1
Overview

TABLE OF CONTENTS

§1.1. Where the Book Fits

This book is an introduction to the analysis of nonlinear elastic structures by the Finite Element Method (FEM). It embodies five Parts:

I **Overview of Nonlinear Problems**.

II **Formulation of Geometrically Nonlinear Finite Elements**.

III **Solution Methods**.

IV **Application to Stability Analysis**.

V **Nonconservative Problems**.

This Chapter presents an overview of where the book fits within the larger scope of Mechanics. It is assumed that the reader has a good idea of what finite elements are, so this aspect is glossed over.

§1.2. 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.

§1.2.1. 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 interrelated with 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 a 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 related areas are hydrodynamics, aerodynamics, atmospheric physics, and combustion.

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, building, bridges, engines, cars, microchips, radio telescopes, robots, roller skates and garden sprinklers. Biological systems, such as a whale, amoeba or pine tree are included if studied from the viewpoint of biomechanics. Ecological, astronomical and cosmological entities also form systems.[1]

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

§1.2.2. 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 the time dependence is 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 the inertial forces are ignored or neglected. Static problems may be classified into strictly static and quasi-static. For the former time need not

[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?” (Stephen Hawking).

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

be considered explicitly; any historical time-like response-ordering parameter (if one is needed) will do. In quasi-static problems such as foundation settlement, creep deformation, rate-dependent plasticity or fatigue cycling, a more realistic estimation of time is required but inertial forces are still neglected.

§1.2.3. Linear vs. Nonlinear

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

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

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.

§1.2.4. 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 of finite number of degrees of freedom:

$$\text{Spatial discretization method} \begin{cases} \textit{Finite Element Method (FEM)} \\ \textit{Boundary Element Method (BEM)} \\ \textit{Finite Difference Method (FDM)} \\ \textit{Finite Volume Method (FVM)} \\ \textit{Spectral Method} \\ \textit{Mesh-Free Method} \end{cases} \tag{1.4}$$

For *linear* problems finite element methods currently dominate the scene, with boundary element methods posting a strong second choice in specific application areas. For *nonlinear* problems the dominance of finite element methods is overwhelming.

Classical 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 dominant. Finite-volume methods, which address finite volume method conservation laws, are important in highly nonlinear problems of fluid mechanics. Spectral methods are based on transforms that map space and/or time dimensions to spaces where the problem is easier to solve.

A recent newcomer to the scene are the mesh-free methods. These are finite different methods on arbitrary grids constructed by a subset of finite element techniques

§1.2.5. FEM Variants

The term *Finite Element Method* actually identifies a broad spectrum of techniques that share common features outlined in introductory FEM textbooks. Two subclassifications that fit well

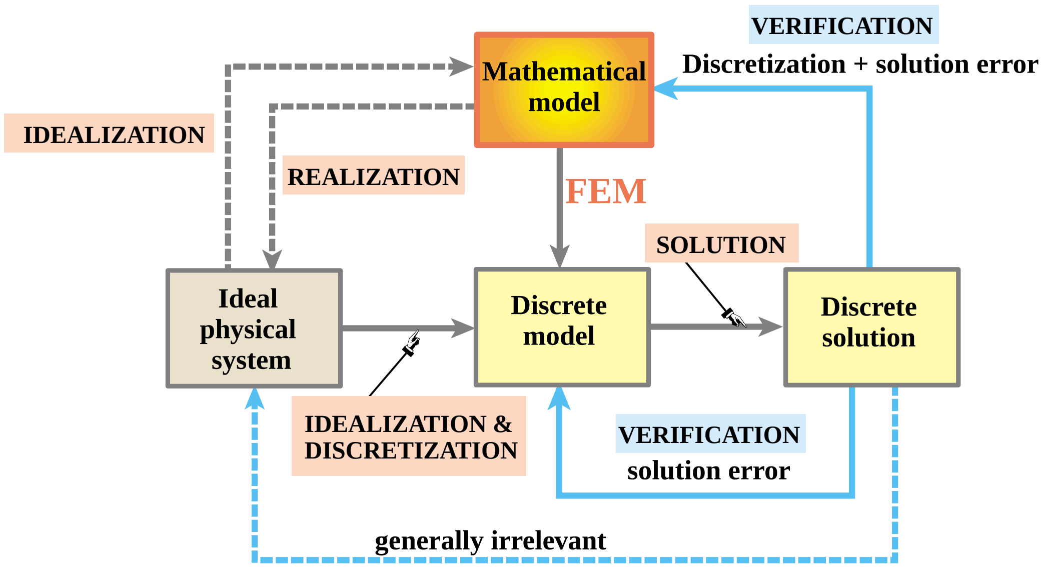

Figure 1.1. The Mathematical FEM. The mathematical model (at top) is the source of the process. Discrete model and solution follow from it. The ideal physical system is inessential.

applications to structural mechanics are

$$\text{FEM Formulation} \begin{cases} \textit{Displacement} \\ \textit{Equilibrium} \\ \textit{Mixed} \\ \textit{Hybrid} \end{cases} \qquad \text{FEM Solution} \begin{cases} \textit{Stiffness} \\ \textit{Flexibility} \\ \textit{Mixed (a.k.a. Combined)} \end{cases} \tag{1.5}$$

Using the foregoing classification, we can state the topic of this book more precisely: the *computational simulation of nonlinear static structural problems* by the Finite Element Method. Of the variants listed in (1.5), emphasis is placed on the *displacement* formulation and *stiffness* solution. This combination is called the *Direct Stiffness Method* or DSM.

§1.3. The FEM Analysis Process

A model-based simulation process using FEM involves doing a sequence of steps. This sequence takes two canonical configurations depending on the environment in which FEM is used. These are reviewed next to introduce terminology.

§1.3.1. The Mathematical FEM

The process steps are illustrated in Figure 1.1. The process centerpiece, from which everything emanates, is 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.[3] This is the *discretization* step. The FEM equations are processed by an equation solver, which delivers a discrete solution (or solutions).

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

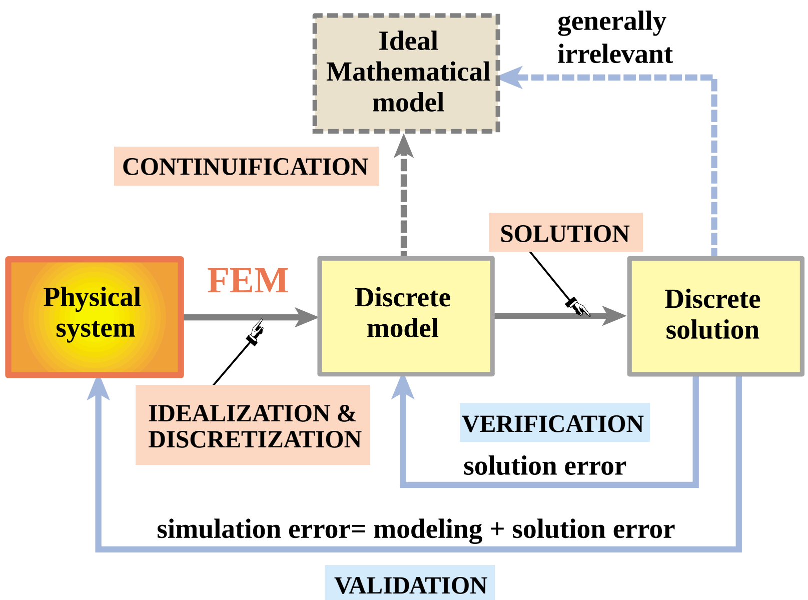

Figure 1.2. The Physical FEM. The physical system is the source of the process. The ideal mathematical model is inessential.

On the left Figure 1.1 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. For example, if the mathematical model is the Poisson's equation, realizations may be a heat conduction or a electrostatic charge distribution problem. This step is inessential and may be left out. Indeed FEM discretizations may be constructed without any reference to physics.

The concept of *error* arises when the discrete solution is substituted in the "model" boxes. This replacement is generically called *verification*. 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.[4] Replacing into the ideal physical system would in principle quantify modeling errors. In the mathematical FEM this is largely irrelevant, however, since the ideal physical system is merely that: a figment of the imagination.

§1.3.2. The Physical FEM

The second way of using FEM is the process illustrated in Figure 1.3. The centerpiece is now the *physical system* to be modeled. Accordingly, this sequence is called the *Physical FEM*. The processes of idealization and discretization are carried out *concurrently* to produce the discrete model. The solution is computed as before.

Just like Figure 1.1 shows an ideal physical system, Figure 1.1 depicts 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

[4] This error can be computed in several ways, the details of which are of little practical importance.

others, notably complex engineering systems, it makes no sense. Indeed 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, known as *verification* and *validation*, respectively. Verification is the same as in the Mathematical FEM: the discrete solution is replaced into the discrete model to get the solution error. As noted above this error is not generally important. Substitution in the ideal mathematical model in principle provides the discretization error. This is rarely useful in complex engineering systems, however, since there is no reason to expect that the mathematical model exists, and if it does, that it is more 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. Since the latter is typically insignificant, the simulation error in practice can be identified with the modeling error.

§1.3.3. Synergy of Physical and Mathematical FEM

The foregoing physical and mathematical sequences are not exclusive but complementary. This synergy[5] 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 very complex systems such as aircraft. 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 differential equation that captures, at a continuum mechanics level,[6] 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. First, separate the structure and view the rest as masses and forces, most of which are time-varying and nondeterministic. Second, consider the aircraft structure as built of *substructures*[7] wings, fuselage, stabilizers, engines, landing gears, and so on. Take each substructure, and continue to decompose it into *components*: rings, ribs, spars, cover, plates, etc, continuing through as many levels as necessary. Eventually those components become sufficiently simple in geometry and connectivity that they can be reasonably well described by continuum 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. 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 expressed in conservation form.

This multilevel decomposition process is diagrammed in Figure 1.3, in which the intermediate substructure level is omitted for simplicity.

---

[5] This 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).

[6] Of course at the atomic and subatomic level quantum mechanics works for everything, from landing gears to passengers. But it would be slightly impractical to model the aircraft by 10^{36} interacting particles.

[7] A *substructure* is a part of a structure devoted to a specific function.

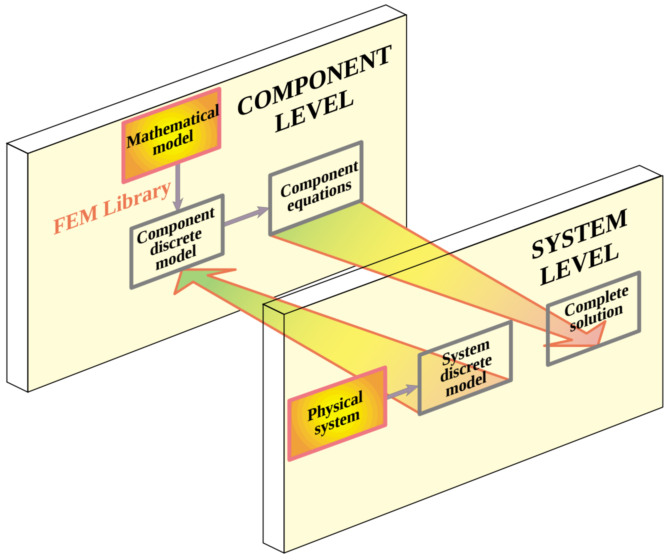

Figure 1.3. Combining physical and mathematical modeling through multilevel FEM. Only two levels (system and component) are shown for simplicity.

Remark 1.2. More intermediate decomposition levels are used in some systems, such as offshore and ship structures, which are characterized by a modular fabrication process. In that case the decomposition mimics the way the system is actually constructed. The general technique, called *superelements*, is outlined in the IFEM book. In nonlinear analysis the analysis procedure is more complex since the static condensation technique of linear analysis no longer works.

Remark 1.3. There is no point in practice in going beyond a certain component level while considering the complete model, since the level of detail can become overwhelming without adding significant information. Further refinement or particular components is done by the so-called global-local analysis techniques. This technique is an instance of multiscale analysis.

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 1.4. Multiple levels are unnecessary here. 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 interpretation 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 is aligned with the view of Figure 1.2. This interpretation has been shaped by the discovery and extensive use of the method in the field of structural mechanics. This relationship is reflected in the use of structural terms such as "stiffness matrix", "force vector" and "degrees of freedom." This terminology carries over to non-structural applications.

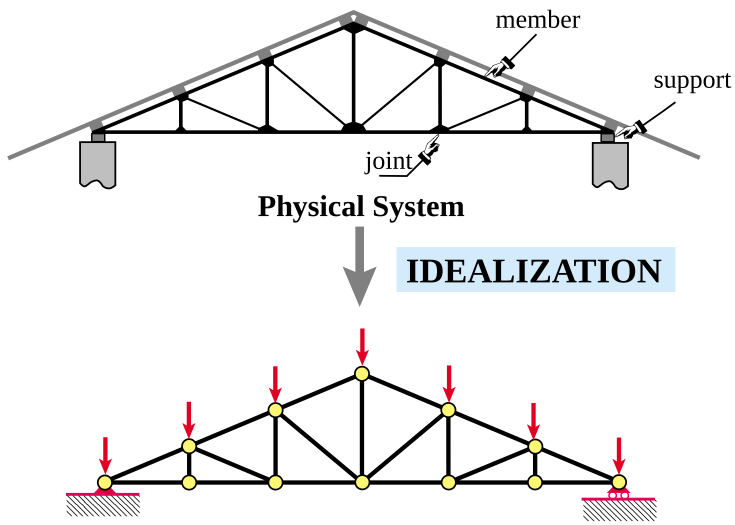

Figure 1.6. The idealization process for a simple structure. The physical system, here a roof truss, is directly idealized by the mathematical model: a pin-jointed bar assembly. This coalesces with the discrete model.

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.

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*. 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 configuration of Figure 1.1. 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.

§1.5. The Solution Morass

In nonlinear analysis the two FEM interpretations are not equal in importance. Nonlinear analysis demands a persistent attention to the underlying physics to avoid getting astray as the "real world" is covered by layer upon layer of mathematics and numerics.

Why is concern for physics of paramount importance? A key component of finite element nonlinear analysis is the solution of the nonlinear algebraic systems of equations that arise upon discretization.

FACT

The numerical solution of nonlinear systems in "black box" mode is *much* more difficult than in the linear case.

The key difficulty is tied to the essentially obscure nature of general nonlinear systems, about which very little can be said in advance. And you can be sure that Murphy's law[8] works silently in the background.

One particularly vexing aspect of dealing with nonlinear systems is the *solution morass*. A determinate system of 1, 1000, or 1000000 *linear* equations has, under mild conditions, one and only one solution. The computer effort to obtain this solution can be estimated fairly accurately if the sparseness (or denseness) of the coefficient matrix is known. Thus setting up linear equation solvers as "black-box" stand-alone functions or modules is a perfectly sensible thing to do.

By way of contrast, a system of 1000 cubic equations has $3^{1000} \approx 10^{300}$ solutions in the complex plane. This is much, much larger than the number of atoms in the Universe, which is merely 10^{50} give or take a few. Suppose just several billions or millions of these are real solutions. Which solution(s) have physical meaning? And how do you compute those solutions without wasting time on the others?

This combinatorial difficulty is overcome by the concept of *continuation*, which engineers also call *incremental analysis*. Briefly speaking, we start the analysis from an easily computable solution — for example, the linear solution — and then try to follow the behavior of the system as actions

[8] If something can go wrong, it *will* go wrong.

applied to it are changed by small steps called *increments*. The previous solution is used as a starting point for the iterative solution-search procedure. The underlying prescription: *follow the physics*. This technique is interwined with the concept of response explained in Chapter 2.

Remark 1.4. Not surprisingly, incremental analysis was used by the aerospace engineers that first used the finite element method for geometrically nonlinear analysis in the late 1950s. Techniques have been considerably refined since then, but the underlying idea remains the same.

We conclude this overview with a historical perspective on nonlinear finite element methods in solid and structural mechanics, along with a succint bibliography.

§1.6. Historical Background

In the history of finite element methods the year 1960 stands out. The name “finite element method” appears for the first time in the open literature in an article by Clough [52]. And Turner, Dill, Martin and Melosh [187] publish a pioneering paper in nonlinear structural analysis. The then-five-year-old “direct stiffness method” (what we now call displacement-assumed finite element method) was applied to

“problems involving nonuniform heating and/or large deflections . . . in a series of linearized steps. Stiffness matrices are revised at the beginning of each step to account for changes in internal loads, temperatures, and geometric configuration.”

Thirty years and several thousand publications later, computerized nonlinear structural analysis has acquired full adult rights, but has not developed equally in all areas.

The first fifteen years (1960-1975) were dominated by *formulation* concerns. For example, not until the late 1960s were correct finite-deflection incremental forms for displacement models rigorously derived. And interaction of flow-like constitutive behavior with the spatial discretization (the so called “incompressibility locking” effects) led to important research into constitutive equations and element formulations.

While the investigators of this period devoted much energy to obtaining correct and implementable nonlinear finite-element equations, the art of solving such equations in a reliable and efficient manner was understandably neglected. This helps to explain the dominance of purely incremental methods. Corrective methods of Newton type did not get much attention until the early 1970s, and then only for geometrically nonlinear problems. At the time of this writing, progress in numerical solution techniques has been uneven: well developed for certain problems, largely a black art in others. To understand the difference, it pays to distinguish between *smooth nonlinearities* and *rough nonlinearities*.

§1.6.1. Smooth Nonlinearities

Problem with smooth nonlinearities are characterized by continuous, path-independent nonlinear relations at the *local* level. Some examples:

1. *Finite deflections* (geometric nonlinearities). Nonlinear effects arise from strain-displacement equations, which are well behaved for all strain measures in practical use.

2. *Nonlinear elasticity*. Stresses are nonlinear but reversible functions of strains.

3. *Follower forces* (e.g., pressure loading). External forces are smooth nonlinear functions of displacements.

A unifying characteristic of this problem is that nonlinearities are of *equality* type, i.e., reversible, and these relations are *continuous* at each point within the structure. Mathematicians call these *smooth mappings*.

It is important to point out, however, that the *overall* structural behavior is not necessarily smooth; as witnessed by the phenomena of buckling, snapping and flutter. But at the local level everything is smooth: nonlinear strain-displacement equations, nonlinear elasticity law, follower pressures.

Methods for solving this class of problems are highly developed, and have received a great deal of attention from the mathematical and numerical analysis community. This research has directly benefitted many areas of structural analysis.

Let us consider finite deflection problems as prototype. Within the finite element community, these were originally treated by purely incremental (step-by-step) techniques; but anomalies detected in the mid-1960's prompted research into consistent linearizations. A good exposition of this early work is given in the book by Oden [123]. Once formulation questions were settled, investigators had correct forms of the "residual" out-of-balance forces and tangent stiffness matrix, and incremental steps began to be augmented with corrective iterations in the late 1960s. Conventional and modified Newton methods were used in the corrective phase. These were further extended through restricted step (safeguarded Newton) and, more recently, variants of the powerful conjugate-gradient and quasi-Newton methods.

But difficulties in detecting and traversing limit and bifurcation points still remained. Pressing engineering requirements for post-buckling and post-collapse analyses led to the development of displacement control, alternating load/displacement control, and finally arclength control. The resultant increment-control methods have no difficulty in passing limit points. The problem of reliably traversing simple bifurcation points without guessing imperfections remains a research subject, while passing multiple or clustered bifurcation points remains a frontier subject. A concerted effort is underway, however, to subsume these final challenges.

These reliable solution methods have been implemented into many special-purpose finite element programs, and incorporation into general-purpose programs is proceeding steadily.

Remark 1.5. As noted above, incremental methods were the first to be used in nonlinear structural analysis. Among the pre-1970 contributions along this line we may cite Argyris and coworkers [9, 10], Felippa [62], Goldberg and Richard [82], Marcal, Hibbitt and coworkers [89,108,109], Oden [122], Turner, Martin and coworkers [110,187,188],

Remark 1.6. The earliest applications of Newton methods to finite element nonlinear analysis are by Oden [122], Mallet and Marcal [107], and Murray and Wilson [117,118]. During the early 1970s Stricklin, Haisler and coworkers at Texas A&M implemented and evaluated self-corrective, pseudo-force, energy-search and Newton-type methods and presented extensive comparisons; see Stricklin *et. al.* [175–178], Tillerson *et. al.* [185], and Haisler *et. al.* [85]. Almroth, Brogan, Bushnell and coworkers at Lockheed began using true and modified Newton methods in the late 1960s for energy-based finite-difference collapse analysis of shells; see Brogan and Almroth [34], Almroth and Felippa [5], Brush and Almroth [37], and Bushnell [38–39]. By the late 1970s Newton-like methods enjoyed widespread acceptance for geometrically nonlinear analysis.

Remark 1.7. Displacement control strategies for finite element post-buckling and collapse analysis were presented by Argyris [11] and Felippa [62] in 1966, and generalized in different directions by Sharifi and Popov [165,166] (fictitious springs), Bergan *et. al.* [25,26], (current stiffness parameter), Powell and Simons [142] and Bergan and Simons [27] (multiple displacement controls). A modification of Newton's method to traverse bifurcation points was described by Thurston [184]. Arclength control schemes for structural problems may be found in the following source papers: Wempner [194], Riks [154], Schmidt [160], Crisfield [48,49], Ramm [146], Felippa [65–66], Fried [73], Park [133], Padovan [130,131], Simo *et.al.* [167], Yang and McGuire [199], Bathe and Dvorkin [20]. Other articles of particular interest are Bathe and Cimento [18], Batoz and Dhatt [22], Bushnell [39], Bergan [27], Geradin *et al.* [79,80]. Meek and Tan [112], Ramm [146,147], Riks [155–157], Sobel and Thomas [170], Zienkiewicz [200,201,203]. Several conferences have been devoted exclusively to nonlinear problems in structural mechanics, for example [12,28,17,125, 179,180,198]. Finite element textbooks and monographs dealing rather extensively with nonlinear problems are by Oden [123], Bathe [19], Bushnell [40], White [196] and Zienkiewicz [202].

Remark 1.8. In the mathematical literature the concept of continuation (also called imbedding) can be traced back to the 1930s. A survey of the work up to 1950 is given by Ficken [70]. The use of continuation by parameter differentiation as a numerical method is attributed to Davidenko [54]. Key papers of this early period are by Freudenstein and Roth [72], Deist and Sefor [58] and Meyer [114], as well as the survey by Wasserstrom [191]. This early history is covered by Wacker [190].

Remark 1.9. Arclength continuation methods in the mathematical literature are generally attributed to Haselgrove [87] and Klopfestein [99] although these papers remained largely unnoticed until the late 1970s. Important contributions to the mathematical treatment are by Abbott [1], Anselone and Moore [8], Avila [14], Brent [31], Boggs [29], Branin [30], Broyden [35,36], Cassel [43], Chow *et. al.* [45], Crandall and Rabinowitz [47], Georg [77,78], Keller and coworkers [44,56,57,93–96], Matthies and Strang [111], Moore [115,116], Pönish [139,140], Rheinboldt and coworkers [59,113, 149–150], Watson [192] and Werner and Spence [195]. Of these, key contributions in terms of subsequent influence are [45,94,149]. For surveys and edited proceedings see Allgower [2,3], Byrne and Hall [42], Küpper [104,105], Rall [145], Wacker [190], and references therein. Textbooks and monographs dealing with nonlinear equation solving include Chow and Hale [46], Dennis and Schnabel [61], Kubíček and Hlaváček [102], Kubíček and Marek [103], Ortega and Rheinboldt [128], Rabinowitz [143], Rall [144], Rheinboldt [153], and Seydel [164]. Of these, the book by Ortega and Rheinboldt remains a classic and an invaluable source to essentially all mathematically oriented work done prior to 1970. The book by Seydel contains material on treatment of conventional and Hopf bifurcations not readily available elsewhere. Nonlinear equation solving is interwined with the larger subject of optimization and mathematical programming; for the latter the textbooks by Gill, Murray and Wright [81] and Fletcher [71] are highly recommended.

§1.6.2. Rough Nonlinearities

Rough nonlinearities are characterized by discontinuous field relations, usually involving *inequality* constraints. Examples: flow-rule plasticity, contact, friction. The *local* response is nonsmooth.

Solution techniques for these problems are in a less satisfactory state, and case-by-case consideration is called for. The local and overall responses are generally path-dependent, an attribute that forces the past response history to be taken into account.

The key difficulty is that conventional solution procedures based on Taylor expansions or similar differential forms may fail, because such Taylor expansions need not exist! An encompassing mathematical treatment is lacking, and consequently problem-dependent handling is presently the rule. For this class of problems incremental methods, as opposed to incremental-iterative methods, still dominate.

Remark 1.10. Earliest publications on computational plasticity using finite element methods are by Gallagher *et. al.* [74], Argyris [9,10], Marcal [108], Pope [138] and Felippa [62]. By now there is an enormous literature on the numerical treatment of inelastic processes, especially plasticity and creep. Fortunately the survey by Bushnell [40], although focusing on plastic buckling, contains over 300 references that collectively embody most of the English-speaking work prior to 1980. Other important surveys are by Armen [13] and Willam [197]. For contact problems, see Oden [126], Bathe and Chaudhary [19], Kikuchi and Oden [97,98], Simo *et. al.* [167] Stein *et. al.* [173], Nour-Omid and Wriggers [121], and references therein.

§1.6.3. Hybrid Approach

What does an analyst do when faced with an unfamiliar nonlinear problem? If the problem falls into the smooth-nonlinear type, there is no need to panic. Robust and efficient methods are available. Even if the whizziest methods are not implemented into one's favorite computer program, there is a wealth of theory and practice available for trouble-shooting.

But what if the problem include rough nonlinearities? A time-honored general strategy is divide and conquer. More specifically, two powerful techniques are frequently available: *splitting* and *nesting*.

Splitting can be used if the nonlinearities can be separated in an additive form:

Smooth + Rough

This separation is usually done at the force level. Then the smooth-nonlinear term is treated by conventional techniques whereas the rough-nonlinear term is treated by special techniques. This scheme can be particularly effective when the rough nonlinearity is *localized*, for example in contact and impact problems.

Nesting may be used when a simple additive separation is not available. This is best illustrated by an actual example. In the early 1970s, some authors argued that Newton's method would be useless for finite-deflection elastoplasticity, as no unique Jacobian exists in plastic regions on account of loading/unloading switches. The argument was compelling but turned out to be a false alarm. The problem was eventually solved by "nesting" geometric nonlinearities within the material nonlinearity, as illustrated in Fig. 1.1.

In the inner equilibrium loop the material law is "frozen", which makes the highly effective Newton-type methods applicable. The non-conservative material behavior is treated in an outer loop where material properties and constitutive variables are updated in an incremental or sub-incremental manner.

Another application of nesting comes in the global function approach (also called Rayleigh-Ritz or reduced-basis approach), which is presently pursued by several investigators. The key idea is to try to describe the *overall* response behavior by a few parameters, which are amplitudes of globally defined functions. The small nonlinear system for the global parameters is solved in an inner loop, while an external loop involving residual calculations over the detailed finite element model is executed occasionally.

Despite its inherent implementation complexity, the global function approach appears cost-effective for smooth, path-independent nonlinear systems. This is especially so when expensive parametric studies are involved, as in structural optimization under nonlinear stability constraints.

Remark 1.11. For geometric-material nesting and subincremental techniques see Bushnell [38–41], and references therein. The global-function approach in its modern form was presented by Almroth, Stern and Brogan [7] and pursued by Noor and coworkers under the name of reduced-basis technique; see Noor and Peters [119] and Noor [120] as well as the chapter by Noor in this volume. For perturbation techniques see the survey by Gallagher [75].

§1.6.4. Summary of Present Status

Solution techniques for smooth nonlinearities are in a fairly satisfactory state. Although further refinements in the area of traversing bifurcation points can be expected, incremental-iterative methods implemented with general increment control appear to be as reliable as an engineer user may reasonably expect.

For rough nonlinearities, case-by-case handling is still necessary in view of the lack of general theories and implementation procedures. Separation or nesting of nonlinearities, when applicable, can lead to significant gains in efficiency and reliability, but at the cost of programming complexity and problem-dependent implementations.