{Finite Difference (FDM)} \\ \textit{Finite Volume (FVM)} \\ \textit{Spectral} \\ \textit{Meshfree} \end{cases} \tag{1.5}$$

In CSM *linear* problems finite element methods currently dominate the scene as regards space discretization.[3] Boundary element methods post a strong second choice in specific application areas. For *nonlinear* problems the dominance of finite element methods is overwhelming.

Space finite difference methods in solid and structural mechanics have virtually disappeared from practical use. This statement is not true, however, for fluid mechanics, where finite difference discretization methods are still important. Finite-volume methods, which directly address the discretization of conservation laws, are important in difficult problems of fluid mechanics, for example high-Re gas dynamics. Spectral methods are based on transforms that map space and/or time dimensions to spaces (for example, the frequency domain) where the problem is easier to solve.

A recent newcomer to the scene are the meshfree methods. These combine techniques and tools of finite element methods such as variational formulation and interpolation, with finite difference features such as non-local support.

§1.1.5. FEM Variants

The term *Finite Element Method* actually identifies a broad spectrum of techniques that share common features outlined in §1.3 and §1.4. Two subclassifications that fit well applications to structural mechanics are[4]

$$\text{FEM Formulation} \begin{cases} \textit{Displacement} \\ \textit{Equilibrium} \\ \textit{Mixed} \\ \textit{Hybrid} \end{cases} \qquad \text{FEM Solution} \begin{cases} \textit{Stiffness} \\ \textit{Flexibility} \\ \textit{Mixed (a.k.a. Combined)} \end{cases} \tag{1.6}$$

Using the foregoing classification, we can state the topic of this book more precisely: the *computational analysis of linear static structural problems* by the Finite Element Method. Of the variants listed in (1.6), emphasis is placed on the *displacement* formulation and *stiffness* solution. This combination is called the *Direct Stiffness Method* or DSM.

---

[3] There are finite element discretizations in time, but they are not so widely used as finite differences.

[4] The distinction between these subclasses require advanced technical concepts, which cannot be covered in an introductory treatment such as this book.

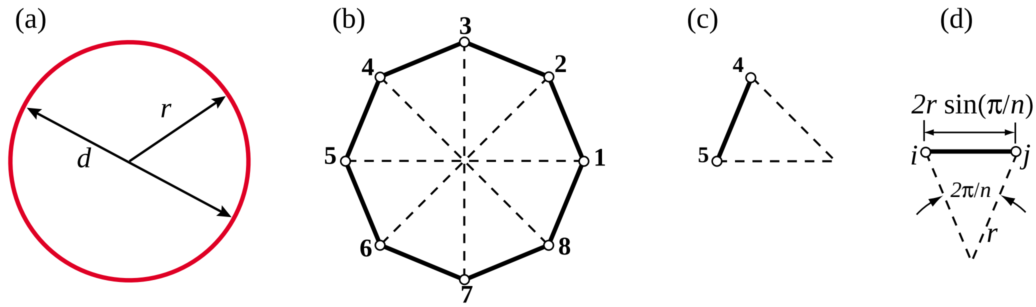

FIGURE 1.1. The “find π” problem treated with FEM concepts: (a) continuum object, (b) a discrete approximation by inscribed regular polygons, (c) disconnected element, (d) generic element.

§1.2. What Does a Finite Element Look Like?

The subject of this book is FEM. But what *is* a finite element? The term admits of two interpretations, as discussed later. For now the underlying concept will be partly illustrated through a truly ancient problem: find the perimeter L of a circle of diameter d. Since $L = \pi\, d$, this is equivalent to obtaining a numerical value for π.

Draw a circle of radius r and diameter $d = 2r$ as in Figure 1.1(a). Inscribe a regular polygon of n sides, where $n = 8$ in Figure 1.1(b). Rename polygon sides as *elements* and vertices as *nodes*. Label nodes with integers $1, \ldots 8$. Extract a typical element, say that joining nodes 4–5, as shown in Figure 1.1(c). This is an instance of the *generic element* i–j pictured in Figure 1.1(d). The element length is $L_{ij} = 2r \sin(\pi/n)$. Since all elements have the same length, the polygon perimeter is $L_n = nL_{ij}$, whence the approximation to π is $\pi_n = L_n/d = n \sin(\pi/n)$.

Table 1.1. Rectification of Circle by Inscribed Polygons (“Archimedes FEM”)

n	$\pi_n = n \sin(\pi/n)$	Extrapolated by Wynn-ϵ	Exact π to 16 places
1	0.000000000000000		
2	2.000000000000000		
4	2.828427124746190	3.414213562373096	
8	3.061467458920718		
16	3.121445152258052	3.141418327933211	
32	3.136548490545939		
64	3.140331156954753	3.141592658918053	
128	3.141277250932773		
256	3.141513801144301	3.141592653589786	3.141592653589793

Values of π_n obtained for $n = 1, 2, 4, \ldots 256$ are listed in the second column of Table 1.1. As can be seen the convergence to π is fairly slow. However, the sequence can be transformed by Wynn's ϵ algorithm[5] into that shown in the third column. The last value displays 15-place accuracy.

Some key ideas behind the FEM can be identified in this example. The circle, viewed as a *source mathematical object*, is replaced by polygons. These are *discrete approximations* to the circle. The sides, renamed as *elements*, are specified by their end *nodes*. Elements can be separated by

[5] A widely used lozenge extrapolation algorithm that speeds up the convergence of many sequences. See, e.g, [273].

disconnecting nodes, a process called *disassembly* in the FEM. Upon disassembly a *generic element* can be defined, *independently of the original circle*, by the segment that connects two nodes i and j. The relevant element property: side length L_{ij}, can be computed in the generic element independently of the others, a property called *local support* in the FEM. The target property: the polygon perimeter, is obtained by reconnecting n elements and adding up their length; the corresponding steps in the FEM being *assembly* and *solution*, respectively. There is of course nothing magic about the circle; the same technique can be be used to rectify any smooth plane curve.[6]

This example has been offered in the FEM literature, e.g. in [172], to aduce that finite element ideas can be traced to Egyptian mathematicians from *circa* 1800 B.C., as well as Archimedes' famous studies on circle rectification by 250 B.C. But comparison with the modern FEM, as covered in following Chapters, shows this to be a stretch. The example does not illustrate the concept of degrees of freedom, conjugate quantities and local-global coordinates. It is guilty of circular reasoning: the compact formula $\pi = \lim_{n\to\infty} n \sin(\pi/n)$ uses the unknown π in the right hand side.[7] Reasonable people would argue that a circle is a simpler object than, say, a 128-sided polygon. Despite these flaws the example is useful in one respect: showing a fielder's choice in the replacement of one mathematical object by another. This is at the root of the simulation process described below.

§1.3. The FEM Analysis Process

Processes using FEM involve carrying out a sequence of steps in some way. Those sequences take two canonical configurations, depending on (i) the environment in which FEM is used and (ii) the main objective: model-based simulation of physical systems, or numerical approximation to mathematical problems. Both are reviewed below to introduce terminology used in the sequel.

§1.3.1. The Physical FEM

A canonical use of FEM is simulation of physical systems. This requires models. Consequenty the methodology is often called *model-based simulation*.

The process is illustrated in Figure 1.2. The centerpiece is the *physical system* to be modeled. Accordingly, this configuration is called the *Physical FEM*. The processes of idealization and discretization are carried out *concurrently* to produce the discrete model. The solution step is handled by an equation solver often customized to FEM, which delivers a discrete solution (or solutions).

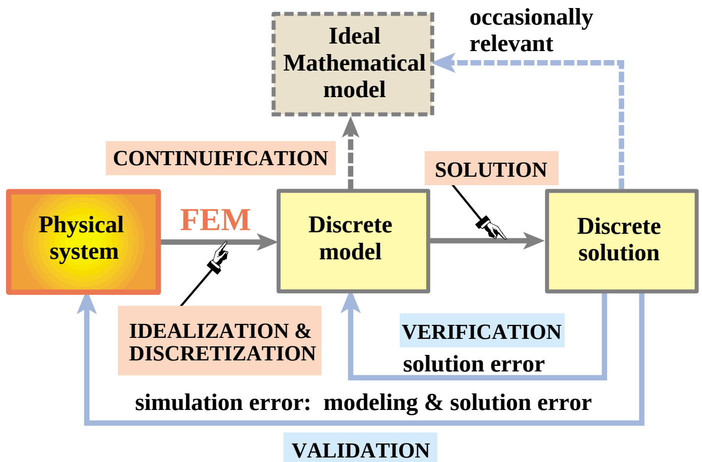

FIGURE 1.2. The Physical FEM. The physical system (left) is the source of the simulation process. The ideal mathematical model (should one go to the trouble of constructing it) is inessential.

[6] A similar limit process, however, may fail in three or more dimensions.

[7] This objection is bypassed if n is advanced as a power of two, as in Table 1.1, by using the half-angle recursion $\sqrt{2}\sin\alpha = \sqrt{1-\sqrt{1-\sin^2 2\alpha}}$, started from $2\alpha = \pi$ for which $\sin\pi = -1$.

Figure 1.2 also shows an *ideal mathematical model*. This may be presented as a *continuum limit* or "continuification" of the discrete model. For some physical systems, notably those well modeled by continuum fields, this step is useful. For others, such as complex engineering systems (say, a flying aircraft) it makes no sense. Indeed Physical FEM discretizations may be constructed and adjusted *without reference to mathematical models*, simply from experimental measurements.

The concept of *error* arises in the Physical FEM in two ways. These are known as *verification* and *validation*, respectively. Verification is done by replacing the discrete solution into the discrete model to get the solution error. This error is not generally important. Substitution in the ideal mathematical model in principle provides the discretization error. This step is rarely useful in complex engineering systems, however, because there is no reason to expect that the continuum model exists, and even if it does, that it is more physically relevant than the discrete model.

Validation tries to compare the discrete solution against observation by computing the *simulation error*, which combines modeling and solution errors. As the latter is typically unimportant, the simulation error in practice can be identified with the modeling error. In real-life applications this error overwhelms the other two.[8]

One way to adjust the discrete model so that it represents the physics better is called *model updating*. The discrete model is given free parameters. These are determined by comparing the discrete solution against experiments, as illustrated in Figure 1.3. Inasmuch as the minimization conditions are generally nonlinear (even if the model is linear) the updating process is inherently iterative.

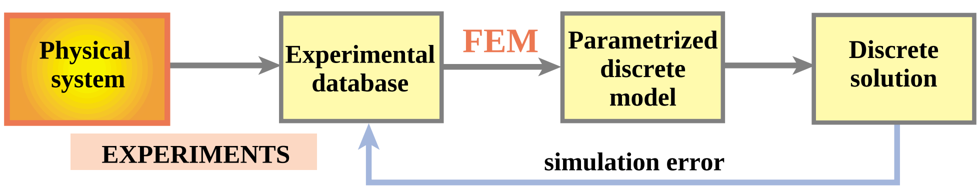

FIGURE 1.3. Model updating process in the Physical FEM.

§1.3.2. The Mathematical FEM

The other canonical way of using FEM focuses on the mathematics. The process steps are illustrated in Figure 1.4. The spotlight now falls on the *mathematical model*. This is often an ordinary or partial differential equation in space and time. A discrete finite element model is generated from a variational or weak form of the mathematical model.[9] This is the *discretization* step. The FEM equations are solved as described for the Physical FEM.

On the left Figure 1.4 shows an *ideal physical system*. This may be presented as a *realization* of the mathematical model. Conversely, the mathematical model is said to be an *idealization* of this system. E.g., if the mathematical model is the Poisson's PDE, realizations may be heat conduction or an electrostatic charge-distribution problem. This step is inessential and may be left out. Indeed Mathematical FEM discretizations *may be constructed without any reference to physics*.

---

[8] "All models are wrong; some are useful" (George Box)

[9] The distinction between strong, weak and variational forms is discussed in advanced FEM courses. In the present book such forms will be largely stated (and used) as recipes.

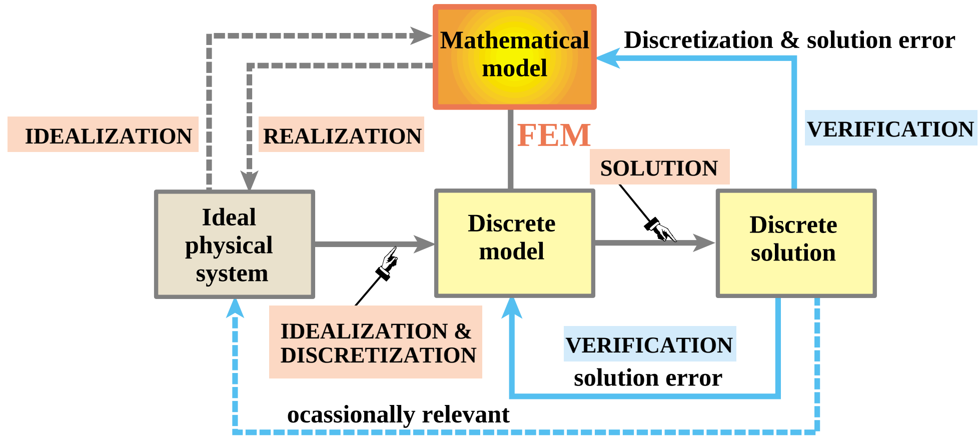

FIGURE 1.4. The Mathematical FEM. The mathematical model (top) is the source of the simulation process. Discrete model and solution follow from it. The ideal physical system (should one go to the trouble of exhibiting it) is inessential.

The concept of *error* arises when the discrete solution is substituted in the "model" boxes. This replacement is generically called *verification*. As in the Physical FEM, the *solution error* is the amount by which the discrete solution fails to satisfy the discrete equations. This error is relatively unimportant when using computers, and in particular direct linear equation solvers, for the solution step. More relevant is the *discretization error*, which is the amount by which the discrete solution fails to satisfy the mathematical model.[10] Replacing into the ideal physical system would in principle quantify modeling errors. In the Mathematical FEM this is largely irrelevant, however, because the ideal physical system is merely that: a figment of the imagination.

§1.3.3. Synergy of Physical and Mathematical FEM

The foregoing canonical sequences are not exclusive but complementary. This synergy[11] is one of the reasons behind the power and acceptance of the method. Historically the Physical FEM was the first one to be developed to model complex physical systems such as aircraft, as narrated in §1.7. The Mathematical FEM came later and, among other things, provided the necessary theoretical underpinnings to extend FEM beyond structural analysis.

A glance at the schematics of a commercial jet aircraft makes obvious the reasons behind the Physical FEM. There is no simple differential equation that captures, at a continuum mechanics level,[12] the structure, avionics, fuel, propulsion, cargo, and passengers eating dinner. There is no reason for despair, however. The time honored *divide and conquer* strategy, coupled with *abstraction*, comes to the rescue.

---

[10] This error can be computed in several ways, the details of which are of no importance here.

[11] Such interplay is not exactly a new idea: "The men of experiment are like the ant, they only collect and use; the reasoners resemble spiders, who make cobwebs out of their own substance. But the bee takes the middle course: it gathers its material from the flowers of the garden and field, but transforms and digests it by a power of its own." (Francis Bacon).

[12] Of course at the (sub)atomic level quantum mechanics works for everything, from landing gears to passengers. But it would be slightly impractical to represent the aircraft by, say, 10^{36} interacting particles modeled by the Schrödinger equations. More seriously, Truesdell and Toupin correctly note that "*Newtonian mechanics, while not appropriate to the corpuscles making up a body, agrees with experience when applied to the body as a whole*, except for certain phenomena of astronomical scale" [253, p. 228].

First, separate the structure out and view the rest as masses and forces. Second, consider the aircraft structure as built up of *substructures* (a part of a structure devoted to a specific function): wings, fuselage, stabilizers, engines, landing gears, and so on.

Take each substructure, and continue to break it down into *components*: rings, ribs, spars, cover plates, actuators, etc. Continue through as many levels as necessary. Eventually those components become sufficiently simple in geometry and connectivity that they can be reasonably well described by the mathematical models provided, for instance, by Mechanics of Materials or the Theory of Elasticity. At that point, *stop*.

The component level discrete equations are obtained from a FEM library based on the mathematical model.

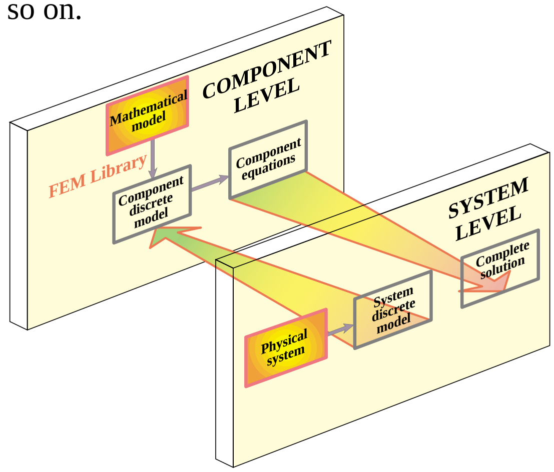

FIGURE 1.5. Combining physical and mathematical modeling through multilevel FEM. Only two levels (system and component) are shown for simplicity.

The system model is obtained by going through the reverse process: from component equations to substructure equations, and from those to the equations of the complete aircraft. This *system assembly* process is governed by the classical principles of Newtonian mechanics, which provide the necessary inter-component "glue." The multilevel decomposition process is diagramed in Figure 1.5, in which intermediate levels are omitted for simplicity

Remark 1.2. More intermediate decomposition levels are used in systems such as offshore and ship structures, which are characterized by a modular fabrication process. In that case multilevel decomposition mimics the way the system is actually fabricated. The general technique, called *superelements*, is discussed in Chapter 11.

Remark 1.3. There is no point in practice in going beyond a certain component level while considering the complete system. The reason is that the level of detail can become overwhelming without adding relevant information. Usually that point is reached when uncertainty impedes further progress. Further refinement of specific components is done by the so-called global-local analysis technique outlined in Chapter 10. This technique is an instance of *multiscale analysis*.

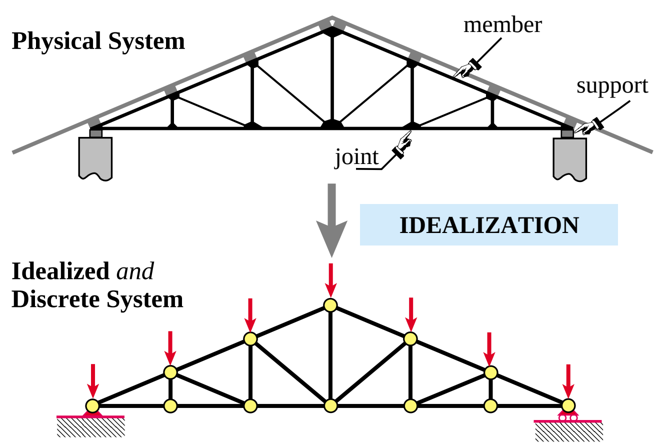

FIGURE 1.6. The idealization process for a simple structure. The physical system — here a roof truss — is directly idealized by the mathematical model: a pin-jointed bar assembly. For this particular structure idealized and discrete models coalesce.

For sufficiently simple structures, passing to a discrete model is carried out in a single *idealization and discretization* step, as illustrated for the truss roof structure shown in Figure 1.6. Other levels are unnecessary in such cases. Of course the truss may be viewed as a substructure of the roof, and the roof as a a substructure of a building.

§1.4. Interpretations of the Finite Element Method

Just like there are two complementary ways of using the FEM, there are two complementary interpretations for teaching it. One stresses the physical significance and is aligned with the Physical FEM. The other focuses on the mathematical context, and is aligned with the Mathematical FEM.

§1.4.1. Physical Interpretation

The physical interpretation focuses on the flowchart of Figure 1.2. This interpretation has been shaped by the discovery and extensive use of the method in the field of structural mechanics. The historical connection is reflected in the use of structural terms such as "stiffness matrix", "force vector" and "degrees of freedom," a terminology that carries over to non-structural applications.

The basic concept in the physical interpretation is the *breakdown* ($\equiv$ disassembly, tearing, partition, separation, decomposition) of a complex mechanical system into simpler, disjoint components called finite elements, or simply *elements*. The mechanical response of an element is characterized in terms of a finite number of degrees of freedom. These degrees of freedoms are represented as the values of the unknown functions as a set of node points. The element response is defined by algebraic equations constructed from mathematical or experimental arguments. The response of the original system is considered to be approximated by that of the *discrete model* constructed by *connecting* or *assembling* the collection of all elements.

The breakdown-assembly concept occurs naturally when an engineer considers many artificial and natural systems. For example, it is easy and natural to visualize an engine, bridge, aircraft or skeleton as being fabricated from simpler parts.

As discussed in §1.3, the underlying theme is *divide and conquer*. If the behavior of a system is too complex, the recipe is to divide it into more manageable subsystems. If these subsystems are still too complex the subdivision process is continued until the behavior of each subsystem is simple enough to fit a mathematical model that represents well the knowledge level the analyst is interested in. In the finite element method such "primitive pieces" are called *elements*. The behavior of the total system is that of the individual elements plus their interaction. A key factor in the initial acceptance of the FEM was that the element interaction can be physically interpreted and understood in terms that were eminently familiar to structural engineers.

§1.4.2. Mathematical Interpretation

This interpretation is closely aligned with the flowchart of Figure 1.4. The FEM is viewed as a procedure for obtaining numerical approximations to the solution of boundary value problems (BVPs) posed over a domain Ω. This domain is replaced by the union $\cup$ of disjoint subdomains $\Omega^{(e)}$ called finite elements. In general the geometry of Ω is only approximated by that of $\cup\Omega^{(e)}$.

The unknown function (or functions) is locally approximated over each element by an interpolation formula expressed in terms of values taken by the function(s), and possibly their derivatives, at a set of *node points* generally located on the element boundaries. The states of the assumed unknown function(s) determined by unit node values are called *shape functions*. The union of shape functions "patched" over adjacent elements form a *trial function basis* for which the node values represent the generalized coordinates. The trial function space may be inserted into the governing equations and the unknown node values determined by the Ritz method (if the solution extremizes a variational principle) or by the Galerkin, least-squares or other weighted-residual minimization methods if the problem cannot be expressed in a standard variational form.

Remark 1.4. In the mathematical interpretation the emphasis is on the concept of *local (piecewise) approximation*. The concept of element-by-element breakdown and assembly, while convenient in the computer implementation, is not theoretically necessary. The mathematical interpretation permits a general approach to the questions of convergence, error bounds, trial and shape function requirements, etc., which the physical approach leaves unanswered. It also facilitates the application of FEM to classes of problems that are not so readily amenable to physical visualization as structures; for example electromagnetics and heat conduction.

Remark 1.5. It is interesting to note some similarities in the development of Heaviside's operational methods, Dirac's delta-function calculus, and the FEM. These three methods appeared as ad-hoc computational devices created by engineers and physicists to deal with problems posed by new science and technology (electricity, quantum mechanics, and delta-wing aircraft, respectively) with little help from the mathematical establishment.[13] Only some time after the success of the new techniques became apparent were new branches of mathematics (operational calculus, distribution theory and piecewise-approximation theory, respectively) constructed to justify that success. In the case of the finite element method, the development of a formal mathematical theory started in the late 1960s, and much of it is still in the making.

§1.5. Keeping the Course

The first Part of this book, covered in Chapters 2 through 10, stresses the physical interpretation of FEM within the framework of the Direct Stiffness Method (DSM). This is done on account of its instructional advantages. Furthermore the computer implementation becomes more transparent because the sequence of operations can be placed in close correspondence with the DSM steps.

Chapters 11 through 19 deal specifically with element formulations. Ingredients of the mathematical interpretation are called upon whenever it is felt proper and convenient to do so. Nonetheless excessive entanglement with the mathematical theory is avoided if it may obfuscate the physics.

In Chapters 2 and 3 the time is frozen at about 1965, and the DSM presented as an aerospace engineer of that time would have understood it. This is not done for sentimental reasons, although that happens to be the year in which the writer began thesis work on FEM under Ray Clough. Virtually all commercial codes are now based on the DSM and the computer implementation has not essentially changed since the late 1960s.[14] What has greatly improved since is "marketing sugar": user interaction and visualization.

§1.6. *What is Not Covered

The following topics are not covered in this book:

1. Elements based on equilibrium, mixed and hybrid variational formulations.
2. Flexibility and mixed solution methods.
3. Plate and shell elements.

[13] Oliver Heaviside took heavy criticism from the lotus eaters, which he returned with gusto. His legacy is a living proof that "England is the paradise of individuality, eccentricity, heresy, anomalies, hobbies and humors" (George Santayana). Paul Dirac was luckier: he was shielded as member of the physics establishment and eventually received a Nobel Prize. Gilbert Strang, the first mathematician to dwelve in the real FEM (the one created by engineers) was kind to the founders.

[14] With the gradual disappearance of Fortran as a "live" programming language, noted in §1.7.7, changes at the implementation level have recently accelerated. E.g., C++, Python, Java and Matlab "wrappers" are becoming more common.

4. Variational methods in mechanics.
5. General mathematical theory of finite elements.
6. Buckling and stability analysis.
7. General nonlinear response analysis.
8. Structural optimization.
9. Error estimates and problem-adaptive discretizations.
10. Non-structural and multiphysics applications of FEM.
11. Designing and building production-level FEM software and use of special hardware (*e.g.* vector and parallel computers)

Topics 1–5 belong to what may be called "Advanced Linear FEM", whereas 6–7 pertain to "Nonlinear FEM". Topics 8–10 fall into advanced applications, whereas 11 is an interdisciplinary topic that interweaves with computer science.

§1.7. *Historical Sketch and Bibliography

This section summarizes the history of structural finite elements since 1950 to date. It functions as a hub for chapter-dispersed historical references.

For exposition convenience, structural "finitelementology" may be divided into four generations that span 10 to 15 years each. There are no sharp intergenerational breaks, but noticeable change of emphasis. The following summary does not cover the conjoint evolution of Matrix Structural Analysis into the Direct Stiffness Method from 1934 through 1970. This was the subject of a separate essay [89], which is also given in Appendix H.

§1.7.1. Who Invented Finite Elements?

Not just one individual, as this historical sketch will make clear. But if the question is tweaked to: who created the FEM in everyday use? there is no question in the writer's mind: M. J. (Jon) Turner at Boeing over the period 1950–1962. He generalized and perfected the Direct Stiffness Method, and forcefully got Boeing to commit resources to it while other aerospace companies were mired in the Force Method. During 1952–53 he oversaw the development of the first continuum based finite elements. In addition to Turner, major contributors to current practice include: B. M. Irons, inventor of isoparametric models, shape functions, the patch test and frontal solvers; R. J. Melosh, who recognized the Rayleigh-Ritz link and systematized the variational derivation of stiffness elements; and E. L. Wilson, who developed the first open source (and widely imitated and distributed) FEM software.

All of these pioneers were in the aerospace industry at least during part of their careers. That is not coincidence. FEM is the confluence of three ingredients, one of which is digital computation. And only large industrial companies (as well as some government agencies) were able to afford mainframe computers during the 1950s.

Who were the popularizers? Four academicians: J. H. Argyris, R. W. Clough, H. C. Martin, and O. C. Zienkiewicz are largely responsible for the "technology transfer" from the aerospace industry to a wider range of engineering applications during the 1950s and 1960s. The first three learned the method from Turner directly or indirectly. As a consultant to Boeing in the early 1950s, Argyris, a Force Method expert then at Imperial College, received reports from Turner's group, and weaved the material into his influencial 1954 serial [8]. To Argyris goes the credit of being the first in constructing a displacement-assumed continuum element [8,p. 62].

Clough and Martin, then junior professors at U.C. Berkeley and U. Washington, respectively, spent "faculty internship" summers at Turner's group during 1952 and 1953. The result of this seminal collaboration was a celebrated paper [255], widely considered the start of the present FEM. Clough baptized the method in 1960 [40] and went on to form at Berkeley the first research group to propel the idea into Civil Engineering applications. Olek Zienkiewicz, originally an expert in finite difference methods who learned the trade from Southwell, was convinced in 1964 by Clough to try FEM. He went on to write the first textbook on the subject [277] and to organize another important Civil Engineering research group in the University of Wales at Swansea.

§1.7.2. G1: The Pioneers

The 1956 paper by Turner, Clough, Martin and Topp [255], henceforth abbreviated to TCMT, is recognized as the start of the current FEM, as used in the overwhelming majority of commercial codes. Along with Argyris' serial [8] they prototype the first generation, which spans 1950 through 1962. A panoramic picture of this period is available in two textbooks [194,205]. Przemieniecki's text is still reprinted by Dover. The survey by Gallagher [108] was influential at the time but is now difficult to access outside libraries.

The pioneers were structural engineers, schooled in classical mechanics. They followed a century of tradition in regarding structural elements as a device to transmit forces. This "element as force transducer" was the standard view in pre-computer structural analysis. It explains the use of flux assumptions to derive stiffness equations in TCMT. Element developers worked in, or interacted closely with, the aircraft industry. (As noted above, only large aerospace companies were then able to afford mainframe computers.) Accordingly they focused on thin structures built up with bars, ribs, spars, stiffeners and panels. Although the Classical Force Method dominated stress analysis during the 1950s [89], stiffness methods were kept alive by use in dynamics and vibration. It is not coincidence that Turner was an world-class expert in aeroelasticity.

§1.7.3. G2: The Golden Age

The next period spans the golden age of FEM: 1962–1972. This is the "variational generation." Melosh showed [179] that conforming displacement models are a form of Rayleigh-Ritz based on the minimum potential energy principle. This influential paper marks the confluence of three lines of research: Argyris' dual formulation of energy methods [8], the Direct Stiffness Method (DSM) of Turner [256–258], and early ideas of interelement compatibility as basis for error bounding and convergence [101,178]. G1 workers thought of finite elements as idealizations of structural components. From 1962 onward a two-step interpretation emerges: discrete elements approximate continuum models, which in turn approximate real structures.

By the early 1960s FEM begins to expand into Civil Engineering through Clough's Boeing-Berkeley connection [48,49] and had been baptized [40,42]. Reading Fraeijs de Veubeke's famous article [102] side by side with TCMT [255] one can sense the ongoing change in perspective opened up by the variational framework. The first book devoted to FEM appears in 1967 [277]. Applications to nonstructural problems had started in 1965 [276], and were treated in some depth by Martin and Carey [172].

From 1962 onwards the displacement formulation dominates. This was given a big boost by the invention of the isoparametric formulation and related tools (numerical integration, fitted natural coordinates, shape functions, patch test) by Irons and coworkers [144–148]. Low order displacement models often exhibit disappointing performance. Thus there was a frenzy to develop higher order elements. Other variational formulations, notably hybrids [195,200], mixed [130,239] and equilibrium models [102] emerged. G2 can be viewed as closed by the monograph of Strang and Fix [228], the first book to focus on the mathematical foundations.

§1.7.4. G3: Consolidation

The post-Vietnam economic doldrums are mirrored during this post-1972 period. Gone is the youthful exuberance of the golden age. This is consolidation time. Substantial effort is put into improving the stock of G2 displacement elements by tools initially labeled "variational crimes" [227], but later justified. Textbooks by Hughes [142] and Bathe [15] reflect the technology of this period. Hybrid and mixed formulations record steady progress [13]. Assumed strain formulations appear [162]. A booming activity in error estimation and mesh adaptivity is fostered by better understanding of the mathematical foundations [237].

Commercial FEM codes gradually gain importance. They provide a reality check on what works in the real world and what doesn't. By the mid-1980s there was gathering evidence that complex and high order elements were commercial flops. Exotic gadgetry interweaved amidst millions of lines of code easily breaks down in new releases. Complexity is particularly dangerous in nonlinear and dynamic analyses conducted by novice users. A trend back toward simplicity starts [164,167].

§1.7.5. G4: Back to Basics

The fourth generation begins by the early 1980s. More approaches come on the scene, notably the Free Formulation [27,28], orthogonal hourglass control [96], Assumed Natural Strain methods [17,224], stress hybrid models in natural coordinates [198,206], as well as variants and derivatives of those approaches: ANDES [76,183], EAS [220,221] and others. Although technically diverse the G4 approaches share two common objectives:

(i) Elements must fit into DSM-based programs since that includes the vast majority of production codes, commercial or otherwise.

(ii) Elements are kept simple but should provide answers of engineering accuracy with relatively coarse meshes. These were collectively labeled "high performance elements" in 1989 [73].

Two more recent trends can be noted: increased abstraction on the mathematical side,[15] and canned recipes for running commercial software on the physical side.

"Things are always at their best in the beginning," said Pascal. Indeed. By now FEM looks like an aggregate of largely disconnected methods and recipes. The blame should not be placed on the method itself, but on the community split noted in the book Preface.

§1.7.6. Precursors

As used today, FEM represents the confluence of three ingredients: Matrix Structural Analysis (MSA), variational approximation theory, and the digital computer. These came together in the early 1950. The reader should not think, however, that they simultaneouly appeared on the table through some alchemy. MSA came on the scene in the mid 1930s when desk calculators became popular, as narrated in Appendix H. And variational approximation schemes akin to those of modern FEM were proposed before digital computers. Three examples:

- The historical sketch of [172] says that "Archimedes used finite elements in determining the volume of solids." The alleged linkage is tenuous. Indeed he calculated areas, lengths and volumes of geometrical objects by dividing them into simpler ones and adding their contributions, passing to the limit as necessary. Where does "variational approximation" come in? Well, one may argue that the volume (area, length) measure of an object is a scalar functional of its geometry. Transmute "measure" into "energy" and "simpler objects" into "elements" and you capture one of the FEM tenets: the energy of the system is the sum of element energies. But for Archimedes to reach modern FEM "long is the way, and hard," since physical energy calculations require derivatives and Calculus would not be invented for 20 centuries.
- In his studies leading to the creation of variational calculus, Euler divided the interval of definition of a one-dimensional functional into finite intervals and assumed a linear variation over each, defined by end values [155, p. 53]. Passing to the limit he obtained what is now called the Euler-Lagrange differential equation of variational calculus. Thus Euler deserves credit for being the first to use a piecewise linear function with discontinuous derivatives at nodes to produce, out of the hat, an ODE with second derivatives. He did not use those functions, however, to obtain an approximate value of the functional.[16]
- In the early 1940s Courant wrote an expository article [55] advocating the variational treatment of partial differential equations. The Appendix of this article contains the first FEM-style calculations on a triangular net for determining the torsional stiffness of a hollow shaft. He used piecewise linear interpolation over each triangle as Rayleigh-Ritz trial functions, and called his idea "generalized finite differences."

---

[15] "If you go too far up, abstraction-wise, you run out of oxygen." (Joel Spolsky).

[16] That would have preceded the invention of direct variational methods (Rayleigh-Ritz) for over one century, while representing also the first FEM-style calculation. A near miss indeed.

- A direct variational approach similar to Courant's was continued by Synge and Prager in the context of functional analysis [203] and exposed in Synge's book [236] as the "hypercircle" method.[17]
- The seminal paper by Turner et al [255] cites two immediate DSM precursors, both dated 1953, by Levy [159] and Schuerch [217]. (Only the former is available as a journal article; both have "delta wings" in the title.) From [255, p. 806]: "In a recent paper Levy has presented a method of analysis for highly redundant structures that is particularly suited to the use of high-speed digital computing machines. ... The stiffness matrix for the entire structure is computed by simple summation of of the stiffness matrices of the elements of the structure."

Precursors prior to 1950 had no influence on the rapid developments of Generation 1 outlined in §1.7.2. Two crucial pieces were missing. First, and most important, was the programmable digital computer. Without computers FEM would be a curiosity, worth perhaps a footnote in an arcane book. Also missing was a *driving application* that could get the long-term attention of scientists and engineers as well as industrial resources to fund R&D work. Aerospace structural mechanics provided the driver because the necessary implementation apparatus of MSA was available since the late 1930s [105].

Matrix procedures had to be moved from desk calculators and punched-tape accounting machines to digital computers, which affluent aerospace companies were able to afford amidst Cold War paranoia. Can you imagine defense funds pouring into hypercircles? Once all pieces were in place, synergy transformed the method into a *product*, and FEM took off.

§**1.7.7. Recommended Books for Linear FEM**

The literature is vast: over 200 textbooks and monographs have appeared since 1967. Some recommendations for readers interested in further studies within *linear* FEM are offered below.

Basic level (reference): Zienkiewicz and Taylor [280]. This two-volume set is a comprehensive upgrade of the previous edition [278]. Primarily an encyclopœdic reference work that gives a panoramic coverage of FEM applications, as well as a comprehensive list of references. Not a textbook or monograph. Prior editions suffered from loose mathematics, largely fixed in this one. A three-volume fifth edition has appeared recently.

Basic level (textbook): Cook, Malkus and Plesha [50]. The third edition is comprehensive in scope although the coverage is more superficial than Zienkiewicz and Taylor. A fourth edition has appeared recently.

Intermediate level: Hughes [142]. It requires substantial mathematical expertise on the part of the reader. Recently (2000) reprinted as Dover edition.

Mathematically oriented: Strang and Fix [228]. Still the most readable mathematical treatment for engineers, although outdated in several subjects. Out of print.

Best value for the $$$: Przemieniecki's Dover edition [205], list price $15.95 (2003). A reprint of a 1966 McGraw-Hill book. Although woefully outdated in many respects (the word "finite element" does not appear except in post-1960 references), it is a valuable reference for programming simple elements. Contains a fairly detailed coverage of substructuring, a practical topic missing from the other books. Comprehensive bibliography in Matrix Structural Analysis up to 1966.

Most fun (if you appreciate British "humor"): Irons and Ahmad [148]. Out of print.

For buying out-of-print books through web services, check the metasearch engine in `www3.addall.com` (most comprehensive; not a bookseller) as well as that of `www.amazon.com`. A newcomer is `www.campusi.com`

---

[17] Curiously this book does not mention, even in passing, the use of digital computers that had already been commercially available for several years. The few numerical examples, all in 2D, are done by hand via relaxation methods.

§1.7.8. Hasta la Vista, Fortran

Most FEM books that include programming samples or even complete programs use Fortran. Those face an uncertain future. Since the mid-1990s, Fortran is gradually disappearing as a programming language taught in USA engineering undergraduate programs. (It still survives in Physics and Chemistry departments because of large amounts of legacy code.) So one end of the pipeline is drying up. Low-level scientific programming is moving to C and C++, mid-level to Java, Perl and Python, high-level to Matlab, Mathematica and their free-source Linux equivalents. How attractive can a book teaching in a dead language be?

To support this argument with some numbers, here is a September-2003 snapshot of ongoing open source software projects listed in `http://freshmeat.net`. This conveys the relative importance of various languages (a mixed bag of newcomers, going-strongs, have-beens and never-was) in the present environment.

Lang	Projects	Perc	Lang	Projects	Perc	Lang	Projects	Perc
Ada	38	0.20%	APL	3	0.02%	ASP	25	0.13%
Assembly	170	0.89%	Awk	40	0.21%	Basic	15	0.08%
C	5447	28.55%	C#	41	0.21%	C++	2443	12.80%
Cold Fusion	10	0.05%	Common Lisp	27	0.14%	Delphi	49	0.26%
Dylan	2	0.01%	Eiffel	20	0.10%	Emacs-Lisp	33	0.17%
Erlang	11	0.06%	Euler	1	0.01%	Euphoria	2	0.01%
Forth	15	0.08%	Fortran	45	0.24%	Haskell	28	0.15%
Java	2332	12.22%	JavaScript	236	1.24%	Lisp	64	0.34%
Logo	2	0.01%	ML	26	0.14%	Modula	7	0.04%
Object Pascal	9	0.05%	Objective C	131	0.69%	Ocaml	20	0.10%
Other	160	0.84%	Other Scripting Engines	82	0.43%			
Pascal	38	0.20%	Perl	2752	14.42%	PHP	2020	10.59%
Pike	3	0.02%	PL/SQL	58	0.30%	Pliant	1	0.01%
PROGRESS	2	0.01%	Prolog	8	0.04%	Python	1171	6.14%
Rexx	7	0.04%	Ruby	127	0.67%	Scheme	76	0.40%
Simula	1	0.01%	Smalltalk	20	0.10%	SQL	294	1.54%
Tcl	356	1.87%	Unix Shell	550	2.88%	Vis Basic	15	0.08%
Xbasic	1	0.01%	YACC	11	0.06%	Zope	34	0.18%

Total Projects: 19079

Notes and Bibliography

Here is Ray Clough's personal account of how FEM and DSM emerged at Boeing in the early 1950s.

> " My involvement with the FEM began when I was employed by the Boeing Airplane Company in Seattle during summer 1952 as a member of their summer faculty program. When I had joined the civil engineering faculty at Berkeley in 1949, I decided to take advantage of my MIT structural dynamics background by taking up the field of Earthquake Engineering. So because the Boeing summer faculty program offered positions with their structural dynamics unit, I seized on that as the best means of advancing my preparation for the earthquake engineering field. I was particularly fortunate in this choice of summer work at Boeing because the head of their structural dynamics unit was Mr. M. J. Turner — a very capable man in dealing with problems of structural vibrations and flutter.
>
> When I arrived for the summer of 1952, Jon Turner asked me to work on the vibration analysis of a delta wing structure. Because of its triangular plan form, this problem could not be solved by procedures based on standard beam theory; so I spent the summer of 1952 trying to formulate a delta wing model built up as an assemblage of one-dimensional beams and struts. However, the results of deflection analyses based on this type of mathematical model were in very poor agreement with data obtained from laboratory tests of a scale model of a delta wing. My final conclusion was that my summer's work was a total failure—however, at least I learned what did not work.
>
> Spurred by this disappointment, I decided to return to Boeing for the summer faculty program in 1953. During the winter, I stayed in touch with Jon Turner so I was able to rejoin the structural dynamics unit in June. The

most important development during the winter was that Jon suggested we try to formulate the stiffness property of the wing by assembling plane stress plates of either triangular or rectangular shapes. So I developed stiffness matrices for plates of both shapes, but I decided the triangular form was much more useful because such plates could be assembled to approximate structures of any configuration. Moreover, the stiffness properties of the individual triangular plates could be calculated easily based on assumptions of uniform states of normal stress in the X and the Y directions combined with an uniform state of shear stress. Then the stiffness of the complete structure was obtained by appropriate addition of the contributions from the individual pieces. The Boeing group called this procedure the direct stiffness method.

The remainder of the summer of 1953 was spent in demonstrating that deflections calculated for structures formed as assemblages of triangular elements agreed well with laboratory measurements on the actual physical models. Also, it became apparent that the precision of the calculated results could be improved asymptotically by continued refinement of the finite element mesh. The conclusions drawn from that summer's work were presented in a paper given by Jon Turner at the annual meeting of the Institute of Aeronautical Sciences in January 1954. However, for reasons I never understood Jon did not submit the paper for publication until many months later. So this paper, which often is considered to be the first published description of the FEM, was not published until September 1956 — more than two years after the verbal presentation.

It is important to note that the basic purpose of the work done by Jon Turner's structural dynamics unit was vibration and flutter analysis. They were not concerned with stress analysis because that was the responsibility of the stress analysis unit. However, it was apparent that the model formed by the direct stiffness method could be used for stress analysis as well as for vibration analysis, and I made plans to investigate this stress analysis application as soon as possible. However, because of my other research responsibilities, I was not able to spend any significant time on the stress analysis question until I went on my sabbatical leave to Trondheim, Norway in September 1956. Then, when I arrived in Norway all I could do was to outline the procedures for carrying out the analysis, and to do calculations for very small systems using a desk calculator because the Norwegian Institute of Technology did not yet have an automatic digital computer.

The presentation of the paper to the Institute of Aeronautical Sciences was the first introduction of the principles of the FEM to a technical audience; although some of the basic concepts of the method were stated a short time later in a series of articles published in Aircraft Engineering by Dr. John H. Argyris during October 1954 to May 1955. However, the rectangular element presented in those articles is only a minor part of that contribution. The Argyris work came to my attention during my sabbatical leave in Norway, and I considered it then (as I still do now) to be the most important series of papers ever published in the field of Structural Mechanics. I credit that work for extending the scope of my understanding of structural theory to the level it eventually attained.

From my personal point of view, the next important event in finite element history was the coining of the name FEM. My purpose in choosing that name was to distinguish clearly the relatively large size pieces of the structure that make up a finite element assemblage as contrasted with the infinitesimal contributions that go into evaluation of the displacements of a structure in a typical virtual work analysis. The name first appeared in a publication that was written to demonstrate the finite element procedure for the civil engineering profession. A much more significant application of the method was presented at the Symposium on the use of Computers in Civil Engineering, held in Lisbon, Portugal in 1962, where it was used to evaluate the stress concentrations developed in a gravity dam that had cracked at its mid-section."

References

Referenced items have been moved to Appendix R. Partly sorted.

Homework Exercises for Chapter 1
Overview

EXERCISE 1.1 [A:15] Work out Archimedes' problem using a circumscribed regular polygon, with $n = 1, 2, 4, \ldots 256$. Does the sequence converge any faster?

EXERCISE 1.2 [D:20] Select one of the following vehicles: truck, car, motorcycle, or bicycle. Draw a two level decomposition of the structure into substructures, and of selected components of some substructures.

EXERCISE 1.3 [D:30] In one of the earliest articles on the FEM, Clough [42] writes:

"When idealized as an assemblage of appropriately shaped two- and three-dimensional elements in this manner, an elastic continuum can be analyzed by standard methods of structural analysis. It should be noted that the approximation which is employed in this case is of physical nature; a modified structural system is substituted for the actual continuum. There need be no approximation in the mathematical analysis of this structural system. This feature distinguishes the finite element technique from finite difference methods, in which the exact equations of the actual physical system are solved by approximate mathematical procedures."

Discuss critically the contents of this paragraph while placing it in the context of time of writing (early 1960s). Is the last sentence accurate?