
BAZANT, Z. P.

Stability of Elastic, Anelastic, and Disintegrating Structures: A Conspectus of Main Results

The article attempts to review the main results in the vast field of stability of structures. The classical field of elastic stability is covered succinctly. The coverage emphasizes the modern problems of anelastic structures exhibiting plasticity and creep, and especially structures disintegrating due to localized fracture and distributed damage. The treatment encompasses thin or slender structures, i.e. the columns, frames, arches, thin-wall beams, plates, and shells, as well as massive but soft bodies buckling three-dimensionally, and includes the static as well as dynamic concepts of stability, dynamic instability of nonconservative systems, energy methods for discrete and continuous structures, thermodynamics of structures, postcritical behavior, and imperfection sensitivity. The legacy of Ludwig Prandtl, who is commemorated by the present Special Issue, is briefly highlighted. The mathematics is kept to the bare minimum, and the derivations as well as the differential equations are omitted. Only the main literature sources to this vast field are cited.

MSC (1991): 73H05, 73H10, 73C99, 73-03

1. Introduction

Stability is a fundamental problem of solid mechanics whose solution determines the ultimate loads at which structures fail or deflect excessively. It is an old problem which has been studied for two and half centuries. The evolution of the theory was for a long time focused on elastic structures for which the loss of stability is the primary cause of failure, the material strength having little or no relevance. Throughout this century, interest gradually expanded to anelastic structures where the stability loss and material failure are intertwined due to plastic behavior and creep. During the last quarter century, the destabilizing effects of material disintegration caused by damage localization and fracture propagation have received enormous attention. Thus, in the present day view, the theory of structural stability may be seen as a very broad field which encompasses most of solid mechanics and intersects with most if its domains, while resting on the same basic concepts and utilizing the same mathematical approaches.

The present article, taking such a very broad viewpoint, will attempt a synthesizing review of the main results in entire field of stability theory as it exists today. The rich spectrum of results in the classical theory of elastic stability, including the difficult but beautiful subject of shell buckling, can be covered in this, admittedly overambitious, review only succinctly. Increased emphasis, relative to the scope or the existing results, will be placed on anelastic structures exhibiting plasticity or creep, and especially on the challenging modern problems of stability of structures losing their integrity because of distributed damage or localized fracture. The discussion will focus on these 'hot' topics in relatively much more detail than usual in stability studies.

To avoid excessive length and to make this review easily accessible to engineers and scientists from other fields desiring to acquaint themselves with structural stability theory, mathematics must be kept to the bare minimum. The differential equations as well as the derivations will be omitted. Nevertheless, to enhance understanding, mentions of the physical causes and mechanisms of various types of instabilities will not be omitted.

Aside from some selected recent contributions, only the main literature sources to the vast field of stability can be cited here. The legacy of the great Ludwig Prandtl (SOMMERFELD 1935), to whose memory the present Special Issue is dedicated, will be briefly highlighted. Extensive literature references and a detailed exposition of most of the material covered here can be found in the book by BAZANT and CEDOLIN (1991), which will simply be cited as [BC]. Excellent reviews of the more classical material, considerably more detailed and much more mathematical, were given by HUTCHINSON (1974) and BUDIANSKY and HUTCHINSON (1972), and a valuable recent review of experimental evidence is found in SINGER et al. (1998).

2. Elastic structures

2.1 Buckling of columns, frames, and arches

The primary cause of failure of slender elastic beams or frames carrying large compressive axial forces due to gravity loads is not failure of the material but attainment of Euler's (1744) first critical load, \( P_{cr1} = (\pi^2/L^2)EI \), where \( E = \) Young's modulus, \( I = \) centroidal moment of inertia of the cross section, and \( L = \) effective length representing the half-wavelength of deflection curve, which depends on the boundary conditions. For the idealized case of a perfect structure, \( P_{cr} \) represents a state of neutral equilibrium (i.e. a state at which the deflections can increase at constant load), and a state of bifurcation of the equilibrium path in the deflection space. The critical stress \( \sigma_{cr} = P_{cr}/A \)
The analysis of beam or column buckling requires that the deflections of the structure be taken into account in writing the differential equations of the beam or frame based on the theory of bending. For reasons of geometry, the problem is nonlinear at large deflections, but linearization for small deflections leads to a linear fourth-order ordinary differential equation for the deflections. The critical load is found from a linear eigenvalue problem.

Real columns inevitably have imperfections such as an initial curvature, axial load eccentricity, or small lateral loads, and often are subjected to large initial bending moments $M_0$. Consequently, the deflections increase rapidly when $P_{\alpha 1}$ is approached. Small deflections $w$ near $P_{\alpha}$ are given by Young's (1807) formula

$$w = \frac{w_0}{1 - (P/P_{\alpha 1})}$$

where constant $w_0$ characterizes the initial imperfection or initial bending moments. While this formula is exact only for initial sinusoidal curvature, it represents an asymptotic approximation near $P_{\alpha 1}$ for all kinds of initial imperfections. Therefore it has recently been adopted as a universal basis of design codes (a less simple formula that is asymptotically equivalent and is exact only for an eccentric load was used before, for steel structures). The deflection given by Young's formula is used to calculate the initial moment magnification caused by $P$ and $w$. The magnification has a factor depending on the initial moment distribution along the column, which is given in design codes by approximate formulae.

Although the basic cases of critical loads of columns with various end conditions were solved by Euler as early as 1744 (long before the theory of bending was completed by Navier), developments in column theory have nevertheless proceeded until modern times. Simple but realistic design formulae taking into account geometrical imperfections and initial bending moments in metallic and concrete columns or frames were incorporated into design codes by the middle of this century.

The curve of finite deflections of very slender elastic columns of constant cross section (a curve called the 'elastica') was solved exactly (Kirchhoff, 1859) in terms of elliptic integrals. Very large deflections of elastic columns and frames can today be easily calculated by nonlinear finite element codes.

Many extensions of the column buckling theory have been studied. The basic cases of spatial buckling of beams subjected to axial force and various kinds of application of torque are amenable to simple analytical solutions. If the column is not slender, or if it is orthotropic, with a low shear modulus (as in fiber composites), the bending theory must be enhanced by taking into account shear deformations, which can substantially decrease the value of critical load (note a later comment on the equivalence of Engesser and Haringx formulae). Consideration of the effect of pressure in compressed fluid-filled pipes, or of axial prestress introduced by embedded tendons, might at first be found tricky but is easy.

The buckling of elastic frames, i.e. assemblages of rigidly or flexibly connected beams, is normally analyzed by the stiffness method, on the basis of the stiffness matrices of beam elements. The stiffness matrix relates the increments of the bending moments and shear and axial forces at the element ends to the increments of the associated displacements and rotations. Because of the second-order moments of the initial axial force on the deflections, the stiffness coefficients depend on the axial force. If the beam has a uniform cross section, this dependence can be described by simple trigonometric, hyperbolic, and polynomial functions of the initial axial load, derived by James (1935) and Livesley and Chandler (1956) and others. Trigonometric expressions for the inverse flexibility matrix coefficients were derived earlier by von Mises and Ratzersdorfer (1926) and Chwalla (1928).

There are two possible approaches to calculating the critical loads of frames: (1) the beam is subdivided into many sufficiently short beam elements, in which case the dependence of the stiffness matrix on the initial axial force can be linearized and the critical load (or parameter of the load system) can then be obtained from a linear matrix eigenvalue problem; or (2) the stiffness matrix of the entire beam is used. In the latter approach, the number of unknowns is one or two orders of magnitude less but the matrix eigenvalue problem is nonlinear because the stiffness coefficients are highly nonlinear functions of the load. The latter approach is computationally far more efficient but requires a more complicated iterative solution based on successive linearizations of the stiffness matrix.

Designers sometimes simplify the frame buckling problem by considering a column within the frame separately, as an elastic column with flexible end restraints, isolated from the frame. However, such a trivialization may often involve large errors on the unsafe side because the dependence of the stiffness of the end restraints on the unknown critical load is neglected.

The redundant internal forces in frames can vary at the critical state while the load is constant. This means that the flexibility matrix of the primary statically determinate structure becomes singular at the critical load. However, the coefficients of this matrix typically become infinite and the matrix loses positive definiteness before the critical load is reached. This makes the flexibility method computationally unsuitable for structures with more than a few redundant internal forces.

Very large regular rectangular frames or lattices lead to linear difference equations. The basic cases can then be solved exactly by the methods of difference calculus ([BC], Sec. 2.9). The simplest modes are the internal buckling of sway and nonsway type, and the boundary buckling. Approximating the difference equations as partial differential
equations, one can solve the long-wave (global) buckling of the frame as a continuum. Because the rotations $\phi$ of the joints are independent of the rotations $\theta$ of the chords of the beams, the proper approximation (proposed in 1970, [BC]), which must be exact up to the second order terms, is the micropolar continuum [several micropolar approximations found in the literature are incorrect because terms such as $\phi_{xx}$ are missing yet contribute to the quadratic terms of energy density ($x =$ spatial coordinate)]. The critical loads of large regular frames or lattices with rectangular boundaries were thus solved in 1973 analytically ([BC]). Various types of built-up or latticed columns can be approximated as a continuous column but the shear deformation must be taken into account.

An intricate case for which a sophisticated special theory has evolved (cf. [BC], Sec. 2.8) is the buckling of high arches and slender rings (or cylindrical shells in the transverse plane). Their buckling is described by a fourth-order linear differential equation for arch deflections. Its coefficients are not constant but vary along the arch with its initial curvature. They also depend on the load, in a way more complicated than in the case of columns. The axial inextensibility of the arch, which is a justifiable simplifying assumption for high (or deep) arches, presents further restrictions. In the case of hinged arches, it excludes the odd numbered critical loads, which correspond to buckling modes with an odd number of waves [noting this restriction, HURLBRINK (1908) corrected the previously accepted Boussinesq’s solution of the lowest critical value of uniform load].

2.2 Dynamic stability analysis and chaos

Up to now, the loads were tacitly implied to be conservative. Such problems can be solved statically and do not necessitate the use of the general criterion of stability, which is dynamic. However, for structures subjected to nonconservative loads, stability may, and often is, lost in a dynamic manner. Important examples of nonconservative loads are (1) loads with prescribed time variation (e.g., pulsating loads), (2) loads generated by the flow of gases (wind loads) or liquids, and (3) reactions from jet or rocket propulsion. An idealization of (2) and (3) are the follower loads, whose orientation follows the rotation of the structure.

The fundamental definition of stability that is generally accepted is due to LIAPUNOV (1893) and may be simply stated as follows. A given solution (or, as a special case, a given equilibrium state) of a dynamic system is stable if all the possible small perturbations of the initial conditions can lead only to small changes of the solution (or response). (In the case of continuous structures, it makes sense to measure the changes in terms of an overall norm of the deflection distribution rather than locally.)

Based on Liapunov's definition, it can be shown that, under certain mild restrictions, stability can be decided by analyzing a linearized system. This, for instance, justifies considering only small strain expression in columns, plates and shells. Stable systems, when perturbed, develop vibrations, either undamped or damped. The vibration frequency depends on the applied load.

It is further easily proved that stability is lost when the frequency of natural vibrations becomes complex (Borchhardt's criterion). In conservative systems, the frequency diminishes with increasing static load and becomes zero at the limit of stability, which represents the critical state of neutral equilibrium (Fig. 1a). Stability is then lost through nonaccelerated (static) motion away from the initial equilibrium state, called the divergence (for aircraft wings) or (more generally) buckling.

![Fig. 1. a) Dependence of free vibration frequency $\omega$ on static load $P$ for conservative systems. b) Typical example for nonconservative systems. c) Strutt diagram for parametric resonance ($p =$ amplitude parameter of pulsating load; $\Omega =$ forcing frequency; $b =$ damping parameter)]
An important property of nonconservative systems is that the frequency at the loss of stability can be nonzero (Fig. 1b), in which case a static stability analysis is inapplicable. Typically, dynamic instability takes the form of vibrations of ever increasing amplitude, during which the structure moves in such a manner that it absorbs an unbounded amount of energy from the nonconservative load (such as wind). The dynamic instability, also called flutter, is an important consideration for aircraft wings, as well as for tall guyed masts, chimneys, and suspension bridges (SIMIU and SCANLAN 1986). A famous example is the aeroelastic instability that destroyed the Tacoma Narrows Bridge in 1940.

Pulsating loads, produced for example by unbalanced rotating machinery or traveling vehicles, can lead to dynamic instability in the form of parametric resonance (RAYLEIGH 1894). The structure moves in such a way that it absorbs an unbounded amount of energy from the load. This kind of resonance occurs at double the natural frequency of lateral vibrations corresponding to the static (average) value of the load (Fig. 1c). The doubling of frequency is explained by the fact that the second-order axial (or in-plane) strains due to lateral deflections of columns (or plates), on which the axial (or in-plane) forces work, have double the frequency of the lateral deflections. Due to higher-order Fourier components of axial strain history, milder parametric resonance may also occur at other integer multiples of the natural frequency. An important aspect of parametric resonance is that the structure gets stabilized by sufficient damping. In conservative systems, by contrast, the damping has no stabilizing influence.

In a rotating coordinate system, forces resulting from apparent accelerations, such as the Coriolis force and gyroscopic moments, can stabilize the structure even though they do no work on the motion of the structure. The Coriolis force, for example, causes an unbalanced rotating shaft to regain stability at supercritical rotation velocities.

An important consequence of Liapunov's definition of stability is the Lagrange-Dirichlet theorem (LAGRANGE 1788): When the total energy is continuous and all the forces are conservative or dissipative, the equilibrium is stable if the potential energy of the structure as a function of all the generalized displacements is positive definite (i.e., has a strict minimum). This greatly simplifies the analysis of conservative systems, making it possible to use static analysis, avoiding dynamic analysis of a structure with symmetry-breaking imperfections. To simplify the analysis of nonconservative systems, it is of great interest to find similar test functions, called Liapunov functions, which make it possible to decide stability without dynamic analysis. Such functions, however, have been discovered only for some special situations.

Nonlinear dynamic systems which cannot be linearized can exhibit a complex dynamic response that is nonperiodic and appears to be random (Fig. 2). Such a response, called chaos, shows nevertheless a certain degree of order and cannot be described by methods of random dynamics. Recently, great attention has been devoted to the trajectories of response of such systems in the phase space (a space whose coordinates are the displacements and velocities). A typical property of damped stable linear oscillators is that the trajectory is attracted, in several characteristic ways, to a single point, called the attractor. For a nonlinear oscillator, the trajectory appears as chaotic but on closer scrutiny is often found to be attracted to something called the strange attractor, which describes a hidden order in the response and normally has a fractal structure. Chaotic systems are inherently unstable — very small perturbations produce trajectories that exponentially diverge from the original trajectory (Fig. 2). This makes the response over longer periods of time unpredictable (THOMPSON 1982, 1989, THOMPSON and STEWART 1986, MOON 1986).

Fig. 2. Chaotic vibration of nonlinear system (divergent response change caused by a very small change in initial conditions)

2.3 Energy analysis of stability of elastic structures, postcritical behavior, and catastrophe theory

By virtue of the Lagrange-Dirichlet theorem, stability of equilibrium of elastic structures under conservative loads can be decided by checking the positive definiteness (existence of a strict minimum) of the potential energy \( \Pi \) as a function of displacement vector \( q \) near the minimum point of \( \Pi \). Aside from \( q \), \( \Pi \) also depends on various control parameters including load \( P \) and, for imperfect structures, also on the imperfection magnitude \( \alpha \) characterizing the deviation from symmetry.

The Taylor series expansion of the potential energy in terms of displacement vector \( q \) about the equilibrium state (defined as \( q = 0 \)) begins with the quadratic term called the second variation, \( \delta^2 \Pi = q^T K q / 2 \), where \( K = \partial^2 \Pi / \partial q^T \partial q \) = tangential stiffness matrix. The structure is stable if \( \delta^2 \Pi \) (or \( K \)) is positive definite. If the potential exists, \( K \) is
symmetric (and real), and so the structure is stable if and only if all the eigenvalues of $K$ are positive; or equivalently, if and only if all the principal minors of $K$ are positive (Sylvester’s criterion).

Of main interest is the lowest critical load, $P = P_{cr1}$, for which $K$ becomes singular and the quadratic form $\delta^2 \Pi$ positive semidefinite. For columns and frames (but not necessarily plates and shells), the lowest critical load is the first, for which the buckling wavelength is the longest. In conservative systems, the lowest critical load represents the limit of stability. For higher loads $P$, the quadratic form is indefinite or negative definite, which implies that conservative systems are unstable for $P > P_{cr1}$.

The behavior after the critical load is reached, called the postcritical behavior (Fig. 3), is determined by the term of the Taylor series expansion of $\Pi$ that comes next after the quadratic term. An important aspect is the postcritical imperfection sensitivity, which describes how the maximum load, $P_{max}$, is affected by the magnitude of small imperfections $\alpha$. Despite tremendous variety of structural forms, all the initial postcritical behavior can take only a few typical forms.

For many structures, $\Pi$ can be considered a function of only one deflection parameter $q$. Typical are symmetric structures which, in absence of imperfections, remain symmetric up to the lowest critical load $P_{cr}$. They possess a potential energy function $\Pi(q)$ that contains the quadratic and quartic terms but misses the cubic terms in the Taylor series expansion. Such structures exhibit postcritical behavior that is symmetric with respect to $q$, termed symmetric bifurcation. Depending on the sign of the quartic term, the critical state may be stable or unstable, which is called the stable or unstable symmetric bifurcation (Fig. 3-I, II). The former (which is typical of columns, symmetric frames, and plates) is imperfection insensitive. The latter is imperfection sensitive (which occurs in some frames; [BC], BA\'\'ANT, and Xiang 1997a); this means that the equilibrium load value of perfect structure ($\alpha = 0$) decreases with $|q|$, and that $P_{max} < P_{cr}$ for $\alpha \neq 0$.

Imperfection sensitivity, i.e. the reduction of $P_{max}$ caused by imperfection, is stronger for elastic structures that exhibit bifurcation and possess the cubic terms in $\Pi(q)$. In that case, which is called asymmetric bifurcation, the critical state is always unstable and the structure is always imperfection sensitive (Figs. 3-III, 4). Such behavior is often exhibited by asymmetric frames (the classical example being the $\Gamma$-frame; Fig. 4), and in an extreme manner by many
shells (e.g. spherical shells, or cylindrical shells under axial compression or bending (but not radial pressure or torsion)).

A famous result of elastic stability theory is that, for all types of imperfection sensitive elastic structures, the postcritical imperfection sensitivity of bifurcation buckling can be only of two types, characterized by KOITER's (1945) power laws:

\[ \Delta P_{\text{max}} / P_{\text{cr}} \propto \alpha^{2/3} \quad \text{or} \quad \alpha^{1/2} \]

where \( \Delta P_{\text{max}} = P_{\text{cr}} - P_{\text{max}} \). The former applies to unstable symmetric bifurcation, and the latter to asymmetric bifurcation. The latter is generally more dangerous because, for sufficiently small \( \alpha \), \( \Delta P_{\text{max}} \) is larger.

Consider now elastic structures that possess no cubic term in \( \Pi(q) \) and exhibit no bifurcation, due to nonexistence of symmetric deflections. In this case, called snapthrough, the deflection curve \( P(q) \) has a limit point (peak, maximum point, \( P_{\text{max}} \)), which represents the stability limit (critical state) if the load is a gravity load (dead load). After the limit point, the response under gravity load becomes dynamic; the structure 'snaps through', in accelerated motion. Such behavior is typical of flat arches or shallow (cylindrical or spherical) shells, in which the failure is caused by shortening of the arch or in-plane normal strains of the shell. The post-peak equilibrium load-deflection curve exhibits postpeak softening, which is unstable for load control (gravity load). For displacement control, the postpeak softening is stable, but only as long as the slope of the curve is negative. If the slope becomes vertical, stability is lost despite displacement control, which is called snapdown. After snapdown, the slope of the curve may become positive again but the states are unstable. Snapdown is exhibited by elastic flat arches or shallow shells loaded through a sufficiently soft spring.

In bifurcation buckling, the deflection curves of imperfect structures exhibit a limit point with snapthrough rather than bifurcation. Since some imperfection is always present, bifurcation buckling is merely an abstraction, albeit a very useful one.

When a structure has two equal or nearly equal critical loads corresponding to different buckling modes, the modes usually interact in a way causing imperfection sensitivity. This occurs in built-up (latticed) columns if the flanges or lattice members buckle locally at the same load as the column as a whole, in stiffened plate girders if the stiffening ribs buckle locally at the same load as the web, in box girders if the stiffeners buckle locally at the same load as the plates or if the plates buckle at the same load as the box, etc. Disregarding postcritical behavior, designs with coincident critical modes may seem to yield optimum weight. But they in fact represent what is called the 'naive' optimal design, which must be avoided.

The type of buckling of a system governed by a potential is qualitatively fully determined by the topological characteristics of the potential surface. The problem is analogous to instabilities encountered in various fields of physics and other sciences and is generally described by the catastrophe theory. Consider the potential \( \Pi \) as a general function of free variables \( q_1, \ldots, q_n \), and of control parameters \( \lambda_1, \ldots, \lambda_m \) corresponding to the load and the imperfections. If \( n = m = 1 \), there exists only one type of catastrophe called the fold, equivalent to snapthrough or to asymmetric bifurcation. If \( n = 1 \) and \( m \leq 2 \), there exist a second catastrophe called the cusp, equivalent to symmetric bifurcation (Fig. 5). If \( n \leq 2 \) and \( m \leq 4 \), there exist seven catastrophes, the five additional ones being called the swallowtail, butterfly, hyperbolic umbilic, elliptic umbilic, parabolic umbilic and double cusp (this remarkable result has been rigorously proven by THOM, 1975). Examples of elastic structures that exhibit each of these seven catastrophes have been given, although some have an air of artificiality. Completely general though the catastrophe theory is usually perceived to be, its present form is nevertheless inapplicable to elasto-plastic, damaging, and fracturing structures — a very important class.

The potential energy concept is useful as the basis of direct variational methods for calculating the critical loads of continuous elastic structures. The deflection field is described as \( u(x) = \sum_{i=1}^{\infty} q_i \phi_i(x) \) where \( \phi_i(x) \) are chosen as linearly independent functions of the coordinate vector \( x \) (1D, 2D, or 3D). In Ritz (or Rayleigh-Ritz) variational method, the potential energy \( \Pi \) is minimized with respect to deflection parameters \( q_i \). This yields the necessary conditions \( \partial \Pi / \partial q_i = 0 \), which represent equations for \( q_i \). If \( \Pi \) is quadratic, the equations are linear, and if they are homogeneous one has a matrix eigenvalue problem for the critical load. The solution converges for \( n \to \infty \) if \( \phi_i(x) \) form a complete system. For a finite \( n \) one has an upper bound approximation on \( P_{\text{cr}} \), which is close enough if \( n \) is large enough or if the selected functions \( \phi_i(x) \) approximate the deflection shape well. The finite element method for conservative problems can be regarded as a special case of the Ritz variational method.

Often it suffices to use only one judiciously chosen function \( \phi_1(x) \); the critical load is then approximated by the RAYLEIGH (1894) quotient

\[ P_R = U / \bar{W} \]

where \( U = \) strain-energy expression (quadratic) in terms of displacement distribution \( u \), and \( \bar{W} = \) expression for work per unit load, \( P = 1 \) (also quadratic). Analyzing \( \delta^2 P_R \), one can prove that \( P_R \) represents an upper bound on \( P_{\text{cr}} \). The Ritz method is equivalent to minimization of \( P_R \) with respect to the parameter of \( \phi_1(x) \) (considered as unknown). In mathematics, the Rayleigh quotient \( P_R \) is equivalently expressed in terms of the differential operators of the boundary value problem (provided the problem is self-adjoint, which must be the case if \( \Pi \) exists).
For statically determinate columns, an upper bound that is always closer to $P_{\alpha 1}$ than $P_R$ is given by the Timoshenko quotient $P_T = \frac{W}{U_1}$ where $U_1 = \text{complementary strain energy calculated from the second-order bending moments}$ $M$ caused by a unit load ($P = 1$) acting on the deflections $w$, expressed in terms of $w$. $P_T$ can be shown to represent nothing else but the Rayleigh quotient as defined in mathematics on the basis of the differential operators of the second-order differential equation for buckling of statically determinate columns.

From the viewpoint of design safety, it would be preferable to have lower rather than upper bounds on $P_{\alpha 1}$; they exist but, unfortunately, are not in most cases close enough to be useful ([BC], Sec. 5.8).

Based on the calculus of variations, from $\Pi$ as a functional of deflection $w(x)$, one can derive the differential equation of the problem. This approach, which is useful for more complicated problems such as thin-wall bars, also yields the boundary conditions that are compatible with the existence of a potential. They are of two kinds: kinematic (essential) or static (natural).

### 2.4 Shells, plates, thin-wall beams, and sandwiches

The design of shells and plates is dominated by stability. Shell buckling is a problem with fascinating history. After the critical loads of externally pressurized spherical shells and axially compressed cylindrical shells were calculated at the dawn of the 20th century (LORENZ 1908, TIMOSHENKO 1910, SOUTHWELL 1914), they were found to be 3 to 8 times larger than the experimental failure loads (Fig. 6b). Despite persistent efforts, the discrepancy remained unexplained until VON KÁRMÁN and TSIEI (1941) (Fig. 6a) found the answer in the extreme imperfection sensitivity of nonlinear postcritical behavior which causes the bifurcation to be asymmetric [their celebrated result was later found to fit the
The critical loads for axisymmetric buckling (Fig. 6d – bottom) represent a one-dimensional problem which can be solved easily. The two-dimensional non-axisymmetric buckling modes (Fig. 6d – top) are harder to solve. Often the problem can be simplified by considering the shell as shallow (which means that the rise of every possible buckle is small compared to the chord of the buckle arc). A famous result for such shells was DONNELL’s (1934) reduction of the critical load problem to one linearized eighth-order partial differential equation for shell deflection. A system of eight linearized partial differential equations, known as the Donnell-Mushtari-Vlasov theory, was obtained for general shallow shells. Based on the shallow shell approximation, critical loads for many modes of various shells have been solved analytically, before the finite element era.
Calculation of the postcritical behavior and failure loads of shells is a difficult problem, even with finite element programs. Although verification of an important design by finite elements is imperative, the design is generally based on critical load solutions reduced by an empirical "knock-down" factor which accounts for imperfection sensitivity and has been tabulated for various practical cases (e.g., KOLLÁR and DULÁCSKA 1984). For instance, for a cylindrical shell of radius $R$ and thickness $h$, the axial internal force at failure is thus expressed as

$$N_{xx}^* = \phi Eh^2/R\sqrt{3(1-\nu^2)}$$

(4)

where $\nu =$ Poisson’s ratio, and $\phi$ depends on $R/h$ (e.g. BUDIANSKY and HUTCHINSON 1972, HUTCHINSON 1974). For combined loadings, a linear interaction diagram may safely be assumed in design (as, e.g., between the maximum $N_{xx}$ values due to bending and to axial loading of a cylindrical shell). Because the imperfections are highly random and have a large influence, a probabilistic estimation of the failure load is appropriate (e.g. BOLOTIN 1969).

The reason for the extreme imperfection sensitivity of shells may be seen in the existence of many buckling modes with nearly equal critical stresses. For example, the critical stresses for many nonaxisymmetric buckling modes of an axially compressed cylindrical shell are only slightly lower than that for the axisymmetric pattern. In postcritical deflections, the buckle patterns may change and the final one is the diamond buckle pattern (Yoshimura pattern, Fig. 6e). When a significant lateral pressure is superimposed on axial compression, the critical loads for different modes are no longer close to each other, and such a shell ceases being imperfection sensitive.

Some nonlinear problems of shells have been approximately solved before the advent of finite elements (e.g., by variational series expansions). Many older analytical solutions, however, have now lost much of their value because of their complexity. But those that are simple nevertheless remain valuable for the understanding they convey and for use as checks on finite element codes, and are invaluable for design optimization and probabilistic modeling.

Let us now discuss the simpler problem of plates, which became well understood about two decades earlier than shells. Fourier series with Ritz variational methods were used before the middle of this century to solve the critical loads of rectangular or circular plates with various edge conditions and diverse combinations of in-plane normal and shear forces (e.g. TIMOSHENKO and GERE 1961). Like shells, plates exhibit many buckling modes. Unlike shells, the critical loads are far apart (and thus they cannot interact to cause imperfection sensitivity). In contrast to columns which are imperfection sensitive and have a large influence, a probabilistic estimation of the failure load is appropriate (e.g. BOLOTIN 1969).

In contrast, the buckling of plates is not imperfection sensitive. This was established early in this century on the basis approximate solutions of the famous von KÁRMÁN (1910)—FÖPPL (1907) nonlinear equations for the initial postcritical behavior (two coupled fourth-order partial differential equations for the deflection and the Airy stress function of in-plane stresses). If large deflections take place, plates supported along their entire boundary show a hefty postcritical reserve. Its source is the ability of a plate to redistribute the in-plane compressive forces into compressed cylindrically buckled strips along the boundaries (for compression parallel to supported edges) and into diagonal tensioned strips for shear loading. When large buckles develop and the cylindrically buckled strips carry most of the in-plane forces, the plate acts essentially as a truss. Ultimately, the buckled strips yield, which means that simple limit analysis can be applied to the truss. Such a truss analogy was exploited by von KÁRMÁN et al. (1932) to deduce stunningly simple approximate formulae for the maximum postcritical load $P_{max}$ of plates subjected to compression and shear; for simply supported rectangular plates compressed in the direction of one side,

$$P_{max} \approx kh^2 \sqrt{E\sigma_Y}$$

(5)

which, remarkably, does not depend on the plate dimensions ($h =$ plate thickness, $\sigma_Y =$ yield stress, and $k \approx$ constant).

Transverse shear deformations are unimportant in thin plates and shells. But they may be significant in composite shells, and dominate the buckling of sandwich plates and shells (e.g. PLANTEMA 1966), that is, plates that consist of stiff but thin face sheets (skins) bonded to a soft core (a foam or honeycomb). Local buckling of the skin, which provokes delamination fracture, in an additional very important mode of instability of sandwiches. In sandwich shells, the critical loads for axisymmetric buckling modes and modes with periodic buckles along the circumference are distinctly separated, which suppresses postcritical imperfection sensitivity (TENNISON and CHAN 1990).

Long thin-wall girders (e.g. metallic cold-formed profiles, concrete or welded steel girders for large bridges and buildings) represent long shells that can be approximately treated according to the theory of thin-wall beams. The first important result in this broad domain was contributed by BRANDTL (1899) in his dissertation on lateral buckling of beams of rectangular cross section subjected to bending, which he obtained in terms of Bessel functions and verified by his experiments. His pioneering work, as well as the simultaneous solution of a special case of lateral buckling by MICHELL (1899), stimulated rapid further progress.

The theory of thin-wall beams can in general be regarded as a semi-variational approach (Kantorovich variational method) in which the basic modes of deformation in the transverse directions are judiciously assumed and energy minimization then yields ordinary differential equations for the deflections and torsional rotations as well as the parameters of these modes as functions of the longitudinal coordinate. The reduction of the problem to ordinary differ-
ential equations greatly simplifies stability analysis. Beside the deformation modes described by Saint-Venant torsion theory and by the theory of bending with plane cross sections, one must consider for beams of open profile a mode in which the cross section warps out of plane. The bimoment produced by the warping mode is an important mechanism in resisting torsion (the resistance being provided by axial normal stresses characterized by the bimoment). In box girders, one must also include modes describing the bending deformation of the cross section within its own plane. One-dimensional finite elements of thin-wall beams of open or closed cross section, incorporating the out-of-plane warping mode and the in-plane cross-section deformation mode, have been formulated and used to analyze critical loads and postcritical behavior.

Solutions based on the warping torsion theory describe the important cases of lateral buckling of thin-wall beams, in which a horizontal beam subjected to bending in the vertical plane twists and bends laterally, and of axial-torsional buckling of an axially compressed column.

Limited though the studies of postcritical behavior have been, some finite element studies indicate that lateral buckling of beams is not imperfection sensitive and has a high postcritical reserve [BC, Chap. 7].

3. Anelastic and disintegrating structures

3.1 Elasto-plastic buckling and Shanley's bifurcation

Elastic structures fail either by exhausting material strength or by instability. Structures that are not elastic may, and usually do, fail by a combination of both. The evolution of material failure becomes a part of process of stability loss of the structure. Consideration of inelastic behavior and material failure is today an essential ingredient in a sound assessment of stability of structures.

A salient property of elasto-plastic behavior is the irreversibility at unloading, manifested by the fact that the unloading modulus $E_u$ is larger that the tangent modulus $E_t$ for further loading (in absence of damage, $E_u = E$). This property causes a type of behavior not seen in elastic structures: The bifurcation of equilibrium path in the load-deflection space can, and in fact does, occur at increasing, rather than constant, load $P$. This phenomenon was discovered by Shanley (1947) in a revolutionizing paper which corrected an erroneous concept lasting over half a century.

The problem has a complicated history. In two subsequent studies, Engesser (1889, 1895) proposed two different formulae for the load at which a perfect elastoplastic column begins to buckle:

$$P_t = \frac{\pi^2}{L^2} E_t I$$
$$P_r = \frac{\pi^2}{L^2} E_r I$$

where $P_t$ is called the tangent modulus load, and $P_r$ the reduced modulus load (Fig. 7a); $E_t$ is the tangent modulus of material for further loading at the stress at initial unbuckled state, and the reduced modulus $E_r$ is calculated as the effective modulus for buckling at constant load [for which $E_t$ applies on the side of neutral axis that undergoes further shortening, while $E_r$ applies on the side that undergoes extension (unloading) during deflection]. The reduced modulus theory, yielding $P_r$, was supported and refined by von Kármán (1910). It had been accepted as valid for five decades, until tests on aluminum alloys and high strength steels, exhibiting a slowly decreasing $E_t$, revealed the buckling to begin at $P_r$, which can be much smaller than $P_t$.

Based on Shanley's (1947) epoch-making discovery, which was immediately accepted by von Kármán and was later generalized to three-dimensional solids by Hill (1962) and others, the first bifurcation of the equilibrium load-deflection path (Fig. 7a) occurs when the tangential stiffness matrix $K_t$ of the structure becomes singular or, in the case of a discontinuous evolution of $E_t$ when the smallest eigenvalue of $K_t$ jumps from positive to negative (the reason will be clarified later in eq. (8)). No unloading occurs anywhere in the structure, and the analysis based on $K_t$ is often called Hill's (1962) method of linear comparison solid. This is a valuable simplification. However, the problem of bifurcation load is still nonlinear because $E_t$ depends on $P_t$, as described by the elastoplastic constitutive law.

In contrast to elastic bifurcation, the Shanley's first bifurcation state does not represent neutral equilibrium; the initial post-bifurcation states are always stable. The symmetry-breaking secondary path (i.e., lateral deflection) is nevertheless the path that actually occurs. These facts can be proven either by analysis of imperfections, or more simply and more generally by calculating the second variation $\delta^2 S$ of the entropy increment (or Helmholtz free energy increment) of a perfect structure and the differences in entropy increment between the primary and secondary paths ([BC], Sec. 10.2; Fig. 7b).

In an elastoplastic column, all the undeflected states between $P_t$ and $P_r$ represent bifurcation states (Fig. 7a). They are stable if the load is controlled, and $P_r$ is the stability limit. If the load-point displacement is controlled, the stability limit is higher than $P_r$. The maximum load, $P_{\text{max}}$, is in practice usually much closer to $P_t$ than to $P_r$ (Fig. 7a). It is for this reason that $P_t$ is so important. The design of metallic columns, frames, thin-wall beams, and shells is now generally based on the tangential modulus, and the concept is also valid for concrete ([BC], Bázant and Xiang, 1997b).

The initial post-bifurcation behavior of perfect and imperfect structures near $P_t$ is stable. In contrast to elastic structures, the deflections to the right and left are not symmetric if the cross section is nonsymmetric. Also, the $P_t$ and $P_{\text{max}}$ values for buckling to opposite sides are different.
When $E_t$ as a function of stress drops suddenly, which occurs for mild steel, $P_r$ coincides with $P_t$. In this light, it might seem strange that tests of hot-rolled steel profiles made of mild steel exhibit a large difference between $P_r$ and $P_t$. The reason is that very large residual thermal stresses get locked in after cooling (OSGOOD 1951, YANG et al. 1952). They cause the cross section parts with different residual stresses to start yielding at very different load values, thus rendering the overall force-deformation diagrams of the cross section smoothly curved. Discovery of this phenomenon resolved several previous decades of groping in efforts to explain the experiments in terms of imperfections alone. Thus Shanley's theory is important not only for alloy steels but also for hot-rolled profiles made of low carbon mild steel which passes from elastic behavior to perfectly plastic yielding almost instantly.

The buckling behavior of reinforced concrete columns is complicated by tensile cracking. Young's formula for the effect of imperfections on the deflection and moment magnification is used in design, complemented by an empirical formula for the effective modulus. The magnified moment $M$ and axial force $P$ are then compared to a semiempirical strength envelope in the $(P, M)$ plane. Consideration of perfect columns is avoided by prescribing a certain minimum load eccentricity for design.

Our discussion so far pertains to small deflection behavior. Large plastic deflections of columns are of interest mainly for predicting the energy absorption capability under blast or impact. If the yield plateau of the material is long enough, the fully plastic softening postpeak response of columns and frames can be determined relatively easily by analyzing a plastic hinge mechanism subjected to axial forces. When the structure is slender and has high initial bending moments or large imperfections, which is often the case, the transition from elastic behavior to the hinge mechanism is so rapid that it may be considered as sudden. A close upper bound for the energy absorption capability of the
structure is then provided by the area under the load-deflection curves for elastic behavior and for the plastic hinge mechanism. The intersection of these curves is an upper bound on the maximum load (provided the elastic load-deflection curve is rising up to the intersection). In the case of buckling of thin plastic plates under in-plane force, one may similarly base a simple calculation on a yield line mechanism.

Difficult three-dimensional problems are the plastic localization instabilities caused by the geometrically nonlinear effects of finite strain (RUDNICKI and RICE 1975, [BC] ch. 13). The best known example is necking (e.g. HUTCHINSON and NEALE 1978) — a narrowing of the cross section of a bar that is initially under uniform tensile strain. The narrowing causes the load to peak and then decrease at increasing displacement. Finite-strain finite element solutions (e.g., NEEDLEMAN 1982, TVERGAARD 1982) are very sensitive to the precise form of the plastic constitutive law. Geometric nonlinearity, however, is not the only cause of necking. A complete analysis needs to take into account also the damage due to the growth of voids or microcracks (which is discussed in a later section).

Analysis of plastic localization instabilities is required to understand the bursting of pipes and other shells caused by internal pressure, bending failure of tubes due to ovalization of cross section, postcritical reserve in plates, and thin-wall beams, propagating buckle in undersea pipelines, etc.

3.2 Importance of normality rule for stability and Ludwig Prandtl’s legacy

When multiaxial stresses and strains arise in buckling, a strong sensitivity to the multiaxial nature of the elasto-plastic constitutive law is encountered. A central concept in the formulation of these laws, which is partly Prandtl’s legacy, is the normality rule. The normality rule requires that, in the nine-dimensional spaces of the components of the stress tensor and the strain rate tensor, the strain rate (plastic flow) vector must be normal to the current yield surface. The first special form of the normality rule, in the context of the $J_2$ flow theory, was proposed by PRANDTL (1924) (the name came later, though). Later, the normality rule was generalized for any type of yield surface and was derived from certain plausible work inequalities (HILL 1950, DRUCKER 1951). It was recognized that adherence to this rule, originated by Prandtl, is essential (DRUCKER 1951, RICE 1971) for ensuring stability of the material, particularly for preventing spurious strain localization instabilities (to be discussed later).

Various experimentally observed phenomena such as the dilatancy of frictional slip were later found to be in apparent violation of the normality rule. One such phenomenon, the so-called vertex effect, was brought to light by GERARD and BECKER’s (1957) tests of torsional buckling of short axially compressed thin-wall cruciform columns (Fig. 7c). The bifurcation load of these columns is proportional to the incremental material stiffness ‘to-the-side’, i.e. the stiffness (in this case the shear modulus) for a stress increment vector that is tangential to the yield surface in the nine-dimensional space of stress components. According to the classical $J_2$-plasticity theory based on the normality rule, as well as all the plasticity theories with a single loading potential surface in the deviatoric stress space, the stiffness ‘to-the-side’ is equal to the initial elastic stiffness. Thus the critical load for torsional buckling of a cruciform column is predicted by these theories to be equal to the elastic critical load, regardless of the axial compressive stress. This can be grossly incorrect and far higher than the measured values (Fig. 7).

This inadequate performance of the plasticity models with a single loading potential in the deviatoric stress space, which still dominate in current computational practice, has not been taken seriously for a long time, and has almost been forgotten in recent research without ever being satisfactorily resolved. Four theories that have been used to deal with this paramount problem of plasticity may be mentioned:

1. Hencky’s deformation theory of plasticity. It predicts the stiffness to-the-side to be the secant stiffness, which can be much smaller than the initial elastic stiffness and gives a relatively good agreement with Gerard and Becker’s buckling experiments. However, Hencky’s is not a fully consistent theory suitable for general finite element programs.
2. Enhancement of the plastic potential surface with a vertex traveling with the current stress point (RICE 1975, RUDNICKI and RICE 1975, HUTCHINSON and NEALE 1978, HUTCHINSON and TVERGAARD 1980, BAZANT 1980; [BC], sec 10.7). This approach, however, does not seem effective for general finite element programs.
3. Multisurface plasticity, proposed by KOTTER (1953), which describes the actual physical source of the vertex effect. The effect is caused by intersection of several loading potential surfaces at the current stress point. When the direction of the load path changes suddenly, some yield surfaces start to unload while others get activated. The important advantage is that the vertex effect can be modeled even while the normality rule, crucially important for soundness of constitutive laws (RICE 1971), can be satisfied separately for the individual surfaces. The multisurface constitutive model, however, is hard to identify from test data and has not been developed for large-scale computations.
4. Taylor-type crystallographic models (TAYLOR 1938), worked out in detail by BATDORF and BUDIANSKY (1949) and refined by others (e.g. RICE 1971, BRONKHORST et al. 1992, BUTLER and McDOWELL 1998), and the related microplane model (BAZANT 1984, BAZANT and PLANAS 1998, ch. 14, CAROL and BAZANT 1998). These constitutive models are defined not in terms of tensorial invariants but in terms of the stress and strain surfaces on surfaces of various orientations in the material (called the microplane), and either the stress components (for Taylor models) or the strain components (for microplane models) are assumed to be the projections of the stress or strain tensor. These are essentially multisurface plasticity models, though not expressed in terms of tensors and their invariants. They can satisfy the normality rules on all the crystallographic planes or the microplanes separately, yet
exhibit the vertex behavior automatically, by describing its physical source directly. The microplane model, with efficient numerical integration over all spatial orientations of microplanes (BAZANT et al. 2000), has been developed for large-scale computation of concrete and metals and has been shown (BROCCA and BAZANT 1999) to automatically reproduce the buckling data of GERARD and BECKER (1957); Fig. 7d. In a generalized sense, these advanced models adhere to Prandtl's original idea.

3.3 Viscoelastic and viscoplastic buckling

Viscoelastic or viscoplastic behavior dissipates energy, which does not destabilize a structure. Thus, according to the Lagrange-Dirichlet theorem, the stability limit of a viscoelastic structure must still be decided by the loss of positivity of the potential for the elastic part of response. However, the stability limit is often not the issue of main practical interest.

The differential or integral equations that govern viscoelastic buckling can be obtained from those that govern elastic buckling by replacing the elastic constants with the corresponding viscoelastic operator. This operator can be of a differential type, based on the Maxwell of Kelvin chain rheological model, or of an integral type, corresponding to a continuous relaxation or retardation spectrum. If there is no aging, Laplace transform can be used to reduce the problem to elastic buckling, and inversion of the Laplace transform then yields the time evolution of buckling deflections. For viscoelastic materials that are solids (i.e. do not possess a purely viscous response), a structure loaded to the stability limit takes an infinite time to develop a finite deflection as a result of an infinitesimal disturbance or imperfection (e.g. FREUDENTHAL 1950, HILTON 1952). Therefore the long-time critical load, \( P_{\infty} \), which is obtained by replacing the instantaneous elastic modulus \( E \) in the critical load solution with the long-time elastic modulus \( E_{\infty} \), is normally unimportant. Important for design is the time to reach, for given imperfections, the maximum tolerable deflection or second-order stress due to buckling. This time must not be less than the required design lifetime.

Viscoplastic buckling is different. In contrast to viscoelastic structures, there exists a finite critical time \( t^* \) at which the deflection triggered by an infinitely small imperfection becomes finite (or that triggered by a finite imperfection becomes infinite according to the geometrically linearized theory); e.g. HOFF (1958). The higher the load, the smaller is \( t^* \). If the viscoplastic material does not have a finite elastic limit (e.g., if the viscoplastic strain rate is a power function of stress magnitude), the critical load according to Liapunov's stability definition is zero. The basis of design is to ensure that \( t^* \) exceeds the required lifetime.

Long-time buckling of concrete structures is a very complicated but important phenomenon which has caused some slender columns and shells to collapse after many years of service. In the low (service) stress range, concrete is viscoelastic but exhibits aging, caused by chemical processes of hydration and by relaxation of microprestress induced by drying and chemical changes — processes going on for many years. The consequence is that the Volterra integral equation for strain history in terms of stress history has a nonconvolution kernel, and the discrete Kelvin or Maxwell chain model has age-dependent viscosities and elastic moduli. The aging makes the Laplace transform methods ineffective even in the viscoelastic range, but numerical time integration of buckling response is easy. Simultaneous drying intensifies creep, and so the problem is coupled with diffusion processes. Furthermore, concrete undergoes cracking during buckling. This and the creep cause a gradual stress transfer from concrete to steel reinforcement. No wonder that crude empirical design procedures are still in use.

Detailed finite element solutions that agree reasonably well with tests have nevertheless been achieved. The design of concrete columns relies on an empirical overconservative reduction of the effective elastic modulus, roughly reflecting creep, aging, and cracking (as well as Shanley effect). For creep with aging, most effective is the use of the age-adjusted effective modulus method, which is based on a theorem stating that if the strain history is linearly dependent on the compliance function, then the stress history is linearly dependent on the corresponding relaxation function ([BC]).

Since the prediction of lifetime is rather uncertain, a statistical approach is appropriate.

3.4 Thermodynamics of structures, inelastic stability, bifurcation, and friction

So far our review of inelastic structures has not dealt with the problem of stability. Stability cannot, in principle, be based on elastic potential energy. This does not mean, however, that stability would have to be analyzed dynamically, according to Liapunov’s stability criterion. An energy approach to stability, which is much simpler, can be based on thermodynamics of structures.

According to Gibbs's form of the second law of thermodynamics, a thermodynamic system is stable if the increment of its entropy calculated for every admissible infinitesimal change of its state is negative. If this increment is zero for some change, the system is critical, and if it is positive for some change, the change must happen and so the initial equilibrium state of the system is unstable. This is a general thermodynamic definition of stability of equilibrium, which is equivalent to Liapunov’s definition (except when dealing with stability of motion). Its beauty is that the motion caused by initial imperfections need not be analyzed. It suffices to analyze the perfect structure, which is much simpler, especially if the structure is inelastic.

An inelastic structure is normally far from a state of thermodynamic equilibrium, at which all the dissipative processes would come to a standstill. In principle, this prohibits using classical thermodynamics, which deals only with
states infinitely close to thermodynamic equilibrium. However, the use of irreversible thermodynamics would cause great complication. It can fortunately be circumvented by the hypothesis of a tangentially equivalent inelastic structure. This is a structure characterized by the tangential moduli, assumed to behave equivalently to the actual inelastic structure for small changes of state. The existence of such a structure is of course tacitly implied whenever the load increment in a computer program is analyzed on the basis of the tangent elastic modulus. At each stage of loading, there are many tangentially equivalent elastic structures corresponding to various possible combinations of tangential moduli for loading and unloading in various parts of the structures. In theory, they all need to be considered.

Deformations of elastic structures are reversible changes, and thus they do not change the entropy. However, the thermodynamic system must include the load, whose changes are defined independently of the equilibrium of the structure. For example, if the structure is in equilibrium under gravity loads \( P \) at initial deflections \( q_0 \), and its deflections \( q \) are then changed by \( dq \) away from equilibrium, the equilibrium values of the reactions \( f(q) \) change but \( P \) does not. The disequilibrium creates entropy change \( \Delta S = \frac{1}{T} \delta^2 W \). Substituting \( f(q) \approx P + K_1 (q - q_0) \) and integrating from \( q_0 \) to \( q_0 + \delta q \), one has

\[
-T \Delta S = \delta^2 W = \frac{1}{2} \delta q^T K_1 \delta q \tag{7}
\]

up to the second-order terms; \( T = \) absolute temperature; \( \delta^2 W = \) second-order work of the reactions, i.e. the equilibrium load values (in the rare case that \( \delta^2 W = 0 \), the fourth variation needs to be analyzed similarly, which we do not discuss here). If the loads vary, the second-order work of loads must be added to the right-hand side of (7). It can be shown that under isothermal conditions, \( \delta^2 W = \delta^2 F = \) second variation of Helmholtz free energy (or total energy) of the structure-load system under isothermal (or adiabatic) conditions. These energies represent the potential energies of the tangentially equivalent elastic structure expressed in terms of the isothermal (or isentropic) tangential moduli of the material.

The difference from elastic stability is that there are many potential energy expressions to consider, each of them valid in a different sector of the space of \( q \), corresponding to different possible combinations of loading and unloading in various parts of the structure. The surfaces of second variation of entropy \( S \) (the negative of potential surfaces) in different sectors are quadratic surfaces, which are joined continuously and with a continuous slope across the sector boundaries (Fig. 7b). The eigenvector for a sectorial quadratic surface may lie inside or outside the corresponding sector. Only if it lies inside, the loss of positive definiteness of the potential surface (or negative definiteness of the entropy surface) in that sector is real and represents stability loss (e.g., in an elasto-plastic column under gravity load, such a situation is first encountered at \( P_t \), not \( P_l \)).

For the basic case of structures with a single load (or load parameter) \( P \) and the associated displacement \( q \), the thermodynamic analysis shows that elastic as well as inelastic structures are stable as long as \( dP/dq > 0 \); the stability limit is reached at the maximum (peak) load, and the postpeak states for which \( dP/dq < 0 \) are unstable. Under displacement control, the postpeak states are stable unless the slope of the descending \( P(q) \) curve becomes vertical, i.e., snapback occurs; if the slope becomes positive, the structure becomes unstable (a stable response can nevertheless be obtained by controlling some internal displacements). A structure that exhibits postpeak softening (\( dP/dq < 0 \)) will exhibit a snapback when it is loaded through a sufficiently soft spring. The snapback instability is reached when \( dP/dq = -C \) where \( C = \) spring stiffness. For elastic structures, postpeak softening can be caused only by buckling. Otherwise, it can also be caused by fracture or damage.

Thermodynamic analysis of bifurcation requires decomposing the infinitesimal loading step into two substeps. The first is a change away from equilibrium, in which the controlled loads or controlled displacements are changed without any change in the stress and deformation of the structure. The second substep is a restoration of equilibrium at constant controls. The equilibrium state approached in the second substep may lie either on the primary (symmetry preserving) or on the secondary (symmetry-breaking) path, depending on which approach to equilibrium produces a larger entropy increment. In this manner one can prove in general, without considering imperfections, that the secondary path must occur. It is called the stable path. This is a different concept than that of stability of equilibrium (for inelastic structures, the postbifurcation equilibrium states on both the primary and the secondary paths may be stable, but only the secondary path is stable).

At small enough loads, the eigenvector for the quadratic entropy surface corresponding to the secondary path lies outside the sector of the displacement space in which this surface is valid. As the load increases, the eigenvector is turning, and at the first bifurcation load this eigenvector first moves into the sector of validity of this surface. At that moment, the incremental equilibrium equations \( K_1 \delta q^{(1)} = \delta f \) and \( K_2 \delta q^{(2)} = \delta f \) for paths (1) and (2) must be valid simultaneously for the same load increment \( \delta f \). So, by subtraction, one gets a homogeneous linear matrix equation for the difference of displacement increments:

\[
K_1 (\delta q^{(2)} - \delta q^{(1)}) = 0. \tag{8}
\]

Hence, the first bifurcation is indicated by singularity of the tangential stiffness matrix \( K_1 \) ([BC] ch. 10; HILL 1958, MAIER et al. 1973, PETRYK 1985, NGUYEN 1987). This matrix corresponds to loading at all points of the structures. Before the first bifurcation, the eigenvector of \( K_1 \) is inadmissible because, in the displacement space, it lies outside the sector for which no point of the structure undergoes unloading. The orientation of this eigenvector rotates during
loading and, at the first bifurcation, it lies at the boundary of this sector, thus making bifurcated path (2) admissible for the first time. After the first bifurcation, the eigenvector for the states on path (1) rotates inside this sector, and so a continuous series of bifurcation states lies on path (1).

The first bifurcation for a structure in which the tangential moduli are changing continuously can be determined by linear analysis of a solid without unloading [known as Hill’s (1962) linear comparison solid], in which the tangent modulus for loading applies everywhere. If the tangential moduli change by a jump, then the first bifurcation occurs when the smallest eigenvalue of \( K_i \) jumps to from a positive to a negative value.

The use of the initial elastic stiffness matrix in load step iterations in finite element programs may be deceptive. Convergence is lost if stability is lost, but not necessarily if the first bifurcation state on the primary loading path has been passed. Thus the iterations based on the initial elastic stiffness may, deceptively, converge for stable states lying on the primary path even if the first bifurcation state has been missed. Thus, if there is any danger of bifurcation, it is important to calculate the tangential stiffness matrix and check its positive definiteness. Imperfections in the sense of the eigenvector of \( K_i \) must be introduced to ensure that the bifurcation response be triggered (e.g. de Borst 1987).

Phenomena such as friction or damage may cause the tangential stiffness matrix \( K \) to be nonsymmetric. If \( K \) is symmetric, positive definiteness is lost when \( \det K = 0 \), which means that the stability loss under load control (gravity load) is characterized by neutral equilibrium. However, if \( K \) is nonsymmetric, the thermodynamic condition of stability limit (critical state), which reads \( \delta q^T K \delta q = 0 \), can be satisfied not only when \( \det K = 0 \) (i.e., when there is neutral equilibrium, \( K \delta q = 0 \)) but also when the vector \( K \delta q = \delta f \) is orthogonal to \( \delta q \), i.e., when \( \delta q^T \delta f = 0 \). This means that, for nonsymmetric \( K \), displacements with non-zero load changes may occur at zero work. Stability is decided by the symmetric part \( \bar{K} \) of matrix \( K \), whereas bifurcation or neutral equilibrium is decided by the singularity of the nonsymmetric matrix \( K \). According to Bromwich theorem, the smallest eigenvalue of the symmetric \( \bar{K} \) is less than or equal to the smallest real part of the eigenvalues of the nonsymmetric \( K \). This means that, in the case of friction or symmetry-breaking damage, positiveness of the real parts of the eigenvalues of the tangential stiffness matrix does not guarantee stability. Stability may be lost before the loading process leads to a bifurcation or a neutral equilibrium state \( (\delta f = 0 \text{ at nonzero } \delta q) \).

An inelastic structure may also be destabilized by load cycles. From the energetic viewpoint, stability in the standard sense (i.e. for one-way deviations from the equilibrium state) does not imply stability for infinitesimal deviation cycles and, vice versa, the latter does not imply the former. If infinitesimal strain cycles of a small material element increase entropy (i.e. produce negative second-order work), the material is locally unstable, which may (although need not) destabilize the structure. Drucker’s (1951) postulate [equivalent to Hill’s (1950) principle of maximum plastic work] prohibits such constitutive laws. This provides useful restrictions such as convexity of the yield surface and the normality rule for the strain rate vector in the nine-dimensional strain space. It also prevents the constitutive law from causing instabilities such as strain localization.

In the case of internal friction, however, the normality rule needs to be in some way relaxed. But then the flow rule, called non-associated, typically produces localization instabilities (such as shear bands, cracking bands) which may or might not be real. It appears that the need for a non-associated flow rule may often arise from adopting a single yield surface where in reality multiple yield surfaces intersecting at the current state point of the nine-dimensional strain space should be considered. The normal strain-rate vectors for all these surfaces may get superposed in a way that makes the resultant strain-rate vector appear not to be normal to the single yield surface defined by the yield limits for radial loadings. Such behavior is captured by the slip theory of plasticity of Taylor (1938) and Batdorf and Budiansky (1949), as well as by later generalization and modification in the form of the microplane constitutive model, which implies the existence of as many loading surfaces as there are microplanes ([BC], Carol and Bažant 1998).

By extensions of Mandel’s (1964) stability analysis of an elastically restrained frictionally sliding block, it has been proven that, in frictional materials, a nonassociated flow rule does not cause instability if the strain-rate vector lies within a certain fan of directions. The fan is bounded by the normal to the yield surface and the normal to the surface of the so-called frictionally-blocked second-order elastic energy density. Simple formulations are available for two cases: (1) direct internal friction, representing the effect of hydrostatic pressure on deviatoric yielding, and (2) so-called inverse friction, representing the effect of hydrostatic stress invariant on the inelastic volume change ([BC], Sec. 10.7).

### 3.5 Finite-strain aspects of stability of three-dimensional bodies

To determine the critical state, the potential energy must be expressed accurately up the quadratic terms. For the strain energy, this condition is satisfied if the small (linearized) strain expressions are used. But the potential energy also includes the work of the initial stresses \( S_{ij} \) which are finite. Therefore, the strains on which \( S_{ij} \) work must be expressed accurately up to their second-order terms (the subscripts label the Cartesian coordinates \( x_i \), \( i = 1, 2, 3 \)). Therefore, a finite strain expression that is correct up to the second-order terms in the displacement gradient must be used. There are, however, infinitely many finite strain tensors that can be chosen as the strain measure. A very general second-order finite strain expression is

\[
\varepsilon_{ij} = \varepsilon_{ij} - \frac{m}{2} \epsilon_{ik} \epsilon_{kj} \quad (9)
\]
where \( e_{jk} \) is the small (linearized) strain tensor, and \( m \) is an arbitrary real number; \( m = 2 \) gives Green's Lagrangian strain tensor, while \( m = 1 \) and \( m = 0 \) gives the second-order approximations to Biot's (1965) strain tensor and Hencky's (logarithmic) strain tensor, respectively.

For beams, plates, and shells, discussed so far, the choice of \( m \) makes no difference because the second-order work of the initial axial or in-plane stresses depends only on the rotations of the cross sections or normals. For massive bodies, however, the choice of \( m \) does make a difference. Historically, beginning with Southwell (1914) and Biezeno and Hencky (1928), various stability formulations corresponding \( m = 2, 1, 0, -1, -2 \) have been proposed. The differential equations of equilibrium for deviations from the initial state, the quadratic variational principles, and the critical load solutions for these formulations seemed to differ and be in conflict. This generated an unnecessary long-running controversy. As it turned out, however, all these formulations are equivalent because the tangential moduli tensors \( C^{(m)}_{ijkl} \) are not the same for different \( m \); according to the requirement of equality of second variation of work, they are related as

\[
C^{(m)}_{ijkl} = C_{ijkl} + \frac{2 - m}{4} \left( S_{ik}\delta_{jm} + S_{jk}\delta_{im} + S_{lm}\delta_{jk} + S_{jm}\delta_{ik} \right)
\]

where \( C_{ijkl} \) are the tangential moduli associated with Green's Lagrangian strain tensor \((m = 2)\), and \( \delta_{ij} \) is Kronecker's delta (Bážant 1971; [BC], Sec. 11.4). (In some formulations, the objective stress increments did not correspond to the same \( m \) as the moduli, but this is incorrect.)

Three-dimensional buckling instabilities are important only if some of the principal values of the tangential elastic moduli tensor is of the same order of magnitude as the initial stress. This can occur for (1) highly anisotropic materials such as fiber composites, (2) composite structures with very soft components such as the core of sandwich plates and shells, (3) continuum approximations of lattice structures and springs, and (4) materials that undergo a drastic reduction of tangential stiffness due to plasticity or damage and develop shear bands, cracking bands, or crushing bands. Simple analytical solutions have been obtained for periodic internal buckling of a compressed massive orthotropic body, periodic surface buckling of a compressed orthotropic halfspace, and bulging or buckling of a compressed thick orthotropic rectangular specimen (Fig. 8), as well as for fiber buckling in composites and shear buckling of thick columns, plates, and helical springs.

Shear buckling is important not only for sandwiches and fiber composites, but also for springs and for built-up (laced or battened) columns treated as continuous. The reason is that the spring pitches, or the lattices and battens, are often too soft to prevent large shear strains. There has been a long-lasting (but unnecessary) controversy between two different formulae for critical stress, one derived by Engesser (1889) for built-up columns, and another derived by Haringx (1942) for helical springs:

\[
\sigma_{\alpha} = \sigma_{E} / (1 + \sigma_{E}/G) \quad \text{and} \quad \sigma_{\alpha} = \bar{G}(\sqrt{1 + 4\bar{G}G - 1})/2.
\]

Here \( \sigma_{E} \) is Euler's critical stress of a slender column buckling without shear, and \( G \) is the effective shear modulus. It has been shown ([BC], Sec. 11.6), however, that the \( E \) values, as well as the \( G \) values, in these two formulae cannot be the same. They have different physical meanings and must be measured or calculated differently. \( \sigma_{E} \) and \( G \) are based on the tangential moduli \( C^{(m)}_{ijkl} \) associated with Green's Lagrangian strain tensor \((m = 2)\), while \( \bar{G} \) and \( \bar{E} \) are based on \( C^{(m)}_{ijkl} \) associated with the strain tensor for \( m = -2 \). According to (11), \( \bar{E} = E + 4\sigma_{\alpha} \), \( \bar{G} = G + \sigma_{\alpha} \). Eqs. (11) further made it possible to resolve similar long-lasting disagreements among various critical load solutions for the critical loads for internal buckling, surface buckling and bulging.

Engesser's (1889) formula can also be used for a continuous approximation of latticed (built-up) columns in which shear deformations are normally significant. This formula has a tragic history. As shown by Prandtl (1907) and others, ignorance of this formula by the design code of the time was the main culprit in the collapse in 1907 of the record-span Quebec bridge over St. Lawrence, with a great loss of life. It is one of the examples of a disaster caused by reliance on an outdated design code which lulls the designer to a false sense of security.

The compression strength of fiber composites is often limited by localized internal buckling. For the critical compressive stress of internal buckling in polymer composites reinforced by fabrics, simple expressions depending on
the shear modulus of the matrix and amplitude of fiber undulation in the weave were derived in 1968 ([BC], Sec. 11.9).

For unidirectional fiber reinforcement, microbuckling of fibers is triggered by fiber misalignment and localizes in a transverse kink band (ROSEN 1965, BUDIANSKY 1983, BUDIANSKY et al. 1997, FLECK 1997). Since the kink is accompanied by microcracks along fibers and causes a drop of axial stress, the propagation of kink band should be treated by methods of fracture mechanics, which is a subject of current research (BAZANT et al. 1999).

3.6 Stability problems of fracture mechanics

Plasticity and hardening damage of the material profoundly affects the stability of structures but is not per se, the cause of instability. The cause lies solely in the nonlinear geometric effect of deformations. On the other hand, fracture and softening damage (such as microcracking or plastic micro-void growth) can destabilize a structure, even in absence of the nonlinear geometric effect.

The energy balance condition of crack propagation condition is $G = R(c)$, where $G$ = energy release rate, and $R(c) = R$-curve = critical value of $G$ depending in general on the crack extension $c$. This condition plays the role of an equilibrium condition — the crack can grow statically. For $G > R(c)$ the growth is dynamic, and for $G < R(c)$ no growth is possible. Under load control, the limit of stability of a structure with a statically growing crack is reached at maximum load, which occurs when $\left[ \frac{\partial G}{\partial c} \right]_p = \frac{\partial R(c)}{\partial c}$. Replacing $=$ with $>$ yields the stability condition. The stability limit for loading through a spring, which characterizes the ductility of a structure, can be determined by calculat-

![Fig. 9. a) Double-edge cracked specimen — diagram of average stress versus displacement, with bifurcations (solid curve: one crack growing, dashed: both growing). b) Longitudinal localization of strain softening in a tensioned bar. c) Corresponding surfaces of incremental potential $\delta^2 \Psi$ (= minus second-order entropy increment) showing bifurcations and instabilities (02 or 05) under controlled axial displacement](image)
ing the load-deflection curve, which can be done on the basis of a solution of the stress intensity factor. Under displace
ment control, fracture growth becomes unstable when the post-peak descending load-deflection curve reaches a
snapback. Some structure geometries never exhibit a snapback instability (unless loaded through a soft enough spring),
while others do. The former is the case, e.g. for notched three-point bend beams. The latter is the case whenever the
ligament is subjected to a normal force (which is called ligament tearing, occurring, e.g. in notched tensile strip speci
mens).

The R-curve is the simplest (and crudest) way to take into account the finiteness of the fracture process zone at
the crack tip. This property, which is more accurately described by the cohesive crack model or the crack band model,
gives rise to a deterministic size effect on the nominal strength of geometrically similar structures with geometrically
similar failure modes. By measuring the size effect in notched specimens, one can deduce the R-curve and the main
parameters of the cohesive crack model.

Simultaneous growth of many cracks (of length $a_i$, $i = 1, 2, \ldots$) typically produces bifurcation as well as stability
loss (Fig. 11a). The incremental potential $\delta^2 F$ (minus the second-order entropy, or free energy) is characterized by a
matrix that involves the partial derivatives $\partial K_i/\partial a_j$ where $K_i$ is the stress intensity factor of crack $i$. The potential
surface is a patch-up of many quadratic sectors joined continuously and smoothly (Fig. 11b). In multi-crack systems,
the advance of one crack tip often causes unloading of the nearby crack tip, with the result that only one of the cracks
can grow. This for instance happens for symmetrically edge-cracked and center-cracked tensioned strip specimens,
in which there are two interacting crack tips.

Another example is the system of parallel equidistant cracks in an elastic halfplane, driven by cooling or drying
shrinkage (Fig. 10a; [BC] Sec. 11.2). As the cooling front advances into the halfplane, the cracks first grow while keep-
ing equal lengths (primary path). At a certain moment (point A in Fig. 10b), a stable bifurcation of the equilibrium
path is reached. For the bifurcated (secondary) path, which is the stable path (i.e. must occur), every other crack stops
growing (Fig. 10a) and gradually closes, while the growing cracks gradually open wider. If the cracks are somehow kept
equally long after the bifurcation, a stability limit (B in Fig. 10b) is eventually reached. The behavior is analogous to
elasto-plastic columns. The parallel crack system, however, may be stable if the cooling temperature profile is rather
flat and has a sufficiently steep front, or if the halfplane is reinforced by bars near the surface.

Similar bifurcations and instabilities are exhibited by parallel cracks in the tensile zone of beams subjected to
bending. Such bifurcations determine the spacing of deep cracks and the ductility of the structure. The initial crack
spacing can be determined (Li et al. 1995) by considering a sudden formation of cracks of length $a_0$ and stipulating
three conditions: (1) The material tensile strength is reached before the cracks appear, (2) the energy released during
the finite crack jump is equal to the total energy dissipated by the initial cracks of length $a_0$, and (3) the energy release
rate at length $a_0$ is critical. The bifurcations in a system of parallel cracks are important because they decide the width
of cracks, which controls the overall effective permeability and the rate of ingress of corrosive agents into structures.

In three-dimensions, an approximate analysis of the entropy increment confirms that a hexagonal pattern of
cooling or drying cracks on a halfspace surface (as seen, e.g., on a drying lake bed or a cooling lava flow) should be
favored over triangular and square patterns as well as parallel planar cracks. However, accurate solutions of the bifur-
cation and stability of three-dimensional cracks seem unavailable.

### 3.7 Damage, localization instabilities, and size effect

Continuum damage mechanics provides a smeared description of the growth of microcracks in brittle materials or voids
in plastic materials. Of main interest is the damage that causes strain-softening — a phenomenon manifested by a
negative slope of the stress-strain curve and generally by a loss of positive definiteness of the matrix of tangential
moduli. Such a material behavior violates Drucker’s postulate and leads to localization instabilities and bifurcations.
Formation of softening hinges in beams and frames causes also localization instabilities and bifurcations (Maier et al. 1973), not only in static but also in dynamic (seismic) loading (Bazant and Jirasek 1996), and so does softening in friction (particularly the frictional stress drop from static friction to dynamic friction); Rice and Ruina (1982).

Although strain-softening stress-strain relations have been used in concrete engineering since the 1950’s and in finite element analysis since 1968 (Bazant 1986), until about 1985 most mechanicians regarded all strain-softening studies with contempt (research into this ‘dubious’ concept was in some totalitarian countries even banned; curiously, the continuum damage mechanics, which exhibits, without a characteristic length, the same localization instabilities as any strain-softening model, nevertheless escaped condemnation, perhaps because the strain-softening was disguised as a separate damage variable while the ‘true’ stress exhibited only hardening). Today, the strain-softening damage is a well-established and indispensable concept — but of course only within the context of some nonlocal material model possessing a characteristic length.

In absence of a characteristic length, one objectionable property of strain softening is that if the elastic moduli matrix ceases being positive definite, the material cannot propagate waves and the dynamic initial-boundary value problem changes its type from hyperbolic to elliptic (Hadamard 1903). Since the unloading modulus is always positive (as experimentally discovered for concrete in the 1960’s), the material can nevertheless propagate unloading waves.

That the concept is not mathematically meaningless was demonstrated by analyzing step waves in a rod caused by suddenly starting to move the ends inward at constant velocity. After the waves meet, two kinds of responses can occur: (1) If the strain front magnitude is less than one half of the strength limit (onset of strain-softening), the waves get superimposed and the stress front doubles; but otherwise (2) the stress drops instantly to zero, the bar splits, and unloading waves emanate from the split. The dynamic problem is ill-posed (and the solution is unstable) because an infinitely small change in the velocity imposed at the ends of the rod can cause a finite change in the response. The solution nevertheless exists. But there is a problem: The splitting of the bar occurs with zero energy dissipation, which makes the strain-softening concept physically unacceptable. Similar unstable and physically unrealistic solutions have been mathematically demonstrated for waves in a rod in which the strain-softening is followed by rehardening, and for radically converging waves in a strain-softening sphere.

Rudnicki and Rice (1975) and Rice (1976) pioneered analysis of static localization of plastic (non-softening) strain into a planar band (e.g. a shear band) in an infinite body. Such localization is caused by the nonlinear geometric effect of finite strain and occurs very near the yield stress value. If strain-softening takes place, its destabilizing effect is normally much stronger than that of geometric nonlinearity. Bifurcation of the localization type is decided merely by the constitutive law; it occurs when the so-called acoustic tensor $A = n \cdot C^f \cdot n$ becomes singular (where $C^f =$ fourth-order tangential moduli tensor, and $n =$ unit normal of the localization band). Various damage constitutive laws have been analyzed with regard to the singularity of $A$ (e.g. Willam and Carol).

Localization can also be triggered by a lack of normality of plastic flow (de Borst 1988, Leroy and Ortiz 1989). In the case of infinite space, this condition represents the stability limit as well. But for an infinite planar localization band of finite thickness $h$ located within an infinite layer of finite thickness $L > h$, the stability limit occurs later and is characterized by singularity of the tensor

$$Z = n \cdot [C^f + C^s h/(L - h)] \cdot n$$

where $C^s =$ material stiffness tensor for unloading. For $L/h \rightarrow \infty$, the condition for the acoustic tensor is recovered.

When a localization band much longer than its thickness cannot develop, because of a finite size of the body, a better, analytically solvable, model is damage localization into an ellipsoidal region. This has been solved with the help of Eshelby’s theorem giving the strain of an ellipsoidal plug that has been fitted into a different ellipsoidal hole in an infinite elastic body ([BC], sec. 13.4). The bifurcations and instabilities are found to depend on the aspect ratios of the ellipsoid and occur later in the loading process than they do for an infinite band.

The usefulness of the bifurcation studies of constitutive laws, however, is rather limited — they reveal only the onset of localization, while often it is much more important to determine the postbifurcation behavior of a body with a damage localization region. In this regard, the concept of strain-softening (or continuum damage) requires a characteristic length (or material length) in order to overcome certain fundamental difficulties.

In a strain-softening bar, a uniform axial strain cannot occur. The strain localizes into a band (fracture process zone) of the smallest possible length. Consequently, the longer the bar, the steeper is the post-peak curve of stress versus average strain (Fig. 9b). A nonlocalized strain distribution loses stability when the softening load-deflection slope reaches a certain magnitude (Fig. 9a) depending on the structure size and the stiffness of the loading device (the stability limit is the only objective and sound characteristic of the ductility of a structure).

An analytical solution of a tensioned bar subdivided into finite elements exhibits bifurcations. Their consequence is that the strain-softening must always localize into a single finite element, no matter how small it is. Numerical finite element computations of strain-softening damage under static loading may converge, but to a wrong solution. They reveal spurious mesh sensitivity — the postpeak load-deflection curves of structures with strain softening are unobjective because very different peak loads and post-peak responses are obtained for different element sizes. The energy dissipated by breaking the structure converges to zero as the finite element size is reduced to zero. These are physically unrealistic features.
The remedy consists in introducing some form of a nonlocal continuum model possessing a characteristic length, of which there are basically three types -- (1) the crack band model, (2) the integral-type model, and (3) the second-gradient type model. The first and third can be regarded as approximations to the second.

The crack band model, proposed in 1976 ([BC], ch. 13; BAŽANT and PLANAS 1998), is the crudest but simplest (and thus preferred by concrete design firms). Its salient characteristic is to impose a certain minimum size $h$ of the finite element. This size must be regarded as a material property, representing a characteristic length $\ell$ of the material; $\ell$ corresponds to the width of the crack band (or strain softening damage band). In the case that localization of strain-softening damage is not prevented by some restraints such as reinforcing bars, the crack band model is essentially equivalent to the cohesive (or fictitious) crack model, in which the crack opening is equal to the cracking strain accumulated over the width of the crack band. Various semiempirical energy-based numerical corrections need to be used when the band does not propagate along the mesh lines (ČERVENKA and ČERVENKA 1998).

In the nonlocal continuum concept, which was introduced for elasticity by ERingen (1965), the stress at a given point depends not only on the strain at that point but also on a certain average of the strain field in a neighborhood of the point, whose effective size corresponds to the characteristic length $\ell$ of the continuum. The average is weighted by a bell-shaped weight function. When the nonlocal concept was introduced in 1984 for the purpose of limiting the localization of strain-softening, the total strain was averaged over the neighborhood. However, this caused two problems: the numerical implementation was cumbersome (requiring imbrication of finite elements), and there existed zero-energy periodic modes of instability which had to be suppressed.

The problems of the total strain averaging were overcome in 1987 by the nonlocal damage model, in which the spatial averaging is done on the damage variable or the inelastic part of strain. With this concept, localization of strain to form a displacement discontinuity line is impossible [reason: if the strain profile became a function with $C_0$ continuity (Dirac delta function), the averaging integral would convert the profile in the next iteration to a function with $C_1$ continuity (Heaviside step function)]. The nonlocality of damage can be physically justified by the smoothing of an array of discrete cracks and by crack interactions (the latter, however, also points to a more complicated nonlocal model which distinguishes the directions of amplifying and shielding crack interactions; BAŽANT 1994, BAŽANT and PLANAS 1998). Finite element studies demonstrate that the nonlocal damage model is free of zero-energy periodic modes of instability, eliminates spurious sensitivity, effectively limits excessive damage localization, and correctly simulates the size effect on the nominal strength as well postpeak behavior.
The nonlocal concept makes sense only if the finite elements are at least three-times smaller than the characteristic length $\ell$, and this is also required to make the propagation direction of a damage band independent of the mesh layout. If the number of elements of this size would be excessive, one may use the crack band model, for which the finite element size should roughly be $\ell$. If the number would still be excessive, one may use a variant of the crack band or cohesive crack model, the idea of which is to embed in the finite element either a crack band (a strain-softening strip limited by strain discontinuity lines) or a cohesive crack (a line of displacement discontinuity); (ORTIZ et al. 1987), BELVYTSCHKO et al. (1998). There are two basic simple forms of such models — kinematically consistent and statically consistent. But it is their combination that is found to be optimal (JIRÁSEK 1998). The embedded discontinuity models are best to model the final stage of damage in which the distributed cracking coalesces into a distinct fracture. But there is a mesh bias for the propagation direction — when the fracture does not run along a mesh line, the correct orientation of the discontinuity is not obtained. As shown by JIRÁSEK (1998), it is best to use the nonlocal model at the beginning, in order to capture the correct discontinuity orientation, and later switch to an embedded discontinuity model.

In the nonlocal damage model, the need to calculate averages over a group of finite elements increases the bandwidth, and the use of asymmetric averaging near the boundaries makes the stiffness matrix nonsymmetric. These features, inconvenient for some types of numerical algorithm, are avoided by the second-gradient damage model, in which the nonlocal variable is obtained from a linear combination of the local variable and its Laplacian (as proposed in 1984; [BC], eq. 13.10.25). This model is derived (and physically justified) by Taylor series expansion of the weight kernel in the nonlocal averaging integral and truncation after the quadratic term. The truncation, however, is physically questionable because crack interactions, from which a nonlocal model can be derived (BAZANT 1994), have a rather slow decay with distance.

A computational drawback of the gradient model is that the boundary conditions become more complicated and that $C_1$ continuity is required for the strain field. This can, however, be circumvented by expressing the local variable as a linear combination of the nonlocal variable and its Laplacian, which provides an additional partial differential equation (Helmholtz equation) from which the nonlocal variable is solved (PEERLINGS et al. 1998). Very promising results have recently been achieved with this approach by de Borst and co-workers.

The main practical consequence of the damage localization instabilities and the existence of a characteristic length of the continuum is the size effect, both on the nominal strength of structures and on their postpeak response. Numerical studies of nonlocal and gradient models of geometrically similar structures of different sizes, exhibiting similar damage band paths, show that the maximum loads approximately follow the simple size effect law proposed in 1984 by BAZANT (e.g. [BC], chs. 12, 13, BAZANT and PLANAS 1998, BAZANT and CHEN 1997). This law has been amply verified experimentally, and has been derived (BAZANT 1997, BAZANT and CHEN 1997, BAZANT et al. 1999) from fracture mechanics by applying the technique of asymptotic matching, whose basic idea was advanced by PRANDTL (1904, 1952) in his analysis of boundary layer flow.

The size effect has been shown important for the design of large concrete structures, for assessments of failure and seismic response of large bodies in geotechnical engineering and geophysics, for Arctic ice studies, etc. The size effect law provides an important check on the validity of a numerical model, and the size effect measurements can be exploited to calibrate the main model parameters. Large size effects due to damage localization and stable growth of large fractures are observed not only in concrete (and mortar) — the problem so far studied most, but also in tensile and compression failures of fiber-polymer composites (BAZANT et al. 1999), sea ice (BAZANT and KIM 1998), rocks and ceramics (BAZANT and PLANAS 1998), wood, and probably in all the other quasi-brittle materials including toughened ceramics, polymer and asphalt concretes, particulate composites, wood particle board, bone, biological shells, stiff clays, cemented sands, grouted soils, coal, paper, various refractories, some special tough alloys, rigid foams, and filled elastomers.

4. Closing remark

While there are many examples of theories that have become complete or almost complete, the theory of stability of structures has not yet reached that point at the end of the second millennium. As far as elastic stability is concerned, it is now understood quite well, although even here rapid progress has recently been happening in certain particular directions, for instance those of chaos or general catastrophes. Almost the same could be said of stability of anelastic structures exhibiting plasticity, viscoelasticity and viscoplasticity. However, as far as stability of structures disintegrating because of damage and fracture is concerned, this is a challenge for the future. Major advances are still to be expected.

Acknowledgement

Partial financial support under grants from ONR (number N00014-91-J-1109) and NSF (number CMS-9713944) to Northwestern University is gratefully appreciated.
YOUNG, T.: A course of lectures on natural philosophy and the mechanical arts. 2 volumes. London 1807.

Received September 23, 1999, revised and accepted April 20, 2000

Address: ZDENĚK P. BAŽANT, Walter P. Murphy Professor of Civil Engineering and Materials Science, Northwestern University, Evanston, Illinois 60208, USA, e-mail: z-bazant@northwestern.edu.