Milestones of Direct Variational Calculus and its Analysis from the 17th Century until today and beyond – Mathematics meets Mechanics – with restriction to linear elasticity

Milestones of Direct Variational Calculus
and its Analysis from the 17th Century
until today and beyond
– Mathematics meets Mechanics –
with restriction to linear elasticity

Erwin Stein

Leibniz Universität Hannover
e-mail: stein@ibnm.uni-hannover.de

Introductory remarks

This treatise collects and reflects the major developments of direct (discrete) variational calculus since the end of the 17th century until about 1990, with restriction to classical linear elastomechanics, such as 1D-beam theory, 2D-plane stress analysis and 3D-problems, governed by the 2nd order elliptic Lamé-Navier partial differential equations.

The extension of the historical review to non-linear elasticity, or even more, to inelastic deformations would need an equal number of pages and, therefore, should be published separately.

A comprehensive treatment of modern computational methods in mechanics can be found in the Encyclopedia of Computational Mechanics [83].

The purpose of the treatise is to derive the essential variants of numerical methods and algorithms for discretized weak forms or functionals in a systematic and comparable way, predominantly using matrix calculus, because partial integrations and transforming volume into boundary integrals with Gauss's theorem yields simple and vivid representations. The matrix $\boldsymbol{D}$ of 1st partial derivatives is replaced by the matrix $\boldsymbol{\mathcal{N}}$ of direction cosines at the boundary with the same order of non-zero entries in the matrix; $\partial/\partial x_i$ corresponds to $\cos(\boldsymbol{n}, \boldsymbol{e}_i)$, $\boldsymbol{x} = x_i \boldsymbol{e}_i$, $\boldsymbol{n} = \cos(\boldsymbol{n}, \boldsymbol{e}_i)\boldsymbol{e}_i$, $i = 1, 2, 3$ for $\Omega \subset \mathbb{R}^3$.

A main goal is to present the interaction of mechanics and mathematics for getting consistent discrete variational methods and from there – in a comparative way – the properties of primal, dual and dual-mixed finite element methods in their historical development. Of course the progress of proper mathematical analysis since the 1970s and 1980s is outlined, concerning consistency, convergence and numerical stability as well as a priori and a posteriori error estimates in the frame of Sobolev spaces for the mostly C^0-continuous test and trial ansatz functions at element interfaces. This mathematical research followed after the more intuitive engineering developments since the 1950s, using the principle of virtual work and the principle of minimum of total potential energy.

A proper finite element method has to regard the motto of this treatise: "mathematics meets mechanics".

Comparative numerical results are not included because numerous new calculations would have been necessary for getting usable comparisons for the various cited articles.

1. Networked thinking in computational sciences

Computation Mechanics needs the networked interaction of mechanics, mathematics and computer science, which is symbolized on the graphics by M.C. Escher from 1956, called "Bond of Union" and expressing "creativity by communication and reproduction". It symbolizes the networked thinking in Computational Sciences, Fig. 1.
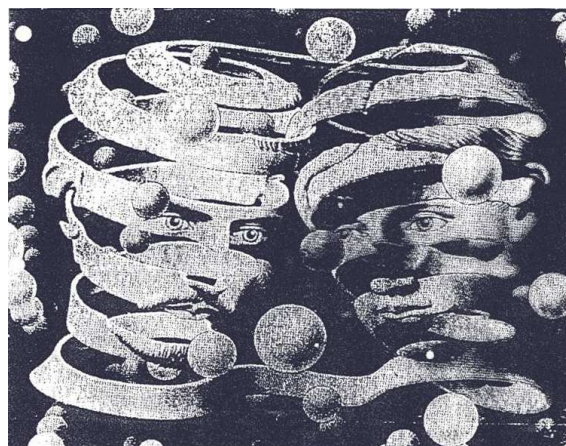
Fig. 1. "Bond of Union" by M.C. Escher, 1956, expressing "creativity by communication and reproduction".

John H. Argyris stated in the early 1960s:

"The computer shapes the theory",

i.e., not only the numerical methods and the solution algorithms are highly influenced by the computer facilities but also the physical and mathematical modeling. The complexity and theoretical depth of multi-physical and multi-scale mathematical models and – of course – the number of unknowns which have to be solved in reasonable time, crucially depend on the computer architecture, the processor speed and the available data storage. Their growth still follows Moore's law, [64].

There is no doubt that verification (bounded error control of the discretized solution) and validation (coupled bounded error control of mathematical modeling in conjunction with relevant experimental data, coupled with numerical and experimental verification) are major tasks for reliability and trustworthyness of computational methods and results in application to real-life problems.

This requires holistic thinking, various interactions and the collaboration of specialists from many disciplines and subjects in science, engineering development and design, of developers and distributors of general purpose software systems and within industrial developments and production of modern technological objects.

2. Eminent scientists in numerical and structural analysis of elasto-mechanics

A rough overview of the historical development of numerical analysis in elasto-mechanics is shown in Fig. 2. We have to decide between the mathematical treatment of numerical solutions for boundary value problems of ordinary and partial differential equations on the left side of Fig. 2, and mechanical modeling and numerical calculation of structures on the right side, which – of course – have the same mathematical modeling in the background.
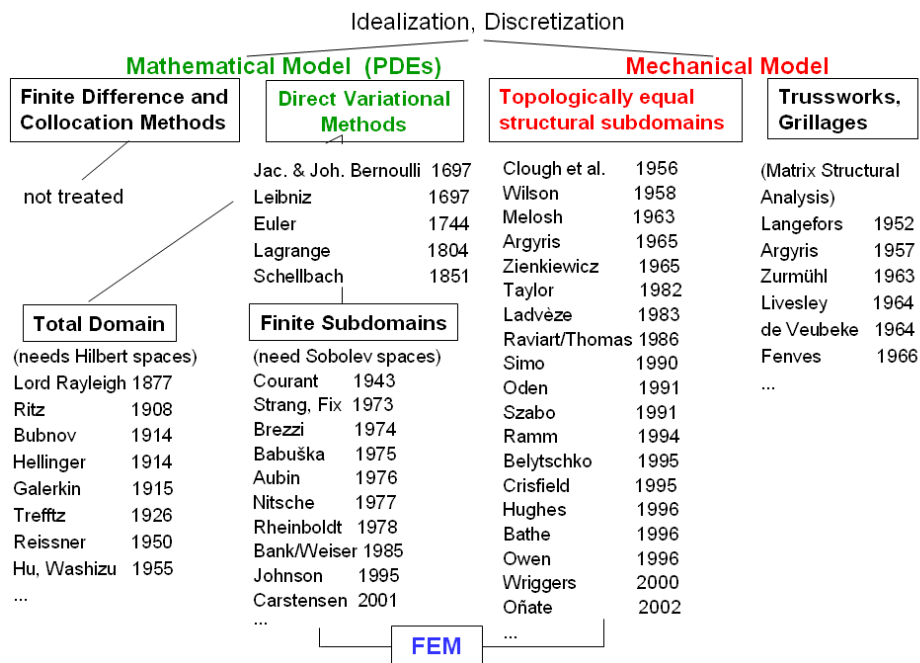
Fig. 2. Eminent scientists in numerical and structural analysis of elasto-mechanics since the 17th century

Left-hand side of Fig. 2: We do not consider here finite difference and collocation methods but direct variational methods only. This begins in 1697 with the first seminal variational calculus by Jacob Bernoulli and the first idea of a direct variational method by Gottfried Wilhelm Leibniz in

Acta Eruditorum with seven solutions for the so-called brachistochrone problem (named by Johann Bernoulli). A strong development took place since the end of the 19th and especially in the 20th century with test and trial functions in the whole domain of a problem with restricted applications. Important contributions came from Lord Rayleigh, Walter Ritz and Boris Grigoryevich Galerkin and some others. The real FEM, id est with test and trial functions in regular subdomains which are topologically equal for the same given operator in the whole domain – the finite elements –, were developed in the middle of the 20th century like a ping-pong match between mechanics and mathematics with joint and stormy international efforts until today.

For this we have to mention primarily Ivo M. Babuška, Werner C. Rheinboldt, Franco Brezzi, Claes Johnson and Rolf Rannacher.

Right-hand side of Fig. 2: Mechanical modeling and numerical simulation also started in the middle of the 20th century – according to the upcoming facilities of digital computers – but only with direct simplified modeling of a continuous 3D structure by 1D truss-works and grillages. In this frame matrix structural analysis with systematic application of the principle of virtual work and its generalizations was developed by Langefors, Argyris and many others. The deciding step to FEM was the use of topologically equal subdomains for discretization with respect to the whole domain and using the already known methods of matrix structural analysis, first published by Ray Clough et al. in 1956. Then a plenty of finite elements for a great variety of engineering problems followed until the end of the 20th century by Wilson, Argyris, Zienkiewicz, Taylor, Szabo and many others.

There was never before and never after a joint photo opportunity for the "Big Three", as it happened in Munich in 1999, Fig. 3.
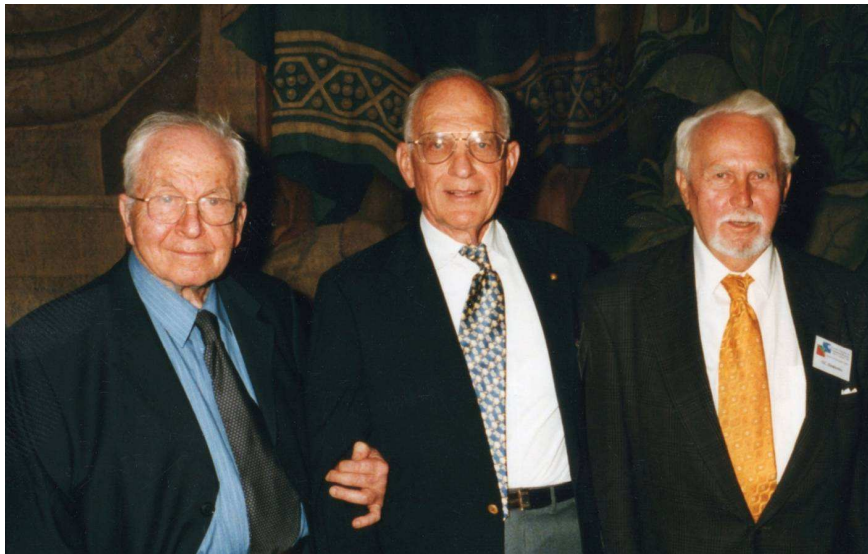
Fig. 3. Three eminent pioneers, the "Big Three" of the finite element method based on structural mechanics: John H. Argyris, Ray W. Clough and Olgierd C. Zienkiewicz (from left to right) at the first European Conference on Computational Mechanics (ECCM) 1999 in Munich, photo by E. Ramm.

3. THE BEGINNING OF CYBERNETIC AND HOLISTIC THINKING IN PHILOSOPHY AND NATURAL SCIENCES

The 17th century can be considered as the cradle of modern natural sciences and technology. In condensed form these developments are, [81, 82]:

- New (mechanistic) natural philosophy at the beginning of the Age of Enlightenment by F. Bacon, T. Hobbes, R. Descartes, B. Spinoza, J. Locke, G.W. Leibniz, . . .

Descartes: *cogito ergo sum*, dualism of mind and body,
Leibniz: *nihil sine ratione, theoria cum praxi, harmonia est unitas in multitudinae*

- The laws of the planetary orbits, by J. Kepler.
  Theory-guided systematic experiments for the mathematical formulation of physical laws, by G. Galilei: *Measure what can be measured, make measurable which is not yet measurable!*

- Analytical geometry by R. Descartes

- Infinitesimal calculus by I. Newton and G.W. Leibniz;
  binary number system by T. Harriot and G.W. Leibniz

- Axiomatic mechanics (Mechanica Rationalis) by I. Newton in his *Philosophia Naturalis Principia Mathematica*

- Conservation and extremum principles in mechanics by E. Torricelli, P. de Fermat and G.W. Leibniz

- Construction of the first mechanical calculating machines by W. Schickard, B. Pascal and G.W. Leibniz

- Many further technical inventions by C.H. Huygens (watches), G.W. Leibniz (improvements of technical equipment for mining), O. von Guerricke (technical applications of air pressure and vacuum, especially *the Magdeburg hemispheres*), . . .

4. PRE-HISTORY OF THE FINITE ELEMENT METHOD (FEM) IN THE 17TH AND 19TH CENTURY

4.1. Torricelli's principle of minimum potential energy of a system of rigid bodies under gravity loads in stable static equilibrium

The first minimum principle in mechanics was established by Evangelista Torricelli (1608–1647), secretary to Galileo Galilei, about 1630, Fig. 4. He found that a mechanical system of two rigid bodies finds its stable static equilibrium in a configuration that the total center of gravity S takes the lowest possible position, Fig. 5, [93].

Fig. 4. Evangelista Torricelli (1608–1647).
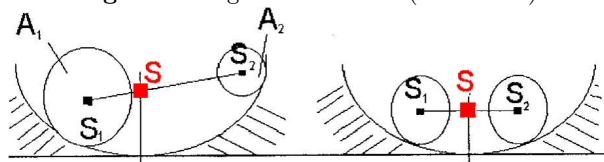
Fig. 5. Two spheres or discs, connected by a hinged bar and moving within a larger spherical or circular foundation. The gravity center S takes the lowest possible position.

4.2. Galilei's approximated solution of minimal time of a frictionless downgliding mass under gravity load

The next important step originates from Galileo Galilei's first formulation of the following optimization problem: Find the curve of a mass which is frictionlessly gliding down a guided path due to gravity in least time, published in his *Discorsi*, Leiden (1638), [39].
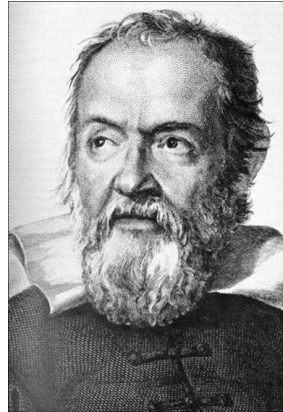
Fig. 6. Galileo Galilei (1564–1642).
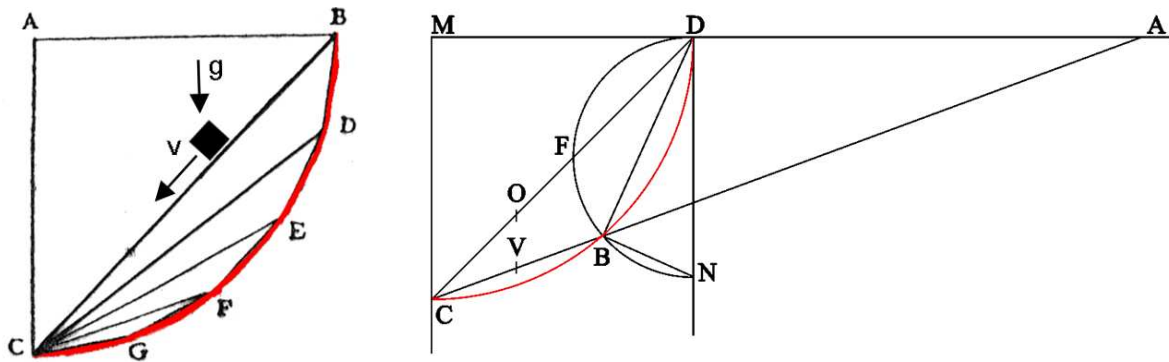
Fig. 7. (left) Galilei's empirical approximation of the optimal curve: a quarter of a circle, after testing different polygons as improvements of the straight line, $t(BC) > t(BDC) > t(BDEC) > t(BDEFC) > t(BDEFGC)$;
(right) Velocity of the gliding mass, $v = \sqrt{2gh}$.

According to the important results of Galilei's problem from 1697, the optimal path of the gliding mass is the common cycloid. This will be treated a little bit later. But it is advantageous to show some important properties of the cycloid in advance.

The cycloid and its evolute (which is again a congruent cycloid) is shown in Fig. 8. It can be generated by a rolling wheel on a straight line as shown in Fig. 9. Important properties of the cycloid are: normal $\overline{SK} = n$ and curvature radius $\overline{SM} = \rho = 2n$.

In the 1660s and 1670s tautochrony or isochrony properties of the cycloid were discovered by Christiaan Huygens, Isaac Newton and Gottfried Wilhelm Leibniz. It has the astonishing property
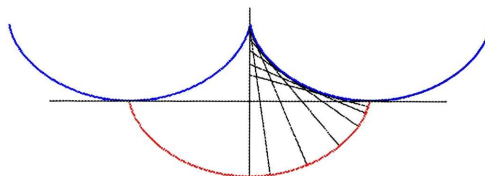
Fig. 8. The cycloid and its evolute are congruent cycloids.
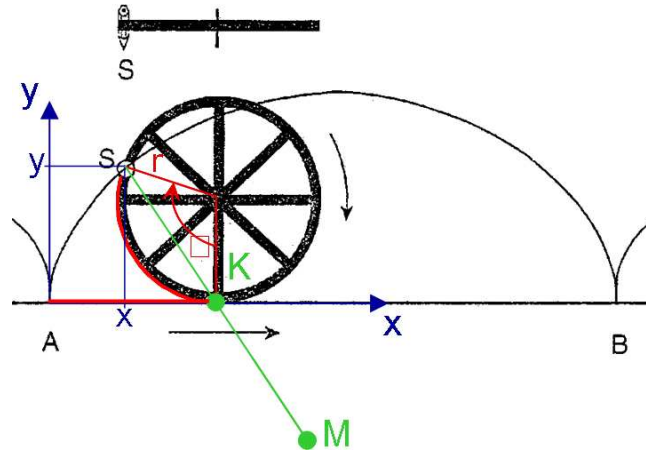
Fig. 9. Generation of a cycloid by a rolling wheel on a straight line.

that two masses A and C', Fig. 10, need the same time for frictionless gliding to the lowest point B_0, as $t_{AB_0} = t_{C'B_0}$ and $\dfrac{t_{AB_0}}{t_{MB_0}} = \dfrac{\widehat{AB_0}}{\overline{AA'}} = \dfrac{4r}{2r} = 2$.
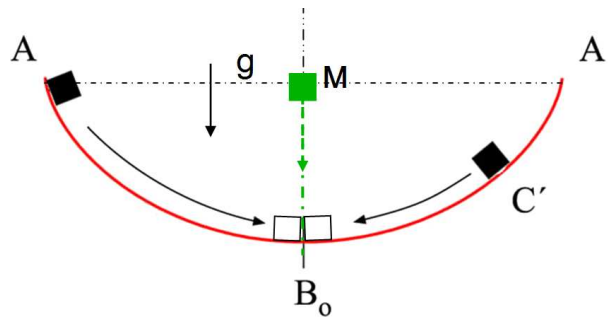
Fig. 10. Tautochrony or isochrony property of the common cycloid.

4.3. Snell's law of light refraction and Fermat's principle of least time for the optical path length

Willebrord van Royen Snel (1580–1626) published in 1621 the law of light refraction which was better reasoned later by Christiaan Huygens with the principle that every point of a wave is the source of a new wave, [48]. This law reads, Fig. 11a,

$$\sin \alpha_1 = \frac{c_1 \Delta t}{AB}, \qquad \sin \alpha_2 = \frac{c_2 \Delta t}{AB},$$

yielding

$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{c_1}{c_2} = n, \qquad \frac{\sin \alpha_1}{c_1} = \frac{\sin \alpha_2}{c_2} = \text{const},$$

n is the refraction coefficient.

Pierre de Fermat (1601 or 1607/8–1665) established the principle of the light path in minimal time using Cartesian coordinates, [33]:

$$T = s_1/c_1 + s_2/c_2 = n_1/s_1 + n_2/s_2, \qquad T = n_1 \left[(x - x_1)^2 + y_1^2 \right]^{1/2} + n_2 \left[(x_2 - x)^2 + y_2^2 \right]^{1/2}.$$

The stationarity condition written with the derivative in the later formulation by Newton and Leibniz reads $dt/dx(P_1, P_2; x) = 0 \rightarrow n_1(x - x_1)/s_1 - n_2(x_2 - x)/s_2 = 0$ or

$$\sin\alpha_1/c_1 = \sin\alpha_2/c_2. \tag{1}$$

In case of linearly varying density from a more to a less dense medium, Fig. 11b, the light path fulfills the optimality (stationarity) condition:

$$T_{AB} = \int ds/c(y)\sqrt{y} = \min; \quad dt = ds/c(y), \;\; \text{yelding} \;\; \sin\alpha(y)/c(y) = \text{const}, \tag{2}$$

which describes a common cycloid.
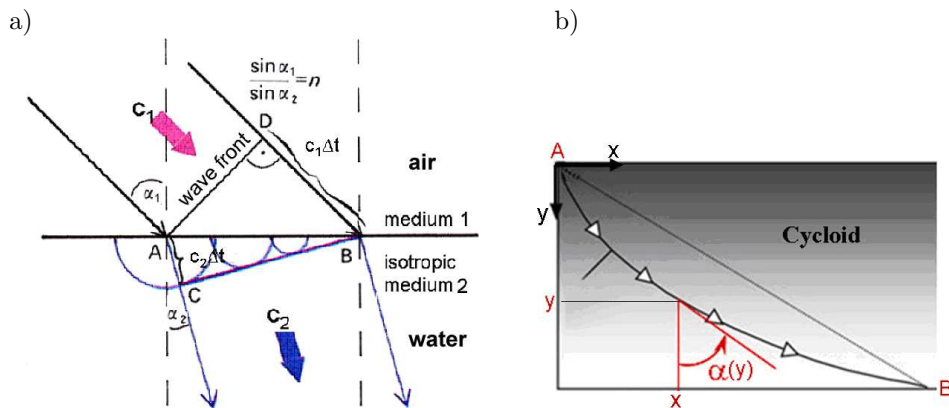
Fig. 11. a) refraction of light at the transition from a less dense medium (air) to a denser medium (water) with the velocities c_1 and c_2; b) the common cycloid as the light path at least time in a medium with linearly varying density.

Fermat's principle had significant influence on the finding of conservation principles for physical problems in the 17$^\text{th}$ and 18$^\text{th}$ century. It will be shown in Sec. 3 that the famous problem of a guided frictionless down-gliding mass due to gravity in shortest time, first posed by Galilei in his *Discorsi* from 1638, also has the solution of the common cycloid.

4.4. Johann I Bernoulli's call for solutions of Galilei's problem now called brachistochrone problem in 1696 and the solutions published by Leibniz in 1697

Johann Bernoulli, Fig. 12b, from Basel published in 1696 in *Acta Eruditorum* a call for the solution of Galilei's problem: to find the function of the curve with minimal time which a frictionless guided downgliding mass needs. The problem should be solved in one year's time and published in the same journal. Johann Bernoulli called this searched curve *brachistochrone* (curve of descent in least time).

A total of seven solutions were submitted and published by Leibniz in 1697, [58]:

(1) one by Johann's older brother Jacob Bernoulli, Fig. 12a, the mathematically most important with the first version of variational calculus,

(2,3) two by Johann Bernoulli himself with elegant geometrical and analytical ideas. One solution concerned directly the brachistochrone problem, and the other treated the corresponding problem of the shortest time of the light path through a medium with linearly changing density,

(4) one graphical solution by Leibniz with a geometrical integration of the ODE of the cycloid by using his transmutation theorem from 1674,

a)   b)
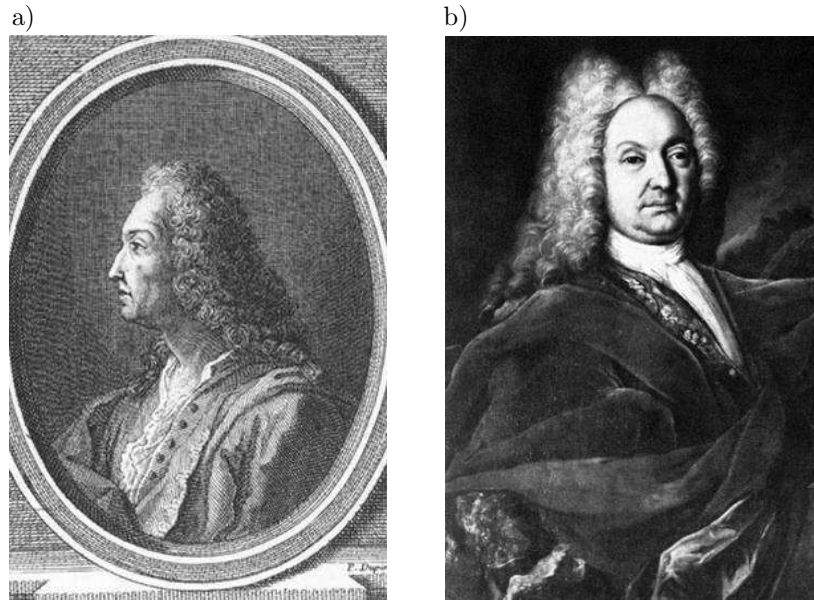
Fig. 12. a) Jacob Bernoulli (1655–1705), the inventor of probability calculus and variational calculus and b) Johann Bernoulli (1667–1748) who contributed a lot to infinitesimal and variational calculus and to mechanics.

(5) one approximated discrete solution by Leibniz, using the same primal idea as Jacob Bernoulli but proposing then in a rough draft a discretized numerical method with parametrized equal finite time steps,

(6) one anonymous solution by Newton without proof, which was only provided in 1724,

(7) one incomplete solution by L'Hôpital and Tschirnhaus.

The mathematical solution of the "brachistochrone problem" was the origin of variational calculus, especially by Jacob Bernoulli, in [58], about 45 years before the seminal work of Leonhard Euler (1707–1783).

Jacob Bernoulli's ingenious derivation of the ODE for the stationary solution of the brachistochrone problem – the curve of a down-gliding mass in shortest time – in anticipation of Leonhard Euler's piece-wise discrete test functions for the first variation of a functional (1743), [37].

Jacob Bernoulli's solution method follows the idea to reduce first the problem to a finite number of problems of infinitesimal calculus: a discrete variation of the extremal curve by triangular test functions between equidistant points $y - h$; y; $y + h$, with the distance h, is introduced, Fig. 13. Then, the discrete stationarity condition reads for each discrete point:

$$t_{CG} + t_{GD} \overset{!}{=} t_{CL} + t_{LD}. \tag{3}$$

The continuous limit for the step $h \to 0$ yields the ordinary differential equation of the cycloid,

$$\frac{ds}{dx} \sim \frac{k}{\sqrt{y}}, \qquad \frac{dx}{dy} = \tan \alpha = \sqrt{\frac{y}{k^2 - y}}, \qquad k^2 = 2r. \tag{4}$$

Jacob Bernoulli thus anticipated Leonhard Euler's discrete concept for the derivation of the differential equation of the extremal function in his famous book *Methodus inveniendi lineas curvas ...* from 1743/44, as can be seen in Euler's original drawing, Fig. 14.

In distinction from Jacob Bernoulli, Euler solved the more general extremal problem of the functional $\int\limits_{x=A}^{Z} F(y(x), y'(x), x) dx = \min\limits_{y(x)}$ with the unknown extremal function $y = f(x)$.
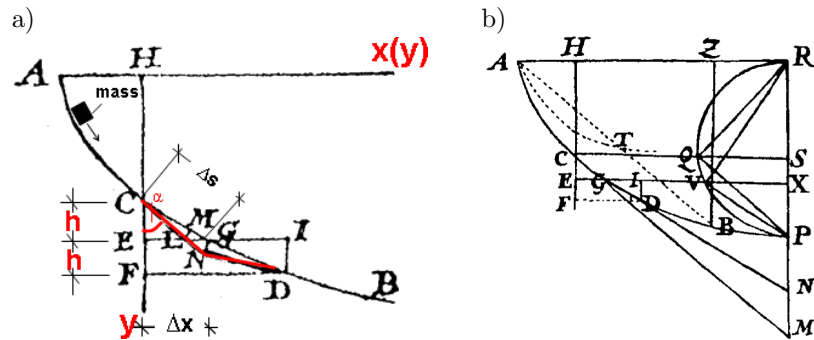
Fig. 13. a) figure by Jacob Bernoulli with discrete triangular test functions providing the stationarity condition; b) figure by Johann Bernoulli for his derivation of the cycloid as minimal curve.
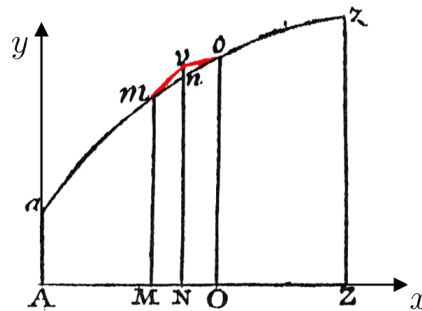
Fig. 14. Original drawing in L. Euler's derivation of the differential equation for the extremal function of isoperimetric variational problems with one unknown, from [37].

As can be seen in Fig. 14, Euler also introduced triangular test functions on equidistant discrete abscissa points x and analyzed the difference quotients at points M, N, D with distance h. Instead of envisaging a special problem, e.g. the brachistochrone problem, he performs the infinitesimal limit for $h \to 0$ for the more general functional $F(y, y', x)$ and gets the ordinary differential equation of the extremal function $y(x)$ as $F_y - \dfrac{\mathrm{d}}{\mathrm{d}x} F_{y'} = 0$.

4.5. Leibniz's discovery of the kinetic energy of a mass as a conservation quantity

In his article *Brevis demonstratio erroris memorabilis Cartesii et aliorium circa legem natura* in *Acta Eruditorum* in 1686 Gottfried Wilhelm Leibniz (1646–1716) falsificates the solution of Réné Descartes (1596–1650) for an exciting problem of this time, namely finding the "true measure of the living force". Descartes, Fig. 15a, assumed in [35] that the product of mass and velocity is a conservation quantity. This can not be true because the velocity is a vector. Furthermore, Descartes derived on this basis seven laws for impacting bodies which are all wrong, see Szabo, [91]. In the application of his "principle" to the postulate that the whole quantity of motion in the universe is conserved, he uses the scalar quantity $m \cdot |\boldsymbol{v}|$ instead of $\boldsymbol{v}^2$, but this is formulated in a mathematically non-understandable way. Leibniz, Fig. 15b, introduced the expression $m \cdot \boldsymbol{v}^2$ for the "true measure", Fig. 16, and thus discovered the kinetic energy of a moved body in the quasi-static case but without the factor 1/2, expressing it as a law of proportionalities, [57].

In the treatise of J. S. Koenig (1712–1757) from 1751, [53], he claims by referring to a copy of a letter from Leibniz to J. Hermann (1678–1733), dated from October 1708, that Leibniz had discovered the principle of least action (in terms of today: the extremum principle for the total potential and kinetic energy of a moved mass-system)which is a minimum in the quasi-static case and a minimum or a maximum in the kinetic case. With this statement Koenig declares that Leibniz had discovered this principle before Maupertuis, President of the new Preussische Akademie

Fig. 15. a) Réné Descartes (1596–1650); b) Gottfried Wilhelm Leibniz (1646–1716).
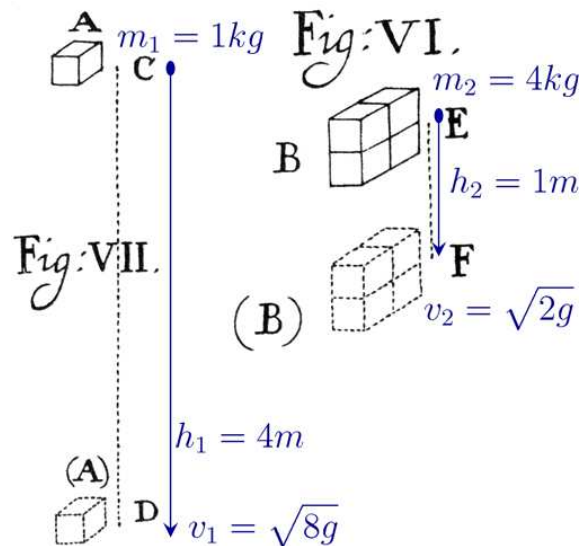
Fig. 16. Leibniz's original drawing in [57] for the derivation of the kinetic energy of a falling body under the action of gravity. With the added measures in blue color the law takes the form $1/2m_1v_1^2 = 1/2m_2v_2^2$, i.e. $4g = 4g$.

der Wissenschaften in the years 1746–1756, who published his "Principe général" in 1746 in the "Memoirs of the Academie des Sciences" in Paris; it postulates in a speculative way that the quantity of actions (defined by the product of velocity and path) is always a minimum – which is not true in the kinetic case of a moved mass. His proofs for some applications violate Newton's laws, especially the inertia axiom (already formulated by Galilei in his "Discorsi"). He saw his principle as a law of economy of nature. Maupertuis was criticized after his publication, and famous scientists like Euler did not regard it as a serious work. But nevertheless Maupertuis fought for his priority of the principle of least action against Koenig who had become a member of the Preussische Akademie (the Prussian Academy), and even the Prussian King Frederick II was involved and supported Maupertuis. The problem was, that Koenig could not provide the original letter of Hermann and was blamed of forgery; he finally drew the consequence and quitted the membership in the Preussische Akademie.

In Immanuel Kant's inglorious first treatise, [51], translated: "Thougths on the true estimation of the Living Forces", he tries to prove that Descartes was right and Leibniz was wrong. His arguments

violate Newton's laws of motion, and in his philosophical arguments he claims that mathematics does not provide the ultimate arguments for the validity of a theory. It is well-known that Kant did not understand Newton's "Principia" because he had severe problems with advanced mathematics and mechanics of his time.

Remark

All the attempts in the 17th and the first half of the 18th century for finding a general law for energy balances of moved masses – the principle of least action – were incomplete and had inconsistencies. From the today's view we only need to consider the integration in time and in space of Newton's fundamental law $\boldsymbol{F} = m\dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}$ for constant mass in time.

Integration along a path s of the force $\boldsymbol{F}$ yields $\displaystyle\int_s \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{s} = \int_s m\dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t}\right)\cdot\mathrm{d}\boldsymbol{s} = \int_v m\boldsymbol{v}\cdot\mathrm{d}\boldsymbol{v} = \dfrac{m\boldsymbol{v}^2}{2}.$

Integration over the time yields $\displaystyle\int_t \boldsymbol{F}\ \mathrm{d}t = \int_t m\dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\ \mathrm{d}t;\ \boldsymbol{F}\ \mathrm{d}t = m\ \mathrm{d}\boldsymbol{v}.$

Leibniz's law for the kinetic energy of a rigid body together with Torricelli's law of minimum potential energy of a system of bodies yield the principle of total potential energy of rigid bodies in the 18th century with severe quarrels about the authorship as outlined above. Based on this, the variational calculus of rigid and elastic bodies in static and kinetic equilibrium, i.e., the analytical mechanics, was developed in the 18th and 19th century by Leonhard Euler, especially by Joseph Louis Lagrange, Lejeune Dirichlet, William Rowan Hamilton and Lord Rayleigh.

4.6. Leibniz's draft of a discrete (direct) solution of the brachistochrone problem

Leibniz's second contribution to the brachistochrone problem, [58], is of special interest for the history of the finite element method and presented as follows.

He pursues the same idea as Jacob Bernoulli by reducing the continuous problem to a discrete one with equidistant support points and triangular test functions between three subsequent points, Figs. 17a and 17b.
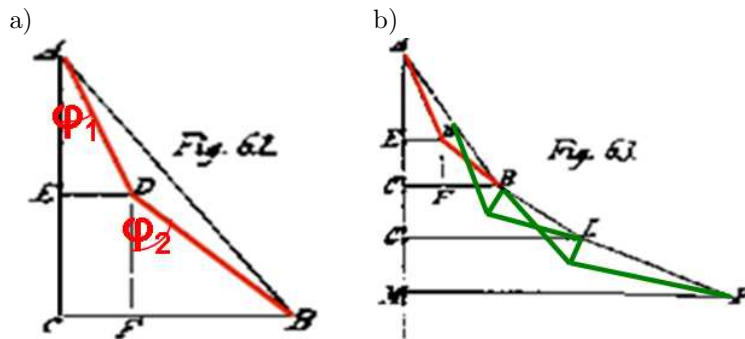
Fig. 17. Leibniz's original figures for his discrete solution: a) discrete solution with linear trial and test functions for one unknown, $\overline{ED}$ at point E; b) representation of piecewise triangular trial and test functions at three equidistant points E, C', C''.

Leibniz did not give the analysis of the method which only was done in the 19th century by K.H. Schellbach, [78], see (7). Even after a request by Johann Bernoulli to continue this research, Leibniz responded that he would not have time and interest.

Due to the importance of Leibniz's first idea for a discrete variational method, the analysis for one discrete unknown is presented hereafter by the author, Fig. 18.
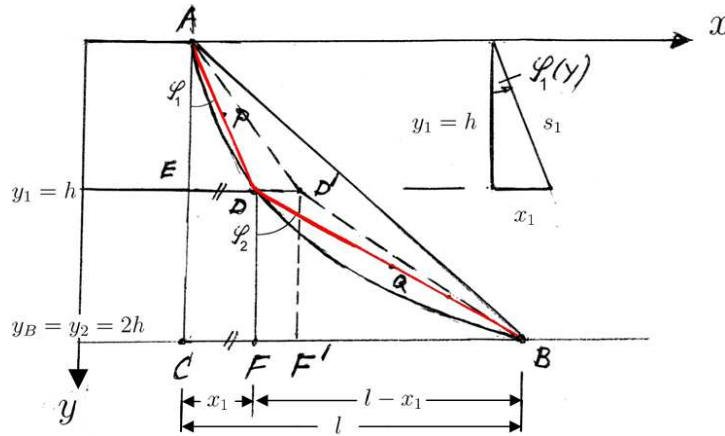
Fig. 18. Measures for the derivation of the stationarity condition for a path in minimal time with one discrete unknown.

The minimum problem reads

$$T_{AB} = \int_A^B dt = \int_{y_A}^{y_B} \frac{ds(y)}{v(y)} = \int_{y_A}^{y_B} \frac{\sqrt{1 + (x'(y))^2}}{\sqrt{2gy}} dy = \min_{x(y)}.$$ (5)

The first approximation with one discrete unknown $x_1(y_1) = x(h)$ yields

$$T_{AB} = \frac{\overline{AD}}{v_{AD}} + \frac{\overline{DB}}{v_{DB}} = \frac{s_1}{v(h)} + \frac{s_2}{v(2h)} = \frac{\sqrt{h^2 + x^2}}{\sqrt{2gh}} + \frac{\sqrt{h^2(l-x)^2}}{\sqrt{2g \cdot 2h}} = \min_{x_h(y)}.$$ (6)

The discrete stationarity condition, not given by Leibniz, is

$$\frac{\partial T_{AB}}{\partial x} \overset{!}{=} 0 \rightsquigarrow \frac{x}{\sqrt{h^2 + x^2}} \cdot \frac{1}{\sqrt{2gh}} - \frac{l-x}{\sqrt{h^2 + (l-x)^2}} \cdot \frac{1}{\sqrt{2g2h}} \overset{!}{=} 0,$$ (7)

yielding

$$\frac{\sin \varphi_1}{v_1} - \frac{\sin \varphi_2}{v_2} \overset{!}{=} 0.$$ (8)

This is the same condition as for the minimum time of a light path moving through a medium with linearly varying density which has the same variational formulation.

The above stationarity condition yields the 4th order polynomial

$$f(x) = x^4 - 2lx^3 + x^2(l^2 + h^2) + 2h^2lx - h^2l^2 \overset{!}{=} 0,$$ (9)

with the linear approximation $x_{lin} = 0,5\ l$ and the exact solution

$$x = 0,69\ l.$$ (10)

The discrete stationarity conditions for n equidistant nodal points read

$$\sin \varphi_i = \frac{x_i - x_{i-1}}{s_i} = \text{const} \ ; \quad i = 1, 2, ..., n.$$ (11)

This representation shows that Leibniz's idea and draft was the first precursor of the finite element method.

4.7. Schellbach's discrete formulation of direct variational stationarity conditions according to Leibniz

Following the ideas of Jacob Bernoulli and Leibniz in [58], Karl Heinrich Schellbach (1805–1892) published in 1851 in the Journal of A.L. Crelle, [78], the discrete algebraic equations of the brachistochrone and other optimization problems by analytical calculation of the integrals for equidistant support points, and in total the discrete algebraic equations in analytical form of 12 variational problems for different boundary and field conditions of the related brachistochrone problem.

In order to point out the status of variational calculus in the middle of the 19th century, we cite from the introduction of [78]: "The reasons for Bernoulli's, Euler's, and Lagrange's methods [for the variational calculus] can not be clearly understood yet ... the variational calculus is the most abstract and most sublime area of all mathematics."

Discrete analytical representations of 12 variational problems using equidistant nodal points are given:

1. Minimum area of a polygon with fixed ends with given length and extensions (Fig. 19.1),

2. Minimum area of a rotational surface with given meridian arc length and boundary conditions, with extended versions (Fig. 19.1),

3. Brachistochrone problem with generalized boundary conditions in B and B' (Fig. 19.2),

4. Brachistochrone problem in a resisting medium (Fig. 19.2),

5. Problem similar to 3. and 4., but with the condition of largest or smallest final velocity in B' (Fig. 19.2).
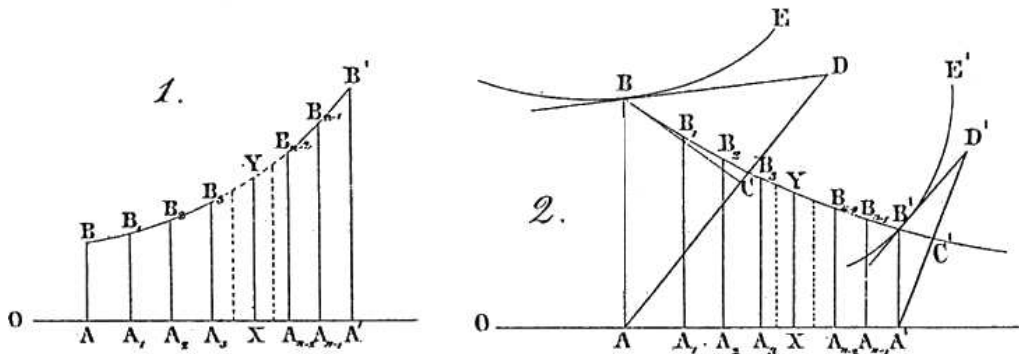
Fig. 19. Figures in Schellbach's article [78] for the discrete (direct) variational solution of geometrical and mechanical optimization problems.

5. THE DEVELOPMENT OF FINITE ELEMENT METHODS IN THE 19TH AND 20TH CENTURY

5.1. Lord Rayleigh's variational approach to eigenvalue problems of differential equations – the Rayleigh quotient

The first variational formulation of mechanical eigenvalue problems and their approximated (discretized) solution in form of the Rayleigh quotient were given by John William Strutt, the 3rd Baron Rayleigh (1842–1919), Fig. 20, in his famous book *Theory of Sound* from 1877, [50].

Fig. 20. Lord Rayleigh.

5.1.1. Linear algebraic eigenvalue problem of a Hermitian matrix

The eigenvalue problem of a Hermitian matrix $\boldsymbol{A}$ reads

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} = \lambda\boldsymbol{I}\boldsymbol{x}; \quad A_{ij} \in \mathbb{R}, i, j = 1, 2, ..., n; \quad \boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}; \quad \det \boldsymbol{A} > 0, \tag{12}$$

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = 0; \quad x \neq 0 \rightsquigarrow \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0, \tag{13}$$

yielding the characteristic polynomial

$$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0; \quad -a_{n-1} = \operatorname{tr}\boldsymbol{A}; \quad (-1)^n a_0 = \det \boldsymbol{A} \tag{14}$$

with n real-valued eigenvalues

$$\lambda_1, \lambda_2, \ldots, \lambda_n \neq 0, \quad \text{with} \quad \lambda_i > \lambda_{i+1}; \quad \lambda_1 = \max \lambda_i, \lambda_n = \min \lambda_i. \tag{15}$$

5.1.2. Orthogonality of the eigenvectors

From the symmetry and positive definiteness of Matrix $\boldsymbol{A}$ we get the orthogonality of n linear independent eigenvectors for n real-valued eigenvalues, as

$$\left.\begin{array}{rcl} \boldsymbol{x}_k^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}_i &=& \lambda_i\boldsymbol{x}_i), \\ (\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{A} &=& \lambda_k\boldsymbol{x}_k^{\mathrm{T}})\boldsymbol{x}_i^{\mathrm{T}} \end{array}\right\} \rightsquigarrow 0 = (\lambda_i - \lambda_k)\boldsymbol{x}_k^{\mathrm{T}}\overbrace{\boldsymbol{A}\boldsymbol{x}_i}^{\boldsymbol{y}_i}; \ \lambda_i \neq \lambda_k \tag{16}$$

$$\rightsquigarrow \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{y}_i = 0 \text{ i.e. } \boldsymbol{x}_k \perp \boldsymbol{y}_i. \tag{17}$$

5.1.3. Rayleigh quotient for λ_i

This yields the Rayleigh quotient for the eigenvalue λ_i:

$$\boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i) \rightsquigarrow \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_i, \tag{18}$$

$$\lambda_i = \frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_i}{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_i} \ ; \ \lambda_i = R(\boldsymbol{x}_i) \tag{19}$$

with orthonormal eigenvectors

$$[\overset{*}{\boldsymbol{x}}_1 \ \overset{*}{\boldsymbol{x}}_2 \ \ldots \ \overset{*}{\boldsymbol{x}}_n] = \overset{*}{\boldsymbol{X}}, \text{ i.e. } \overset{*}{\boldsymbol{x}}_i^{\mathrm{T}}\overset{*}{\boldsymbol{x}}_j = \delta_{ij}, \tag{20}$$

$$[\operatorname{diag} \lambda_i] = \overset{*}{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{A}\overset{*}{\boldsymbol{X}}. \tag{21}$$

5.1.4. Variational (weak) form of equilibrium and minimum principle of total potential energy of a linear elastic solid body

The static deformations of a linear elastic solid body – kinematically a C^1 manifold – are described in the primal form by the three Navier-Lamé's partial differential equations for the displacements in Cartesian coordinates, extended here for elastodynamics.

5.1.5. Lamé's equations for linear elastodynamics with mixed BCs

(i) Kinematic relations:

$$\boldsymbol{u}(\boldsymbol{x},t) = u_i(\boldsymbol{x},t)\boldsymbol{e}_i; \quad \boldsymbol{x} = x_i\boldsymbol{e}_i \in \Omega \subset \mathbb{R}^3; \quad i = 1,2,3, \tag{22}$$

$$\boldsymbol{\epsilon}(\boldsymbol{x},t) = \epsilon_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j := \mathbf{grad}_{\mathrm{s}}\boldsymbol{u}(\boldsymbol{x},t), \tag{23}$$

(ii) Constitutive equations:

$$\boldsymbol{\sigma}(\boldsymbol{x},t) = \mathbb{C} : \boldsymbol{\epsilon}(\boldsymbol{u}(\boldsymbol{x},t)); \quad \boldsymbol{\sigma} = \sigma_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j; \quad \mathbb{C}_{ijkl} \in \mathbb{R}, \quad \mathbb{C} = \mathbb{C}^{\mathrm{T}}, \quad \det\mathbb{C} > 0, \tag{24}$$

(iii) Equilibrium conditions:

$$\operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}(\boldsymbol{x},t)) + \boldsymbol{f} - \rho\ddot{\boldsymbol{u}} = \boldsymbol{0}; \rightsquigarrow \operatorname{div}\left[\mathbb{C} : \mathbf{grad}_{\mathrm{s}}\boldsymbol{u}(\boldsymbol{x},t)\right] + \boldsymbol{f} - \rho\ddot{\boldsymbol{u}} = \boldsymbol{0}. \tag{25}$$

By eliminating the stresses and physical strains we get the three coupled hyperbolic differential equations of 2nd order for the displacements, depending on the position vector $\boldsymbol{x}$ and time t, with the symmetric differential operator matrix $\boldsymbol{L}$, Eq. (26). We restrict ourselves to Dirichlet boundary conditions. Furthermore, the boundary tractions $\boldsymbol{t}(\boldsymbol{x},t,\boldsymbol{n})$ are given by Cauchy's theorem.

$$\left\|\begin{array}{l} \boldsymbol{L}[\boldsymbol{u}(\boldsymbol{x},t)] + \boldsymbol{f}(\boldsymbol{x},t) - \rho\ddot{\boldsymbol{u}}(\boldsymbol{x},t) = \boldsymbol{0} \\ \boldsymbol{u} = \boldsymbol{0} \text{ at } \Gamma_D; \ \Gamma_D \cup \Gamma_N = \partial\Omega \\ \boldsymbol{B}[\boldsymbol{u}(\boldsymbol{x},t)] = \boldsymbol{\sigma}(\boldsymbol{x},t)\cdot\boldsymbol{n}(\boldsymbol{x},t)|_{\Gamma_N} = \bar{\boldsymbol{t}}(\boldsymbol{x},t,\boldsymbol{n}) \text{ at } \Gamma_N. \end{array}\right. \tag{26}$$

In vectorial operator form, these PDEs read

$$(\mu + \lambda)\mathbf{grad}(\operatorname{div}\boldsymbol{u}(\boldsymbol{x},t)) + \mu\operatorname{div}(\mathbf{grad}(\boldsymbol{u}(\boldsymbol{x},t))) + \rho\boldsymbol{b} = \rho\ddot{\boldsymbol{u}}(\boldsymbol{x},t) \ \forall\boldsymbol{x} \in \Omega \subset \mathbb{R}^3 \tag{27}$$

or

$$(\mu + \lambda)\mathbf{grad}\ e + \mu\Delta\boldsymbol{u} + \rho\boldsymbol{b} = \rho\ddot{\boldsymbol{u}}; \quad e = \operatorname{div}\boldsymbol{u}, \quad \Delta\boldsymbol{u} = \operatorname{div}(\mathbf{grad}\boldsymbol{u}). \tag{28}$$

5.1.6. Variational form and minimum principle with test functions $v(\boldsymbol{x},\tau)$

Introducing admissible test functions

$$\boldsymbol{v}(\boldsymbol{x},\tau) \in \mathcal{V}; \quad \mathcal{V} = \{\boldsymbol{v} \in [\mathcal{H}^1(\Omega)]^3; \quad \boldsymbol{v} = \boldsymbol{0} \text{ at } \Gamma_D\} \quad \text{at time} \quad t = \tau, \tag{29}$$

multiplying with Eq. (25) and integrating over the spatial domain Ω

$$\int_\Omega \left[(\operatorname{div}\boldsymbol{\sigma})\cdot\boldsymbol{v} + \boldsymbol{f}\right]\cdot\boldsymbol{v} \ \mathrm{d}\Omega - \int_\Omega \rho\ddot{\boldsymbol{u}}\cdot\boldsymbol{v} \ \mathrm{d}\Omega \stackrel{!}{=} 0 \tag{30}$$

as well as partial integration and applying the divergence theorem to the 1st term, using the properties of the test function

$$\int_\Omega \operatorname{div}(\boldsymbol{\sigma})\cdot\boldsymbol{v} \ \mathrm{d}\Omega = \int_S \cancel{\operatorname{div}(\boldsymbol{\sigma}\cdot\boldsymbol{v})\ \mathrm{d}S} - \int_\Omega \boldsymbol{\sigma} : \mathbf{grad}_{\mathrm{s}}\boldsymbol{v} \ \mathrm{d}\Omega \tag{31}$$

yields the weak form, also introducing the bilinear form $a(\boldsymbol{u}, \boldsymbol{v})$, multiplied with -1

$$
\left[\underbrace{- \int_\Omega [\mathbb{C} : \boldsymbol{\epsilon}(\boldsymbol{u})] : \boldsymbol{\epsilon}(\boldsymbol{v}) \, \mathrm{d}\Omega}_{+a(\boldsymbol{u},\boldsymbol{v})=a(\boldsymbol{v},\boldsymbol{u})} + \underbrace{\int_\Omega \boldsymbol{f} \cdot \boldsymbol{v} \, \mathrm{d}\Omega + \int_{\Gamma_N} \bar{\boldsymbol{t}} \cdot \boldsymbol{v} \, \mathrm{d}S}_{-F(\boldsymbol{v})} + \underbrace{\int_\Omega \rho \ddot{\boldsymbol{u}} \cdot \boldsymbol{v} \, \mathrm{d}\Omega}_{-\rho(\ddot{\boldsymbol{u}},\boldsymbol{v})} = 0 \right] \cdot (-1), \tag{32}
$$

and with the further property

$$
\int_\Omega \ddot{\boldsymbol{u}} \cdot \boldsymbol{v} \, \mathrm{d}\Omega = \cancel{\int_S (\dot{\boldsymbol{u}} \cdot \boldsymbol{v}) \, \mathrm{d}S} - \int_\Omega \dot{\boldsymbol{u}} \cdot \dot{\boldsymbol{v}} \, \mathrm{d}\Omega, \tag{33}
$$

we arrive at the extremum principle for the functional $J[\boldsymbol{v}]$ by L. Dirichlet (1857)

$$
\Pi + K = J[\boldsymbol{v}] = \underbrace{\frac{1}{2} a(\boldsymbol{v}, \boldsymbol{v})}_{\Pi_{int}(\boldsymbol{v})} \underbrace{-F(\boldsymbol{v})}_{\Pi_{ext}(\boldsymbol{v})} \underbrace{- \frac{\rho}{2}(\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}})}_{K(\dot{\boldsymbol{v}})} ; \quad J[\boldsymbol{u}] = \min_{\boldsymbol{v} \in \mathcal{V}} J[\boldsymbol{v}]. \tag{34}
$$

5.1.7. Stationarity condition as variational form or principle of virtual work

The stationarity condition of the functional $J[\boldsymbol{v}]$ yields the weak variational form for the searched extremal problem, or in mechanical terms, the principle of virtual work in space and time

$$
\partial\Pi + \partial K = \partial J[\boldsymbol{v}] = a(\boldsymbol{v}, \boldsymbol{v}) - F(\boldsymbol{v}) - \rho(\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}}) \overset{!}{=} 0. \tag{35}
$$

5.1.8. Proof of the minimum property of $J[\boldsymbol{u}]$

Theorem:

$$
J[\boldsymbol{u} + \kappa\boldsymbol{v}] \geq J[\boldsymbol{u}], \boldsymbol{u}, \boldsymbol{v} \in \mathcal{V} \tag{36}
$$

Proof:
Assume a solution $\widetilde{\boldsymbol{u}} = \boldsymbol{u} + \kappa\boldsymbol{v}, \; \kappa \in \mathbb{R}, \; \widetilde{\boldsymbol{u}} \in \mathcal{V}$

$$
J[\boldsymbol{u} + \kappa\boldsymbol{v}] = \frac{1}{2} a(\boldsymbol{u} + \kappa\boldsymbol{v}, \boldsymbol{u} + \kappa\boldsymbol{v}) - F(\boldsymbol{u} + \kappa\boldsymbol{v}) - \frac{\rho}{2}(\dot{\boldsymbol{u}} + \kappa\dot{\boldsymbol{v}}, \dot{\boldsymbol{u}} + \kappa\dot{\boldsymbol{v}}) \tag{37}
$$

$$
= J[\boldsymbol{u}] + \kappa \left[a(\boldsymbol{u}, \boldsymbol{v}) - F(\boldsymbol{v}) + \frac{1}{2}\kappa^2 a(\boldsymbol{v}, \boldsymbol{v}) - \frac{\rho}{2}\kappa^2(\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}}) \right] \tag{38}
$$

$$
\text{for } \kappa = 1: \quad = J[\boldsymbol{u}] + \underbrace{\frac{1}{2} a(\boldsymbol{v}, \boldsymbol{v})}_{\geq 0} \underbrace{- \frac{\rho}{2}(\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}})}_{\geq 0} \rightsquigarrow J[\boldsymbol{u}] = \min_{\boldsymbol{v} \in \mathcal{V}} J[\boldsymbol{v}] ; \; \frac{\rho}{2}(\dot{\boldsymbol{v}}, \dot{\boldsymbol{v}}) \leq \frac{1}{2} a(\boldsymbol{v}, \boldsymbol{v}). \tag{39}
$$

The kinetic energy is limited by the elastic energy in case of Dirichlet boundary conditions.

5.1.9. Rayleigh-Ritz direct variational method for the smallest eigenvalue

With $F(\boldsymbol{v}) \equiv 0$ and the time-harmonic ansatz

$$
\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{v}(\boldsymbol{x}) \cos(\omega t); \; \dot{\boldsymbol{v}}(\boldsymbol{x}, t) = -\boldsymbol{v}(\boldsymbol{x})\omega \sin(\omega t) \tag{40}
$$

we get the stationarity conditions in time and space

$$a(\boldsymbol{v}(\boldsymbol{x},t),\boldsymbol{v}(\boldsymbol{x},t)) = a(\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{v}(\boldsymbol{x})) \int_{2\pi} \cos^2(\omega t) \mathrm{d}t; \tag{41}$$

$$-\rho(\dot{\boldsymbol{v}}(\boldsymbol{x},t),\dot{\boldsymbol{v}}(\boldsymbol{x},t)) = -\rho(\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{v}(\boldsymbol{x}))\,(-\omega)^2 \int_{2\pi} \sin^2(\omega t) \mathrm{d}t\,;\ \lambda := \omega^2, \tag{42}$$

yielding in total the 1st variation of the functional J with the eigenvalue λ

$$\delta J[\boldsymbol{v}(\boldsymbol{x})] = a(\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{v}(\boldsymbol{x})) - \lambda(\boldsymbol{v})\rho(\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{v}(\boldsymbol{x})) \stackrel{!}{=} 0. \tag{43}$$

5.1.10. Rayleigh-Ritz quotient

The above condition (43) directly yields the Rayleigh-Ritz quotient with the important minimum property of the eigenvalue $\lambda(\boldsymbol{v})$

$$\lambda(\boldsymbol{v}) = \frac{1}{\rho}\frac{a(\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{v}(\boldsymbol{x}))}{(\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{v}(\boldsymbol{x}))} = \frac{\displaystyle\int_{\Omega} \boldsymbol{\epsilon}(\boldsymbol{v}) : \mathbb{C} : \boldsymbol{\epsilon}(\boldsymbol{v})\ \mathrm{d}\Omega}{\displaystyle\int_{\Omega} \sqrt{\rho}\boldsymbol{v} \cdot \sqrt{\rho}\boldsymbol{v}\ \mathrm{d}\Omega}\,; \quad \lambda(\boldsymbol{u}) = \min_{\boldsymbol{v}\in\mathcal{V}} \lambda(\boldsymbol{v}). \tag{44}$$

The n-parametric Rayleigh-Ritz ansatz in the total spatial domain

$$\widetilde{\boldsymbol{v}}(\boldsymbol{x}) = \sum_{I=1}^{N} C_I \varphi_I(\boldsymbol{x})\ \forall\ \varphi_I(\boldsymbol{x}) \in \widetilde{\mathcal{V}} \subset \mathcal{V} \tag{45}$$

with Ritz parameters C_I results in a minimal sequence for $\lambda(\boldsymbol{v})$

$$\min \lambda = \lambda_1; \quad \lambda_{1,N} \quad \text{for} \quad N \quad \text{test functions}; \quad \lambda_{1,N} \geq \lambda_1. \tag{46}$$

The proof of the minimum property of the approximated smallest eigenvalues was given by Walter Ritz (1878–1909) in 1909, [74].

5.2. Ritz's first mathematical foundation of direct variational methods for the Kirchhoff plate equation

Walter Ritz (1878–1909), at last professor of mathematics at the Universität Göttingen, Fig. 21, published his famous article, [74], translated: "On a New Method for the Solution of Certain Problems in Mathematical Physics", in 1909.

We cite from the introduction with translation into English:

The boundary value problems in mathematical physics usually require the representation of finite, continuous functions in prescribed finite domains. Only exceptionally, the expansion in power series is possible, and even more seldom this is numerically usable in the total domain...

Thus, there is a request for an approximated representation of the integrals in the total prescribed domain by a polynomial of given degree n...in a way that for growing n the accuracy is growing unlimited, finally resulting in polynomial expansions of the integrals ... Also Fourier series can be useful. In general, one can use most of those functions $\psi_1, \psi_2, \ldots, \psi_n$ which can be chosen according to qualitative observation.

Fig. 21. Walter Ritz (1878–1909).

In the following a condensed representation of Ritz's article is presented. The power series of linear independent and relative complete chosen functions $\psi_i(x,y)$ for 2D problems reads

$$w_n = \psi_0 + a_1\psi_1 + a_2\psi_2 + \cdots + a_n\psi_n, \tag{47}$$

with the unknown so-called Ritz-parameters $a_i, i = 1 \ldots n$.

Ritz presents in his article the numerical method and its convergence proof for the determination of the a_i, under the presumption that the variational problem

$$J = \int_a^b f(x, w, w', w'', \ldots w^{(0)}) \, \mathrm{d}x = \min_{a_i} \tag{48}$$

has to be solved. The parameters a_i are determined through the n stationarity conditions

$$\frac{\partial J_n}{\partial a_1} = 0, \quad \frac{\partial J_n}{\partial a_2} = 0, \quad \ldots \quad \frac{\partial J_n}{\partial a_n} = 0, \tag{49}$$

which yield a linear system of algebraic equations for the a_i.

5.2.1. Solution for the Kirchhoff biharmonic elastic plate equation for rectangular clamped domains

The Kirchhoff plate equation for lateral load $f(x,y) \in \mathcal{L}_2(R)^2$ reads

$$\Delta\Delta w \equiv \frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4} = f(x,y), \tag{50}$$

where at the total boundary $L = \partial R$ clamped boundary conditions

$$\overline{w} = 0, \quad \frac{\partial \overline{w}}{\partial n} = 0 \tag{51}$$

have to be fulfilled.

Ritz postulates the finiteness and continuity of the lateral displacement of the plate middle surface w and its derivatives up to order 4 in R and at L, which can be reduced by using today the adequate Sobolev functional analysis with weak derivatives at the boundary.

For the 2D clamped plate problem the integral (48), which is only given for 1D problems, takes the form

$$J = \iint\limits_R \left[\frac{1}{2}(\Delta w)^2 - f(x,y)w \right] \mathrm{d}S. \tag{52}$$

Partial integration yields

$$\iint\limits_R \Delta U \Delta V \, \mathrm{d}S = \iint\limits_R V \Delta\Delta U \, \mathrm{d}S + \int\limits_L \left[\Delta U \frac{\partial V}{\partial n} - \Delta V \frac{\partial U}{\partial n} \right] \mathrm{d}s. \tag{53}$$

The first variation of J reads

$$\delta J = \iint\limits_R \left[\Delta w \delta \Delta w - f \delta w \right] \mathrm{d}S, \tag{54}$$

with $U = w$, $V = \partial w$ and regarding $w = 0$, $\frac{\partial w}{\partial n} = 0$ and $\delta w = 0$, $\frac{\partial \delta w}{\partial n} = 0$ at L, resulting in

$$\delta J = \iint\limits_R \left[\Delta\Delta w - f \right] \delta w \, \mathrm{d}S. \tag{55}$$

One can show that each solution of this equation yields a minimum of J.

Existence and convergence proofs for the numerical solution of the Kirchhof plate PDE with convex domains and clamped edges are following in Ritz's article.

5.2.2. Remarks on Trefftz method from 1926

A remarkable counterpart to Ritz's direct variational calculus is Trefftz's method, Erich Trefftz (1888–1937) – see E. Trefftz: Ein Gegenstück zum Ritzschen Verfahren, in: Verhandlungen des 2. Internationalen Kongresses für technische Mechanik, 131–137, Zürich 1926. Herein, the homogeneous PDEs are fulfilled a priori in the whole domain, either by analytical solutions for the displacements or by stress functions, fulfilling the bipotential Beltrami PDEs. Opposite to Ritz method, mixed boundary conditions are fulfilled approximately by the discrete variational method.

For second order elliptic self-adjoint and well-posed BVPs, this yields symmetric positive definite system matrices for admissible test and trial functions at the boundaries, and convergence is assured for complete test and trial polynomials at the boundary – in Trefftz original work considering the whole boundary – as in Ritz method.

In case of the fourth order bipotential Kirchhoff plate equation or of corresponding shell bending theories, Trefftz's method is only consistent for pure displacement boundary conditions (BCs), i.e., for clamped edges. In case of so-called Navier BCs and free edges (with pure Neumann conditions) an additional least-squares term for the integral of the errors of the kinematical BCs, multiplied with a penalty factor, has to be inserted for achieving unambiguous and thus converging results of Trefftz approximations.

These problems were treated in the author's doctoral thesis from 1964, in the doctoral thesis of P. Weidner from 1967 and also in the habilitation thesis of the author from 1969. From the 1970s to the 1990s Zienkiewicz et al., Stein & Peters, Jirousek, Piltner, Zielinski and some others generalized Trefftz's method for discrete boundary element techniques.

By generalizing Trefftz method to finite (discrete) boundary elements there arose a competition with the important discrete boundary integral equation method (BIEM or simply BEM) with the advantageous kernel functions derived from integral transformations based on the Signorini identity.

However, there are recent improvements of Trefftz methods by choosing hybrid mixed variational formulations and special test and trial functions, e.g., applying a domain decomposition method for solving the algebraic equations.

In this treatise, Trefftz method was not presented as a whole chapter because then also BIEM had to be presented with comparisons what would have increased the size of this work considerably; this might be the subject for a separate article.

5.3. Galerkin's discrete variational method for linear elliptic differential equations

Boris Grigoryevich Galerkin (1871–1945), Fig. 22, substantially contributed to the foundation of discrete variational methods for linear elastic problems, i.e. for linear elliptic boundary value problems, in his famous article on beams and plates [38] from 1915, Fig. 23.

Fig. 22. Boris Grigoryevich Galerkin (1871–1945).

Стержни и пластинки.

Ряды въ нѣкоторыхъ вопросахъ упругаго равновѣсія стержней и пластинокъ.

Б. Г. Галеркинъ.

1. Нѣкоторые вопросы упругаго равновѣсія стержней и пластинокъ какъ статическаго, такъ и динамическаго, приводятъ къ дифференціальнымъ уравненіямъ, преимущественно 2-го—4-го порядковъ. Общее рѣшеніе этихъ уравненій не всегда легко найти, но и въ тѣхъ случаяхъ, когда рѣшеніе найти нетрудно или оно извѣстно, рѣшеніе не всегда удовлетворяетъ условіямъ задачи. Приходится найденное рѣшеніе дифф. уравненія приспособлять къ условіямъ физической задачи, что не всегда удается выполнить съ достаточной степенью точности. Напр., рѣшеніе, данное Морисомъ Леви для тонкихъ прямоугольныхъ пластинокъ [1]), если и удовлетворяетъ уравненію Кирхгофа, то должно быть въ каждомъ отдѣльномъ случаѣ приспособлено къ заданнымъ для наружнаго обвода пластинки условіямъ, — въ результатѣ получимъ точное рѣшеніе дифф. ур—ія, [2]) но въ общемъ приближенное рѣшеніе задачи теоріи упругости. Въ отдѣльныхъ случаяхъ, напр., для пластинки, свободно опертой по краямъ, рѣшеніе можетъ быть взято сравнительно легко съ любой степенью точности, но для пластинки съ закрѣпленными краями полученіе хотя бы приближенныхъ результатовъ требуетъ громадной затраты труда, на что указываютъ произведенныя работы [3]).

Есть и другой путь. Можно подобрать рѣшеніе въ видѣ ряда съ неопредѣленными коэффиціентами, каждый членъ котораго удовлетворяетъ условіямъ на концахъ или по краямъ, а коэффиціенты опредѣлить такъ, чтобы, въ общемъ, рѣшеніе являлось приближеннымъ рѣшеніемъ дифф. ур. (рѣшеніе Леви можетъ быть сведено къ рядамъ съ неопредѣленными коэффиціентами, каждый членъ котораго удовлетворяетъ дифф. ур., коэффиціенты же подбираются такъ, чтобы рѣшеніе удовлетворяло условіямъ по краямъ пластинки).

Изъ приближенныхъ методовъ рѣшеній широкое при-

мѣненіе получилъ въ послѣднее время методъ Ritz'a [1]). Методъ этотъ сводится вкратцѣ къ слѣдующему.

Задаемся уравненіемъ упругой линіи стержня или упругой поверхности пластинки:

$$w = \Sigma A_n \varphi_n, \quad \ldots \ldots \ldots (a),$$

гдѣ φ_n — функція, зависящая отъ координатъ и удовлетворяющая условіямъ на концахъ стержня или по краямъ пластинки, $A_1, A_2 \ldots A_n$ — неопредѣленные коэффиціенты. Пользуясь ур. (a), составляемъ выраженіе для потенціальной энергіи системы V, затѣмъ находимъ T-работу внѣшнихъ силъ по взятой кривой оси или поверхности. Какъ V, такъ и T, очевидно, будутъ заключать неопредѣленные коэффиціенты A_n, которые опредѣляются затѣмъ изъ условія:

$$\frac{dV}{dA_n} - \lambda \frac{dT}{dA_n} = 0, \quad \text{гдѣ } \lambda - \text{постоянный коэффи-}$$

ціентъ. Если $V = T$, λ въ вопросахъ теоріи упругости равенъ 2.

Профессоръ С. П. Тимошенко разработалъ и примѣнилъ этотъ методъ къ цѣлому ряду вопросовъ устойчивости упругихъ системъ [2]); имъ разсмотрѣно громадное количество этихъ вопросовъ, многіе изъ которыхъ впервые получили свое блестящее освѣщеніе въ его работахъ. Этотъ методъ примѣненъ проф. Hager'омъ [3]), Лоренцемъ [4]) и друг. къ изгибу пластинокъ [5]).

---

[1]) Maurice Lévy. Sur l'équilibre élastique d'une plaque rectangulaire. C. R. 1899 CXXIX p. 535—539.
Estanave. Thèses. 1900.
[2]) Б. М. Кояловичъ. Объ одномъ уравненіи съ частными производными четвертаго порядка. 1902.
[3]) И. Бубновъ. Строительная механика корабля. Ч. II. 1914.
H. Hencky, Der Spannungszustand in rechteckigen Platten. 1913.
В. Г. Галеркинъ. Прямоугольныя пластинки, опертыя по краямъ. 1915. «Изв. Петрогр. Политехн. Инст.», т. XXIV.

[1]) W. Ritz. Ueber eine neue Methode zur Lösung gewisser Variationsproblem der mathematischen Physik. «J. f. reine u. angewandte Mathematik.» 1909.
W. Ritz. Theorie der Transversalschwingungen einer quadratischen Platte mit freien Rändern. «Annalen der Physik.» 1909. B. 28.
[2]) С. П. Тимошенко. Объ устойчивости упругихъ системъ. Кіевъ. 1910.
[3]) Hager. Berechnung ebener, rechteckiger Platten mittels trigonometrischer Reihen. 1911.
[4]) Lorenz. Technische Elastizitätslehre. 1913.
[5]) Здѣсь необходимо указать, что въ работѣ проф. Hager'а имѣется рядъ существенныхъ ошибокъ, лишавшихъ нѣкоторые его выводы какого-либо значенія [напр., для плиты опертой по 3-мъ сторонамъ (стр. 60) и опертой въ 4-хъ точкахъ (стр. 72) выбранные ряды не удовлетворяютъ необходимымъ условіямъ по краямъ]. Но, помимо этого, въ ней имѣется одинъ основной недостатокъ, подрывающій значеніе всѣхъ его выводовъ. Этотъ недостатокъ работы проф. Hager'а состоитъ въ способѣ примѣненія имъ метода Ritz'а—Тимошенко: составляя выраженіе для работы внутрен-

Fig. 23. First page of Galerkin's article in Wjestnik Ingenerow, Petrograd 1915, entitled "Application of series to the problems of equilibrium of elastic beams and plates", [38].

Galerkin's discrete variational method for linear elliptic differential equations with ansatz functions in the whole domain, starts directly from the variational form (weighted residual or 1st variation of the corresponding functional) instead of beginning with the energy functional in Ritz's work, or, in terms of mechanics, using the principle of virtual work as the weak formulation of equilibrium for admissible strains.

In the following the original representation of Galerkin in Cartesian coordinates is adapted to modern terminology, using vector notation.

The mixed boundary value problem reads, e.g. of the Lamé PDEs of linear elasticity, Eqs. (27) and (28),

$$\mathcal{L}[\mathbf{u}] = \mathbf{f} \text{ in } \mathbf{\Omega} \in \mathbb{R}^2, \tag{56}$$

with $\mathcal{L}$ as the linear elliptic differential operator of order $2m$, $m = 1$ for the Lamé equations, and the kinematic (Dirichlet) boundary conditions

$$\mathcal{B}_m[\mathbf{u}] = \mathbf{0} \text{ on } \partial\mathbf{\Omega}. \tag{57}$$

The ansatz functions

$$\{\boldsymbol{u}_j(x,y)\}_{j=1}^n, \text{s.t. } \mathcal{B}_m[u_j] = \mathbf{0} \text{ on } \partial\mathbf{\Omega} \ \forall \ \mathbf{j} \tag{58}$$

are chosen such that the essential (kinematic) boundary conditions are fulfilled.

The discrete approximation of solution $\boldsymbol{u}$ with admissible ansatz functions $\boldsymbol{u}_j(x,y)$ and the unknown scalar parameters α_j reads

$$\widetilde{\boldsymbol{u}} = \sum_{j=1}^n \alpha_j \boldsymbol{u}_j(x,y), \tag{59}$$

with the necessary conditions

$$\mathcal{B}_m[\widetilde{\boldsymbol{u}}] = \mathbf{0} \text{ on } \partial\mathbf{\Omega}. \tag{60}$$

The main new idea of Galerkin is to postulate the orthogonality of the $\boldsymbol{u}_j$ with respect to the strong residuum $\mathcal{R}[\widetilde{\boldsymbol{u}}] = \mathcal{L}[\widetilde{\boldsymbol{u}}] - \boldsymbol{f}$ in Ω

$$\int_\Omega \mathcal{L}[\widetilde{\boldsymbol{u}}] \cdot \boldsymbol{u}_j \, \mathrm{d}A = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{u}_j \, \mathrm{d}A \ \forall \ j, \quad \mathrm{d}A = \mathrm{d}x\mathrm{d}y, \tag{61}$$

or

$$\int_\Omega \underbrace{(\mathcal{L}[\widetilde{\boldsymbol{u}}] - \boldsymbol{f})}_{\mathcal{R}[\widetilde{\boldsymbol{u}}]} \cdot \boldsymbol{u}_j, \quad \mathrm{d}A = \mathbf{0}, \ j = 1, 2, \ldots, n. \tag{62}$$

Thus, Galerkin's approximation method is an inner product projection – not a $\mathcal{L}^2$ projection – of the whole analytical solution space $\mathcal{V}(\Omega)$ of the BVP into the reduced discretized function space $\mathcal{V}_h \subset \mathcal{V}$. In case of well-posed selfadjoint elliptic BVPs, the approximated inner product yields symmetric and positive definite global stiffness matrices with the presumption that the discrete ansatz functions in the whole domain are kinematically admissible and complete.

Replacing $\boldsymbol{f} = \mathcal{L}[\widetilde{\boldsymbol{u}}]$ in (62) yields

$$\int_\Omega (\mathcal{L}[\widetilde{\boldsymbol{u}}] - \mathcal{L}[\boldsymbol{u}]) \cdot \boldsymbol{u}_j \, \mathrm{d}A = \mathbf{0} \tag{63}$$

and furthermore for the whole approximated solution, which is not given in Galerkin's article,

$$\int_{\Omega} (\mathcal{L}[\widetilde{u}] - \mathcal{L}[u]) \cdot \widetilde{u} \, dA = 0, \tag{64}$$

and using the linearity of operator $\mathcal{L}$, we finally get for the discretization error $e := \widetilde{u} - u$

$$\int_{\Omega} \mathcal{L}[\underbrace{\widetilde{u} - u}_{e}] \cdot \widetilde{u} \, dA = 0. \tag{65}$$

Integration by parts and applying the divergence theorem yields the variational or weak form of the approximated problem

$$a(\widetilde{u} - u, \widetilde{u}) = a(e, \widetilde{u}) = 0, \tag{66}$$

which is the important orthogonality condition, also addressed as Galerkin orthogonality, which states that the Galerkin method yields the orthogonality of the error e with respect to the approximation $\widetilde{u}$. This is a system of linear symmetric algebraic equations (for symmetric operators $\mathcal{L}$) for determining the coefficients α_j, $j = 1 \ldots n$.

Remark

In case of the finite element method with ansatz and test functions in finite subdomains of the whole domain, the variational condition in the whole domain must be fulfilled in each subdomain Ω_e of each finite element e, yielding the algebraic system for the whole domain $\Omega = \bigcup_e \Omega_e$

$$\bigcup_{\Omega_e} a_e(\underbrace{u_h - u}_{e_h}, u_h) = 0 \ \forall \ u_h \in \mathcal{V}_h \subset \mathcal{V}; \quad u \in \mathcal{V}. \tag{67}$$

However, due to jumps of the derivative $\frac{\partial u_h}{\partial n}$ and thus of the tractions at element interfaces, convergence cannot be shown at this level, but requires the introduction of Sobolev approximation spaces, as shown in Sec. 7.2.

Also the engineer I. Bubnov introduced in his book on structural mechanics of ships from 1914 with 650 pages variational approximations and applications to structural systems, [23].

5.4. Courant's first introduction of finite elements

The famous applied mathematician Richard Courant (1888–1972), Fig. 24, first conceptually introduced trial and test functions in finite subdomains of the 2D boundary value problem of St. Venant's torsion of prismatic beams without warping resistance, [30]. We cite from his introductory remarks:

> As Henri Poincaré once remarked, "solution of a mathematical problem" is a phrase of indefinite meaning. Pure mathematicians sometimes are satisfied with showing that the non-existence of a solution implies a logical contradiction, while engineers might consider a numerical result as the only reasonable goal.

> This address will deal with a topic in which such a synthesis of theoretical and applied mathematics has become particularly convincing. Since Gauss and W. Thompson, the equivalence between boundary value problems of partial differential equations on the one hand and problems of the calculus of variations on the other hand has been a central point in analysis. At first, the theoretical interest in existence proofs dominated and only much later

Fig. 24. Richard Courant (1888–1972). He was the first mathematician who introduced finite subdomains for discrete variational methods.

were practical applications envisaged by two physicists, Lord Rayleigh and Walter Ritz; they independently conceived the idea of utilizing this equivalence for numerical calculation of the solutions, by substituting for the variational problems simpler approximating extremum problems in which only a finite number of parameters need be determined.

In his so-called address he outlines his idea for the St. Venant's torsion of prismatic bars with arbitrary cross section, within a domain B and "finite elements", i.e., regular subdomains with contours $C_1, C_2, C_3, \ldots$ and areas $A_1, A_2, A_3, \ldots$. The multiply connected domain between C and $C_1, C_2, C_3, \ldots$ may be called B^*, Fig. 25.
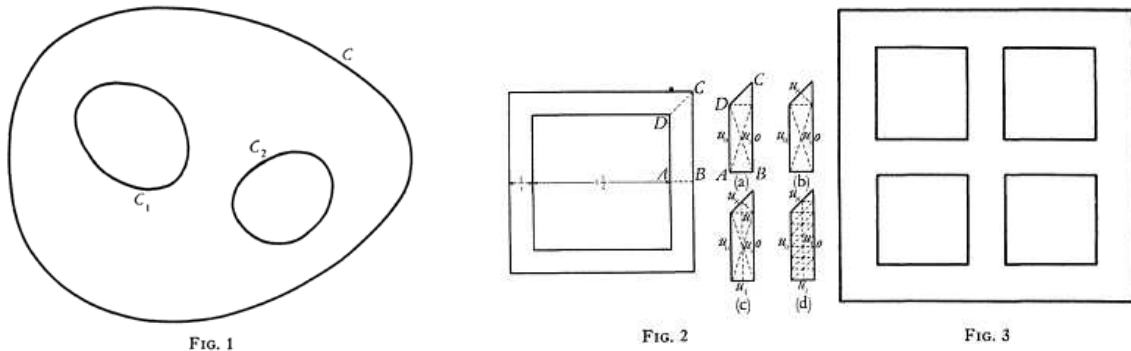
Fig. 25. Figures of the St. Venant's torsion problem of prismatic bars with arbitrary multi-connected cross sections.

We further cite from Courant's article:

Then the adequate variational formulation of the torsion problem in proper units is: To find a function $\phi = u$ continuous in $B + C$, having piecewise continuous first derivatives in B, having the boundary values zero on C and constant, but not prescribed values c_i in the holes B_i, such that for the whole domain B

$$D(\phi) = \iint \left[(\phi_x^2 + \phi_y^2) + 2\phi \right] \, \mathrm{d}x \mathrm{d}y \tag{68}$$

attains its least value d for $\phi = u$. The function u then will give the stresses in the cross section by differentiation.

The paper denoted as an "address" is motivating the essential generalization of Ritz's method to ansatz functions in finite subdomains and presents in the appendix the example in Fig. 25, herein Figs. 2 and 3 for a quadratic cross section with 4 quadratic cutouts, using linear triangular elements. Also some roots of the analysis are mentioned. Courant is considered by the mathematical community to be the founder of primal FEM.

5.5. The seminal first engineering derivation and application of the finite element method in structural mechanics by Clough et al.

In 1956, M.J. Turner, R.W. Clough, H.C. Martin and L.J. Topp published the article on the stiffness and deflection analysis of complex structures, [94], which was a joint work of the structural mechanics group at the University of California at Berkeley with Professor Ray W. Clough (*1920), Fig. 26, as the leading scientist and engineers from the Boeing Aircraft Company in Seattle. The underlying severe technical problem was the sufficiently accurate analysis of stresses and strains in skew stiffened box girder wings of aircrafts which became very important during World War II and thereafter, Fig. 27.

Fig. 26. Ray W. Clough (*1920) who developed the first engineering-based derivation and application of the finite element method for structural problems of aircraft wings.
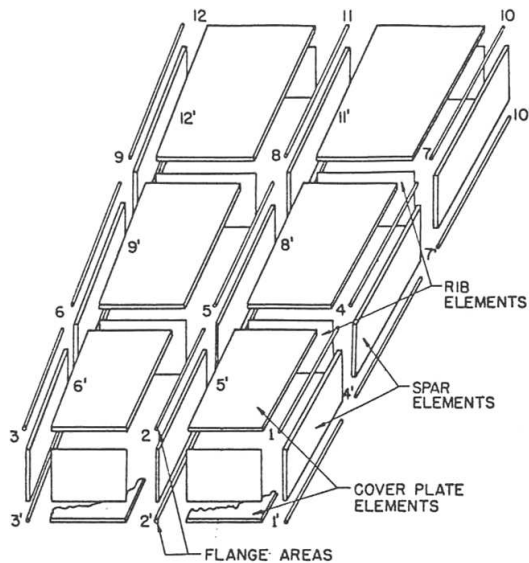
FIG. 3. Wing structure breakdown.
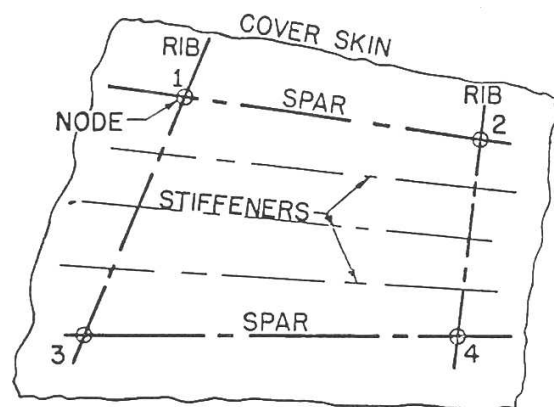
FIG. 6. Stiffened cover skin element.

Fig. 27. Figures of the St. Venant's torsion problem of prismatic bars with arbitrary multi-connected cross sections. Original Fig. 3: Numerical simulation of wing structure breakdown; original Fig. 6: Stiffened cover skin element. The quadrilateral plate element $1 - 2 - 3 - 4$ is assumed to possess in-plane stiffness only. Since two independent displacement components can occur at each node, the order of the K-matrix (stiffness matrix $\boldsymbol{K}$) for this (plane stress) plate element will be 8×8.

Remark

As pointed out in Fig. 2, structural engineers approximated plates, folded plates and shells, especially as stiffened structures, by systems of beams and rods which could be analyzed in the 1930s and 1940s using matrix calculus and electrical calculating machines for solving the algebraic equations. Also the discretization by finite difference methods was further developed, especially in the form of generalized finite difference methods (Mehrstellenverfahren), but herein there arises the problem of developing finite difference stars at boundaries which is a special problem at skew boundaries.

Therefore, the idea of using topologically equal subdomains (finite elements) of the whole domain for equal trial and test functions for all elements in a variational setting was the striking idea and the birth of modern computational mechanics. The main advantage is that the element stiffness matrices only need the topology and the metric of the finite elements themselves but not of neighboured elements and not at and outside the boundary. The second advantage is gained by the fact that the resulting displacements, strains and stresses are the corresponding approximated values of the mathematical model of the real structure with the same topology and metric.

5.5.1. The stiffness matrix for a triangular plane stress plate element with a linear displacement trial and test functions

With the assumption of a linear elastic plane stress problem, the trial and test functions for a linear displacement ansatz are used for deriving the stiffness matrix of such an element, Fig. 28. In matrix

a)
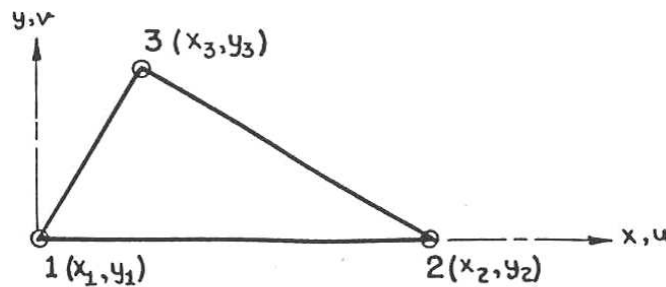
FIG. 8. Node designation for triangular plate element.

b)

Fig. 28. a) Triangular plane stress element with six nodal degrees of freedom $u_{x1}, u_{y1}, u_{x2}, \ldots, u_{y3}$; b) Symmetric stiffness matrix of order 8×8 of the plane stress triangle.

notation, the relation between the fictitious (energy-equivalent) nodal forces F_j and corresponding nodal displacements δ_j is derived, Fig. 28 (b), as, [94],

$$\{F\} = [K]\{\delta\}; \quad [K] = [K]^{\mathrm{T}}, \tag{69}$$

$$\{\delta\}^{\mathrm{T}} = \{u_{x1}\ u_{y1}\ u_{x2}\ \ldots\ u_{y3}\}, \tag{70}$$

$$\{F\}^{\mathrm{T}} = \{F_{x1}\ F_{y1}\ F_{x2}\ \ldots\ F_{y3}\}. \tag{71}$$

$[K]$ has the correct rank 5, i.e. $8 - 3$, where 3 is the number of admitted rigid body modes (two displacements and one rotation).

5.5.2. The Berkeley School of Computational Mechanics

The first strong development of finite element methods for rods, plates and shells took place in the Department of Structural Engineering at the University of California at Berkeley with the eminent Professors Ray Clough, Ed Wilson, R.L. Taylor and their students and postdocs, later professors, such as Jürgen Bathe, Peter Wriggers and Juan Simo, who unfortunately died very early. They developed numerous important finite elements for all types of structures, the algorithms and solvers. R.L. Taylor created especially the program system FEAP with a macro language, which was further continuously extended and is used today in academia all over the world.

5.6. The primal finite element method (FEM) for linear theory of elasticity with matrix notation since the 1960th

It is useful to present at this point the primal FEM for 3D linear elastic systems, based on the principle of virtual work, using matrix notation, [2, 3, 5, 26, 34, 40, 60, 68, 97], as it will be also used for the dual (or stress) FEM as well as for the hybrid stress FEM and the dual mixed FEM.

The mixed BVP of the Lamé's equations is given as follows:

(1a) Definition of the displacement vector and its components as a column vector

$$\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{x}) = \big\{u_1(\boldsymbol{x})\quad u_2(\boldsymbol{x})\quad u_3(\boldsymbol{x})\big\}; \quad \boldsymbol{x}^{\mathrm{T}} = \big\{x_1\quad x_2\quad x_3\big\}, \tag{72}$$

(1b) with the geometric boundary conditions $\boldsymbol{u} - \overline{\boldsymbol{u}} = \boldsymbol{0}$ at Γ_D, $\Gamma_D \cup \Gamma_N = \Gamma = \partial\Omega$, $\Gamma_D \cap \Gamma_N = \varnothing$.

(2) The geometrically linear strain tensor is presented by the column vector of the strain components

$$\boldsymbol{\epsilon}^{\mathrm{T}}(\boldsymbol{x}) = \big\{\epsilon_{11}(\boldsymbol{x})\quad \epsilon_{22}(\boldsymbol{x})\quad \epsilon_{33}(\boldsymbol{x})\quad 2\epsilon_{12}(\boldsymbol{x})\quad 2\epsilon_{23}(\boldsymbol{x})\quad 2\epsilon_{31}(\boldsymbol{x})\big\} \tag{73}$$

and defined as

$$\boldsymbol{\epsilon}(\boldsymbol{x}) := \boldsymbol{D}\boldsymbol{u}(\boldsymbol{x}) \text{ in } \Omega, \tag{74}$$

with the differential operator matrix

$$\boldsymbol{D} = \begin{bmatrix} \partial_1 & & \\ & \partial_2 & \\ & & \partial_3 \\ \partial_1 & \partial_2 & \\ & \partial_3 & \partial_2 \\ \partial_3 & & \partial_1 \end{bmatrix}; \quad \partial_i = \partial/\partial x_i; \tag{75}$$

$\boldsymbol{D}\boldsymbol{u}$ is equivalent to the symmetric gradient operator ∇_{sym}.

(3a) Definition of stresses:

$$\boldsymbol{\sigma}^{\mathrm{T}}(\boldsymbol{x}) = \{\sigma_{11}(\boldsymbol{x}) \quad \sigma_{22}(\boldsymbol{x}) \quad \sigma_{33}(\boldsymbol{x}) \quad \sigma_{12}(\boldsymbol{x}) \quad \sigma_{23}(\boldsymbol{x}) \quad \sigma_{31}(\boldsymbol{x})\}, \tag{76}$$

(3b) Equilibrium conditions:

$$\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma} + \rho\boldsymbol{b} = \boldsymbol{0} \text{ in } \Omega \tag{77}$$

(3c) Tractions and their equilibrium conditions at Neumann boundary, using the Gauss's divergence theorem

$$\boldsymbol{t}(\boldsymbol{x}) = \boldsymbol{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\sigma}(\boldsymbol{x}); \ \boldsymbol{t}(\boldsymbol{x}) = \bar{\boldsymbol{t}}(\boldsymbol{x}) \text{ at } \Gamma_N, \tag{78}$$

with the direction matrix $\boldsymbol{\mathcal{N}}$ which has the same structure as the operator matrix $\boldsymbol{D}$, (75),

$$\boldsymbol{\mathcal{N}} = \begin{bmatrix} \cos(\boldsymbol{n}, \boldsymbol{e}_1) & & \\ & \cos(\boldsymbol{n}, \boldsymbol{e}_2) & \\ & & \cos(\boldsymbol{n}, \boldsymbol{e}_3) \\ \cos(\boldsymbol{n}, \boldsymbol{e}_2) & \cos(\boldsymbol{n}, \boldsymbol{e}_1) & \\ & \cos(\boldsymbol{n}, \boldsymbol{e}_3) & \cos(\boldsymbol{n}, \boldsymbol{e}_2) \\ \cos(\boldsymbol{n}, \boldsymbol{e}_3) & & \cos(\boldsymbol{n}, \boldsymbol{e}_1) \end{bmatrix} \tag{79}$$

(4) The constitutive equation – elasticity law for isotropic material – is

$$\boldsymbol{\sigma} = \mathbb{C}\boldsymbol{\epsilon}, \tag{80}$$

with the symmetric and positive definite matrix of the elasticity tensor

$$\mathbb{C} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & & & \\ \nu & 1-\nu & \nu & & & \\ \nu & \nu & 1-\nu & & & \\ & & & 1-2\nu & & \\ & & & & 1-2\nu & \\ & & & & & 1-2\nu \end{bmatrix}, \tag{81}$$

and

$$\begin{aligned} \mathbb{C} &= \mathbb{C}^{\mathrm{T}}; \ \det\mathbb{C} \neq 0; \\ \boldsymbol{\epsilon}^{\mathrm{T}}\mathbb{C}\boldsymbol{\epsilon} &= \begin{cases} \neq 0 \text{ for } \boldsymbol{\epsilon} \neq \boldsymbol{0} \\ = 0 \text{ for } \boldsymbol{\epsilon} = \boldsymbol{0} \end{cases} \end{aligned} \tag{82}$$

and the inverse or complicance matrix $\overset{*}{\mathbb{C}}$ according to $\boldsymbol{\epsilon}_{phys} = \overset{*}{\mathbb{C}}\boldsymbol{\sigma}; \ \overset{*}{\mathbb{C}} = \mathbb{C}^{-1}$.

The principle of virtual work (weak form of equilibrium) reads for the continuous system

$$\delta\mathcal{A} = \int_{\Omega} \delta\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma}(\boldsymbol{u}) + \rho\boldsymbol{b}) \, \mathrm{d}\Omega + \int_{\Gamma_N} \delta\boldsymbol{u}^{\mathrm{T}}(\bar{\boldsymbol{t}} - \boldsymbol{t}(\boldsymbol{u})) \, \mathrm{d}\Gamma \overset{!}{=} 0, \tag{83}$$

under the conditions that $\boldsymbol{\epsilon}_{phys} = \boldsymbol{\epsilon}_{geom} := \boldsymbol{D}\boldsymbol{u}$ and $\delta\boldsymbol{\epsilon}(\boldsymbol{u})$ are $\mathcal{C}^1$-continuous in Ω, $\delta\boldsymbol{u} = \boldsymbol{0}$ at Γ_D, $\boldsymbol{b}$ is square-integrable in Ω and $\bar{\boldsymbol{t}}$ is quare-integrable at Γ_N.

Partial integration of the first term in the first integral and applying the Gauss's divergence theorem results in

$$\delta\mathcal{A} = \int_{\Gamma_N} (\delta\boldsymbol{u})^{\mathrm{T}} (\underbrace{\boldsymbol{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\sigma}}_{\boldsymbol{t}(\boldsymbol{u})}\ \mathrm{d}\Gamma - \int_{\Omega} \underbrace{((\delta\boldsymbol{u}^{\mathrm{T}})\boldsymbol{D}^{\mathrm{T}})}_{(\boldsymbol{D}\delta\boldsymbol{u})^{\mathrm{T}}=\delta\boldsymbol{\epsilon}_{geom}^{\mathrm{T}}}\ \mathbb{C}(\boldsymbol{D}\boldsymbol{u})\ \mathrm{d}\Omega$$

$$+ \int_{\Omega} \delta\boldsymbol{u}^{\mathrm{T}}\rho\boldsymbol{b}\ \mathrm{d}\Omega + \int_{\Gamma_N} \delta\boldsymbol{u}^{\mathrm{T}}(\bar{\boldsymbol{t}} - \boldsymbol{t}(\boldsymbol{u}))\ \mathrm{d}\Gamma \overset{!}{=} 0, \qquad (84)$$

and finally in the variational form or the principle of virtual work

$$\underbrace{\int_{\Omega} \delta\boldsymbol{\epsilon}^{\mathrm{T}}(\delta\boldsymbol{u})\mathbb{C}\boldsymbol{\epsilon}(\boldsymbol{u})\ \mathrm{d}\Omega}_{a(\boldsymbol{u},\delta\boldsymbol{u})} = \underbrace{\int_{\Omega} (\delta\boldsymbol{u}^{\mathrm{T}}\rho\boldsymbol{b}\ \mathrm{d}\Omega + \int_{\Gamma_N} (\delta\boldsymbol{u})^{\mathrm{T}}\bar{\boldsymbol{t}}\ \mathrm{d}\Gamma}_{l(\delta\boldsymbol{u})} \ , \qquad (85)$$

where a and l are the bilinear and linear forms of the variational problem.

The same result is obtained from the principle of minimum of total potential energy, i.e. the functional $\mathcal{F}(\boldsymbol{u})$

$$\mathcal{F}(\boldsymbol{u}) = \mathcal{U}(\boldsymbol{u}) + \Pi(\boldsymbol{b},\bar{\boldsymbol{t}}) = \frac{1}{2}\int_{\Omega} \boldsymbol{\epsilon}^{\mathrm{T}}(\boldsymbol{u})\mathbb{C}\boldsymbol{\epsilon}(\boldsymbol{u})\ \mathrm{d}\Omega - \int_{\Omega} \boldsymbol{u}^{\mathrm{T}}\rho\boldsymbol{b}\ \mathrm{d}\Omega - \int_{\Gamma_N} \boldsymbol{u}^{\mathrm{T}}\bar{\boldsymbol{t}}\ \mathrm{d}\Omega \overset{!}{=} \min_{u}, \qquad (86)$$

with the stationarity condition

$$\delta\mathcal{F}(\boldsymbol{u}) = \int_{\Omega} \delta\boldsymbol{\epsilon}^{\mathrm{T}}\mathbb{C}\boldsymbol{\epsilon}(\boldsymbol{u})\ \mathrm{d}\Omega - \int_{\Omega} \delta\boldsymbol{u}^{\mathrm{T}}\rho\boldsymbol{b}\ \mathrm{d}\Omega - \int_{\Gamma_N} \delta\boldsymbol{u}^{\mathrm{T}}\bar{\boldsymbol{t}}\ \mathrm{d}\Gamma \overset{!}{=} 0. \qquad (87)$$

Partial integration and divergence theorem, applied to the first term, yields

$$\delta\mathcal{F}(\boldsymbol{u}) = \int_{\Gamma_N} \delta\boldsymbol{u}^{\mathrm{T}}\underbrace{\boldsymbol{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\sigma}(\boldsymbol{u})}_{\boldsymbol{t}(\boldsymbol{u})}\ \mathrm{d}\Gamma - \int_{\Gamma_N} \delta\boldsymbol{u}^{\mathrm{T}}\bar{\boldsymbol{t}}\ \mathrm{d}\Gamma - \int_{\Omega} \delta\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma}(\boldsymbol{u}))\ \mathrm{d}\Omega - \int_{\Omega} \delta\boldsymbol{u}^{\mathrm{T}}\rho\boldsymbol{b}\ \mathrm{d}\Omega \overset{!}{=} 0, \qquad (88)$$

and finally the weak form of equilibrium

$$\delta\mathcal{F}(\boldsymbol{u}) = \int_{\Gamma_N} \delta\boldsymbol{u}^{\mathrm{T}}[\boldsymbol{t}(\boldsymbol{u}) - \bar{\boldsymbol{t}}]\ \mathrm{d}\Gamma - \int_{\Omega} \delta\boldsymbol{u}^{\mathrm{T}}[(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma}(\boldsymbol{u})) + \rho\boldsymbol{b}]\ \mathrm{d}\Omega \overset{!}{=} 0. \qquad (89)$$

Direct variational approximation with finite elements – FEM – using trial and virtual displacements is done as

$$\boldsymbol{u}_{h,e}(\boldsymbol{x}) = \boldsymbol{N}(\boldsymbol{x})\widehat{\boldsymbol{u}}_e;\ \ \delta\boldsymbol{u}_{h,e} = \boldsymbol{N}\delta\widehat{\boldsymbol{u}}_e \text{ in } \Omega_e,\ \text{ with } \boldsymbol{u}_{h,e}(\boldsymbol{x}) \in \mathcal{H}^1(\Omega) \subset \mathbb{R}^3, \qquad (90)$$

with the trial and test shape functions

$$\boldsymbol{N}(\boldsymbol{x}) = \begin{bmatrix} \phi_{u_1}^{\mathrm{T}}(\boldsymbol{x}) & & \\ & \phi_{u_2}^{\mathrm{T}}(\boldsymbol{x}) & \\ & & \phi_{u_3}^{\mathrm{T}}(\boldsymbol{x}) \end{bmatrix}, \qquad (91)$$

where the column vectors $\boldsymbol{\phi}_{u_{i;i=1,2,3}}(\boldsymbol{x})$ are complete polynomials, usually chosen as Lagrange or Legendre polynomials, describing unit displacement states for the introduced nodal degrees of freedom.

The discretized strain takes the form

$$\epsilon_{h,e}(\boldsymbol{x}) = \underbrace{\boldsymbol{D}\boldsymbol{N}(\boldsymbol{x})}_{\boldsymbol{B}(\boldsymbol{x})}\widehat{\boldsymbol{u}}_e; \ \delta\epsilon_{h,e} = \boldsymbol{B}(\boldsymbol{x})\delta\widehat{\boldsymbol{u}}_e, \tag{92}$$

under the conditions that the strains $\epsilon_{h,e} = \boldsymbol{D}\boldsymbol{u}_{h,e}$ and $\delta\epsilon_{h,e}$ are $\mathcal{C}^1$-continuous in Ω_e, the $\boldsymbol{u}_{h,e}$ and $\delta\boldsymbol{u}_{h,e}$ are $\mathcal{C}^1$-continuous at all element interfaces Γ_e, i.e., $[[\boldsymbol{u}_{h,e}]] = \boldsymbol{0}$ at Γ_e; $[[\delta\boldsymbol{u}_{h,e}]] = \boldsymbol{0}$ at Γ_e and $\delta\boldsymbol{u}_{h,e} = \boldsymbol{0}$ at $\Gamma_{D,e}$.

The extension of the principle of virtual work, (85), to the discretized system then reads

$$\bigcup_e\left\{\delta\widehat{\boldsymbol{u}}_e^{\mathrm{T}}\underbrace{\int_{\Omega_e}\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{x})\mathbb{C}\boldsymbol{B}(\boldsymbol{x})\ \mathrm{d}\Omega}_{\boldsymbol{k}_e=\boldsymbol{k}_e^{\mathrm{T}};\ \det\boldsymbol{k}_e=0}\widehat{\boldsymbol{u}}_e\right\} = \bigcup_e\left\{\delta\widehat{\boldsymbol{u}}_e^{\mathrm{T}}\underbrace{\left[\int_{\Omega_e}\boldsymbol{N}^{\mathrm{T}}(\boldsymbol{x})\rho\boldsymbol{b}\ \mathrm{d}\boldsymbol{x} + \int_{\Gamma_{N,e}}\boldsymbol{N}^{\mathrm{T}}(\boldsymbol{x})\overline{\boldsymbol{t}}\ \mathrm{d}\Gamma\right]}_{\widehat{\overline{\boldsymbol{p}}}_e}\right\}. \tag{93}$$

Assembling of the elements to the system by Boolean matrices $\boldsymbol{a}_e$, according to the unknown global reduced nodal displacement vector $\widehat{\boldsymbol{U}}$, where *global* means the geometric assembling of the elements at the nodes and *reduced* means the elimination of nodal displacements at Dirichlet boundaries $\Gamma_{D,e}$, at least avoiding rigid body displacements without linear dependencies, results in

$$\widehat{\boldsymbol{u}}_e = \boldsymbol{a}_e\widehat{\boldsymbol{U}}; \ \delta\widehat{\boldsymbol{u}}_e = \boldsymbol{a}_e\delta\widehat{\boldsymbol{U}}; \ \widehat{\overline{\boldsymbol{p}}}_e = \boldsymbol{a}_e\widehat{\overline{\boldsymbol{P}}}. \tag{94}$$

Of course, these large rectangular incidence matrices $\boldsymbol{a}_e$, which are mainly occupied with zeros, are not directly used but realized by index lists attached as column vectors to the element stiffness matrices (stored as upper triangular matrices), from which they are inserted into the global reduced stiffness matrix.

Inserting (94) into (93) yields the kinematic assembling process

$$\underbrace{\delta\widehat{\boldsymbol{U}}^{\mathrm{T}}}_{\neq 0}\{\bigcup_e[\underbrace{\boldsymbol{a}_e^{\mathrm{T}}\boldsymbol{k}_e\boldsymbol{a}_e}_{\boldsymbol{K}=\boldsymbol{K}^{\mathrm{T}};\ \det\boldsymbol{K}\neq 0}]\widehat{\boldsymbol{U}}\} = \delta\widehat{\boldsymbol{U}}^{\mathrm{T}}\{\underbrace{\bigcup_e\boldsymbol{a}_e^{\mathrm{T}}\widehat{\overline{\boldsymbol{p}}}_e}_{\widehat{\overline{\boldsymbol{P}}}}\} \tag{95}$$

and thus the linear algebraic equation system for solving the global nodal displacement vector

$$\boldsymbol{K}\widehat{\boldsymbol{U}} = \widehat{\overline{\boldsymbol{P}}} \rightsquigarrow \widehat{\boldsymbol{U}} = \boldsymbol{K}^{-1}\widehat{\overline{\boldsymbol{P}}}. \tag{96}$$

The nodal displacements of element e are

$$\boldsymbol{u}_{h,e}(\boldsymbol{x}) = \boldsymbol{N}(\boldsymbol{x})\boldsymbol{a}_e\widehat{\boldsymbol{U}} \tag{97}$$

and the stresses

$$\boldsymbol{\sigma}_{h,e}(\boldsymbol{x}) = \mathbb{C}\boldsymbol{B}(\boldsymbol{x})\boldsymbol{a}_e\widehat{\boldsymbol{U}}. \tag{98}$$

5.7. h- and p-test and trial functions as lagrangian interpolation polynomials in 1D representation

In the 1960s node- and element-wise h-discretizations (h is the characteristic element length) based on Lagrangian polynomials were first used.

For a given set of distinct points ξ_j and numbers η_j, the Lagrangian polynomial is the one of least degree that at each ξ_j assumes the given value η_j at this point. The interpolating polynomial of least degree is unique, but changing the coordinate ξ_j requires a complete recalculation of the interpolant.

For a given set of $k+1$ data points $(\xi_0, \eta_0), \ldots (\xi_j, \eta_j) \ldots (\xi_k, \eta_k)$ with disjoint points j, the interpolation polynomial is a linear combination

$$L(\xi) := \sum_{j=0}^{k} l_j(\xi)$$

of the Lagrangian basis polynomial

$$l_j(\xi) := \prod_{\substack{0 \leq m \leq k \\ m \neq j}} \frac{\xi - \xi_m}{\xi_j - \xi_m} = \frac{\xi - \xi_0}{\xi_j - \xi_0} \frac{\xi - \xi_1}{\xi_j - \xi_1} \cdots \frac{\xi - \xi_{j-1}}{\xi_j - \xi_{j-1}} \frac{\xi - \xi_{j+1}}{\xi_j - \xi_{j+1}} \cdots \frac{\xi - \xi_k}{\xi_j - \xi_k}, \tag{99}$$

and thus

$$l_i(\xi_i) = \prod_{m \neq i} \frac{\xi_i - \xi_m}{\xi_i - \xi_m} = 1 \tag{100}$$

and the orthogonality $l_j(\xi_i) = \delta_{ji} = \delta_{ij}$.

The standard Lagrangian finite element ansatz has a nodal basis, for the polynomial order $p = 1$ with equidistant nodes k, Fig. 29.
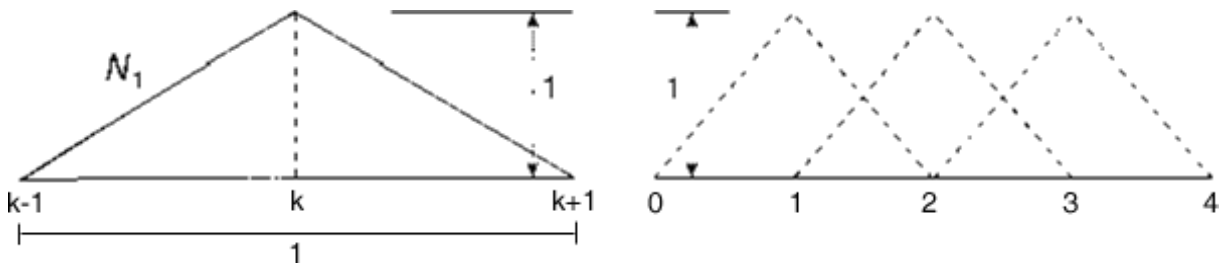
Fig. 29. 1D nodal basis of linear lagrangian shape functions.

The piecewise linear node-based Lagrangian polynomials for $p = 1$ at an inner nodal point k read, using the dimensionless coordinate $\xi = \dfrac{l}{2}$; $0 \leq \xi \leq 1$,

$$N_k(\xi) = \begin{cases} 2\xi, & 0 \leq \xi \leq \dfrac{1}{2} \\ 2\left(\dfrac{1}{2} - \xi\right), & \dfrac{1}{2} \leq \xi \leq 1 \end{cases}. \tag{101}$$

Correspondingly, hierarchical linear shape functions are depicted in Fig. 30.
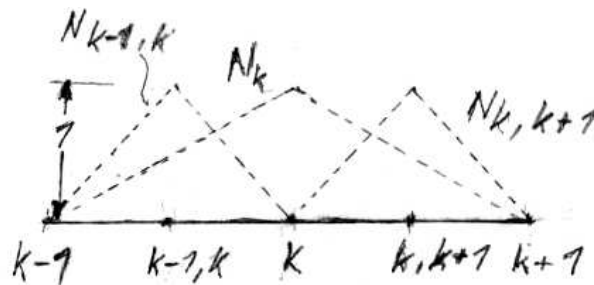
Fig. 30. 1D hierarchical nodal basis of linear Lagrangian shape functions.

Element-wise linear shape functions as lagrangian polynomials are shown in Fig. 31.

The element-wise linear Lagrangian polynomials read

$$N_e(\xi) = \left\{ \begin{array}{l} N_{e,k}(\xi) = 1 - \xi \\ N_{e,k+1}(\xi) = \xi \end{array} \right\} 0 \leq \xi \leq 1. \tag{102}$$
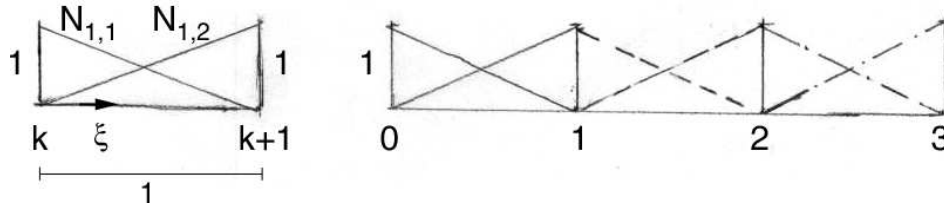
Fig. 31. 1D element basis of linear Lagrangian polynomials.

For higher p orders the element shape functions for $p = 1; 2; 3$ are shown in Fig. 32.
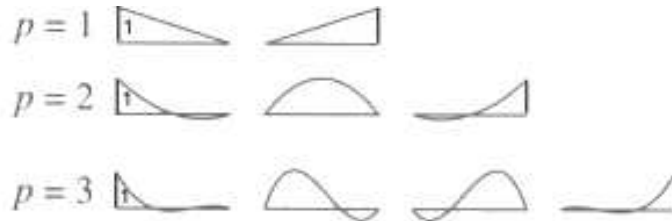
Fig. 32. Lagrangian shape functions of a 1D element for $p = 1; 2; 3$.

The p-version of finite element spaces is constructed – in contrast to the nodal basis of Lagrangian type polynomials – by a hierarchical basis such that the lower order shape functions are contained in the higher order basis, Fig. 33.
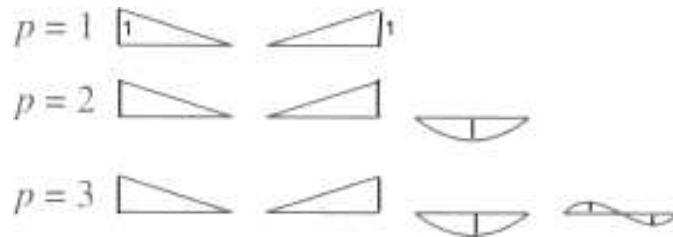
Fig. 33. 1D hierarchical basis functions of the p-version for $p = 1; 2; 3$.

This p-version was first published by Szabó, Babuška et al. in 1978 and 1981, [90] and [12] and then in 1991, [89].

The ansatz spaces for 2D and 3D problems with quadrilateral and hexagonal elements are built by tensor products of the shape functions of Fig. 33.

5.7.1. Cubic Hermitian polynomials for beams

The Euler-Bernoulli static theory of thin elastic beams yields the 4th order differential equation

$$\frac{\mathrm{d}^4 w(x)}{\mathrm{d}x^4} = \frac{p(x)}{EI}. \tag{103}$$

The solutions are in Hilbert spaces $\mathcal{H}^4(\Omega) \in \mathbb{R}^1$. The primal FEM then requires test and trial functions $v_h(x) \in \mathcal{H}_0^2(\Omega_e)$ for finite elements Ω_e, and furthermore the shape functions have to be C^1-continuous at element interfaces, Fig. 34.
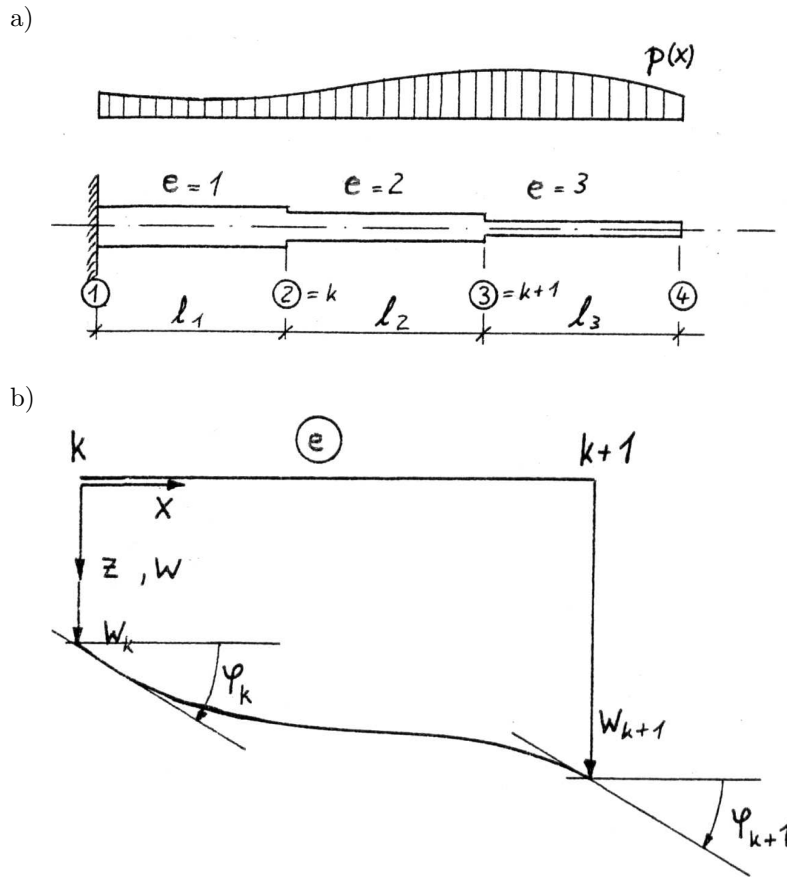
Fig. 34. a) Clamped beam with discretization into three beam elements, each with constant inertia moment I; b) beam element e with the four nodal kinematic unknowns w_k, φ_k, w_{k+1} and φ_{k+1} and the deflection $w(x)$.

Using the dimensionless coordinate $\xi_e = x/l_e$ and introducing the monomial $1; \xi; \xi^2; \xi^3, \xi \equiv \xi_e$, the four cubic Hermitian polynomials (unit displacement states)

$$N_1(\xi) = 1 - 3\xi^2 + 2\xi^3, \tag{104}$$

$$N_2(\xi) = l(\xi - 2\xi^2 + \xi^3), \tag{105}$$

$$N_3(\xi) = 3\xi^2 - 2\xi^3, \tag{106}$$

$$N_4(\xi) = l(-\xi^2 + \xi^3) \tag{107}$$

are derived, [43].

5.7.2. Legendre polynomials

For implicit residual error estimators enhanced h- or p-test spaces are required. Therefore, it is suitable to introduce orthogonal polynomial bases as p- and h-versions. Legendre polynomials as linear combinations of Lagrange polynomials are adequate for this technique, [55, 56].

The Legendre polynomials are defined as

$$P_n(x) := \frac{1}{2^n n!} \frac{d^n}{dx^n} [(x^2 - 1)^n], \tag{108}$$

with the orthogonality property

$$\int_{-1}^{1} P_n P_m \ \mathrm{d}x = \frac{2}{2m+1} \delta_{nm} \ , \ m \geq n; \qquad \delta_{nm} = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases} ;$$

$$P_n(x) := \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}x^n} [(x^2 - 1)^n]. \tag{109}$$

A hierarchical basis $L_n(r)$ based on Legendre polynomials is defined as

$$L_0(r) := \frac{1}{2}(1 - r) \ \ L_1(r) := \frac{1}{2}(1 + r), \tag{110}$$

and for $n \geq 2$ we get

$$L_n(r) := \int_{-1}^{r} P_{n-1}(x) \ \mathrm{d}x = \int_{-1}^{r} \frac{1}{2n-1} \frac{\mathrm{d}}{\mathrm{d}x} (P_n - P_{n-2}) \ \mathrm{d}x \tag{111}$$

and finally

$$L_n(r) := \left[\frac{1}{2n-1} (P_n - P_{n-2}) \right]_{-1}^{r} = \frac{1}{2n-1} [P_n(r) - P_{n-2}(r)]. \tag{112}$$

These hierarchical polynomials have the boundary properties

$$L_0(-1) = 1 \ ; \quad L_0(+1) = 0 \ ; \quad L_1(-1) = 0 \ ; \quad L_1(+1) = 1 \ ; \quad L_n(\pm 1) = 0 \ \ \text{for} \ \ n \geq 2. \tag{113}$$

If n is the maximal order of the actual approximation space then the hierarchically expanded test space is derived by the orthogonality condition

$$\int_{-1}^{+1} [L_n(x) \cdot L_k(x)] \ \mathrm{d}x = 0 \quad \text{for} \quad k < n \quad \text{and} \quad k \neq n - 2. \tag{114}$$

5.7.3. Remark on 2D and 3D ansatz spaces

The extension of 1D shape functions to those for 2D and 3D problems, using Cartesian coordinates, polar-, cylinder- or spherical coordinates is usually done be tensor products of the 1D shape functions.

For the special case of triangular elements area coordinates are favorable, see Subsec. 5.8.

5.8. h- and p-test and trial functions as Lagrangian interpolation polynomials for triangular 2D elements

Suitable shape functions for triangles as nodal unit displacement states are presented with homogeneous area coordinates $\lambda_1, \lambda_2, \lambda_3$, Fig. 35. A point P within the triangle is given by $\lambda_1 = A_1/A, \lambda_2 = A_2/A, \lambda_3 = A_3/A$, with $\lambda_1 + \lambda_2 + \lambda_3 = 1$, i.e. $\lambda_3 = \widehat{\lambda_3}(\lambda_1, \lambda_2)$, [60].

The relation between the area coordinates and the cartesian coordinates of an arbitrary point $P(x_P, y_P)$ is

$$\begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{Bmatrix} = \frac{1}{2A} \begin{bmatrix} A_1^0 & \alpha_1 & \beta_1 \\ A_2^0 & \alpha_2 & \beta_2 \\ A_3^0 & \alpha_3 & \beta_3 \end{bmatrix} \begin{Bmatrix} 1 \\ x_P \\ y_P \end{Bmatrix}, \tag{115}$$
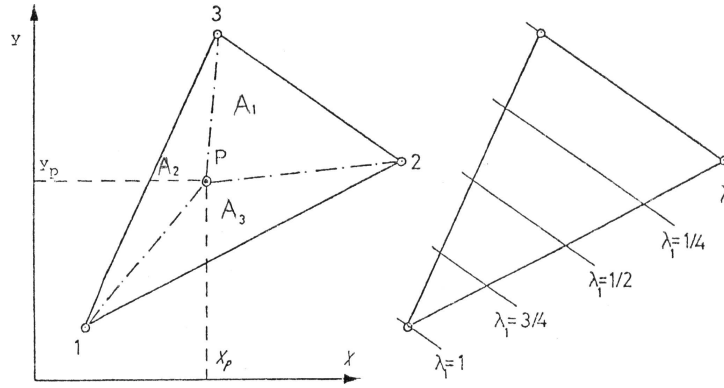
Fig. 35. Unit displacement states using area coordinates λ_i for the nodes 1, 2, 3.

with $A_i^0 = \det \begin{bmatrix} x_{i+1} & y_{i+1} \\ x_{i+2} & y_{i+2} \end{bmatrix}$ and $\begin{aligned} \alpha_i &= y_{i+1} - y_{i+2} \\ \beta_i &= -x_{i+1} + x_{i+2} \end{aligned}$, with $i = 1, 2, 3$ cyclic.

The inverse relation easily reads

$$\begin{Bmatrix} 1 \\ x_P \\ y_P \end{Bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} \begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{Bmatrix}. \tag{116}$$

The bilinear shape functions for the displacements u_x, u_y of an element with six nodal degrees of freedom, Fig. 36, read

$$u_{x,h} = \sum_{k=1}^{3} \lambda_k \widehat{u}_{x,k}, \ u_{y,h} = \sum_{k=1}^{3} \lambda_k \widehat{u}_{y,k}. \tag{117}$$
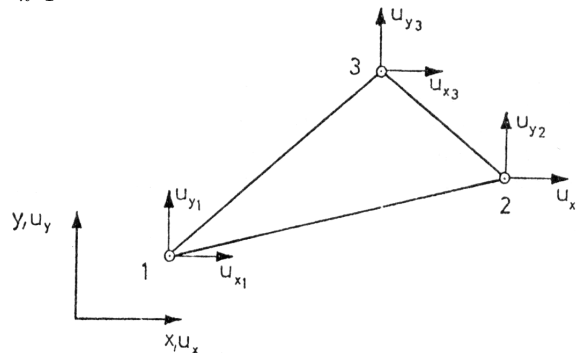
Fig. 36. Plane linear triangular element for 2D linear elastic problems with six nodal degrees of freedom in the three corner points.

The finite element trial functions then become

$$\begin{Bmatrix} u_{x,h} \\ u_{y,h} \end{Bmatrix}_e = \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & \lambda_2 & \lambda_3 \end{bmatrix} \begin{Bmatrix} \widehat{u}_{x,1} & \widehat{u}_{x,2} & \widehat{u}_{x,3} & \widehat{u}_{y,1} & \widehat{u}_{y,2} & \widehat{u}_{y,3} \end{Bmatrix}_e^{\mathrm{T}}. \tag{118}$$

In the same fashion quadratic shape functions can be used as unit displacement states, introducing the three midside points additionally to the corner points.

5.9. Important early finite elements for plane stress and plate bending

It is notable that already in the 1950s and 1960s a lot of displacement elements were derived in the engineering community, especially by Clough, Wilson, Melosh et al. at the University of California

at Berkeley and by Argyris and his collaborators at the University of Stuttgart (using the so-called "natural mode technique", [4]). The following Figs. 37, 38 and 39, elaborated by Buck and Scharpf, [24], show the systematic creation of finite elements with growing polynomial orders.
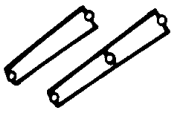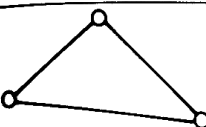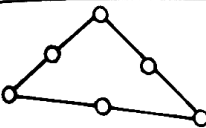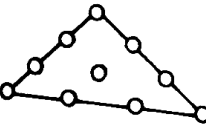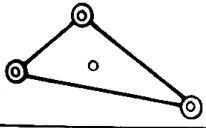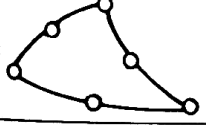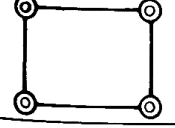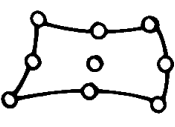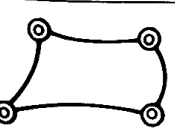
Element geometry	Nodal degrees of freedom (DOFs)	Number of DOFs	Displacement ansatz
	u_x u_y (u_z)	4 resp. 6	linear or quadratic
	u_x u_y	6	complete linear polynomial
	u_x u_y	12	complete quadratic polynomial
	u_x u_y	20 resp. 18	complete cubic polynomial
	u_x u_y $u_{x,x}$ $u_{x,y}$ $u_{y,x}$ $u_{y,y}$	20 resp. 18	complete cubic polynomial
	u_x u_y	12	complete quadratic polynomial in curvilinear coordinates
	u_x u_y	8	bilinear polynomial
	u_x u_y $u_{x,x}$ $u_{x,y}$ $u_{x,xy}$ $u_{y,x}$ $u_{y,y}$ $u_{y,xy}$	32	incomplete 6th order polynomial (Hermite interpolation)
	u_x u_y	8	bilinear polynomial in skew coordinates
	u_x u_y	16 resp. 18	incomplete 4th order polynomial in curvilinear coordinates (Lagrange interpolation)
	u_x u_y $u_{x,x}$ $u_{x,y}$ $u_{y,x}$ $u_{y,y}$	32	incomplete 6th order polynomial in curvilinear coordinates (Hermite interpolation)

Fig. 37. Early rod and plane stress finite elements from the 1950s and 1960s, from [24].

The derivation of element stiffness matrices, especially of plates in bending, by variational methods was given by Stein and Wunderlich, [86].
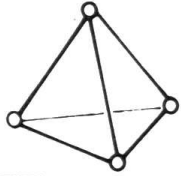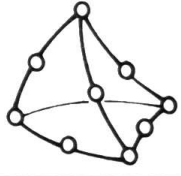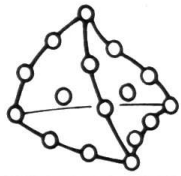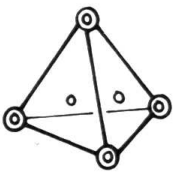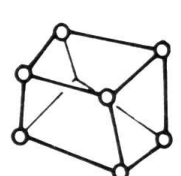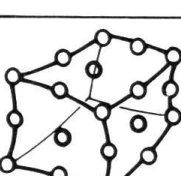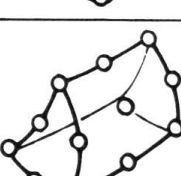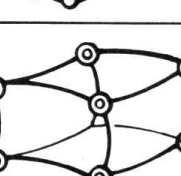
Element geometry	Nodal degrees of freedom (DOFs)			Number of DOFs	Displacement ansatz
	u_x	u_y	u_z	12	complete linear polynomial
	u_x	u_y	u_z	30	complete quadratic polynomial in curvilinear coordinates
	u_x	u_y	u_z	60	complete cubic polynomial in curvilinear coordinates
	u_x $u_{x,x}$ $u_{y,x}$ $u_{z,x}$	u_y $u_{x,y}$ $u_{y,y}$ $u_{z,y}$	u_z $u_{x,z}$ $u_{y,z}$ $u_{z,z}$	60 (resp. 48)	complete cubic polynomial
	u_x	u_y	u_z	24	incomplete cubic polynomial in skew coordinates
	u_x	u_y	u_z	81 resp. 60	incomplete 6th order polynomial in curvilinear coordinates (Lagrange interpolation)
	u_x	u_y	u_z	54	incomplete 6th order polynomial in curvilinear coordinates (Lagrange interpolation)
	u_x $u_{x,x}$ $u_{y,x}$ $u_{z,x}$	u_y $u_{x,y}$ $u_{y,y}$ $u_{z,y}$	u_z $u_{x,z}$ $u_{y,z}$ $u_{z,z}$	96	incomplete 6th order polynomials in curvilinear coordinates (Hermite interpolation)

Fig. 38. Early 3D finite elements for elastic continua from the 1950s and 1960s, from [24].

Element geometry	Nodal degrees of freedom (DOFs)	Number of DOFs	Displacement ansatz
	$w \; w_{,x} \; w_{,y}$	9	incomplete cubic polynomial
	E: $w \; w_{,x} \; w_{,y}$ **M:** $w \; w_{,n}$	15	4th order polynomial with 3 additional higher terms
	E: $w \; w_{,x} \; w_{,y}$ $w_{,xx} \; w_{,xy} \; w_{,yy}$ **M:** $w_{,n}$	21	complete 5th order polynomial
	$w \; w_{,x} \; w_{,y}$ $w_{,xx} \; w_{,xy} \; w_{,yy}$	18	incomplete 5th order polynomial
	E: $w \; w_{,x} \; w_{,y}$ $w_{,xx} \; w_{,xy} \; w_{,yy}$ **M:** $w_{,n}$ **Z:** $w_{,x} \; w_{,y}$ **I:** w	35	complete 7th order polynomial
	$w \; w_{,x} \; w_{,y}$	9	incomplete cubic polynomial in the 3 subdomains
	$w \; w_{,x} \; w_{,y} \; w_{,xy}$	16	incomplete bicubic 6th order polynomial
	$w \; w_{,\xi} \; w_{,\eta} \; w_{,\xi\eta}$	16	bicubic polynomial in skew coordinates
	E: $w \; w_{,x} \; w_{,y}$ **M:** $w_{,n}$	12 resp. 19	cuibc polynomial in each of the 12 subdomains
	E: $w \; w_{,x} \; w_{,y}$ **M:** $w_{,n}$	16	complete cubic polynomial in each of the 4 subdomains

Fig. 39. Early plate bending finite elements for the Kirchhoff-Love plate equation with compatible displacement ansatz functions from the 1950s and 1960s, from [24].

5.10. Isoparametric linear and quadratic quadrilateral elements by Irons and Zienkiewicz

In 1968 B. M. Irons and O.C. Zienkiewicz published the so-called isoparametric 2D elements by bijective mapping unit square elements to arbitrary quadrangles with bilinear shape functions for

the geometry and the elastic displacements, Fig. 40a, with biquadratic shape functions to curved quadrangles in Fig. 40b, and higher polynomials in the same fashion, [49].
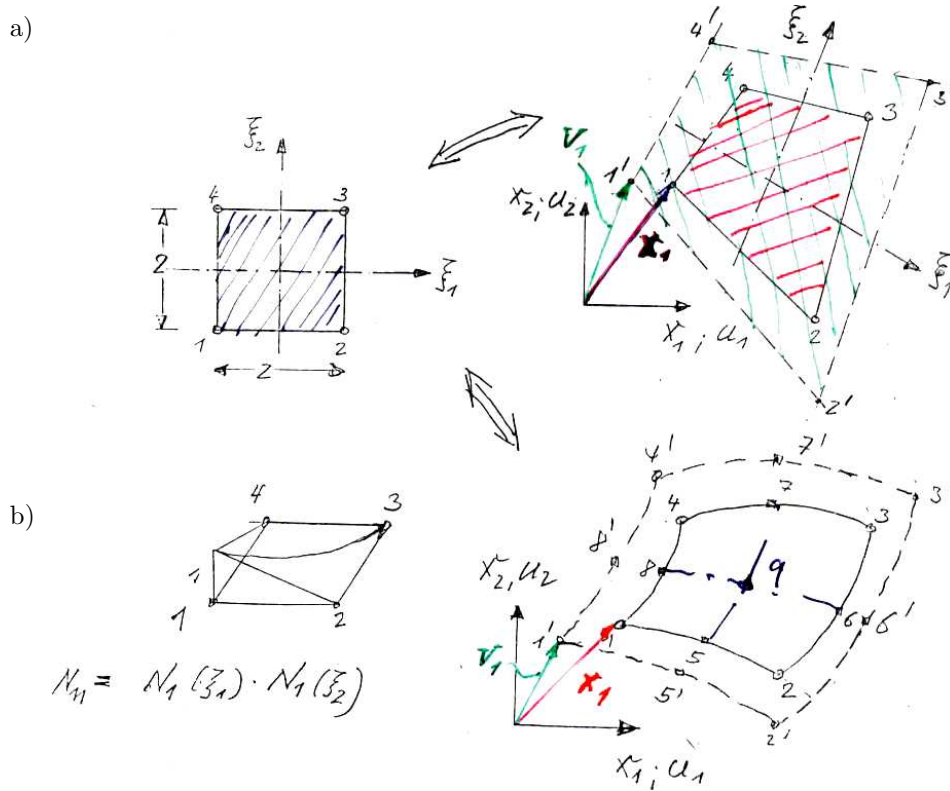
Fig. 40. Isoparametric finite element shape functions with a) bilinear and b) biquadratic polynomials.

The derivatives of the shape functions $N_i(\xi, \eta)$ with their dimensionless coordinates ξ and η are required with respect to the coordinates x and y of the given structural problem, as

$$\frac{\partial N_i}{\partial \xi} = \frac{\partial N_i}{\partial x}\frac{\partial x}{\partial \xi} + \frac{\partial N_i}{\partial y}\frac{\partial y}{\partial \xi}. \tag{119}$$

The shape functions for the bilinear Q1-element read

$$N_i(\xi, \eta) = \frac{1}{4}(1 + \xi_i\xi)(1 + \eta_i\eta). \tag{120}$$

For both derivatives we get in matrix notation

$$\begin{Bmatrix} \dfrac{\partial N_i}{\partial \xi} \\ \dfrac{\partial N_i}{\partial \eta} \end{Bmatrix} = \begin{bmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial y}{\partial \xi} \\ \dfrac{\partial x}{\partial \eta} & \dfrac{\partial y}{\partial \eta} \end{bmatrix} \begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \end{Bmatrix} = \boldsymbol{J} \begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \end{Bmatrix}, \tag{121}$$

with the Jacobian matrix $\boldsymbol{J}$ and the inverse relation

$$\begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \end{Bmatrix} = \boldsymbol{J}^{-1} \begin{Bmatrix} \dfrac{\partial N_i}{\partial \xi} \\ \dfrac{\partial N_i}{\partial \eta} \end{Bmatrix}, \tag{122}$$

with the inverse mapping condition $\det \boldsymbol{J} > 0$.

The strains

$$
\epsilon(x,y) = \left\{ \begin{matrix} \epsilon_x \\ \epsilon_x \\ \gamma_{xy} \end{matrix} \right\} = \underbrace{\begin{bmatrix} \dfrac{\partial}{\partial x} & 0 \\ 0 & \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} \end{bmatrix}}_{\boldsymbol{D}} \left\{ \begin{matrix} u_x(x,y) \\ u_y(x,y) \end{matrix} \right\} = \boldsymbol{D}\boldsymbol{u}(x,y) \tag{123}
$$

are approximated as

$$
\epsilon_h(x,y) = \underbrace{\boldsymbol{D}(x,y)\boldsymbol{J}^{-1}\boldsymbol{\mathcal{N}}(\xi,\eta)}_{\boldsymbol{B}(\xi,\eta)} \widehat{\boldsymbol{u}}_h, \tag{124}
$$

with the vector of nodal displacements $\widehat{\boldsymbol{u}}_h$.

Isoparametric elements for 2D and 3D problems are available in all general purpose finite element programs.

6. THE COMPLEMENTARY FINITE ELEMENT METHOD WITH DISCRETIZED STRESS APPROACHES, THE HYBRID STRESS METHOD AND DUAL MIXED METHODS

In the development of computational mechanics we first have to consider systematic energy-based *matrix structural analysis* of the complementary so-called *force method* and the displacement method for static analysis of statically overdetermined beam systems, truss-works and grillages referring to the names in the right column of Fig. 2, Langefors, Argyris, ..., Fenves from the 1950s to the 1970s.

First the *force method* (based on elastic flexibilities) was developed since the beginning of the 20th century, e.g. by Müller-Breslau, [63], and secondly the displacement method (based on stiffnesses) in the 1920s by Ostenfeld, [67], and many others.

The finite element displacement method (primal FEM) for rods and beams can be easily interpreted as a generalization of the classical displacement method; herein all adjacing beams in a node are rotated together, whereas in primal FEM the connections of the beams with a node are seperated and only reconnected by kinematic assembling of the finite elements to the global system.

In the complementary fashion, the finite element force (or stress) method (complementary FEM) can be seen as a generalization of the classical force method for calculating statically overdetermined beam systems, e.g., frames and trusses; therein the statically overdetermined system has to be weakened by dissolving kinematic continuities, e.g., introducing hinges in 1D continuous beams at the supports and in the same moment related statically overdetermined unknowns. The introduction of linear independent statically determined systems with the related overdetermined unknowns is rather easy for trained engineers but it is complicated in case of automatic generalization for arbitrary systems, realized in a computer program. For this purpose, one needs topology matrices for the recognition of statically determined systems, e.g. in [52] and [62]. Many attempts have been made from the 1950s to the 1970s to automize the traditional *force method* towards a rather general finite element stress method, assigned to John H. Argyris, [2, 3, 5], Fraijs de Veubeke, [34], and Robinson, [76].

6.1. The theorems of Betti and Maxwell, and the theorems of Castigliano and Menabrea

A systematic basis for strucutral analysis of linear elastic static beam and truss systems, governed by ordinary differential equations of 2nd and 4th order was given by Carlo Alberto Castigliano

(1847–1884), [29], and an amendment by Federico Luigi Menabrea (1809–1896), [61], based on the reciprocity theorems of Enrico Betti (1823–1892), [15], and James Clerk Maxwell (1831–1879), [59].

The stored elastic strain energy of a linear elastic solid system reads

$$\mathcal{U}(\boldsymbol{\epsilon}) = \int_\Omega \frac{1}{2}\sigma_{ij}(\boldsymbol{\epsilon})\epsilon_{ij} \ \mathrm{d}\Omega = \frac{1}{2}\mathcal{C}_{ijkl}\epsilon_{kl}\epsilon_{ij}, \ \Omega \subset \mathbb{R}^3, \tag{125}$$

with the symmetric stress components σ_{ij} and the symmetric strain tensor $\boldsymbol{\epsilon} = \epsilon_{ij}\boldsymbol{e}_i \otimes \boldsymbol{\epsilon}_j$, $\epsilon_{ij} := \frac{1}{2}(u_{i,j} + u_{j,i})$; $\boldsymbol{u}(\boldsymbol{x}) = u_i(\boldsymbol{x})\boldsymbol{e}_i$, in matrix notation

$$\mathcal{U}(\boldsymbol{\epsilon}) = \int_\Omega \frac{1}{2}\boldsymbol{\sigma}^\mathrm{T}(\boldsymbol{\epsilon})\boldsymbol{\epsilon} \ \mathrm{d}\Omega = \frac{1}{2}\int_\Omega \boldsymbol{\epsilon}^\mathrm{T}\mathbb{C}\boldsymbol{\epsilon} \ \mathrm{d}\Omega. \tag{126}$$

The constitutive equation $\boldsymbol{\sigma} = \mathbb{C}\boldsymbol{\epsilon}$, $\mathbb{C} = \mathbb{C}^\mathrm{T}$; $\boldsymbol{\epsilon}^\mathrm{T}\mathbb{C}\boldsymbol{\epsilon} > 0$ for $\boldsymbol{\epsilon} > 0$; $\boldsymbol{\epsilon}^\mathrm{T}\mathbb{C}\boldsymbol{\epsilon} = 0$ for $\boldsymbol{\epsilon} = \boldsymbol{0}$, and the complementary constitutive equation $\boldsymbol{\epsilon} = \mathbb{C}^{-1}\boldsymbol{\sigma}$, then yield two conjugate forms of the stored elastic energy, namely the primal form or strain energy

$$\mathcal{U}(\boldsymbol{\epsilon}) = \int_\Omega \frac{1}{2}\boldsymbol{\epsilon}^\mathrm{T}\mathbb{C}\boldsymbol{\epsilon} \ \mathrm{d}\Omega \tag{127}$$

and the dual form or stress energy

$$\overset{*}{\mathcal{U}}(\boldsymbol{\sigma}) = \int_\Omega \frac{1}{2}\boldsymbol{\sigma}^\mathrm{T} \overset{*}{\mathbb{C}} \boldsymbol{\sigma} \ \mathrm{d}\Omega. \tag{128}$$

The work of the given physical forces $\boldsymbol{F}_k$ for static loading at points k along displacements $\boldsymbol{u}_k$, using the superposition law for linear elastic systems, i.e. an arbitrary sequence of loading, is

$$\mathcal{W} = \frac{1}{2}\sum_k \boldsymbol{F}_k \cdot \boldsymbol{u}_k = \frac{1}{2}\sum_k F_k u_k \cos \alpha_k = \frac{1}{2}\sum_k F_k f_{kk}, \tag{129}$$

presuming that the solution spaces of the structural members (stretched or bended or torsioned beams) admit point loads. Otherwise, e.g., for plane stress problem or plates in bending, distributed loads acting within a finite support have to be applied in order to avoid singularities.

The static energy theorem for the isothermal state reads

$$\mathcal{U} = \mathcal{W} \text{ and } \overset{*}{\mathcal{U}} = \mathcal{W} = \overset{*}{\mathcal{W}}, \tag{130}$$

which takes the form, e.g., for stretching of prismatic beams

$$\mathcal{U}_{stretch} = \frac{1}{2}\int_l EA\epsilon^2(x) \ \mathrm{d}x; \quad \overset{*}{\mathcal{U}}_{stretch} = \frac{1}{2}\int_l \frac{N^2(x)}{EA} \ \mathrm{d}x \tag{131}$$

and for bending of beams with Bernoulli-Euler theory

$$\mathcal{U}_{bend} = \frac{1}{2}\int_l EI\kappa^2(x) \ \mathrm{d}x; \quad \overset{*}{\mathcal{U}}_{bend} = \frac{1}{2}\int_l \frac{M^2(x)}{EI} \ \mathrm{d}x. \tag{132}$$

6.1.1. The theorems of Betti and Maxwell

The reciprocity theorem of Betti, mathematician, physicist, and professor at the University of Pisa, is the basis of the force method for statically overdetermined linear elastic beam systems using symmetry properties of the work done by exterior forces, [15].

It is important to state again that the solution spaces of the given 1D-4th order-differential equations and their boundary conditions with Sobolev spaces $\mathcal{H}^4(\Omega)$ as test and solution spaces admit single forces and moments acting at points of a structure without generating singular stresses. This holds for beams, loaded by axial forces (yielding axial normal stresses, modeled by a differential equation of 2nd order), transverse forces (yielding axial bending stresses and transverse shear stresses, approximately modeled by a differential equation of 4th order of Euler-Bernoulli type without transverse shear deformations or of Timoshenko type with constant transverse shear deformations), and also for eccentric transverse forces, leading to shear stresses due to torsion, developed by de St. Venant.

The admissibility of single forces and moments does not hold for all other 2D and 3D linear and nonlinear elastic systems, e.g., for 2D plane stress analysis where stress singularities appear in the vicinity of an acting single force.

For a given force $\boldsymbol{F}_i$ acting on a linear elastic beam at point i, Fig. 41, we get the "eigen-work" for static loading along the displacement $\boldsymbol{u}_i$

$$\mathcal{W}_{ii} = \frac{1}{2}\boldsymbol{F}_i \cdot \boldsymbol{u}_i = \frac{1}{2}F_i(v_i \cos \alpha_i) = \frac{1}{2}F_i f_{ii}, \tag{133}$$

with the projection

$$f_{ii} = u_i \cos \alpha_i, \tag{134}$$

and further introducing

$$f_{ii} = \delta_{ii} F_i; \quad \delta_{ii} = \frac{f_{ii}}{F_i}. \tag{135}$$

Fig. 41. Force $\boldsymbol{F}_i$, acting at point i of a beam, causing the displacement (projection) f_{ii} in the direction of $\boldsymbol{F}_i$.

Next, two forces $\boldsymbol{F}_i$ and $\boldsymbol{F}_k$ are acting one after the other at the points i and k, Fig. 42.
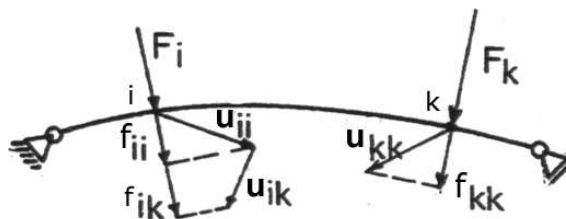
Fig. 42. First $\boldsymbol{F}_i$ is acting at point i and then additionally $\boldsymbol{F}_k$ at point k.

First, only $\boldsymbol{F}_i$ is acting, yielding

$$\mathcal{W}_{ii} = \frac{1}{2}\boldsymbol{F}_i \cdot \boldsymbol{u}_{ii} = \frac{1}{2}F_i f_{ii} = \frac{1}{2}F_i \delta_{ii} F_i. \tag{136}$$

Then, additionally $\boldsymbol{F}_k$ is acting statically at point k, causing the displacement $\boldsymbol{u}_{kk}$ and the "eigenwork"

$$\mathcal{W}_{kk} = \frac{1}{2}\boldsymbol{F}_k \cdot \boldsymbol{u}_{kk} = \frac{1}{2}F_k f_{kk} = \frac{1}{2}F_k \delta_{kk} F_k, \tag{137}$$

with $\delta_k = \dfrac{f_{kk}}{F_k}$.

But $\boldsymbol{F}_i$ is acting in full size during the loading with $\boldsymbol{F}_k$ along $\boldsymbol{u}_{ik}$ (index i marks the point and the index k the cause), and thus does the so-called "foreign work"

$$\mathcal{W}_{ik} = 1\boldsymbol{F}_i \cdot \boldsymbol{u}_{ik} = F_i f_{ik} = F_i \delta_{ik} F_k, \tag{138}$$

with $f_{ik} = \boldsymbol{u}_{ik} \cos\alpha_i; \ \delta_{ik} = \frac{f_{ik}}{F_k}$.

In total we get the work

$$\mathcal{W}^{(1)} = \mathcal{W}_{ii} + \mathcal{W}_{kk} + \mathcal{W}_{ik}. \tag{139}$$

Changing the sequence of loading, i.e., first $\boldsymbol{F}_k$ and then $\boldsymbol{F}_i$ yields the work

$$\mathcal{W}^{(2)} = \mathcal{W}_{kk} + \mathcal{W}_{ii} + \mathcal{W}_{ki}. \tag{140}$$

Because the loading sequence of a linear elastic system is arbitrary according to the superposition theorem, it holds

$$\mathcal{W}^{(1)} = \mathcal{W}^{(2)} \text{ and thus } \mathcal{W}_{ik} = \mathcal{W}_{ki}, \tag{141}$$

which is **Betti's reciprocity theorem**, further yielding

$$\delta_{ik} = \delta_{ki}, \tag{142}$$

Maxwell's theorem, which is Betti's theorem for unit loads; Maxwell was a great physicist, mathematician and famous professor at the Universities of Edinburgh and Cambridge.

For n forces $\boldsymbol{F}_i, \boldsymbol{F}_k$ acting with arbitrary sequence we get the total work by superposition

$$\mathcal{W} = \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}\underbrace{F_i F_k \delta_{ik}}_{\mathcal{W}_{ik}}. \tag{143}$$

Differentiating the total complementary work $\overset{*}{\mathcal{W}}$ with respect to the nominal value F_j of force $\boldsymbol{F}_j$ at point j, we get

$$\frac{\partial \mathcal{W}}{\partial F_j} = \frac{1}{2}\sum_{k=1}^{n}F_k \delta_{jk} + \frac{1}{2}\sum_{i=1}^{n}F_i \delta_{ij}; \ \delta_{ij} = \delta_{ji} \tag{144}$$

$$\frac{\partial \mathcal{W}}{\partial F_j} = \sum_{i=1}^{n}F_i \delta_{ij} = f_i, \tag{145}$$

which is **the 2nd theorem of Castigliano**, also often addressed as **the complementary 1st theorem**.

Using the energy theorem $\mathcal{W} = \overset{*}{\mathcal{U}}$, Eq. (130), we arrive at

$$\frac{\partial \mathcal{W}}{\partial F_j} = \frac{\partial \overset{*}{\mathcal{U}}}{\partial F_j} = \frac{\partial}{\partial F_j} \left[\frac{1}{2} \left(\int_l \frac{N^2(x)}{EA} \, \mathrm{d}x + \int_l \frac{M^2(x)}{EI} \, \mathrm{d}x + \dots \right) \right], \tag{146}$$

with

$$\frac{\partial N(x)}{\partial F_j} = \overline{N}_j(x), \tag{147}$$

which is the function of normal forces due to the unit load $F_j = 1$ at point j, and further

$$\frac{\partial M(x)}{\partial F_j} = \overline{M}_j(x), \tag{148}$$

correspondingly. This finally yields

$$\frac{\partial \overset{*}{\mathcal{U}}}{\partial F_j} = \int_l \underbrace{\frac{N(x)}{EA}}_{\epsilon(x)} \overline{N}_j(x) \, \mathrm{d}x + \int_l \underbrace{\frac{M(x)}{EI}}_{\kappa(x)} \overline{M}_j(x) \, \mathrm{d}x + \dots. \tag{149}$$

6.1.2. The 1st and 2nd theorem of Castigliano and the theorem of Menabrea

The **1st Castigliano's theorem** states, [29],

$$\frac{\partial \mathcal{U}}{\partial u_j} = |\boldsymbol{F}_j| \cos \alpha_j = F_j \tag{150}$$

and the complementary **2nd Castigliano's theorem** reads

$$\frac{\partial \overset{*}{\mathcal{U}}}{\partial F_j} = |\boldsymbol{u}_j| \cos \alpha_j = f_{jj}, \tag{151}$$

where the given forces $\boldsymbol{F}_j$ and the displacements $\boldsymbol{u}_j$ naturally have different directions.

A natural extension of the 2nd Castigliano's theorem to statically overdetermined (unknown) forces X_j was given in the **theorem of Menabrea**, [61], Frederico Luigi Conte di Menabrea (1809–1896), professor of mechanics at the military academy of the University of Torino. It states that the derivative of the complementary energy with respect to static overdetermined unknown X_j is zero,

$$\frac{\partial \overset{*}{\mathcal{U}}}{\partial X_j} = 0, \tag{152}$$

because the X_j have to realize the kinematic continuity which was released by introducing a static determined system and the X_j as exterior forces, according to the cutting principle by Leonard Euler.

For a statically overdetermined linear elastic static system with n unknown forces or/and moments X_i; $i = 1, 2, \dots, n$, introduced in an admissible statically determined system, and m acting forces F_k, the theorem of Menabrea yields the linear algebraic system of elasticity equations for calculating the static unknowns X_i, with a symmetric positive definite equation matrix of the unit deformations δ_{ik}.

With the work of the X_i and F_k,

$$\mathcal{W} = \widehat{\mathcal{W}}(X_i, F_k); \quad i = 1, 2, \dots, n; \quad k = 1, 2, \dots, m \tag{153}$$

the Menabrea theorem postulates

$$\frac{\partial \mathcal{W}}{\partial X_j} = \frac{\partial \overset{*}{\mathcal{U}}}{\partial X_j} = f_j = \sum_{i=1}^{n} \delta_{ij} X_j + \sum_{i=1}^{n} \delta_{i0} \overset{!}{=} 0, \tag{154}$$

with

$$\delta_{ij} = \int\limits_l \frac{\overline{N}_i(x)}{EA} \overline{N}_j(x) \ \mathrm{d}x + \int\limits_l \frac{\overline{M}_i(x)}{EI} \overline{M}_j(x) \ \mathrm{d}x + \ldots \tag{155}$$

and

$$\delta_{i0} = \int\limits_l \frac{\overline{N}_i(x)}{EA} \underbrace{\sum_{k=1}^{m} \overline{N}_k(x) F_k}_{N_0(x)} \ \mathrm{d}x + \int\limits_l \frac{\overline{M}_i(x)}{EI} \underbrace{\sum_{k^*=1}^{m} \overline{M}_{k^*}(x) F_{k^*}}_{M_0(x)} \ \mathrm{d}x + \ldots \ . \tag{156}$$

Thus, the "force method" leads to a linear algebraic system of equations for determing the X_j, in matrix notation

$$[\delta_{ij}]_{(n,n)} \{X_j\}_{(n,1)} = -\{\delta_{i0}\}_{(n,1)}. \tag{157}$$

These equations realize the kinematic continuity of the system at the virtual discontinuities in connection with introducing the X_i, according to Euler's cutting principle.

Extension: The applied loads can also be distributed. The only condition is that the energy expressions are square-integrable.

Remark

As mentioned above, this classical "force method" of matrix structural analysis from the 19th century is hard to generalize for automatic generation of the δ_{ij}-matrices. This needs a topological description and decomposition of the given system into a statically determined one including the choice of the admissible statical unknowns, [52].

Important contributions to the finite element force method originate from de Veubeke, [34].

Thus, a computer based automized finite element force method, applicable for a wide range of structural engineering systems, is scarcely to elaborate, especially in the manner of a finite element method, where the whole structure is virtually dissolved into single beam elements.

However, the complementary force method namely the "displacement method" for beam systems, which was developed by Asger Skovgaard Ostenfeld (1866–1931) since the 1930s, [67], can be organized very efficiently, e.g., for stiff frame systems. One can see the primal FEM with displacement discretizations as a generalization of Ostenfeld's method. Herein, the rotation angle of a node of a frame system with all ending beams is an unknown kinematic quantity which has to be determined by the equilibrium conditions at the node. More effort is necessary for unknown rotation angles of beams. It is important to do this with the principle of virtual work in order to get – as in the force method – symmetric equation systems for the unknown node and beam rotation angles. This was systematically done by John Argyris (1913–2004) in his important papers, [2, 3, 5].

It is evident that it does not provide topological problems in defining the kinematically determined (stiff) systems within the displacement method – in contrast to finding an admissible statically determined (soft) system in the force method.

And there is a straightforward transition from Ostenfeld's displacement method to primal FEM by first dissolving all kinematic connections of beams at the nodes, choosing displacement discretizations with unknown nodal rotations for each single beam (or even parts of it) as finite elements and then postulating all kinematic connectivities by the equilibrium conditions at the introduced nodes.

6.2. Dual or stress finite element method in matrix notation

The principle of complementary virtual work (weak form of the kinematic equations) of a 3D elastic system with pure Dirichlet boundary conditions reads, compare the principle of virtual work (83) and (85),

$$\delta \overset{*}{\mathcal{A}} = \int_\Omega (\delta\boldsymbol{\sigma})^{\mathrm{T}} \, (\underbrace{\overset{*}{\mathbb{C}}\,\boldsymbol{\sigma}}_{\boldsymbol{\epsilon}_{phys}} - \underbrace{\boldsymbol{Du}}_{\boldsymbol{\epsilon}_{geom}}) \, \mathrm{d}\Omega + \int_{\Gamma_D} \delta\boldsymbol{t}^{\mathrm{T}} \underbrace{(\boldsymbol{u}(\boldsymbol{\sigma}) - \overline{\boldsymbol{u}})}_{\overset{!}{=}\mathbf{0}} \, \mathrm{d}\Gamma \overset{!}{=} 0, \tag{158}$$

where the under-brace below the first integrand is $\overset{!}{=}\mathbf{0}$,

under the conditions that the trial and test stresses fulfill the equilibrium conditions

$$\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma} + \rho\boldsymbol{b} = \mathbf{0} \quad \text{in} \quad \Omega \quad \text{and} \quad \boldsymbol{D}^{\mathrm{T}}(\delta\boldsymbol{\sigma}) = \mathbf{0} \quad \text{in} \quad \Omega, \tag{159}$$

and the tractions satisfy the equilibrium conditions at Neumann boundaries,

$$\boldsymbol{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\sigma} =: \boldsymbol{t} = \overline{\boldsymbol{t}} \quad \text{at} \quad \Gamma_N \quad \text{and} \quad \delta(\boldsymbol{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\sigma}) = \delta\boldsymbol{t} = \mathbf{0} \quad \text{at} \quad \Gamma_N, \tag{160}$$

and the physical strains are $\boldsymbol{\epsilon}_{phys} = \overset{*}{\mathbb{C}}\,\boldsymbol{\sigma}$.

Then applying partial integration to the second term of the first integral of (158) and the divergence theorem to the first resulting term, the principle reads

$$\delta \overset{*}{\mathcal{A}} = \int_\Omega (\delta\boldsymbol{\sigma})^{\mathrm{T}} \overset{*}{\mathbb{C}}\,\boldsymbol{\sigma} \, \mathrm{d}\Omega - \int_{\Gamma_D} \underbrace{(\delta\boldsymbol{\sigma}^{\mathrm{T}}\boldsymbol{\mathcal{N}})}_{\delta\boldsymbol{t}^{\mathrm{T}}} \boldsymbol{u} \, \mathrm{d}\Gamma + \int_\Omega \underbrace{(\delta\boldsymbol{\sigma}^{\mathrm{T}}\boldsymbol{D})}_{=\mathbf{0}} \boldsymbol{u} \, \mathrm{d}\Omega + \int_{\Gamma_D} \delta\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{u} - \overline{\boldsymbol{u}}) \, \mathrm{d}\Gamma \overset{!}{=} 0, \tag{161}$$

which yields the principle of virtual stresses or of complementary virtual work as a basis for stress FEM

$$\delta \overset{*}{\mathcal{A}} = \underbrace{\int_\Omega \delta\boldsymbol{\sigma}^{\mathrm{T}} \overset{*}{\mathbb{C}}\,\boldsymbol{\sigma} \, \mathrm{d}\Omega}_{\overset{*}{a}(\boldsymbol{\sigma},\delta\boldsymbol{\sigma})} - \underbrace{\int_{\Gamma_D} \delta\boldsymbol{t}^{\mathrm{T}}\overline{\boldsymbol{u}} \, \mathrm{d}\Gamma}_{\overset{*}{l}(\delta\boldsymbol{\sigma})} \overset{!}{=} 0. \tag{162}$$

This result is also obtained by the stationarity condition for the total complementary energy functional

$$\overset{*}{\mathcal{F}} = \overset{*}{\mathcal{U}} + \overset{*}{\mathit{\Pi}}_{ext} = \frac{1}{2} \int_\Omega \boldsymbol{\sigma}^{\mathrm{T}}\overset{*}{\mathbb{C}}\boldsymbol{\sigma} \, \mathrm{d}\Omega - \int_{\Gamma_D} \boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\sigma})\overline{\boldsymbol{u}} \, \mathrm{d}\Gamma \overset{!}{=} \min_{\boldsymbol{\sigma}},$$

$$\overset{*}{\mathit{\Pi}}_{int} = \overset{*}{\mathcal{U}}, \quad \overset{*}{\mathit{\Pi}}_{ext} = -2 \overset{*}{\mathcal{W}} \quad \text{are potential energies,} \tag{163}$$

under the conditions: $\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma} + \rho\boldsymbol{b} = \mathbf{0}$ in Ω, $\boldsymbol{D}^{\mathrm{T}}\delta\boldsymbol{\sigma} = \mathbf{0}$ in Ω, $\boldsymbol{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\sigma} := \boldsymbol{t} = \overline{\boldsymbol{t}}$ at Γ_N. The stationarity condition reads, subtracting and adding the term $\int_\Omega \delta\boldsymbol{\sigma}^{\mathrm{T}}(\boldsymbol{Du}) \, \mathrm{d}\Omega$,

$$\delta\overset{*}{\mathcal{F}} = \delta\overset{*}{\mathcal{U}} + \delta\overset{*}{\mathit{\Pi}}_{ext} = \int_\Omega \delta\boldsymbol{\sigma}^{\mathrm{T}}\overset{*}{\mathbb{C}}\boldsymbol{\sigma} \, \mathrm{d}\Omega - \int_{\Gamma_D} \delta\boldsymbol{t}^{\mathrm{T}}\overline{\boldsymbol{u}} \, \mathrm{d}\Gamma - \int_\Omega \delta\boldsymbol{\sigma}^{\mathrm{T}}(\boldsymbol{Du}) \, \mathrm{d}\Omega + \int_\Omega \delta\boldsymbol{\sigma}^{\mathrm{T}}(\boldsymbol{Du}) \, \mathrm{d}\Omega, \tag{164}$$

and applying partial integration and the divergence theorem to the last term yields

$$\delta \overset{*}{\mathcal{F}} = \int_\Omega (\delta\boldsymbol{\sigma})^{\mathrm{T}} \, (\underbrace{\overset{*}{\mathbb{C}}\,\boldsymbol{\sigma}}_{\boldsymbol{\epsilon}_{phys}} - \underbrace{\boldsymbol{Du}}_{\boldsymbol{\epsilon}_{geom}}) \, \mathrm{d}\Omega + \int_{\Gamma_D} \delta\boldsymbol{t}^{\mathrm{T}} \underbrace{(\boldsymbol{u}(\boldsymbol{\sigma}) - \overline{\boldsymbol{u}})}_{\overset{!}{=}\mathbf{0}} \, \mathrm{d}\Gamma - \int_\Omega \underbrace{(\delta\boldsymbol{\sigma}^{\mathrm{T}}\boldsymbol{D})}_{\overset{!}{=}\mathbf{0}} \boldsymbol{u} \, \mathrm{d}\Omega \overset{!}{=} 0, \tag{165}$$

which is the principle of virtual stress, i.e. the weak form of kinematic compatibility.

6.2.1. Derivation of the complementary FEM with pure stress discretization

The principle of complementary virtual work for a discretization of the whole domain $\Omega = \bigcup_e \Omega_e$ into n elements Ω_e with equilibrated stresses and $\mathcal{C}^0$-continuous tractions for 3D BVPs reads, [69], see also [34, 40], as well as [52] and [62],

$$\delta \overset{*}{\mathcal{A}}_h = \bigcup_e \left\{ \int_{\Omega_e} \delta \boldsymbol{\sigma}_h^{\mathrm{T}}(\boldsymbol{x}) \overset{*}{\mathbb{C}} \boldsymbol{\sigma}_h(\boldsymbol{x}) \, \mathrm{d}\Omega - \int_{\Gamma_{De}} \delta \boldsymbol{t}_h^{\mathrm{T}}(\boldsymbol{x}) \overline{\boldsymbol{u}} \, \mathrm{d}\Gamma \right\} \overset{!}{=} 0, \ \boldsymbol{x}^{\mathrm{T}} = \left\{ x \ \ y \ \ z \right\} \text{ in } \Omega_e, \tag{166}$$

with the equilibrium conditions for the discretized stresses

$$\boldsymbol{D}^{\mathrm{T}} \boldsymbol{\sigma}_h + \rho \boldsymbol{b} = \boldsymbol{0} \quad \text{in} \quad \Omega_e, \qquad \boldsymbol{D}^{\mathrm{T}} (\delta \boldsymbol{\sigma}_h) = \boldsymbol{0} \quad \text{in} \quad \Omega_e,$$

$$[[\boldsymbol{t}_h]] = \boldsymbol{0} \quad \text{at} \quad \Gamma_e, \qquad [[\delta \boldsymbol{t}_h]] = \boldsymbol{0} \quad \text{at} \quad \Gamma_e \quad \text{and} \quad \delta \boldsymbol{t}_h = \boldsymbol{0} \quad \text{at} \quad \Gamma_{N,e}, \tag{167}$$

yielding the discretized principle of complementary work

$$\delta \overset{*}{\mathcal{A}}_h = \bigcup_e \left\{ \int_{\Omega_e} (\delta \boldsymbol{\sigma}_h)^{\mathrm{T}} (\overset{*}{\mathbb{C}} \boldsymbol{\sigma}_h - \boldsymbol{D} \boldsymbol{u}) \, \mathrm{d}\Omega + \int_{\Gamma_e} \delta \boldsymbol{t}_h^{\mathrm{T}} [[\boldsymbol{u}]] \, \mathrm{d}\Gamma + \int_{\Gamma_{D,e}} \delta \boldsymbol{t}_h^{\mathrm{T}} (\boldsymbol{u} - \overline{\boldsymbol{u}}) \, \mathrm{d}\Gamma \right\} \overset{!}{=} 0. \tag{168}$$

The discrete stress ansatz and its equilibrium conditions read

$$\boldsymbol{\sigma}_h(\boldsymbol{x}) = \boldsymbol{N}_\sigma(\boldsymbol{x}) \widehat{\boldsymbol{s}} \quad \text{in} \quad \Omega_e; \qquad \delta \boldsymbol{\sigma}_h(\boldsymbol{x}) = \boldsymbol{N}_\sigma(\boldsymbol{x}) \delta \widehat{\boldsymbol{s}} \quad \text{in} \quad \Omega_e, \tag{169}$$

fulfilling

$$\boldsymbol{D}^{\mathrm{T}} \boldsymbol{\sigma}_h = \boldsymbol{0} \quad \text{in} \quad \Omega_e \quad \text{and} \quad [[\boldsymbol{t}_h]] = [[\boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{\sigma}_h]] = \boldsymbol{0} \quad \text{at} \quad \Gamma_e, \tag{170}$$

$$\delta \boldsymbol{t}_h = \boldsymbol{\mathcal{N}}^{\mathrm{T}} \delta \boldsymbol{\sigma}_h \quad \text{at} \quad \Gamma_{De}; \qquad \boldsymbol{t}_h = \boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{\sigma}_h = \overline{\boldsymbol{t}} \quad \text{at} \quad \Gamma_{N_e}, \tag{171}$$

and further the constitutive equation

$$\boldsymbol{\epsilon}_{h,phys} = \overset{*}{\mathbb{C}}_e \boldsymbol{\sigma}_h \quad \text{in} \quad \Omega_e; \qquad \overset{*}{\mathbb{C}}_e = \overset{*}{\mathbb{C}}_e^{\mathrm{T}} = \mathbb{C}_e^{-1}, \quad \text{positive definite.} \tag{172}$$

The introduced matrices are definded as

$$\boldsymbol{\sigma}_h^{\mathrm{T}}(\boldsymbol{x}) = \left\{ \sigma_{11,h}(\boldsymbol{x}) \quad \sigma_{22,h}(\boldsymbol{x}) \quad \sigma_{33,h}(\boldsymbol{x}) \quad \tau_{12,h}(\boldsymbol{x}) \quad \tau_{23,h}(\boldsymbol{x}) \quad \tau_{31,h}(\boldsymbol{x}) \right\}, \tag{173}$$

with the parameter ansatz

$$\boldsymbol{\sigma}_h(\boldsymbol{x}) = \boldsymbol{N}_\sigma(\boldsymbol{x}) \widehat{\boldsymbol{s}}_e \text{ in } \Omega_e \tag{174}$$

and the matrix of shape functions

$$\boldsymbol{N}_\sigma(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{\phi}_{\sigma_{11}}^{\mathrm{T}}(\boldsymbol{x}) & & & \\ & \boldsymbol{\phi}_{\sigma_{22}}^{\mathrm{T}}(\boldsymbol{x}) & & \\ & & \ddots & \\ & & & \boldsymbol{\phi}_{\tau_{31}}^{\mathrm{T}}(\boldsymbol{x}) \end{bmatrix}, \tag{175}$$

$\boldsymbol{\phi}_{\sigma_{ij}}(\boldsymbol{x})$ is the vector of complete, linear independent polynomials for each stress component, fulfilling the equilibrium conditions (170) and (171). The unknown nodal stresses of element e are

$$\widehat{\boldsymbol{s}}_e^{\mathrm{T}} = \left\{ \widehat{\boldsymbol{s}}_{e,\sigma_{11}}^{\mathrm{T}} \quad \widehat{\boldsymbol{s}}_{e,\sigma_{22}}^{\mathrm{T}} \quad \widehat{\boldsymbol{s}}_{e,\sigma_{33}}^{\mathrm{T}} \quad \widehat{\boldsymbol{s}}_{e,\tau_{12}}^{\mathrm{T}} \quad \widehat{\boldsymbol{s}}_{e,\tau_{23}}^{\mathrm{T}} \quad \widehat{\boldsymbol{s}}_{e,\tau_{31}}^{\mathrm{T}} \right\}. \tag{176}$$

The matrix of direction cosines at element interfaces with the unit normal vector $\boldsymbol{n}^{\mathrm{T}} = \left\{\cos(\boldsymbol{n}, \boldsymbol{e}_1) \quad \cos(\boldsymbol{n}, \boldsymbol{e}_2) \quad \cos(\boldsymbol{n}, \boldsymbol{e}_3)\right\}$ is given in (79). The virtual interface and boundary tractions are defined as

$$\delta \boldsymbol{t}_h^{\mathrm{T}} = \left\{ t_{h,1} \quad t_{h,2} \quad t_{h,3} \right\} \text{ at } \Gamma_e; \ \bar{\boldsymbol{t}}^{\mathrm{T}} = \left\{ \bar{t}_1 \quad \bar{t}_1 \quad \bar{t}_3 \right\} \text{ at } \Gamma_{N_e}. \tag{177}$$

Equation (166) then takes the form

$$\bigcup_e \{ \delta \widehat{\boldsymbol{s}}_e^{\mathrm{T}} \underbrace{\int_{\Omega_e} \boldsymbol{N}_\sigma^{\mathrm{T}} \overset{*}{\mathbb{C}} \boldsymbol{N}_\sigma \, \mathrm{d}\Omega}_{\boldsymbol{f}_e = \boldsymbol{f}_e^{\mathrm{T}}; \ \det \boldsymbol{f}_e \neq 0; \ \boldsymbol{f}_e \text{ pos. def.;}} \widehat{\boldsymbol{s}}_e - \delta \widehat{\boldsymbol{s}}_e^{\mathrm{T}} \underbrace{\int_{\Gamma_{D_e}} \boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{N}_\sigma \overline{\boldsymbol{u}} \, \mathrm{d}\Gamma}_{\widehat{\overline{\boldsymbol{u}}}_{D_e}} \} \overset{!}{=} 0. \tag{178}$$

The element flexibility matrix $\boldsymbol{f}_e$ is regular because the shape functions $\boldsymbol{N}_\sigma$ are linearly independent.

This formulation is favorable for pure inhomogeneous Dirichlet boundary conditions.

In case of Neumann boundaries the prescribed tractions $\bar{\boldsymbol{t}}$ at Γ_{N_e} enter (178) with terms

$$\int_{\Gamma_{N_e}} \boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{N}_\sigma \widehat{\boldsymbol{s}}_e \, \mathrm{d}\Gamma = \bar{\boldsymbol{t}}_e \text{ at all } \Gamma_{N_e}, \tag{179}$$

which are directly eliminated on element level and added to the right-hand side.

Assembling of the elements to the system – i.e., realizing the static connectivities – can be formally expressed by Boolean matrices $\boldsymbol{b}_e$ and introducing the global unknown nodal stress vector $\widehat{\boldsymbol{S}}$ as

$$\widehat{\boldsymbol{s}}_e = \boldsymbol{b}_e \widehat{\boldsymbol{S}} \ \forall \ \Omega_e. \tag{180}$$

This yields the global discretized weak form for the kinematic equations
$\boldsymbol{\epsilon}_{h,phys,e} = \boldsymbol{\epsilon}_{h,geom,e} \ \forall \ \Omega_e$

$$\underbrace{\delta \widehat{\boldsymbol{S}}^{\mathrm{T}}}_{\neq \boldsymbol{0}} [\underbrace{(\bigcup_e \boldsymbol{b}_e^{\mathrm{T}} \boldsymbol{f}_e \boldsymbol{b}_e)}_{\boldsymbol{F} = \boldsymbol{F}^{\mathrm{T}}, \ \det \boldsymbol{F} \neq 0} \widehat{\boldsymbol{S}} - \underbrace{\bigcup_e \boldsymbol{b}_e^{\mathrm{T}} \widehat{\overline{\boldsymbol{u}}}_{D_e}}_{\widehat{\overline{\boldsymbol{U}}}_D}] = 0, \tag{181}$$

with the global flexibility matrix $\boldsymbol{F}$ and the global given displacement vector $\widehat{\overline{\boldsymbol{U}}}_D$, and thus we get the algebraic equation system and its solution for well-posed BVPs

$$\boldsymbol{F} \widehat{\boldsymbol{S}} = \widehat{\overline{\boldsymbol{U}}}_D \rightsquigarrow \widehat{\boldsymbol{S}} = \boldsymbol{F}^{-1} \widehat{\overline{\boldsymbol{U}}}_D. \tag{182}$$

Remark

It is important to observe that the element flexibility matrix $\boldsymbol{f}_e$ has the full rank if the stress shape functions are linearly independent.

However, the element stiffness matrix $\boldsymbol{k}_e$ in the primal FEM, namely

$$\boldsymbol{k}_e = \int_{\Omega_e} \underbrace{(\boldsymbol{D} \boldsymbol{N}_u)^{\mathrm{T}}}_{\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{x})} \mathbb{C} \, \boldsymbol{B}(\boldsymbol{x}) \, \mathrm{d}\Omega; \ \det \boldsymbol{k}_e = 0, \tag{183}$$

has a rank deficiency which is equal to the number of admitted rigid body motions of the finite element. This is due to the fact that the linearly independent shape functions, $\boldsymbol{N}_u(\boldsymbol{x})$, are differentiated with the operator matrix $\boldsymbol{D}(\boldsymbol{x})$.

Only by implementing the Dirichlet boundary conditions into the assembling process – ensuring that the total system does not have rigid body motions – the assembled reduced global stiffness matrix becomes regular, formulated here with reduced Boolean kinematic connectivity matrices $\boldsymbol{a}^{\mathrm{T}}_{e,red}$, yielding

$$\boldsymbol{K}_{red} = \bigcup_e \boldsymbol{a}^{\mathrm{T}}_{e,red}\boldsymbol{k}_e\boldsymbol{a}_{e,red}; \quad \boldsymbol{K}_{red} = \boldsymbol{K}^{\mathrm{T}}_{red}; \quad \det \boldsymbol{K}_{red} \neq 0. \tag{184}$$

The global algebraic equation system, which weakly fulfills all equilibrium conditions in all elements Ω_e, at element interfaces Γ_e and at Neumann boundaries Γ_{N_e}, reads

$$\boldsymbol{K}_{red}\widehat{\boldsymbol{U}} = \widehat{\widehat{\boldsymbol{P}}} \rightsquigarrow \widehat{\boldsymbol{U}} = \boldsymbol{K}^{-1}_{red}\widehat{\widehat{\boldsymbol{P}}}. \tag{185}$$

6.2.2. The simplest triangular pure stress element

It is evident that the construction of pure finite stress elements is very restricted. The following example shows a triangular plane stress element using complete, rotationally invariant polynomials of 5th order. It is easy to see that polynomials with 3rd and 4th order are not possible.

For the three stresses σ_x, σ_y and σ_{xy} we get $3 \cdot 21 = 63$ parameters. These are reduced to 33 independent parameters according to the equilibrium conditions which have to be pre-fulfilled.

The number of stress parameters is

$$N = 3\left\{\sum_{j=1}^{j_e}[(j+1)+1]\right\} = 3\frac{1}{2}(j_e+1)(j_e+2) \text{ minus } \Delta N = 2\frac{1}{2}j_e(j_e+1), \; j_e = 5.$$

From the reduced number of parameters

$$N_{red} = N - \Delta N = \sum_{j=1}^{j_e}(j+3) = \frac{1}{2}(j_e+1)(j_e+6),$$

we get for $j_e = 5$ the reduced number of degrees of freedom $N_{red} = 33$ and choose the following unknown stresses: 3 stress components at each of the 3 corner nodes, which needs 9 parameters in total. Then $33 - 9 = 24$ unknown parameters remain. They are used for the tractions (with two parameters each) at 4 equidistant nodes of the 3 element sides, yielding $3 \cdot 8 = 24$ parameters, Fig. 43.

For $j_e = 4$ or $j_e = 3$, again choosing the stresses in the corner nodes as unknown parameters, one does not get a divisible equal number of parameters and nodes at element sides.
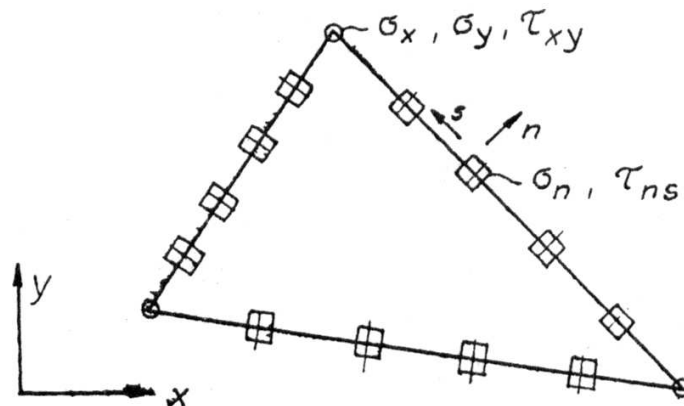
Fig. 43. Finite stress element for plane stress with 33 unknown parameters.

6.3. The so-called hybrid stress method by Pian

It is obvious that the pure stress method is not applicable for more complicated structural systems, especially 3D problems. Therefore, Theodore Pian (1919–2009) had the idea in 1964, [70], to weaken the equilibrium conditions for the tractions at element interfaces and to add these conditions, multiplied with Lagrangian parameters, to the complementary energy functional, resulting in a saddle point problem. We can anticipate the result: the Lagrangian parameters become the displacements at element interfaces. One can reduce the element matrices to pseudo-stiffness matrices and assemble them to the system matrix like in the primal displacement finite element method.

In generalization of the minimal stress functional, (163), the hybrid stationary stress functional $\overset{*}{\mathcal{U}}_{hyb} + \overset{*}{\Pi} =$ stat. with relaxed interface and boundary conditions for the discretized system reads

$$\overset{*}{\mathcal{F}}_{hyb,h} = \overset{*}{\mathcal{U}}_{hyb,h} + \overset{*}{\Pi}_h = \overset{*}{\mathcal{U}}_h + \int\limits_{\cup_e \Gamma_e} \boldsymbol{\lambda}^{\mathrm{T}} [[\boldsymbol{t}_h]] \, \mathrm{d}\Gamma - \int\limits_{\cup_e \Gamma_{N,e}} \boldsymbol{\lambda}^{\mathrm{T}} \overline{\boldsymbol{t}}_h \, \mathrm{d}\Gamma + \overset{*}{\Pi} \to \mathrm{stat.}, \tag{186}$$

with

$$\overset{*}{\mathcal{U}}_h = \frac{1}{2} \int\limits_{\cup_e \Omega_e} \boldsymbol{\sigma}_h^{\mathrm{T}}(\boldsymbol{x}) \, \overset{*}{\mathbb{C}} \, \boldsymbol{\sigma}_h(\boldsymbol{x}) \, \mathrm{d}\Omega; \quad \overset{*}{\Pi}_h = - \int\limits_{\cup_e \Gamma_{D,e}} \boldsymbol{t}_h^{\mathrm{T}}(\boldsymbol{\sigma}_h) \overline{\boldsymbol{u}} \, \mathrm{d}\Gamma. \tag{187}$$

The equilibrium conditions within the element domains are $\boldsymbol{D}^{\mathrm{T}} \boldsymbol{\sigma}_h = \boldsymbol{0}$ in Ω_e; $\boldsymbol{D}^{\mathrm{T}} \delta\boldsymbol{\sigma}_h = \boldsymbol{0}$ in Ω_e; $\boldsymbol{\sigma}_h = \overset{*}{\mathbb{C}} \boldsymbol{\epsilon}_{phys,h}$, whereas the interface and boundary conditions for the tractions, i.e. $[[\boldsymbol{t}_h]] \neq \boldsymbol{0}$ at Γ_e; $\boldsymbol{t}_h - \overline{\boldsymbol{t}} \neq \boldsymbol{0}$ at $\Gamma_{N,e}$, which are fulfilled approximately by the discrete variational process. Due to the Lagrangian extension only a stationary value can be achieved. The necessary stationarity condition reads

$$\delta \overset{*}{\mathcal{F}}_{hyb,h} = \bigcup_e \left\{ \int\limits_{\Omega_e} \delta\boldsymbol{\sigma}_h^{\mathrm{T}} \overset{*}{\mathbb{C}} \, \boldsymbol{\sigma}_h^{\mathrm{T}} \, \mathrm{d}\Omega + \int\limits_{\Gamma_e} \delta\boldsymbol{\lambda}^{\mathrm{T}} [[\boldsymbol{t}_h(\boldsymbol{\sigma}_h)]] \, \mathrm{d}\Gamma + \int\limits_{\Gamma_{N,e}} \delta\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{t}_h \, \mathrm{d}\Gamma \right.$$
$$\left. + \int\limits_{\Gamma_e} \boldsymbol{\lambda}^{\mathrm{T}} [[\delta\boldsymbol{t}_h]] \, \mathrm{d}\Gamma - \int\limits_{\Gamma_{N,e}} \delta\boldsymbol{\lambda}^{\mathrm{T}} \overline{\boldsymbol{t}} \, \mathrm{d}\Gamma - \int\limits_{\Gamma_{D,e}} \delta\boldsymbol{t}_h^{\mathrm{T}}(\boldsymbol{\sigma}_h) \overline{\boldsymbol{u}} \, \mathrm{d}\Gamma \right\} = 0. \tag{188}$$

Subtracting and adding $\int\limits_{\cup_e \Omega_e} \delta\boldsymbol{\sigma}_h^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{u}_h) \, \mathrm{d}\Omega$ and applying partial integration to the positive term yields $\int\limits_{\Gamma_e} \boldsymbol{u}_h^{\mathrm{T}} [[\delta\boldsymbol{t}_h]] \, \mathrm{d}\Gamma + \int\limits_{\Gamma_e} \boldsymbol{\lambda}^{\mathrm{T}} [[\delta\boldsymbol{t}_h]] \, \mathrm{d}\Gamma$, and thus the result for the Lagrangian parameters

$$\boldsymbol{\lambda}_h = -\boldsymbol{u}_h \text{ at } \Gamma_e. \tag{189}$$

Then, the discrete stationarity condition finally reads

$$\delta \overset{*}{\mathcal{F}}_{hyb,h} = \bigcup_e \left\{ \int\limits_{\Omega_e} \delta\boldsymbol{\sigma}_h^{\mathrm{T}} [\overset{*}{\mathbb{C}} \, \boldsymbol{\sigma}_h - \boldsymbol{D}\boldsymbol{u}(\boldsymbol{\sigma}_h)] \, \mathrm{d}\Omega + \int\limits_{\Gamma_e} [[\delta\boldsymbol{t}_h^{\mathrm{T}}]][\boldsymbol{\lambda} - \boldsymbol{u}_h] \, \mathrm{d}\Gamma \right.$$
$$\left. + \int\limits_{\Gamma_e} \delta\boldsymbol{u}_h^{\mathrm{T}} [[\boldsymbol{t}_h]] \, \mathrm{d}\Gamma + \int\limits_{\Gamma_{N,e}} \delta\boldsymbol{u}^{\mathrm{T}} [\boldsymbol{t}_h - \overline{\boldsymbol{t}}] \, \mathrm{d}\Gamma - \int\limits_{\Gamma_{D,e}} \delta\boldsymbol{t}_h^{\mathrm{T}} [\boldsymbol{u}_h - \overline{\boldsymbol{u}}] \, \mathrm{d}\Gamma \right\} = 0. \tag{190}$$

All the terms in the square brackets are approximated to be weakly zero.

From (188) to (190), the stationarity condition as the basis for hybrid stress FEM then results in

$$
\delta \overset{*}{\mathcal{F}}_{hyb,h} = \bigcup_{e} \left\{ \int_{\Omega_e} \delta\boldsymbol{\sigma}_h^{\mathrm{T}} \overset{*}{\mathbb{C}} \, \boldsymbol{\sigma}_h \, \mathrm{d}\Omega - \int_{\Gamma_e} (\delta\boldsymbol{\sigma}_h^{\mathrm{T}} \boldsymbol{\mathcal{N}}) \boldsymbol{u}_h \, \mathrm{d}\Gamma - \int_{\Gamma_e} \delta\boldsymbol{u}_h^{\mathrm{T}} \boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{\sigma}_h \, \mathrm{d}\Gamma \right.
$$

$$
\left. + \int_{\Gamma_{N,e}} \delta\boldsymbol{u}_h^{\mathrm{T}} \overline{\boldsymbol{t}}_h \, \mathrm{d}\Gamma - \int_{\Gamma_{D,e}} \delta(\boldsymbol{\sigma}_h^{\mathrm{T}} \boldsymbol{\mathcal{N}}) \overline{\boldsymbol{u}} \, \mathrm{d}\Gamma \right\} \overset{!}{=} 0. \quad (191)
$$

The trial and test functions for the stresses are

$$
\begin{aligned}
\boldsymbol{\sigma}_h &= \boldsymbol{N}_\sigma(\boldsymbol{x})\widehat{\boldsymbol{s}}_e, \quad \text{with} \quad \boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma}_h = \boldsymbol{0} \quad \text{in} \quad \Omega_e; \\
\delta\boldsymbol{\sigma}_h &= \boldsymbol{N}_\sigma(\boldsymbol{x})\delta\widehat{\boldsymbol{s}}_e, \quad \text{with} \quad \boldsymbol{D}\delta\boldsymbol{\sigma}_h = \boldsymbol{0} \quad \text{in} \quad \Omega_e,
\end{aligned} \quad (192)
$$

with

$$
\boldsymbol{N}_\sigma(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{\phi}_{\sigma_{11}}^{\mathrm{T}}(\boldsymbol{x}) & & & \\ & \boldsymbol{\phi}_{\sigma_{22}}^{\mathrm{T}}(\boldsymbol{x}) & & \\ & & \ddots & \\ & & & \boldsymbol{\phi}_{\sigma_{31}}^{\mathrm{T}}(\boldsymbol{x}) \end{bmatrix}, \quad (193)
$$

and the trial and test displacements (Lagrangian parameters) at element interfaces

$$
-\boldsymbol{\lambda} = \boldsymbol{u}_h = \boldsymbol{N}_u(\boldsymbol{x})\widehat{\boldsymbol{u}}_e; \quad \text{with} \quad [\![\boldsymbol{u}_h]\!] = \boldsymbol{0} \quad \text{at} \quad \Gamma_e;
$$

$$
\delta\boldsymbol{u}_h = \boldsymbol{N}_u(\boldsymbol{x})\delta\widehat{\boldsymbol{u}}_e, \quad \text{with} \quad [\![\delta\boldsymbol{u}_h]\!] = \boldsymbol{0} \quad \text{at} \quad \Gamma_e,
$$

$$
\text{with} \quad \boldsymbol{N}_u(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{\phi}_{u_1}^{\mathrm{T}}(\boldsymbol{x}) & & \\ & \boldsymbol{\phi}_{u_2}^{\mathrm{T}}(\boldsymbol{x}) & \\ & & \boldsymbol{\phi}_{u_3}^{\mathrm{T}}(\boldsymbol{x}) \end{bmatrix}. \quad (194)
$$

The stationarity condition, Eq. (191), reads in discrete form

$$
\bigcup_{e} \left\{ \delta\widehat{\boldsymbol{s}}_e^{\mathrm{T}} \left[\underbrace{\int_{\Omega_e} \boldsymbol{N}_\sigma^{\mathrm{T}}(\boldsymbol{x}) \overset{*}{\mathbb{C}} \boldsymbol{N}_\sigma(\boldsymbol{x}) \, \mathrm{d}\Omega}_{\boldsymbol{f}_e} \widehat{\boldsymbol{s}}_e - \underbrace{\int_{\Gamma_e} (\boldsymbol{N}_\sigma^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\mathcal{N}})\boldsymbol{N}_u(\boldsymbol{x}) \, \mathrm{d}\Gamma}_{\boldsymbol{l}_e} \widehat{\boldsymbol{u}}_e - \underbrace{\int_{\Gamma_{D,e}} \boldsymbol{\mathcal{N}}\boldsymbol{N}_\sigma^{\mathrm{T}}(\boldsymbol{x})\overline{\boldsymbol{u}} \, \mathrm{d}\Gamma}_{\widehat{\overline{\boldsymbol{u}}}_{D,e}} \right] \right.
$$

$$
\left. + \delta\widehat{\boldsymbol{u}}_e^{\mathrm{T}} \left[-\underbrace{\int_{\Gamma_e} (\boldsymbol{D}\boldsymbol{N}_u(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{N}_\sigma(\boldsymbol{x}) \, \mathrm{d}\Gamma}_{\boldsymbol{l}_e^{\mathrm{T}}} \widehat{\boldsymbol{s}}_e - \underbrace{\int_{\Gamma_{N,e}} \boldsymbol{N}_u^{\mathrm{T}}(\boldsymbol{x})\overline{\boldsymbol{t}} \, \mathrm{d}\Gamma}_{\widehat{\overline{\boldsymbol{t}}}_{N,e}} \right] \right\} \overset{!}{=} 0, \quad (195)
$$

with the row-regular Lagrangian matrices $\boldsymbol{l}_e$ and $\boldsymbol{l}_e^{\mathrm{T}}$ with integrals on element interfaces.

This yields in condensed form

$$
\bigcup_{e} \left\{ \underbrace{\{\delta\widehat{\boldsymbol{s}}_e^{\mathrm{T}} \quad \delta\widehat{\boldsymbol{u}}_e^{\mathrm{T}}\}}_{\neq \boldsymbol{0}} \left(\underbrace{\begin{bmatrix} \boldsymbol{f}_e & -\boldsymbol{l}_e \\ -\boldsymbol{l}_e^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{Bmatrix} \widehat{\boldsymbol{s}}_e^{\mathrm{T}} \\ \widehat{\boldsymbol{u}}_e^{\mathrm{T}} \end{Bmatrix} - \begin{Bmatrix} \widehat{\overline{\boldsymbol{u}}}_{D,e} \\ \widehat{\overline{\boldsymbol{t}}}_{N,e} \end{Bmatrix}}_{\overset{!}{=}\boldsymbol{0}} \right) \right\} \overset{!}{=} 0. \quad (196)
$$

The unknown nodal stresses can be eliminated on element level. The first equation yields $\widehat{\boldsymbol{s}}_e = \boldsymbol{f}_e^{-1}\boldsymbol{l}_e\widehat{\boldsymbol{u}}_e + \boldsymbol{f}_e^{-1}\widehat{\boldsymbol{u}}_{D,e}$, and inserted in the second one results in

$$\underbrace{\boldsymbol{l}_e^{\mathrm{T}}\boldsymbol{f}_e^{-1}\boldsymbol{l}_e}_{\overset{*}{\boldsymbol{k}}_e=\overset{*}{\boldsymbol{k}}_e^{\mathrm{T}}}\widehat{\boldsymbol{u}}_e + \underbrace{\boldsymbol{l}_e^{\mathrm{T}}\boldsymbol{f}_e^{-1}\widehat{\boldsymbol{u}}_{D,e} + \widehat{\overline{\boldsymbol{t}}}_{N,e}}_{\widehat{\overline{\boldsymbol{t}}}^{*}} = \boldsymbol{0}. \tag{197}$$

The elimination of $\widehat{\boldsymbol{s}}_e$ is only possible by the regularity of the element flexibility matrix $\boldsymbol{f}_e$; $\overset{*}{\boldsymbol{k}}_e$ can be called a pseudo-element stiffness matrix.

The hybrid stress method has the benefit of kinematic assembling of the given system like in primal FEM because the remaining unknown degrees of freedom are nodal displacements.

6.3.1. A hybrid rectangular plane stress element

Pian published his first hybrid stress element in 1964, [69, 70], Fig. 44.
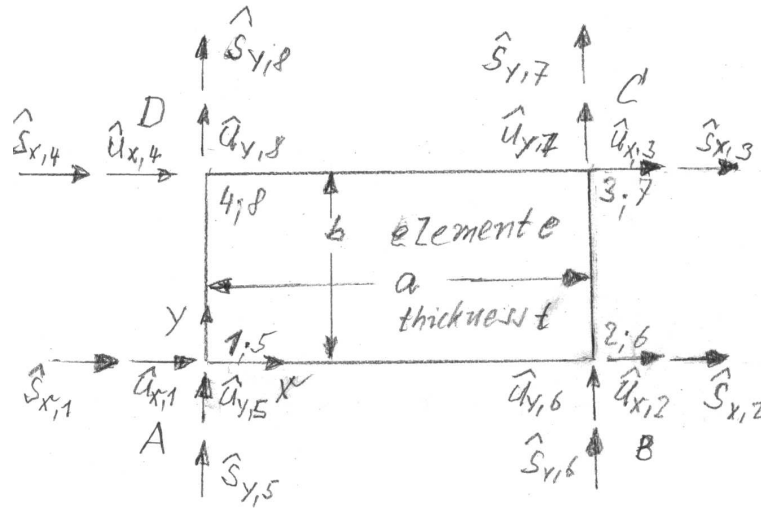
Fig. 44. Hybrid rectangular plane stress element by T.H.H. Pian with bilinear shape functions for stresses and displacements.

The stress ansatz is

$$\begin{aligned}
\sigma_x &= \beta_1 + \beta_2 y & +\beta_6 x & & +\beta_8 y^2 + \beta_{10}x^2 & +\ldots \\
\sigma_y &= \beta_3 + \beta_4 x & +\beta_7 y & & +\beta_9 x^2 + \beta_{10}y^2 & +\ldots \\
\tau_{xy} &= \beta_5 & -\beta_6 y - \beta_7 x & \quad -2\beta_{10}xy & & +\ldots,
\end{aligned} \tag{198}$$

which fulfills the homogeneous equilibrium conditions $\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma} = \boldsymbol{0}$ in Ω_e.

Using the first five parameters for a linear stress approximation with the nodal values

$$\widehat{\boldsymbol{\beta}}_e = \left\{ \beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5 \right\}^{\mathrm{T}}, \tag{199}$$

the matrix of stress shape functions is

$$\phi_\sigma = \begin{bmatrix} 1 & y & & & \\ & & y & x & \\ & & & & 1 \end{bmatrix}. \tag{200}$$

With the ansatz (198)–(200) the boundary tractions are

$$
\boldsymbol{t}_{\Gamma_e}(x,y) = t \left\{ \begin{array}{c} -\tau_{AB}(x) \\ -\sigma_{y,AB}(x) \\ \sigma_{x,BC}(y) \\ \tau_{BC}(y) \\ \tau_{DC}(x) \\ \sigma_{y,DC}(x) \\ -\sigma_{x,AD}(y) \\ -\tau_{AD}(y) \end{array} \right\} = \underbrace{\left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & -x & 0 \\ 1 & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & x & 0 \\ 1 & -y & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right]}_{\boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{N}_\sigma} \underbrace{\left(\begin{array}{c} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{array} \right)}_{\boldsymbol{\beta}_e},
\tag{201}
$$

in closed form $\boldsymbol{t}_{\Gamma_e} = \boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{N}_\sigma \boldsymbol{\beta}_e$, and introducing the energy equivalent fictitious nodal forces $\widehat{\boldsymbol{s}}_e := \{\widehat{s}_{x1}, \ \widehat{s}_{x,2} \ \ \widehat{s}_{x,3} \ \ \widehat{s}_{x,4} \ \ \widehat{s}_{y,5} \ \ \widehat{s}_{y,6} \ \ \widehat{s}_{y,7} \ \ \widehat{s}_{y,8}\}$ with $\widehat{\boldsymbol{s}}_e = \boldsymbol{\mathcal{N}}^{\mathrm{T}} \boldsymbol{t}_{\Gamma_e}(x,y)$ as the input for the Lagrangian matrix $\boldsymbol{l}_e$.

The displacements at element interfaces are also approximated linearly as

$$
u_{x,h}(x, y = 0) = \widehat{u}_{x,1}\left(1 - \frac{x}{a}\right) + \widehat{u}_{x,2}\frac{x}{a}
$$

$$
u_{y,h}(x, y = 0) = \widehat{u}_{y,1}\left(1 - \frac{x}{a}\right) + \widehat{u}_{y,2}\frac{x}{a}
\tag{202}
$$

$$
\dots,
$$

and in total the boundary displacements are

$$
\boldsymbol{u}_{\Gamma_e}(x,y) = \boldsymbol{N}_u(x,y)\widehat{\boldsymbol{u}}_e,
\tag{203}
$$

$$
\underbrace{\left\{ \begin{array}{c} u_{x,AB}(x) \\ u_{y,AB}(y) \\ u_{x,BC}(y) \\ u_{y,BC}(y) \\ u_{x,DC}(x) \\ u_{y,DC}(x) \\ u_{x,AD}(y) \\ u_{y,AD}(y) \end{array} \right\}}_{\boldsymbol{u}_{\Gamma_e}(x,y)} = \underbrace{\left[\begin{array}{cccccccc} 1 - \dfrac{x}{a} & \dfrac{x}{a} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 - \dfrac{x}{a} & \dfrac{x}{a} & 0 & 0 \\ 0 & 1 - \dfrac{y}{b} & \dfrac{y}{b} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 - \dfrac{y}{b} & \dfrac{y}{b} & 0 \\ 0 & 0 & \dfrac{x}{a} & 1 - \dfrac{x}{a} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dfrac{x}{a} & 1 - \dfrac{x}{a} \\ 1 - \dfrac{y}{b} & 0 & 0 & \dfrac{y}{b} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 - \dfrac{y}{b} & 0 & 0 & \dfrac{y}{b} \end{array} \right]}_{\boldsymbol{N}_u(x,y)} \underbrace{\left\{ \begin{array}{c} \widehat{u}_{x,1} \\ \widehat{u}_{x,2} \\ \widehat{u}_{x,3} \\ \widehat{u}_{x,4} \\ \widehat{u}_{y,5} \\ \widehat{u}_{y,6} \\ \widehat{u}_{y,7} \\ \widehat{u}_{y,8} \end{array} \right\}}_{\widehat{\boldsymbol{u}}_e},
\tag{204}
$$

where $\boldsymbol{N}_u(x,y)$ is the matrix of unit displacement states of the element boundaries for unit displacements of the four corner nodes.

This leads us to the requested Lagrangian coupling matrix $\boldsymbol{l}_e$, with integrals over the element boundaries, according to (195), yielding

$$\boldsymbol{l}_e = \int\limits_{\Gamma_e} (\boldsymbol{N}_\sigma^{\mathrm{T}}(x,y)\boldsymbol{\mathcal{N}})\boldsymbol{N}_u(x,y) \, \mathrm{d}s \tag{205}$$

$$= \begin{bmatrix} -b/2 & b/2 & b/2 & -b/2 & 0 & 0 & 0 & 0 \\ -b^2/6 & b^2/6 & b^2/3 & -b^2/3 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & -a/2 & -a/2 & a/2 & a/2 \\ 0 & 0 & 0 & 0 & -a^2/2 & -a^2/2 & a^2/3 & a^2/6 \\ -a/2 & -a/2 & a/2 & a/2 & -b/2 & b/2 & b/2 & -b/2 \end{bmatrix} \tag{206}$$

The element flexibility matrix $\boldsymbol{f}_e = \int\limits_{\Omega_e} \boldsymbol{N}_\sigma^{\mathrm{T}}\mathbb{C}^{-1}\boldsymbol{N}_\sigma \, \mathrm{d}\Omega$ is calculated according to Subsec. 6.2, (178), and equivalently the right-hand side vectors, (195).

This element has the analytically expected convergence properties, i.e., linear for stresses and displacements. However, the extension to general quadrilaterals is practically impossible, and also the generalization to 3D problems was not realized effectively.

6.4. Dual mixed FEM for linear elastic problems

6.4.1. The Hellinger-Prange-Reissner two-field functional

The (one-field) displacement FEM has the significant drawback that the approximation of the stress field usually has jumps on element interfaces, since it is obtained from derivatives of $\mathcal{C}^0$-continuous displacements at element interfaces.

Moreover, low order displacement elements may provide robustness problems with non-existing global interpolation constants, showing up as numerical instabilities, e.g., as locking phenomena in case of nearly incompressible elastic materials or in case of Timoshenko beam theory and Reissner-Mindlin plate theory with approximated constant transverse shear deformations along the thickness. It can be shown that in general robustness problems require mixed stress and displacement approximations resulting in saddle point problems with global infsup-conditions for achieving numerical stability.

It has to be mentioned that many engineering attempts have been made for overcoming the global infsup-conditions by reduced integration and by local counting and balancing of the nodal stress and displacement quantities, also using various patch tests. But these techniques can not replace the infsup-condition for the mixed functional in general.

As already shown in the previous Subsec. 6.3 for the dual and hybrid dual FEM, a mixed FEM based on primal FEM, i.e., the principle of virtual work, is not advantageous for mixed generalizations because the element stiffness matrices have reduced rank and thus can not be inverted and eliminated on element level.

Thus, the dual mixed FEM can be understood as a generalization of the hybrid stress method or as a method sui generis, to approximate both stresses and displacements in the element domains via the Legendre transformation of the specific complementary energy, where the elements of the Lagrangian coupling matrices are integrals over the finite element domains – i.e., not only on element interfaces as for the hybrid stress method, [42, 71–73].

The complementary volume-specific strain-energy functional $\overset{*}{u}(\boldsymbol{\sigma})$, according to $\overset{*}{\mathcal{U}}(\boldsymbol{\sigma}) = \int_{\Omega} \overset{*}{u}(\boldsymbol{\sigma}) \, \mathrm{d}\Omega$ is defined – again in matrix notation – as the Legendre transformation

$$\overset{*}{u}(\boldsymbol{\sigma}) := \boldsymbol{\sigma}^{\mathrm{T}} \boldsymbol{\epsilon}_{phys} - u(\boldsymbol{\epsilon}_{geom}) \quad \text{with} \quad \boldsymbol{\epsilon}_{geom} := \boldsymbol{D}\boldsymbol{u}, \tag{207}$$

for which the total differential

$$\mathrm{d}u(\boldsymbol{\epsilon}_{geom}) = \mathrm{d}\boldsymbol{\sigma}^{\mathrm{T}} \frac{\partial u}{\partial \boldsymbol{\epsilon}_{geom}} = \mathrm{d}\boldsymbol{\sigma}^{\mathrm{T}} \mathbb{C}\boldsymbol{\epsilon} \tag{208}$$

exists. Then also $\overset{*}{u}$ has the total differential

$$\mathrm{d}\overset{*}{u} = \mathrm{d}\boldsymbol{\sigma}^{\mathrm{T}} \boldsymbol{\epsilon}_{phys}; \quad \boldsymbol{\epsilon}_{phys} = \frac{\partial \overset{*}{u}}{\partial \boldsymbol{\sigma}}, \tag{209}$$

and recalling (207), the specific stress energy functional takes the form

$$\overset{*}{u}(\boldsymbol{\sigma}) = \frac{1}{2}\boldsymbol{\sigma}^{\mathrm{T}} \mathbb{C}^{-1} \boldsymbol{\sigma}; \quad \mathbb{C}^{-1} =: \overset{*}{\mathbb{C}}. \tag{210}$$

The Hellinger-Reissner functional for the test stresses $\boldsymbol{\tau}(\boldsymbol{x})$ and the test displacements $\boldsymbol{v}(\boldsymbol{x})$ is defined for a continuous system with mixed boundary conditions as

$$\mathcal{F}_{HR}(\boldsymbol{\tau}, \boldsymbol{v}) := \int_{\Omega} \left[\boldsymbol{\tau}^{\mathrm{T}} \boldsymbol{\epsilon}(\boldsymbol{v}) - \frac{1}{2}\boldsymbol{\tau}^{\mathrm{T}} \mathbb{C}^{-1} \boldsymbol{\tau} \right] \mathrm{d}\Omega + \Pi_{ext}(\boldsymbol{v});$$

$$\Pi_{ext}(\boldsymbol{v}) = -\int_{\Omega} \boldsymbol{v}^{\mathrm{T}} \rho \boldsymbol{b} \, \mathrm{d}\Omega - \int_{\Gamma_N} \boldsymbol{v}^{\mathrm{T}} \bar{\boldsymbol{t}} \, \mathrm{d}\Gamma. \tag{211}$$

Different from the previous Subsec. 6.3.1 we chose the shorter notation $\boldsymbol{\tau} \equiv \delta\boldsymbol{\sigma}$ for the test stresses (virtual stresses) and $\boldsymbol{v} \equiv \delta\boldsymbol{u}$ for the test displacements (virtual displacements) in the domain Ω. They have to fulfill the following conditions:

$\boldsymbol{\tau} \in \mathcal{T}; \ \boldsymbol{\tau}(\boldsymbol{x}) = \boldsymbol{\tau}^{\mathrm{T}}(\boldsymbol{x})$ is square-integrable in Ω, $\boldsymbol{\tau} = \boldsymbol{0}$ at $\Gamma_N;$ $\boldsymbol{v} \in \mathcal{V}$, $\boldsymbol{v}$ is $\mathcal{C}^1$-continous in Ω and zero at Dirichlet boundaries, $\boldsymbol{v} = \boldsymbol{0}$ at Γ_D.

The stresses $\boldsymbol{\sigma}(\boldsymbol{x})$ and displacements $\boldsymbol{u}(\boldsymbol{x})$ are determined for the condition that the functional (211) becomes stationary for the saddle-point problem

$$\mathcal{F}_{HR}(\boldsymbol{\sigma}, \boldsymbol{u}) = \inf_{\boldsymbol{\tau} \in \mathcal{T}} \sup_{\boldsymbol{v} \in \mathcal{V}} \mathcal{F}_{HR}(\boldsymbol{\tau}, \boldsymbol{v}). \tag{212}$$

The necessary condition is that the first variation becomes zero as

$$\delta\mathcal{F}_{HR} = \delta_{\boldsymbol{\tau}}\mathcal{F}_{HR} + \delta_{\boldsymbol{v}}\mathcal{F}_{HR} \overset{!}{=} 0, \qquad \delta\mathcal{F}_{HR} := \boldsymbol{\tau}^{\mathrm{T}} \frac{\partial \mathcal{F}_{HR}}{\partial \boldsymbol{\tau}}\bigg|_{\boldsymbol{\tau}=\boldsymbol{\sigma}} + \boldsymbol{v}^{\mathrm{T}} \frac{\partial \mathcal{F}_{HR}}{\partial \boldsymbol{v}}\bigg|_{\boldsymbol{v}=\boldsymbol{u}} \overset{!}{=} 0 \tag{213}$$

yielding

$$\delta_{\boldsymbol{\tau}}\mathcal{F}_{HR} = \int_{\Omega} \boldsymbol{\tau}^{\mathrm{T}} [\underbrace{\boldsymbol{D}\boldsymbol{u}}_{\boldsymbol{\epsilon}_{geom}} - \underbrace{\mathbb{C}^{-1}\boldsymbol{\sigma}}_{\boldsymbol{\epsilon}_{phys}}] \, \mathrm{d}\Omega \overset{!}{=} 0 \quad \forall \boldsymbol{\tau} \in \mathcal{T}, \tag{214$_1$}$$

$$\delta_{\boldsymbol{v}}\mathcal{F}_{HR} = \int_{\Gamma_N} \boldsymbol{v}^{\mathrm{T}} (\underbrace{\mathcal{N}^{\mathrm{T}}\boldsymbol{\sigma}}_{\boldsymbol{t}}) \, \mathrm{d}\Gamma - \int_{\Omega} (\boldsymbol{v}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}})\boldsymbol{\sigma} \, \mathrm{d}\Omega - \int_{\Omega} \boldsymbol{v}^{\mathrm{T}} \rho \boldsymbol{b} \, \mathrm{d}\Omega - \int_{\Gamma_N} \boldsymbol{v}^{\mathrm{T}}\bar{\boldsymbol{t}} \, \mathrm{d}\Gamma$$

$$= -\int_{\Omega} \boldsymbol{v}^{\mathrm{T}} [\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma} + \rho\boldsymbol{b}] \, \mathrm{d}\Omega + \int_{\Gamma_N} \boldsymbol{v}^{\mathrm{T}} [\boldsymbol{t} - \bar{\boldsymbol{t}}] \, \mathrm{d}\Gamma \overset{!}{=} 0 \quad \forall \boldsymbol{v} \in \mathcal{V}, \tag{214$_2$}$$

where in $(214)_2$ partial integration and the divergence theorem were applied to the first term $\int_\Omega \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma})\,\mathrm{d}\Omega$.

The square brackets in $(214)_1$ and $(214)_2$ are the weak conditions for the kinematic compatibility, i.e., $\boldsymbol{\epsilon}_{phys} = \boldsymbol{\epsilon}_{geom}$, but in difference to the dual functional $\overset{*}{\mathcal{F}}$ and the hybrid dual functional $\overset{*}{\mathcal{F}}_{hyb}$, the test stresses do not need to fulfill the equilibrium conditions in the domain; they are fulfiulled weakly as can be seen from $(214)_2$. The same holds for the equilibrium condition at Neumann boundary. Thus, the directly approximated fulfillment of the equilibrium conditions by the Hellinger-Reissner functional is obtained together with the fact that the kinematic conditions are also approximately fulfilled only. As a 1D example, using linear shape functions for displacements and stresses, the dual mixed FEM (Hellinger-Reissner) yields linear displacements and linear stresses but constant geometrical and linear physical strains; and using linear displacements for the primal FEM we get linear displacements and constant stresses, combined with constant geometrical and physical strains.

6.4.2. Hellinger-Reissner dual mixed FEM for linear elastic boundary value problems

As in the previous subsections we proceed from the continuous problem to a discretized finite element system. Then, from (211) we arrive at

$$\mathcal{F}_{HR,h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h)$$
$$= \bigcup_e \left\{ \int_{\Omega_e} \boldsymbol{\tau}_h^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{v}_h)\,\mathrm{d}\Omega - \frac{1}{2}\int_{\Omega_e} \boldsymbol{\tau}_h^{\mathrm{T}}\overset{*}{\mathbb{C}}\boldsymbol{\tau}_h\,\mathrm{d}\Omega - \int_{\Omega_e} \boldsymbol{v}_h^{\mathrm{T}}\rho\boldsymbol{b}\,\mathrm{d}\Omega - \int_{\Gamma_{N,e}} \boldsymbol{v}_h^{\mathrm{T}}\overline{\boldsymbol{t}}\,\mathrm{d}\Gamma \right\} \to \underset{\boldsymbol{\tau}_h, \boldsymbol{v}_h}{\mathrm{stat.}}, \qquad (215)$$

with $\boldsymbol{\tau}_h \in \mathcal{T}_h \subset \mathcal{T}$, $\boldsymbol{\tau}_h(\boldsymbol{x}) \in L_2(\Omega_e)$ and $\boldsymbol{v}_h \in \mathcal{V}_h \subset \mathcal{V}$, $\boldsymbol{v}_h \in H^1(\Omega_e)$, $[\![\boldsymbol{v}_h]\!] = \boldsymbol{0}$ at Γ_e, $\boldsymbol{v}_h = \boldsymbol{0}$ at $\Gamma_{D,e}$.

The first variation results in, compare $(214)_1$,

$$\delta_{\boldsymbol{\tau}_h}\mathcal{F}_{HR} = \bigcup_e \left[\underbrace{\int_{\Omega_e} \boldsymbol{\tau}_h^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{u}_h)\,\mathrm{d}\Omega}_{-b_e(\boldsymbol{\tau}_h, \boldsymbol{u}_h)} - \underbrace{\int_{\Omega_e} \boldsymbol{\tau}_h^{\mathrm{T}}\overset{*}{\mathbb{C}}\boldsymbol{\sigma}_h\,\mathrm{d}\Omega}_{-a_e(\boldsymbol{\tau}_h, \boldsymbol{\sigma}_h)} \right] \overset{!}{=} 0, \qquad (216)_1$$

$$\delta_{\boldsymbol{v}_h}\mathcal{F}_{HR} = \bigcup_e \left[\underbrace{\int_{\Omega_e} \boldsymbol{v}_h^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{\sigma})\,\mathrm{d}\Omega}_{-b_e^{\mathrm{T}}(\boldsymbol{v}_h, \boldsymbol{\sigma}_h)} - \underbrace{\int_{\Omega_e} \boldsymbol{v}_h^{\mathrm{T}}\rho\boldsymbol{b}\,\mathrm{d}\Omega - \int_{\Gamma_{N,e}} \boldsymbol{v}_h^{\mathrm{T}}\overline{\boldsymbol{t}}\,\mathrm{d}\Gamma}_{F_e(\boldsymbol{v}_h)} \right] \overset{!}{=} 0. \qquad (216)_2$$

With the introduced bilinear forms $a_e(\boldsymbol{\tau}_h, \boldsymbol{\sigma}_h)$, $b_e(\boldsymbol{\tau}_h, \boldsymbol{u}_h)$ and $b_e^{\mathrm{T}}(\boldsymbol{v}_h, \boldsymbol{\sigma}_h)$ as well as the linear form $F_e(\boldsymbol{v}_h)$ the two stationarity conditions read

$$\bigcup_e \{a_e(\boldsymbol{\tau}_h, \boldsymbol{\sigma}_h) + b_e(\boldsymbol{\tau}_h, \boldsymbol{u}_h)\} = 0 \quad \forall \boldsymbol{\tau}_h \in \mathcal{T}_h \subset \mathcal{T}, \qquad (217)_1$$

$$\bigcup_e \{b_e^{\mathrm{T}}(\boldsymbol{v}_h, \boldsymbol{\sigma}_h) + F_e(\boldsymbol{v}_h)\} = 0 \quad \forall \boldsymbol{v}_h \in \mathcal{V}_h \subset \mathcal{V}. \qquad (217)_2$$

Discrete test and trial functions are introduced as tensor products of Lagrangian or Legendre polynomials with unknown nodal values, as

$$\boldsymbol{\tau}_h(\boldsymbol{x}) = \boldsymbol{N}_\sigma(\boldsymbol{x})\delta\widehat{\boldsymbol{s}}_e, \ \boldsymbol{\sigma}_h(\boldsymbol{x}) = \boldsymbol{N}_\sigma(\boldsymbol{x})\widehat{\boldsymbol{s}}_e \text{ in } \Omega_e, \tag{218}_1$$

$$\boldsymbol{v}_h(\boldsymbol{x}) = \boldsymbol{N}_u(\boldsymbol{x})\delta\widehat{\boldsymbol{u}}_e, \ \boldsymbol{u}_h(\boldsymbol{x}) = \boldsymbol{N}_u(\boldsymbol{x})\widehat{\boldsymbol{u}}_e \text{ in } \Omega_e, \tag{218}_2$$

with the shape functions

$$\boldsymbol{N}_\sigma(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{\phi}^{\mathrm{T}}_{\sigma_{11}}(\boldsymbol{x}) & & & \\ & \boldsymbol{\phi}^{\mathrm{T}}_{\sigma_{22}}(\boldsymbol{x}) & & \\ & & \ddots & \\ & & & \boldsymbol{\phi}^{\mathrm{T}}_{\sigma_{31}}(\boldsymbol{x}) \end{bmatrix}, \tag{219}$$

and

$$\boldsymbol{N}_u(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{\phi}^{\mathrm{T}}_{u_1}(\boldsymbol{x}) & & \\ & \boldsymbol{\phi}^{\mathrm{T}}_{u_2}(\boldsymbol{x}) & \\ & & \boldsymbol{\phi}^{\mathrm{T}}_{u_3}(\boldsymbol{x}) \end{bmatrix}. \tag{220}$$

We then get the system of algebraic equations, multiplying $(216)_1$ with -1

$$\bigcup_e \left\{ \left\{ \underbrace{\delta\widehat{\boldsymbol{s}}^{\mathrm{T}}_e}_{\neq 0} \quad \underbrace{\delta\widehat{\boldsymbol{u}}^{\mathrm{T}}_e}_{\neq 0} \right\} \left(\begin{bmatrix} \underbrace{\int_{\Omega_e} \boldsymbol{N}^{\mathrm{T}}_\sigma \overset{*}{\mathbb{C}} \boldsymbol{N}_\sigma \, \mathrm{d}\Omega}_{\boldsymbol{f}_e = \boldsymbol{f}^{\mathrm{T}}_e; \ \det \boldsymbol{f}_e \neq 0} & \underbrace{-\int_{\Omega_e} \boldsymbol{N}^{\mathrm{T}}_\sigma \boldsymbol{D} \boldsymbol{N}_u \, \mathrm{d}\Omega}_{\boldsymbol{l}_e} \\ \underbrace{-\int_{\Omega_e} \boldsymbol{N}^{\mathrm{T}}_u \boldsymbol{D}^{\mathrm{T}} \boldsymbol{N}_\sigma \, \mathrm{d}\Omega}_{\boldsymbol{l}^{\mathrm{T}}_e} & \boldsymbol{0} \end{bmatrix} \begin{Bmatrix} \widehat{\boldsymbol{s}}_e \\ \widehat{\boldsymbol{u}}_e \end{Bmatrix} \right. \right.$$
$$\left. \left. - \begin{Bmatrix} \boldsymbol{0} \\ \underbrace{\int_{\Omega_e} \boldsymbol{N}^{\mathrm{T}}_u \rho\boldsymbol{b} \, \mathrm{d}\Omega + \int_{\Gamma_{N,e}} \boldsymbol{N}^{\mathrm{T}}_u \overline{\boldsymbol{t}} \, \mathrm{d}\Gamma}_{\widehat{\overline{\boldsymbol{p}}}_e} \end{Bmatrix} \right) \right\} = 0, \tag{221}$$

in which the terms within the large round brackets have to become zero. Using the above notations for the matrices and vectors we arrive at the global mixed FE system

$$\bigcup_e \left\{ \begin{bmatrix} \boldsymbol{f}_e & \boldsymbol{l}_e \\ \boldsymbol{l}^{\mathrm{T}}_e & \boldsymbol{0} \end{bmatrix} \begin{Bmatrix} \widehat{\boldsymbol{s}}_e \\ \widehat{\boldsymbol{u}}_e \end{Bmatrix} - \begin{Bmatrix} \boldsymbol{0} \\ \widehat{\overline{\boldsymbol{p}}}_e \end{Bmatrix} \right\} = \begin{Bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{Bmatrix}, \tag{222}$$

from where we eliminate the nodal stresses on element level according to $\widehat{\boldsymbol{s}}_e = \boldsymbol{f}^{-1}_e \boldsymbol{l}_e \widehat{\boldsymbol{u}}_e$ and resulting in

$$\underbrace{\boldsymbol{l}^{\mathrm{T}}_e \boldsymbol{f}^{-1}_e \boldsymbol{l}_e}_{\overset{*}{\boldsymbol{k}}_e} \widehat{\boldsymbol{u}}_e = \widehat{\overline{\boldsymbol{p}}}_e. \tag{223}$$

Then the assembling of the elements to the global system is purely kinematic, as in primal FEM, using the Boolean matrices $\boldsymbol{a}_e$ and the global reduced displacement vector $\widehat{\boldsymbol{U}}$ as

$$\widehat{\boldsymbol{u}}_e = \boldsymbol{a}_e \widehat{\boldsymbol{U}}, \tag{224}$$

yielding

$$\underbrace{\sum_e (\boldsymbol{a}_e^{\mathrm{T}} \overset{*}{\boldsymbol{k}}_e \boldsymbol{a}_e)}_{\overset{*}{\boldsymbol{K}} = \overset{*}{\boldsymbol{K}}^{\mathrm{T}} \; ; \; \det \overset{*}{\boldsymbol{K}} \neq 0;} \widehat{\boldsymbol{U}} = \underbrace{\sum_e \boldsymbol{a}_e \widehat{\overline{\boldsymbol{p}}}_e}_{\widehat{\overline{\boldsymbol{P}}}}, \tag{225}$$

from which the unknown global nodal displacement vector $\widehat{\boldsymbol{U}}$ is calculated by solving the linear algebraic equation system as

$$\widehat{\boldsymbol{U}} = \overset{*}{\boldsymbol{K}}^{-1} \widehat{\overline{\boldsymbol{P}}}. \tag{226}$$

It has to be noted that in order to prove the existence and uniqueness of the solution $(\boldsymbol{\sigma}, \boldsymbol{u}) \in \mathcal{T} \subset \mathcal{V}$, apart from the ellipticity of the bilinear form a, the numerical stability condition of the saddle-point problem has to be proven; this is the so-called BB-inf-supcondition (BB: Brezzi and Babuška), [9, 19, 20], for the bilinear form $b_e(\boldsymbol{\tau}_h, \boldsymbol{u}_h)$, (216) and (217), i.e., the Lagrangian coupling matrix for the dual mixed stress and displacement ansatz. This condition reads

$$\inf_{\boldsymbol{v} \in \mathcal{V}} \sup_{\boldsymbol{\tau} \in \mathcal{T}} \frac{b(\boldsymbol{\tau}, \boldsymbol{v})}{\|\boldsymbol{\tau}\|_{\mathcal{T}} \|\boldsymbol{v}\|_{\mathcal{V}}} \geq \beta, \tag{227}$$

with a positive constant β from Korn's second inequality.

Note that alternatively, the Hellinger-Reissner functional (215) can be obtained from the primal energy functional by enforcing the non-fulfilled kinematic compatibility equation with the Lagrangian multiplier $\boldsymbol{\lambda}$, yielding

$$\mathcal{F}_{HR,h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) = \bigcup_e \left\{ \frac{1}{2} a_e(\boldsymbol{v}, \boldsymbol{v}) - F(\boldsymbol{v}) - \int_{\Omega_e} \boldsymbol{\lambda}^{\mathrm{T}} [\boldsymbol{\epsilon}_{geom}(\boldsymbol{v}) - \mathbb{C}^{-1} \boldsymbol{\tau}] \, \mathrm{d}\Omega \right\}, \tag{228}$$

and

$$\boldsymbol{\lambda} = \boldsymbol{\tau} \tag{229}$$

by the first variation of this functional.

Remark

The Hellinger-Reissner-type mixed finite element formulation is of particular importance for rigorous stability proofs of finite element discretizations, especially for low order trial and test functions. This gives rise to locking phenomena and/or spurious strains in case of shear-elastic beam-, plate- and shell-theories of Timoshenko and Reissner-Mindlin type with approximated constant transverse shear strains over the thickness as well as spurious strains in case of nearly incompressible elastic deformations. In the first case the differential equations of 4th order are reduced to 2nd order with independent variables for displacements and rotations of the cross-section normals. Choosing both with linear shape functions yields bending-locking and thus the wrong rank of the stiffness matrix. Choosing numerical underintegration on element level results in the correct rank of the element stiffness matrix, [92, 99], and this gives rise to many engineering attempts for constructing element patch tests. Also the construction of enhanced assumed strain (EAS) methods by Simo and Rifai (1990), [79], for low order shape functions has to be addressed, see also [16].

But from the mathematical point of view the mentioned nearly incompressible elastic deformations and shear-elastic bending theories provide non-robust BVPs where a global interpolation constant of the discretization error does not exist. It was shown, e.g. by Brezzi and Braess, that

those non-robust problems need the fulfillment of global inf-sup-conditions because this prevents non-regular global stiffness matrices by controlling softening of the energy due to relaxed requirements on the kinematic compatibility. It also should be mentioned that a proof of the applicability of a choosen stress and displacement ansatz in the element domain by counting and comparing the number of unknown parameters is not sufficient, also not on patches or on global level.

Important papers on the convergence and stability of mixed finite elements and special mixed elements are [6, 7, 21] and [31].

6.4.3. Primal and mixed FEM for the Timoshenko beam

A primal finite element for the Timoshenko beam theory with approximated transverse shear deformations shows locking in case of linear shape functions for the lateral displacements and – independently – for the rotations of the traces of cross-sections. This can be overcome by selective reduced integration of the shear strain energy. But a mathematically sound, i.e., stable discretization method for this non-robust problem requires a mixed formulation by which the locking problem is overcome automatically.

Under the presumption of $w \ll h \ll l$, the hypotheses of the Timoshenko theory are, Fig. 45:

(i)

$$w(x, y, z) \approx w(x) = w_{bend} + w_{shear}, \tag{230}$$

(ii)

$$u(x, z) = -\phi(x)z \text{ for } |\phi(x)| \ll 1, \tag{231}$$

(iii)

$$\kappa(x) = \phi'(x), \ \kappa \text{ the curvature}, \tag{232}$$

(iv)

$$\gamma_{av}(x) = w'(x) - \phi(x), \ \gamma_{av} \text{ the constant transverse shear angle}. \tag{233}$$
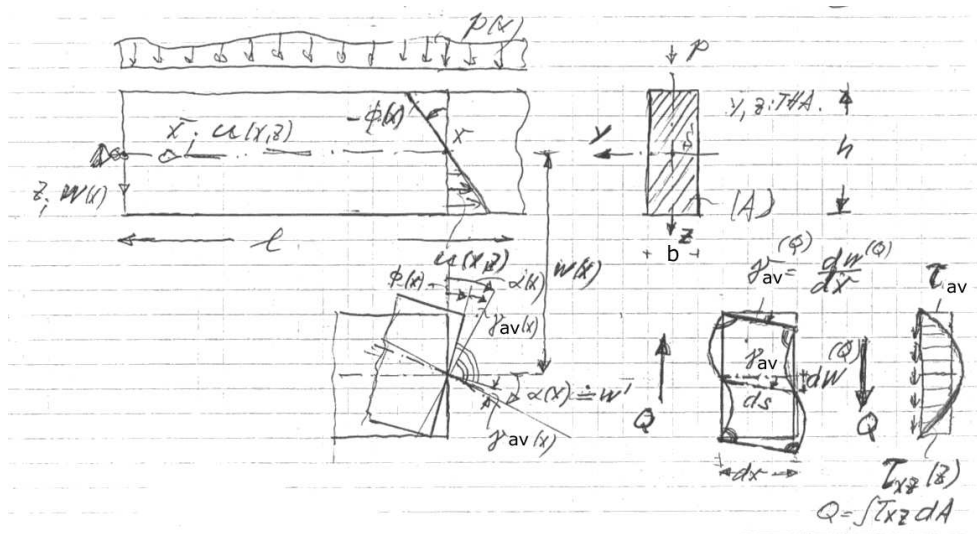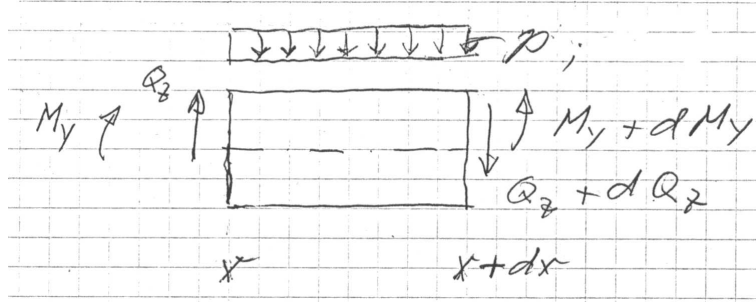

Fig. 45. Approximated constant transverse shear deformations in the Timoshenko beam theory.

Fig. 46. Equilibrium conditions for the bending moments and shear forces.

The equilibrium conditions of bending moments and transverse shear forces read, Fig. 46,

$$M_y'(x) = Q_z(x); \ Q_z'(x) = -p(x), \ \text{yiedling} \ M_y''(x) = -p(x) \tag{234}$$

The bending moment is defined for the linear stress distribution (Bernoulli-Euler hypothesis) as

$$M_y(x) = \int_A \sigma_y(x,z) \ \mathrm{d}A \ z = -EI_y \ \phi'(x), \tag{235}$$

and the elastic curvature is $\phi'(x) = -\dfrac{M_y(x)}{EIy}$. The transverse shear force results from

$$Q_z(x) = \int_A \tau_{xz}(x,z) \ \mathrm{d}A \overset{!}{=} \int_{A_s} \tau_{av}(x) \ \mathrm{d}A_{shear}, \tag{236}$$

yielding $Q_z(x) = GA_s \underbrace{(w'(x) - \phi(x))}_{\gamma_{av}(x)}$, where G is the elastic shear modulus and A_s is the reduced cross-section shear area for constant (averaged) transverse shear stresses $\tau_{av}(x)$, resulting from the postulated equality of the specific deformation energies for Bernoulli-Euler beam theory (with parabolic transverse shear stresses along the thickness) and the Timoshenko approximation (with constant transverse shear stresses in the cross-section).

The approximated cross section area A_s then results form the condition

$$\int_A \underbrace{\tau_{xz}(x,z)}_{\text{Euler-Bern. th.}} \ \mathrm{d}A \overset{!}{=} \underbrace{\frac{Q_z(x)}{A_s}}_{\tau_{av}(x)}$$

as

$$A_s = \frac{I_y^2}{\displaystyle\int_A \left(\frac{S_y(z)}{b}\right)^2 \ \mathrm{d}A}. \tag{237}$$

For rectangular cross sections it follows $A_s = \dfrac{A}{1.2}$.

The differential equation of the Timoshenko beam theory is obtained from the equilibrium condition $Q_z'(x) = -p(x)$ and the constitutive equation for the shear force $Q_z'(x) = GA_s(w''(x) - \phi'(x))$ as

$$w''(x) - \phi'(x) = -\frac{p(x)}{GA_s}. \tag{238}$$

Energy formulation

The bending energy reads

$$\mathcal{U}_{bend} = \frac{1}{2} \int_l -M_y(x) \underbrace{\kappa_y(x) \, \mathrm{d}x}_{\mathrm{d}\phi(x)} = \frac{1}{2} \int_l EI_y \phi'^2(x) \, \mathrm{d}x, \tag{239}$$

and the transverse shear energy correspondingly

$$\mathcal{U}_{shear} = \frac{1}{2} \int_l GA_s(w'(x) - \phi(x))^2 \, \mathrm{d}x, \tag{240}$$

in total

$$\mathcal{U} = \mathcal{U}_{bend} + \mathcal{U}_{shear}. \tag{241}$$

The loss of potential of line forces $\overline{p}(x)$ is

$$\Pi_{ext} = -\int_l \overline{p}(x)w(x) \, \mathrm{d}x. \tag{242}$$

Primal finite element approach

1D-elements with linear shape functions for $\phi_h(x)$ and $w_h(x)$ are used as $N_1(\xi) = 1 - \xi$, $\xi = \dfrac{x}{h}$, and $N_2(\xi) = \xi$ for the left and right nodes 1 and 2. The crucial point – providing locking – is the choice of equal test and trial functions for the lateral displacement w_h and the rotation of the cross-section traces ϕ_h as

$$w_h(\xi) = \sum_{I=1}^{2} N_I(\xi)w_I; \;\; \phi_h(\xi) = \sum_{I=1}^{2} N_I(\xi)\phi_I, \tag{243}$$

although ϕ_h has a one order higher derivative compared with w_h.

The exact integration of the discretized deformation energy, Fig. 47, reads

$$\mathcal{U}_h(w_h, \phi_h) = \underbrace{\frac{1}{2} EI_y \left(\frac{\phi_2 - \phi_1}{l}\right)^2 l}_{\mathcal{U}_{h,bend}} + \underbrace{\frac{1}{2} GA_s \left[\left(\frac{w_2 - w_1}{l} - \frac{\phi_2 + \phi_1}{2}\right)^2 + \frac{1}{3}\left(\frac{\phi_2 - \phi_1}{2}\right)^2\right] l}_{\mathcal{U}_{h,shear}}. \tag{244}$$

For thin beams it holds $GA_s l_s^2 \gg EI_y$, i.e. $\gamma_{h,av}(x) \ll \phi_h(x)$ and $\mathcal{U}_{h,shear} \ll \mathcal{U}_{h,bend}$. In the limit case $\gamma_{h,av}(x) = w'_h(x) - \phi_h(x) = 0$ we get, see last drawing in Fig. 47, $\dfrac{w_2 - w_1}{l} - \dfrac{\phi_2 + \phi_1}{2} = 0$ and further

$$\gamma_{h,av}|_{\xi=\frac{1}{2}} = (w' - \phi)|_{\xi=\frac{1}{2}} = 0 \rightsquigarrow w'_h = \phi_h. \tag{245}$$

The normality rule at the midpoint of the beam element yields

$$\frac{\phi_2 - \phi_1}{2} = 0 \rightsquigarrow \phi'_h = 0, \tag{246}$$

which means zero curvature. This is a spurious constraint with the consequence of shear locking for exact integration of the deformation energy.
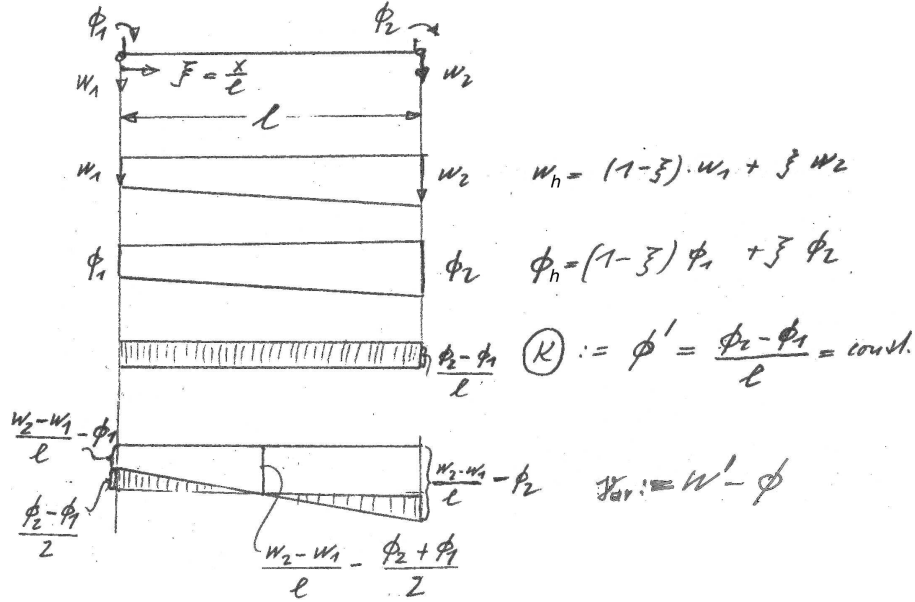
Fig. 47. Linear finite element shape functions for the Timoshenko beam theory.

The shear energy has quadratic terms which would need two Gauss points for full integration. For avoiding locking, a selective reduced one-point integration (in the middle) is chosen, which results in the following element stiffness matrix $\boldsymbol{k}_e$ with the correct rank for the displacement vector $\widehat{\boldsymbol{u}}_e^{\mathrm{T}} = \{ w_1 \quad \phi_1 \quad w_2 \quad \phi_2 \}$

$$
\boldsymbol{k}_e = \begin{bmatrix}
\dfrac{GA_s}{l} & \dfrac{GA_s}{2} & -\dfrac{GA_s}{l} & \dfrac{GA_s}{2} \\[2mm]
\dfrac{GA_s}{2} & \dfrac{GA_s l}{4} + \dfrac{EI_y}{l} & -\dfrac{GA_s}{2} & \dfrac{GA_s l}{4} - \dfrac{EI_y}{l} \\[2mm]
-\dfrac{GA_s}{l} & -\dfrac{GA_s}{2} & \dfrac{GA_s}{l} & -\dfrac{GA_s}{2} \\[2mm]
\dfrac{GA_s}{2} & \dfrac{GA_s l}{4} - \dfrac{EI_y}{l} & -\dfrac{GA_s}{2} & \dfrac{GA_s l}{4} + \dfrac{EI_y}{l}
\end{bmatrix} . \tag{247}
$$

The underlined terms are equal, such that they vanish within the elimination process. This can be seen from the drawing at the bottom of Fig. 47.

For full (two-point) integration, the underlined expressions in (247) are $\dfrac{GA_s l}{3}$ and the dotted line terms are $\dfrac{GA_s l}{6}$ which cause a rank deficiency for large bending stiffness, resulting in nearly zero lateral displacements.

A dual mixed FEM for the Timoshenko beam (i.e., with relaxed kinematic conditions) is based on the functional for an element e

$$
\mathcal{F}_{HR,h,e}[M_{y,h}, Q_{z,h}; w_h, \phi_h] = \int\limits_{l_e} \left[-M_{y,h}\phi_h' + Q_{z,h}(w_h' - \phi_h) \right] \mathrm{d}x
$$

$$
+ \frac{1}{2} \int\limits_{l_e} M_{y,h} EI_y M_{y,h} \, \mathrm{d}x + \frac{1}{2} \int\limits_{l_e} Q_{z,h} GA_s Q_{z,h} \, \mathrm{d}x - \int\limits_{l_e} \overline{p} w_h \, \mathrm{d}x, \tag{248}
$$

in which the kinematic conditions $\kappa_h(\phi_h) = -\dfrac{M_{y,h}}{EI_y}$ and $\gamma_{s,h}(w'_h, \phi_h) = Q_{z,h}/GA_s$ are relaxed and fulfilled approximately in the weak sense.

The linear trial and test functions read, with the left and right nodes 1 and 2 of the beam element

$$
\begin{aligned}
w_{h,w} &= w_1(1-\xi) + w_2\xi; \quad \xi = x/l_e, \\
\phi_{h,e} &= \phi_1(1-\xi) + \phi_2\xi, \\
M_{y,h,e} &= M_1(1-\xi) + M_2\xi, \\
Q_{z,h,w} &= Q_1(1-\xi) + Q_2\xi.
\end{aligned}
\tag{249}
$$

Introducing this ansatz into the stationarity condition $\delta\mathcal{F}_{HR,h} \overset{!}{=} 0$ and eliminating $M_{y,h,e}$ and $Q_{z,h,e}$ on element level yields the element stiffness matrix (247) with the correct rank, i.e., without the dotted line terms.

6.4.4. A mixed plate element for Kirchhoff-Love plate bending by Herrmann

L.R. Herrmann published the following dual mixed element already in 1967, [44]. The related discrete Hellinger-Reissner functional for an element Ω_e reads in Cartesian coordinates x, y for the plate middle surface

$$
\begin{aligned}
\mathcal{F}_{HR,h,e} = \int\limits_{\Omega_e} \{m_{x,x} \quad m_{y,y} m_{xy,y} + m_{yx,x}\}_h \left\{\begin{matrix} w_{,x} \\ w_{,y} \\ \frac{1}{2}(w_{,y} + w_{,x}) \end{matrix}\right\}_h \mathrm{d}x\mathrm{d}y \\
- \frac{1}{\dfrac{Et^3}{12(1-\nu)^2}} \frac{1}{2} \int\limits_{\Omega_e} \{m_x \quad m_y \quad m_{xy}\}_h \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & 2(1+\nu) \end{bmatrix} \left\{\begin{matrix} m_x \\ m_y \\ m_{xy} \end{matrix}\right\}_h \mathrm{d}x\mathrm{d}y \\
- \int\limits_{\Omega_e} \overline{p}w \, \mathrm{d}x\mathrm{d}y, \tag{250}
\end{aligned}
$$

and the whole functional of a clamped plate with Dirichlet boundary conditions reads

$$
\mathcal{F}_{HR,h} = \sum_e \mathcal{F}_{HR,h,e} \to \text{stat.},
\tag{251}
$$

with the stationarity condition

$$
\delta\mathcal{F}_{HR,h} \overset{!}{=} 0.
\tag{252}
$$

The mixed ansatz for a triangular plate element with the corner nodes 1, 2 and 3 is

$$
w_h = C_1 + C_2 x + C_3 y,
\tag{253}
$$

which can be expressed by the nodal displacements w_1, w_2, w_3, and

$$
m_{x,h} = C_4; \ m_{y,h} = C_5; \ m_{xy,h} = C_6.
\tag{254}
$$

Again, as a benefit of the dual mixed method, the bending moments C_4, C_5, C_6 can be eliminated on element level such that the assembling of elements has only to be done for the nodal displacements.

6.4.5. A three-field variational formulation based on the Hu-Washizu functional

As we have seen, however, the mixed approach presented does not yield a continuous stress field in case of low-order shape functions. Furthermore, with the Hellinger-Reissner type mixed method we still encounter difficulties in nearly incompressible elastic materials. To cope with these problems, another pairing of test and solution spaces is required, namely $\mathcal{H}(\mathrm{div}, \Omega)$ and $\mathcal{L}_2(\Omega)$ instead of $\mathcal{L}_2(\Omega)$ and $\mathcal{V}$. Therefore, a three-field variational formulation was proposed by Hu and Washizu in 1955, [96].

Rather than two independent unknown fields $\boldsymbol{\sigma} \in \mathcal{T}$ and $\boldsymbol{u} \in \mathcal{V}$, as used in the Hellinger-Reissner functional, now an additional unknown strain field $\boldsymbol{\epsilon} \in \mathcal{E}$, which leads to the Hu-Washizu functional

$$\mathcal{F}_{HW} : \mathcal{E} \times \mathcal{V} \times \mathcal{T} \to \mathbb{R}, \tag{255}$$

with $\mathcal{E} = \mathcal{T} = [\mathcal{L}_2(\Omega)]^{3\times3}$ in the simplest case, whereas furthermore $\mathcal{V} = [\mathcal{H}^1(\Omega)]^{3\times3}$. The field equations follow from three variational conditions for the existence of a saddle-point problem, namely $\mathrm{D}_{\boldsymbol{\epsilon}}\mathcal{F}_{HW}(\boldsymbol{\epsilon}, \boldsymbol{u}, \boldsymbol{\sigma}) \cdot \boldsymbol{\eta} = 0$, $\mathrm{D}_{\boldsymbol{u}}\mathcal{F}_{HW}(\boldsymbol{\epsilon}, \boldsymbol{u}, \boldsymbol{\sigma}) \cdot \boldsymbol{v} = 0$ and $\mathrm{D}_{\boldsymbol{\sigma}}\mathcal{F}_{HW}(\boldsymbol{\epsilon}, \boldsymbol{u}, \boldsymbol{\sigma}) \cdot \boldsymbol{\tau} = 0$, or alternatively $\delta_{\boldsymbol{\epsilon}}\mathcal{F}_{HW} = 0$, $\delta_{\boldsymbol{u}}\mathcal{F}_{HW} = 0$ and $\delta_{\boldsymbol{\sigma}}\mathcal{F}_{HW} = 0$.

Thus, in the associated mixed variational problem, we seek a solution $(\boldsymbol{\epsilon}, \boldsymbol{u}, \boldsymbol{\sigma}) \in \mathcal{E} \times \mathcal{V} \times \mathcal{T}$, satisfying in the weak continuous version,
the constitutive equations

$$\int_{\Omega} (\mathbb{C} : \boldsymbol{\epsilon} - \boldsymbol{\sigma}) : \boldsymbol{\eta} \; \mathrm{d}\Omega = 0 \quad \forall \boldsymbol{\eta} \in \mathcal{E}, \tag{256$_1$}$$

the equilibrium conditions

$$\int_{\Omega} \boldsymbol{\sigma} : (\mathrm{grad}_{sym}\boldsymbol{v}) \; \mathrm{d}\Omega = \int_{\Omega} \boldsymbol{f} \cdot \boldsymbol{v} \; \mathrm{d}\Omega + \int_{\Gamma_N} \bar{\boldsymbol{t}} \cdot \boldsymbol{v} \; \mathrm{d}\Gamma \quad \forall \boldsymbol{v} \in \mathcal{V} \tag{256$_2$}$$

and the kinematic compatibility conditions

$$\int_{\Omega} [(\mathrm{grad}_{sym}\boldsymbol{u}) - \boldsymbol{\epsilon}] : \boldsymbol{\tau} \; \mathrm{d}\Omega = 0 \quad \forall \boldsymbol{\tau} \in \mathcal{T}. \tag{256$_3$}$$

The solution of the mixed problem is a saddle-point

$$\mathcal{F}_{HW}(\boldsymbol{\epsilon}, \boldsymbol{u}, \boldsymbol{\sigma}) = \sup_{\boldsymbol{\eta} \in \mathcal{E}} \sup_{\boldsymbol{v} \in \mathcal{V}} \inf_{\boldsymbol{\tau} \in \mathcal{T}} \mathcal{F}_{HW}(\boldsymbol{\eta}, \boldsymbol{v}, \boldsymbol{\tau}). \tag{257}$$

The introduction of finite element spaces and the calculation of the bilinear forms follow from the methods outlined before in this subsection for the Hellinger-Reissner mixed method.

The discrete variational conditions of the Hu-Washizu mixed finite element method then result in

$$\bigcup_{e} \{a_e(\boldsymbol{\eta}_h, \boldsymbol{\epsilon}_h) + b_e(\boldsymbol{\sigma}_h, \boldsymbol{\eta}_h) = 0 \quad \forall \boldsymbol{\eta}_h \in \mathcal{E}_h\}, \tag{258$_1$}$$

$$\bigcup_{e} \{c_e(\boldsymbol{\sigma}_h, \boldsymbol{v}_h) = F_e(\boldsymbol{v}_h) \quad \forall \boldsymbol{v}_h \in \mathcal{V}_h\}, \tag{258$_2$}$$

$$\bigcup_{e} \{b_e(\boldsymbol{\epsilon}_h, \boldsymbol{\tau}_h) + c_e(\boldsymbol{\tau}_h, \boldsymbol{u}_h) = 0 \quad \forall \boldsymbol{\tau}_h \in \mathcal{T}_h\}. \tag{258$_3$}$$

The associated inf-sup condition reads

$$\inf_{\boldsymbol{\tau} \in \mathcal{T}} \left\{ \sup_{\boldsymbol{\eta} \in \mathcal{E}} \frac{b(\boldsymbol{\tau}, \boldsymbol{\eta})}{\|\boldsymbol{\tau}\|_{\mathcal{T}} \|\boldsymbol{\eta}\|_{\mathcal{E}}} + \sup_{\boldsymbol{v} \in \mathcal{V}} \frac{c(\boldsymbol{\tau}, \boldsymbol{v})}{\|\boldsymbol{\tau}\|_{\mathcal{T}} \|\boldsymbol{v}\|_{\mathcal{V}}} \right\} \geq \beta \tag{259}$$

with a positive constant β.

6.4.6. Finite element discretization

Next, shape functions for the displacement, stress and strain fields have to be chosen under the important restriction that the global inf-sup condition is fulfilled. Similar to the global equation system for the Hellinger-Reissner principle, (222), we arrive at the global algebraic equation system for the Hu-Washizu functional

$$
\bigcup_e \left\{ \begin{bmatrix} \boldsymbol{f}_e & \boldsymbol{0} & \boldsymbol{l}_{\boldsymbol{\epsilon},e} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{l}_{\boldsymbol{u},e} \\ \boldsymbol{l}_{\boldsymbol{\epsilon},e}^{\mathrm{T}} & \boldsymbol{l}_{\boldsymbol{u},e}^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{Bmatrix} \widehat{\boldsymbol{s}}_e \\ \widehat{\boldsymbol{\epsilon}}_e \\ \widehat{\boldsymbol{u}}_e \end{Bmatrix} \right\} = \begin{Bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \widehat{\boldsymbol{p}}_e \end{Bmatrix} . \tag{260}
$$

Since the stress and strain shape functions are in $\mathcal{L}_2$ spaces, we can eliminate $\boldsymbol{f}_e$ on element level, get the fictitious element stiffness matrix $\boldsymbol{k}_e^* = \boldsymbol{l}_{\boldsymbol{u},e}^{\mathrm{T}} \boldsymbol{f}_e^{-1} \boldsymbol{l}_{\boldsymbol{u},e}$ and finally arrive at the system

$$
\bigcup_e \left\{ \begin{bmatrix} \boldsymbol{k}_e^* & \boldsymbol{l}_{\boldsymbol{u},e} \\ \boldsymbol{l}_{\boldsymbol{u},e}^{\mathrm{T}} & \boldsymbol{0}_e \end{bmatrix} \begin{Bmatrix} \widehat{\boldsymbol{\sigma}}_e \\ \widehat{\boldsymbol{u}}_e \end{Bmatrix} \right\} = \begin{Bmatrix} \boldsymbol{0}_e \\ \widehat{\boldsymbol{p}}_e \end{Bmatrix} . \tag{261}
$$

Alternatively, the elimination of $\widehat{\boldsymbol{\epsilon}}_e$ is possible, depending on the ansatz functions.

Remark

A further reduction is not possible because $\boldsymbol{k}_e^*$ is non-regular in general. Stolarski and Belytschko, [88], have shown that for certain choices of pairings, the finite element solution $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)$ of the Hu-Washizu formulation equals the Hellinger-Reissner formulation with the same test spaces $\mathcal{T}_h$ and $\mathcal{V}_h$.

Finally, we remark that other mixed functionals and associated finite element methods, which are balanced on the element level and also provide global numerical stability and well-posedness, have been developed in the 1980s and 1990s, especially the PEERS-element (plane elasticity with reduced symmetry), Arnold, Brezzi, Douglas (1984), [6], and Stenberg (1984), [87], and especially the class of BDM-elements by Brezzi, Douglas and Marini (1985), [22].

The Hu-Washizu-based FEM was used especially for special material properties in order to circumvent numerical stability problems, but it has to be pointed out that special pairings of test and solution spaces are required, fulfilling the global inf-sup condition.

7. FOUNDATION OF THE MATHEMATICAL THEORY OF PRIMAL FEM AND ERROR ANALYSIS IN THE 20TH CENTURY

The mathematical foundation of the finite element method for elliptic partial differential equations of 2nd order began in the 1960s and 1970s with the eminent researchers Bramble, Hilbert, Babuška, Rheinboldt, Strang, Fix, Aubin, Nitsche, Brezzi, Ciarlet, Johnson, Oden and many others.

The simplicity of primal FEM against FDM – i.e. test and trial functions and their operating only within element domains – is accompanied by the severe problem that for 2nd order PDEs the C^0-continuous test and trial functions at element interfaces provide jumps of derivatives, i.e. of strains and stresses, and therefore require Sobolev spaces for the mathematical analysis.

7.1. Dirichlet's principle of minimum of total potential energy

The following convergence theory and error analysis holds for linear elliptic partial differential equations, usually of 2nd order and with self-adjoint operators. An important problem of this type is linear static theory of elasticity of deformable bodies where the displacements $\boldsymbol{u}(\boldsymbol{x}) \in \Omega \subset \mathbb{R}^3$ are C^1 manifolds.

The variational formulation of the 3D-Lamé's theory of linear elasticity was established by Lejeune Dirichlet (1806–1859) in 1835. For mixed boundary conditions, the symmetric variational form with a test function $v(x) \in \mathcal{H}^1(\Omega)$ reads

$$a(v, v) = \int_\Omega \epsilon(v) : \mathbb{C} : \epsilon(v) \, \mathrm{d}V, \tag{262}$$

with the test space $\mathcal{V} = \{v \in [\mathcal{H}^1(\Omega)]^3; v = 0 \text{ at } \Gamma_D\}$, $\epsilon(v)$ the linear symmetric strain tensor and $\mathbb{C}$ the symmetric and positive definite 4th order elasticity tensor.

The linear forms for the given loads are

$$F(v) = \int_\Omega f \cdot v \, \mathrm{d}V; \quad g(v) = \int_{\Gamma_N} \bar{t} \cdot v \, \mathrm{d}S, \tag{263}$$

with the body force $f(x) \in \mathcal{L}_2(\Omega)$ and the conservative given boundary tractions $\bar{t}(v) \in \mathcal{L}_2(\Gamma_N)$.

The variational functional – the total potential energy of the system –

$$J(v) := \underbrace{1/2 a(v, v)}_{\Pi_{int}} \underbrace{-[F(v) + g(v)]}_{\Pi_{ext}} \to \min_{v \in \mathcal{V}} \tag{264}$$

is constructed from the above bilinear and linear forms and yields a minimum for the analytical solution $u(x) \in \mathcal{H}^2(\Omega)$ for well-posed problems in stable static equilibrium.

The necessary stationarity condition requires that the 1st variation of the Dirichlet's functional vanishes, as

$$\delta J(v) = 0 \Rightarrow u, \tag{265}$$

and the necessary condition for the minimal property is that the 2nd variation is larger than zero in the solution point, as

$$\delta^2 J(v) > 0 \text{ for } v = u. \tag{266}$$

Theorem

A unique minimal solution $u(x)$ exists. It is proven by the Lagrangian embedding method as follows:

Assuming a neighboured solution $\widetilde{u} = u + tv$, $t \in \mathbb{R}$, yields the minimal functional

$$J(u + tv) = 1/2 \, a(u + tv, u + tv) - F(u + tv) - g(u + tv) \tag{267}$$

$$= J(u) + t[a(u, v) - F(v) - g(v) + 1/2 \, t^2 a(v, v)] \tag{268}$$

and for $t = 1$

$$J(u + v) = J(u) + 1/2 \, a(v, v) \; \forall \; v \in \mathcal{V}, \tag{269}$$

$$J(u + v) > J(u); \quad J(u) = \min_{v \in \mathcal{V}} J(v). \tag{270}$$

The 2nd variation of J yields the quadratic form

$$\delta^2 J(v) = \int_\Omega \delta\epsilon(v) : \mathbb{C} : \delta\epsilon(v) \, \mathrm{d}\Omega > 0 \text{ for } \mathbb{C} \quad \text{positive definite}, \tag{271}$$

which proves the minimal property of J.

7.2. Test and trial spaces for variational calculus

7.2.1. Hilbert and Sobolev spaces

By the scalar (or inner) product

$$(u,v)_0 := (u,v)_{\mathcal{L}_2(\Omega)} = \int_\Omega u(\boldsymbol{x})\, v(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{272}$$

the space $\mathcal{L}_2(\Omega)$ of quadratically integrable functions in Ω is completed to a Hilbert space (complete vector space with an inner product) with the norm

$$\|u\|_0 = \sqrt{(u,u)_0}. \tag{273}$$

This important function space was introduced by David Hilbert (1862–1943), [45].

In direct variational methods with C^0-continuous displacement ansatz for primal FEM the crucial problem of jumps of the derivatives of displacements at element interfaces arises because the test and trial functions $\boldsymbol{v}(\boldsymbol{x})$ and $\boldsymbol{u}(\boldsymbol{x})$ have to be only C^0-continuous at element interfaces and Neumann boundaries. Therefore, the derivatives (the strains) and thus the stresses as well as their projections with the normal vectors at element interfaces are not C^0-continuous. By this the Hilbert spaces are not sufficient for finite element analysis, and this was the reason that the mathematics of FEM was only developed from the 1970s on, based on Sobolev spaces $\mathcal{H}^m(\Omega)$, discovered by Sergei L. Sobolev (1908–1989) in his famous work from 1963, [80], in the Steklov Institute for Mathematics, Moscow.

The Sobolev space $\mathcal{H}^m(\Omega)$ is the set of all functions $\boldsymbol{u}(\boldsymbol{x}) \in \mathcal{L}_2(\Omega)$ which have weak derivatives $\partial^\alpha u \,\forall\, |\alpha| \le m$.

In $\mathcal{H}^m(\Omega)$ a scalar product

$$(u,v)_m := \sum_{|\alpha|\le m} (\partial^\alpha u, \partial^\alpha v)_0 \tag{274}$$

is defined with the complete norm

$$\|u\|_m := \sqrt{(u,u)_m} = \sqrt{\sum_{|\alpha|\le m} \|\partial^\alpha u\|^2_{\mathcal{L}_2(\Omega)}}; \|\boldsymbol{u}\|_m > 0 \quad \text{for} \quad \boldsymbol{u} \ne \boldsymbol{0} \tag{275}$$

and the seminorm (the internal energy with allowed rigid body modes)

$$|u|_m := \sqrt{\sum_{|\alpha|=m} \|\partial^\alpha u\|^2_{\mathcal{L}_2(\Omega)}}; |\boldsymbol{u}|_m \ge 0 \quad \text{for} \quad \boldsymbol{u} \ne \boldsymbol{0}. \tag{276}$$

Inclusions of the Sobolev spaces are:

$$\begin{array}{ccccccc}
\mathcal{L}_2(\Omega) = & \mathcal{H}^0(\Omega) & \supset & \mathcal{H}^1(\Omega) & \supset & \mathcal{H}^2(\Omega) & \supset \ldots \\
& \| & & \cup & & \cup & \\
& \mathcal{H}^0_0(\Omega) & \supset & \mathcal{H}^1_0(\Omega) & \supset & \mathcal{H}^2_0(\Omega) & \supset \ldots
\end{array} \tag{277}$$

7.3. The Poincaré-Friedrichs' inequality

For estimating the interpolation error of FE discretizations the Poincaré-Friedrichs'-inequaltity is of great importance.

Suppose that the domain Ω is contained in an n-dimensional cube. Then the theorem holds

$$\|\boldsymbol{v}\|_0 \le s \, |\boldsymbol{v}|_1 \ \forall \, \boldsymbol{v} \in \mathcal{H}_0^1(\Omega) \tag{278}$$

with s the edge length of the cube and

$$\boldsymbol{v}_\Omega = \frac{1}{|\Omega|} \int\limits_\Omega \boldsymbol{v}(\boldsymbol{x}) \, \mathrm{d}\Omega \tag{279}$$

the average value of $\boldsymbol{v}$ over Ω, with $|\Omega|$ standing for the Lebesgue measure of the domain Ω.

The proof only requires zero boundary conditions on Γ_D; $\Gamma = \Gamma_D \cup \Gamma_N = \partial\Omega$; $\Gamma_D \cap \Gamma_N = \varnothing$.

This theorem states that the $\mathcal{L}_2$ norm of the test function can be estimated from above by the seminorm in $\mathcal{H}_0^1$, i.e., for test functions which are C^0 continuous at interfaces and zero at Dirichlet boundaries.

These theorems were established by Henri Poincaré (1854-1912) and a little bit later in an alternate form by Kurt Otto Friedrichs (1901–1982). They are important for proving upper bound properties of finite element solutions for growing numbers of degrees of freedom (DOFs).

7.4. The Céa-lemma – optimality of the Galerkin method

The Céa-lemma states:

The Galerkin approximation of the variational $\mathcal{V}$-elliptic form, Eqs. (262) and (263),

$$a(\boldsymbol{u}, \boldsymbol{v}) = F(\boldsymbol{v}) + g(\boldsymbol{v}), \tag{280}$$

with $\boldsymbol{v} \in \mathcal{V}$, namely

$$a(\boldsymbol{u}_h, \boldsymbol{v}_h) = F(\boldsymbol{v}_h) + g(\boldsymbol{v}_h) \text{ with } \boldsymbol{v}_h \in \mathcal{V}_h \subset \mathcal{V}, \tag{281}$$

yielding

$$\boldsymbol{u}_h \to \boldsymbol{u} \text{ for } N \to \infty, \tag{282}$$

is quasi-optimal, i.e., better than any other approximation.

The Céa-lemma reads

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_m \le \frac{C}{\alpha} \inf_{\boldsymbol{v}_h \in \mathcal{V}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_m, \tag{283}$$

with $M = \dfrac{C}{\alpha}$ the ellipticity constant.

The simple proof for this important theorem starts with the subtraction of the variational forms for the true and the approximated solution, i.e. from the error $\boldsymbol{e} := \boldsymbol{u} - \boldsymbol{u}_h$, yielding

$$a(\underbrace{\boldsymbol{u} - \boldsymbol{u}_h}_{=:\boldsymbol{e}}, \boldsymbol{v}) = 0 \ \forall \ \boldsymbol{v} \in \mathcal{V}_h; \ a(\boldsymbol{e}, \boldsymbol{v}) = R \ne 0 \ \forall \ \boldsymbol{v} \in \mathcal{V}, \tag{284}$$

with the message, that the residuum of the weak form of equilibrium is unequal zero for the approximated solution. Postulating the coercivity of the internal energy, i.e. $a(\boldsymbol{v}, \boldsymbol{v}) \Rightarrow \infty$ for $\boldsymbol{v} \Rightarrow \infty$, and introducing another test space $\boldsymbol{w} := \boldsymbol{v}_h - \boldsymbol{u}_h \in \mathcal{V}_h$ and with $a(\boldsymbol{u} - \boldsymbol{u}_h, \boldsymbol{v} - \boldsymbol{v}_h) = 0$ it follows

$$\alpha \|\boldsymbol{u} - \boldsymbol{u}_h\|_m^2 \le a(\boldsymbol{u} - \boldsymbol{u}_h, \boldsymbol{u} - \boldsymbol{u}_h) = a(\boldsymbol{u} - \boldsymbol{u}_h, \boldsymbol{u} - \boldsymbol{v}_h) + a(\boldsymbol{u} - \boldsymbol{u}_h, \boldsymbol{v}_h - \boldsymbol{v}),$$
$$\alpha \|\boldsymbol{u} - \boldsymbol{u}_h\|_m^2 \le C \|\boldsymbol{u} - \boldsymbol{u}_h\|_m \|\boldsymbol{u} - \boldsymbol{v}_h\|_m. \tag{285}$$

This theorem was proven by Jean Céa (*1932) in 1964, [27].

7.5. The Lax-Milgram theorem of the existence and uniqueness of stationary (extremal) solutions

Under the presumptions of the coercivity of the internal energy, i.e. the bilinear form of the test function v, $a(v, v) \geq \alpha \|v\|_{\mathcal{V}}^2$, $a(v, v) \to \infty$ as $|v| \to \infty$, Fig. 48a, and furthermore of the existence of a continuous linear functional $F(v)$, $F(v) \leq C\|v\|_{\mathcal{V}} < \infty$ with a linear growth of $F(v)$ with $\|f\|$, Fig. 48b, and postulating that the displacements are bounded by the load, i.e. equilibrium is stable, the following Lax-Milgram theorem holds which was completely proven by Babuška.
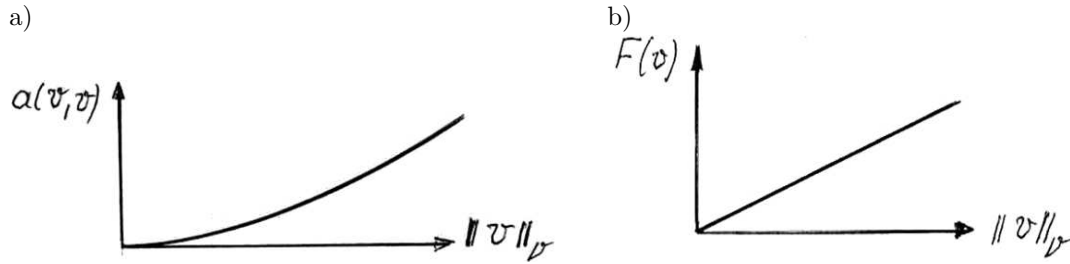

Fig. 48. a) coercivity of strain energy $a(v, v)$, b) continuous linear functional $F(v)$.

Theorem

For the bilinear form $a(v, v) : \mathcal{H}^1 \times \mathcal{H}^1 \to \mathbb{R}$ and $F(v) \in \mathcal{H}_0 \to \mathbb{R}$, the variational problem

$$J(v) = \min_{v} J(v); \;\; J(v) := 1/2 a(v, v) - F(v) = \min_{v} \tag{286}$$

has one solution in $\mathcal{V}$.

The proof uses the Poincaré-Friedrichs'-inequality and yields:

1. Existence:

$$a(v, v) \geq 1/2\alpha \|v\|^2 - \|f\| \|v\| \tag{287}$$

$$\alpha \|\Delta v\|^2 \leq a(\Delta v, \Delta v), \;\; \text{for } \Delta v := v_n - v_m, n > m \tag{288}$$

and

2. Uniqueness: The solution u is unique, because each minimal sequence $a(v, v) \geq \alpha \|v\|^2$ is a Cauchy sequence.

7.6. A priori error estimates in the energy norm

Theorem: Given a quasi-uniform triangulation $\mathcal{T}$ of Ω; then the FE approximation $u_h \in \mathcal{V}_h$ with linear shape functions yields the a priori error estimate

$$\|e\|_1 = \|u - u_h\|_1 \leq Ch \|u\|_2 \leq Ch \|f\|_0 \tag{289}$$

$$\||e\|| = \int_{\Omega} \epsilon(e) : \mathbb{C} : \epsilon(e) \, \mathrm{d}\Omega, \;\; \text{the energy norm} \tag{290}$$

$$\||e\|| = C\mathcal{O}(h^p) \tag{291}$$

$$\|e\|_{\mathcal{L}_2} = C\mathcal{O}(h^{p+1}), \tag{292}$$

With $v_h(x) \in [\mathcal{H}^1(\Omega)]^3$, $f(x) \in [\mathcal{L}_2(\Omega)]^3$ and sufficient regularity of the BCs, for Lamé-BVPs the approximated displacements u_h converge to $u(x) \in [\mathcal{H}^2(\Omega)]^3$.

This was proven by Aubin, [8], and Nitsche, [66].

8. Residual (energy-based) explicit (relative) a posteriori error estimator for mixed BVPs of the Lamé's equations, Babuška, Miller (1987)

The weak form for primal FEM of mixed BVPs reads

$$a(\boldsymbol{u}_h, \boldsymbol{v}_h) = F(\boldsymbol{v}_h) + g(\boldsymbol{v}_h) \ \forall \ \boldsymbol{v}_h \in \mathcal{V}_h. \tag{293}$$

Introducing the discretization error $\boldsymbol{e} := \boldsymbol{u} - \boldsymbol{u}_h$, the residuum of the weak form is

$$a(\boldsymbol{e}, \boldsymbol{v}) = a(\boldsymbol{e}, \underbrace{\boldsymbol{v} - \boldsymbol{w}_h}_{I_h \boldsymbol{v}}) := R(\boldsymbol{v} - \boldsymbol{w}_h) \ \forall \ \boldsymbol{v} \in \mathcal{V}, \ \forall \ \boldsymbol{w}_h \in \mathcal{V}_h \tag{294}$$

and yields the inequality

$$a(\boldsymbol{e}, \boldsymbol{v}) \le \sum_{e=1}^{n_e} \boldsymbol{R}_e(\boldsymbol{u}_h) \cdot \boldsymbol{v} \ dV + \sum_{r=1}^{n_r} \boldsymbol{J}_r(\boldsymbol{u}_h) \cdot \boldsymbol{v} \ dA \ , \tag{295}$$

with the residua of equilibrium $\boldsymbol{R}_e$ in Ω_e and the jumps of tractions $\boldsymbol{J}_r$ at Γ_r.

8.1. Residual (energy-based) explicit a posteriori error estimator with strict upper and lower bounds

The a priori upper bound estimate in the energy norm reads

$$|||\boldsymbol{e}||| \le C \, \mathcal{O}(h^p), \tag{296}$$

where the global interpolation constant C is the product of a proper interpolation and a stability constant, depending on the differential operator and the BVP. It is unknown for explicit residual a posteriori error estimators and can become large, depending on the ellipticity property and the well-posedness of the BVP. For example, it may be large for strong elastic anisotropy.

In double-logarithmic coordinates this estimator yields a straight line with the slope $\tan \alpha(p) = \frac{p}{1}$, i.e. for $p = 1$ the stresses converge linearly, Fig. 49.
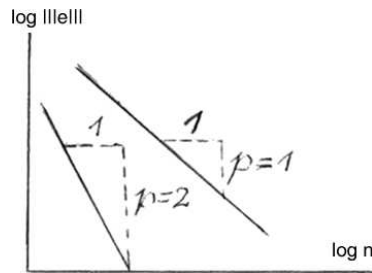
Fig. 49. Linear and quadratic convergence of the stresses for $p = 1$ and $p = 2$.

With the lower interpolation estimate according to the Poincaré-Friedrichs'-inequality

$$\|\boldsymbol{u} - I\boldsymbol{u}\|_{\mathcal{H}_1(\Omega)} \le |||\boldsymbol{u} - I\boldsymbol{u}||| \tag{297}$$

the energy norm of the exact error $\boldsymbol{e} = \boldsymbol{u} - \boldsymbol{u}_h \in \mathcal{V}$ is bounded from above and from below by the important result of Babuška, Rheinboldt (1978), [11],

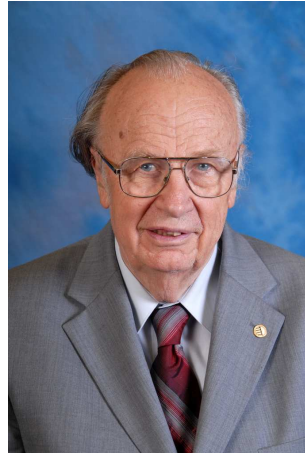$$C_l |||\Pi_h \boldsymbol{e}||| \le |||\boldsymbol{e}||| \le C_u |||\Pi_h \boldsymbol{e}||| \ , \tag{298}$$

where $\Pi_h \boldsymbol{e}$ is the computable projection of $\boldsymbol{e} \in \mathcal{V}$ onto the ansatz space $\mathcal{V}_h$.

The effectivity index quantifies the quality of the estimated error with respect to the real error as, Babuška (1980), Fig. 50,

$$\theta = \frac{\eta}{|||\boldsymbol{e}|||}; \ \ \eta = \left(\sum_{i=1}^{n} \boldsymbol{e}_i^2\right)^{1/2}. \tag{299}$$

Fig. 50. Ivo M. Babuška (*1926), who contributed seminal theoretical results on a priori and a posteriori error estimation of FEM.

In case of explicit residual error estimators the calculation of θ does not make sense because the global constant C is unknown, and the choice $C = 1$ usually yields too large effectivity indices, which – for upper bound esimators – should not be larger than about 2 for efficient adaptivity.

The Babuška-Miller explicit residual a posteriori error estimator, [10], controls the approximated equilibrium conditions in each element, at element interfaces and at Neumann boundaries.

8.2. Residual implicit a posteriori error estimates

These constant-free estimators, developed since the 1980s, are calculated on element patches or on single elements (via improved (averaged) tractions at element interfaces), and they need the solution of local equation systems. Upper and lower bounds can be guaranteed using Cauchy-Schwartz inequality; other versions with different norms yield good effectivity indices with non-guaranteed bounds.

Numerous implicit estimators have been developed, and they are subjects of current research. Their representation in this article would require too much space and – even though the author is strongly active in this field since the 1970s – this relevant topic is not treated here. But some important articles and books are cited, namely by Bank and Weiser, [13], Braess, [18], Verfürth, [95], Ainsworth and Oden, [1], Stein and Rüter, [84], as well as Ladevèze and Pelle, [54]. In the article by Stein, Rüter and Ohnimus, [85], Neumann-type implicit goal-oriented error estimators via improved interface tractions are presented for both verification and validation according to experimental results.

9. Gradient smoothing "ZZ" explicit and implicit a posteriori error estimators

Different from the explicit residual error estimator by Babuška and Miller (1987), [10] and [11], O.C. Zienkiewicz and J.Z. Zhu, [100], published an a posteriori error estimator based on gradient-smoothing of stresses and recalculation of improved stresses for which the shape functions are

chosen equal to those for the primal displacement ansatz functions which are C^0-continuous at element interfaces, Fig. 51.

Fig. 51. Olgierd C. Zienkiewicz (1921–2009), one of the most important developers of computational mechanics in its first period since the 1960s.

9.1. Explicit (relative) "ZZ" smoothing estimator

For the primal FEM with the (e.g. linear) discretized displacements $\boldsymbol{u}_h(\boldsymbol{x}) = \boldsymbol{N}(\boldsymbol{x})\hat{\boldsymbol{u}}_e$ in Ω_e, $\Omega = \cup_e \Omega_e$, and discretized strains $\boldsymbol{\epsilon}_h(\boldsymbol{u}_h) = \boldsymbol{DN}\hat{\boldsymbol{u}}_e = \boldsymbol{B}\hat{\boldsymbol{u}}_e$ yields the discretized stresses $\boldsymbol{\sigma}_h(\boldsymbol{u}_h(\boldsymbol{x})) = \mathbb{C}\boldsymbol{B}(\boldsymbol{x})\hat{\boldsymbol{u}}_e$, according to the symmetric and positive definite elasticity tensor $\mathbb{C}$.

The idea of gradient smoothing is to introduce so-called *recovered* stresses

$$\boldsymbol{\sigma}_h^*(\boldsymbol{x}) \coloneqq \boldsymbol{N}(\boldsymbol{x})\,\hat{\boldsymbol{\sigma}}_e^* \ \forall \ \Omega_e \tag{300}$$

with same shape functions $\boldsymbol{N}(\boldsymbol{x})$ (e.g. linear) as for the displacement ansatz $\boldsymbol{u}_h(\boldsymbol{x})$ and with C^0-continuity at element interfaces, Fig. 52. Of course, it has to be proven that the recovered stresses $\boldsymbol{\sigma}_h^*$ are more accurate than the discretized stresses $\boldsymbol{\sigma}_h$. Then the least squares conditions for all elements read

$$\bigcup_{e=1}^{N_{el}} \int_{\Omega_e} \underbrace{[\boldsymbol{\sigma}_h^*(\boldsymbol{x}) - \boldsymbol{\sigma}_h(\boldsymbol{u}_h(\boldsymbol{x}))]^2}_{\boldsymbol{e}_\sigma^*(\Omega_e)} \, \mathrm{d}\Omega_e \to \min_{\hat{\boldsymbol{\sigma}}_k^*} \hookrightarrow \hat{\boldsymbol{\sigma}}_k^*, \tag{301}$$
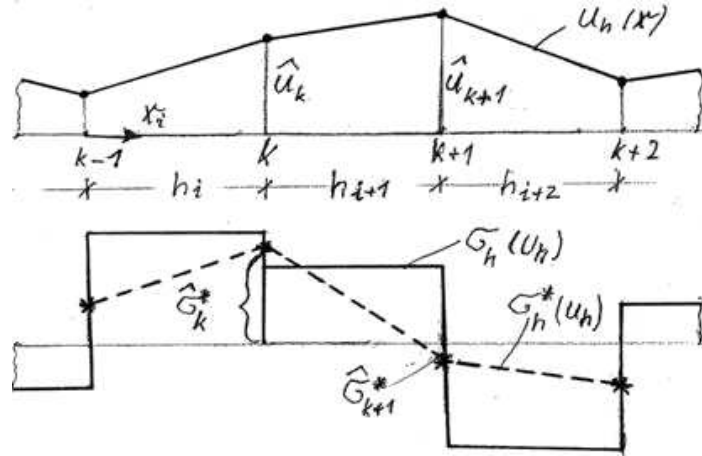
with the a posteriori gradient smoothing stress error

$$\boldsymbol{e}_\sigma^*(\boldsymbol{x}) \coloneqq \boldsymbol{\sigma}_h^*(\boldsymbol{x}) - \boldsymbol{\sigma}_h(\boldsymbol{u}_h(\boldsymbol{x})) \text{ in } \Omega_e \in \Omega, \tag{302}$$

yielding a system of linear equations for the nodal values $\hat{\boldsymbol{\sigma}}_k^*$ of nodes k.

The calculation of nodal values $\hat{\boldsymbol{\sigma}}_k^*$ for the whole system follows by solving the whole symmetric system of linear equations (the stationarity condition) or by considering approximately only the diagonal equation matrix or by nodal averaging for adjacent elements at each node k.

The global error estimator in the complementary energy norm then reads

$$\||\boldsymbol{e}_\sigma^*\||_\Omega = \left[\sum_{e=1}^{N_{el}} \left(\int_{\Omega_e} \boldsymbol{e}_\sigma^* : \mathbb{C}^{-1} : \boldsymbol{e}_\sigma^* \, \mathrm{d}\Omega \right) \right]^{1/2}. \tag{303}$$

Fig. 52. 1D linear C^0-continuous displacement ansatz (top), and discretized stresses $\boldsymbol{\sigma}_h(\boldsymbol{u}_h)$ as well as recovered C^0-continuous stresses $\boldsymbol{\sigma}_h^*(\boldsymbol{u}_h)$ (bottom).

There were numerous debates amongst mathematicians and mechanicians in the 1980s and 1990s whether this estimator was really mathematically sound, yielding upper and lower bounds for well-posed BVPs with sufficient regularity properties, or in other words, whether and how the recovered stresses are more accurate than the discretized ones. On the other hand-numerous comparative convergence studies showed the superiority or at least equal convergence behaviour with respect to application of explicit residual error estimators. Then an upper bound was proven by Rodríguez (1994), [77], as

$$\|\boldsymbol{\sigma}(\boldsymbol{u}) - \boldsymbol{\sigma}^*(\boldsymbol{u}_h)\|_{L_2(\Omega)} \leq C \|\boldsymbol{\sigma}(\boldsymbol{u}) - \boldsymbol{\sigma}(\boldsymbol{u}_h)\|_{L_2(\Omega)}, \tag{304}$$

where C is a global interpolation constant. Upper and lower bounds were proven by Carstensen and Funken (2001), [28], as

$$\frac{1}{1+C} \underbrace{\|\boldsymbol{\sigma}^*(\boldsymbol{u}_h) - \boldsymbol{\sigma}(\boldsymbol{u}_h)\|_{L_2(\Omega)}}_{e^*(\boldsymbol{\sigma}^*)} \leq \underbrace{\|\boldsymbol{\sigma}(\boldsymbol{u}) - \boldsymbol{\sigma}(\boldsymbol{u}_h)\|_{L_2(\Omega)}}_{e(\boldsymbol{\sigma}(\boldsymbol{u}_h))}$$

$$\leq \frac{1}{1-C} \underbrace{\|\boldsymbol{\sigma}^*(\boldsymbol{u}_h) - \boldsymbol{\sigma}(\boldsymbol{u}_h)\|_{L_2(\Omega)}}_{e^*(\boldsymbol{\sigma}^*)}. \tag{305}$$

9.2. Implicit (absolute) "ZZ" SPR smoothing estimator

Zienkiewicz and Zhu published in 1992 an important improvement of their explicit estimator by the so-called *SPR* (superconvergent patch recovery) which is performed with postprocessing on element patches $\bar{\Omega}_P$ around the considered element nodes (instead of nodal averaging of stress components), [101], Fig. 53.

SPR is based on the fact that – in case of sufficient regularity – the discretized stresses $\boldsymbol{\sigma}(\boldsymbol{u}_h)|_{\Omega_e}$ are more accurate at Gauss points (GPs) of elements Ω_e than at element interfaces $\partial\Omega_e$ and nodal points of the elements. Chosing these GPs as optimal sampling points $\boldsymbol{\xi}_n(\bar{\Omega}_P)$ superconvergence can be achieved for the improved nodal stress components $\hat{\sigma}_I^*(\boldsymbol{X}_k)|_{\bar{\Omega}_P}$.

The vector of the searched nodal stress components at patch node k is presented by bilinear monomials for 2D problems, e.g. using $\mathbb{Q}_1$-quadrilateral elements, as

$$\hat{\sigma}_I^*(\boldsymbol{X}_k)|_{\bar{\Omega}_P} := \Pi^T(\boldsymbol{X}_k)\hat{\boldsymbol{a}}_I; \ \Pi = [1 \ x \ y \ xy]^T \text{ for } \Omega_e \subset \mathbb{R}^2. \tag{306}$$
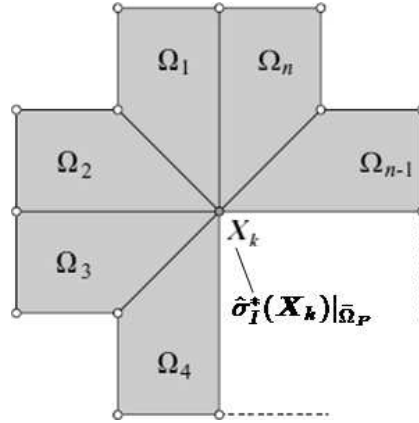
Fig. 53. SPR for element patch $\bar{\Omega}_P$ surrounding node k.

The unknown coefficients $\hat{\boldsymbol{a}}_I$ are determined by the conditions

$$\sum_{n_{GP|\bar{\Omega}_P}} [\sigma_I(\boldsymbol{u}_h(\boldsymbol{\xi}_n)) - \Pi^T(\boldsymbol{\xi}_n)\hat{\boldsymbol{a}}_I] = 0, \tag{307}$$

with the row-regular rectangular matrix Π. In case of choosing more sampling points than coefficients of the vector $\hat{\boldsymbol{a}}_{I,k}$ the least squares method

$$\sum_{n_{GP|\bar{\Omega}_P}} \Pi(\boldsymbol{\xi}_n)\Pi^T(\boldsymbol{\xi}_n)\hat{\boldsymbol{a}}_I = \sum_{n_{GP|\bar{\Omega}_P}} \Pi(\boldsymbol{\xi}_n)\boldsymbol{\sigma}_I(\boldsymbol{u}_h(\boldsymbol{\xi}_n)) \tag{308}$$

yields $\hat{\boldsymbol{a}}_I$ from a small symmetric, positive definite system of linear algebraic equations for each patch. Boundary nodes $\boldsymbol{X}_n(\partial\bar{\Omega})$ need a special treatment.

Superconvergence could be shown for many applications by applying quasi-optimal mesh refinements for the calculated a posteriori error estimators.

Important remark

From the state of the art of the available general purpose finite element programs there is a big advantage of the ZZ estimators compared against residual estimators. The reason is that the smoothing estimator $\||\boldsymbol{e}_{ZZ}^*\||$ only needs data from the considered elements themselves but not from the neighbouring elements – whereas the residual estimator $\||\boldsymbol{e}_{Res}\||$ requires the jumps of the tractions to neighbouring elements in the residuum of equilibrium. However, so far the needed access to element neighbourhoods is not available in all industrial finite element programs, and therefore, the ZZ estimators are predominant in practice. They are available in most program systems, combined with either local mesh refinements, e.g., for triangular and tetrahedral elements, or for global mesh refinements according to the distribution of scaled error indicators of the system, e.g., in Abaqus.

9.3. Techniques and rules for adaptive finite element discretizations

9.3.1. Triangular meshes for 2D problems

Figure 54 shows refinements of triangular elements.

One can easily proof that the condition number of the stiffness matrices of refined element patches depend on the ratio of side lengths and the angles between adjacent element sides. Therefore, bounds for these quantities have to be applied for local mesh refinements, [17, 75].

For global remeshing of triangular elements according to the distribution of scaled error indicators we especially refer to Delaunay-type triangulation algorithms, [17].
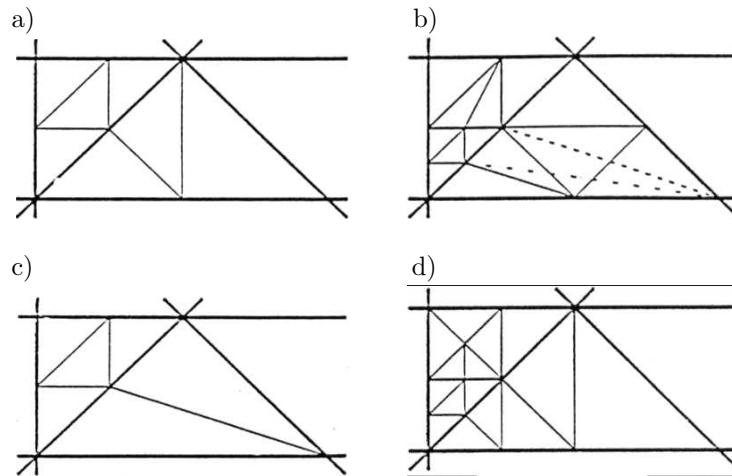
Fig. 54. a) standard refinements of triangles without irregular nodes; b) resolving of elements for next refined step in order to avoid small angles; c) halving of longest side (Rivara); d) no resolving of elements.

9.3.2. Quadrilateral meshes for 2D problems

Figure 55 shows refinements of quadrilateral elements, e.g. $\mathbb{Q}_1$-elements.

Two techniques are indicated: first, avoiding irregular nodes; irregular elements are introduced and resolved in the next refinement step, Fig. 55a,b, and secondly, the so-called hanging node technique where irregular nodes of regular elements are eliminated by proper interpolation of the shape functions, Fig. 55c,d.
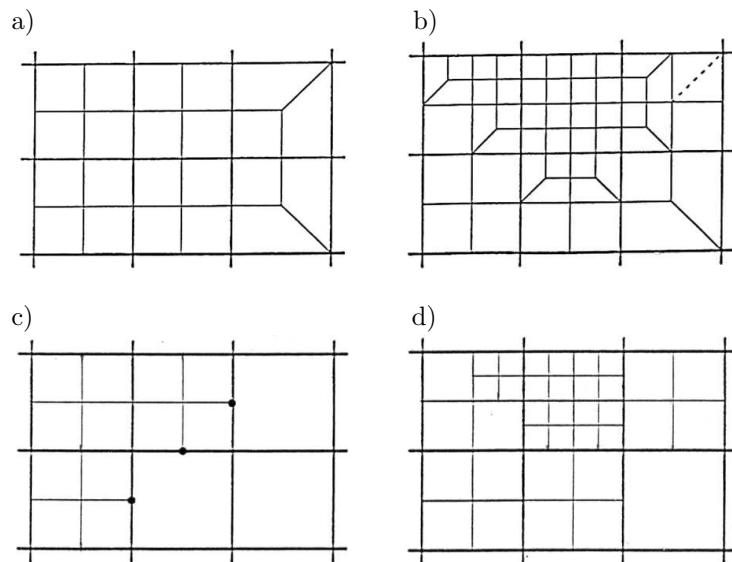
Fig. 55. Local refinement techniques for quadrilateral elements: a) and b) transition elements for avoiding irregular nodes, irregular elements are resolved in the next refinement step; c) and d) preserving regular elements by eliminating irregular nodes via hanging node technique.

It is obvious that the hanging node technique yields much better locality of the refinements and thus is recommendable.

In case of hexahedral elements for 3D problems also hanging node techniques on element surfaces should be applied for locality and simple algorithms, as well as regarding the fact that refinements with regular nodes yield some degenerated elements with hyperbolic surfaces. A local refinement based on preserving regular elements creates 84 new elements, [65].

10. Collection of a posteriori error estimates for primal FEM of Lamé's PDES

Table 1 shows a systematic classification of a posteriori error estimates for primal FEM. Next to residual and gradient smoothing estimators hierarchical estimators by Bank and Smith and gradient free estimators by Huerta, Diez et al. are included. The other arrangement shows explicit and implicit estimators.

Table 1. Survey of a posteriori error estimates for primal FEM.

	residual (Babuška, Rheinboldt, Miller, Bank, Weiser, Ladevèze, Oden, Stein)	**hierarchical** (Bank, Smith)	**gradient free** (Huerta, Diez, et.al.)	**gradient smoothing** (Zienkiewicz, Zhu; Carstensen, Funken)
explicit (relative) dep. from interpolation and stability constants	• upper and lower bounds • estimation of residua of equilib. in all Ω_e and at all Γ_e and at $\Gamma_{N,e}$	• lower and upper bound approx. • approx. in a hierarchically expanded subspace	• with equilibrated test stresses on element level • no data required at Γ_e	• upper bound property • element-wise C^0-cont. stress approximation • data only from Ω_e, not from Γ_e
implicit (absolute), with enhanced test space on el. or patch level	• equilib. of interface tractions, solv. local Neumann problems • const.-free error bounds		• lower bound property • enhanced test space required	• SPR with higher converg. property in Gauss integr.-points compared to nodal points • using higher polynomials for stresses on patches

11. Further important milestones of FEM

So far mainly primal finite element methods were presented here, but beginning with the 1920s with papers by Hellinger, Reissner, Prager and Washizu generalized variational principles were applied especially for parameter-dependent (i.e., non-robust) problems, leading to mixed finite element methods by adding relaxed field and/or interface conditions, multiplied with Lagrangian parameters, to the related variational principle. These methods yield saddle-point problems and need to fulfill global inf-sup stability conditions for the nodal values of displacements and stresses and have wide fields of applications.

Further generalizations of primal FEM for elastic static problems concern elastic dynamic, elasto-plastic and time-dependent visco-elastic and visco-plastic deformations.

Since the 1990s multiscale FEMs for multiphysical problems were developed, and concerning fracture mechanics, the so-called XFEM, was first published by Belytschko and Black (1999), [14]. Herein, the FE mesh is chosen theoretically independent from progressing cracks, and meshless FE discretizations are applied for complex problems. Furthermore, particle methods, [32], which are equivalent to element-free Galerkin methods are developed very fast and get theoretical foundation, [46].

On the other hand, isogeometric analysis, especially by Hughes et al., [47], was developed in the last decade very fast for applications with high continuity requirements for structural surfaces using NURBS and T-splines. These elements are discretizing the undeformed and the deformed bodies and surfaces and thus require C^1-continuity at element interfaces.

Contact problems are an important class of problems with a long tradition. Today, the micro-properties of contact zones, are crucial for modeling and computation, using multi-physical and multi-scale concepts, [98].

Table 2 gives a survey of these new discretization methods.

Table 2. Further important milestones of generalized FEMs.

Parameter-dependent BVPs with locking or spurious modes (numerical instabilities) for elastic problems, since the 1980s	• global inf sup condition has to be fulfilled • mixed FEs required in principle • engineering methods: reduced integr. enhanced assumed strains
FEM for kinematically nonlinear elastic problems, e.g. structural instabilities of beams, plates and shells	• geometrical stiffness matrix • incremental methods, Lagrangian, Eulerian, ALE • consistent numerical tangent matrix • step length control
FEM for inelastic and time-dependent visco-elastic-plastic deformations	• predictor-corrector methods with projections to actual yield surface; continuous and discontinuous time discretization
FEM for elasto-dynamic problems	• linear dynamics in time or in phase space • FEs in time and space, numerical integration in time, problem: stability, conservation of energy and momentum
FE^2 for multiscale FEM	homogenization from micro and meso to macro-scale via Hill's equation and projection methods
XFEM, esp. for fracture mechanics	singular functions from analytical solution at crack tip as enrichment functions
Particle Methods	• equivalent to element-free Galerkin method; problems: stability analysis • use of meshless discretizations
Isogeometric analysis	FEM with real geometry, integration of computer aided design and finite element analysis, using NURBS, T-splines, ...

12. DEVELOPMENT OF EQUATION SOLVERS IN CONJUNCTION WITH HARD- AND SOFTWARE PROGRESS

Not only consistent discretization methods and their achievable accuracy but moreover the adequate mathematical modeling with engineering requirements for accuracy and physical depth as well as – very important – the dimension of the resulting linear or non-linear algebraic equation systems are highly dependent on the power and performance of the available computers.

The numerical solution of boundary value problems with classical finite element methods (primal variational methods) for linear elliptic problems leads to – usually large-dimensional – linear algebraic equation systems with symmetric and positive definite global stiffness matrices. Thus, the computational capacity depends on the computer power and the solution methods. The break even point of efficiency for direct solvers (by Gauss-elimination) and iterative solvers has moved to larger and larger equation systems, according to the fast development of CPUs and storage media.

In 1965, Moore's law was published, [64], assessing in the period from 1958 to 1965 that the number of transistors that can be placed on an integrated circuit doubles approximately every two years. The processing speed, the memory capacity and some more properties of electronic computers

are strongly connected to this law. Moore's law of predicting an exponential growth has proved to be true until today.

Viewing the today's situation, linear equation systems with about 10^6 DOFs can be solved on an average-price laptop using direct elimination methods within less than an hour.

Direct solvers for linear algebraic equation systems

The first choice for solving linear equation systems with symmetric and positive definite system matrices $\boldsymbol{A}$ is Gauss's stepwise elimination by $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^{\mathrm{T}}$ factorization into lower and upper triangular and diagonal matrices, usually using row-pivoting, and following recursive back-calculation of the unknown vector, here the global nodal displacement vector.

Better efficiency is achieved by complete or incomplete Gauss-Cholesky decompositions according to $\boldsymbol{A} = \widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^{\mathrm{T}}$ with $\widehat{\boldsymbol{L}} \coloneqq \boldsymbol{L}\boldsymbol{D}^{1/2}$.

Sparse solution algorithms were developed for weakly occupied diagonally-oriented system matrices, providing considerable savings in storage demands and CPU time, especially for 3D finite element discretizations, see e.g., Duff and Reid, [36].

The program system LINPACK offers various solvers for linear algebraic systems including unsymmetric and sparse system matrices.

An often used algorithm concerns overdetermined linear equation systems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with rectangular matrices $\boldsymbol{A}_{(m,n)}$, $m > n$, with m rows and n columns, $\boldsymbol{A}$ is column-regular. These matrices appear, e.g., in least squares algorithms and overdetermined collocation methods. Then the Gauss transformation $\widehat{\boldsymbol{A}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ yields a symmetric, positive definite matrix which can be eliminated.

Iterative solvers for linear and non-linear algebraic equation systems

Large linear and non-linear algebraic equation systems need iterative solvers, according to the relevant break-even point. According to the i-th iteration step for the residuum $\boldsymbol{r}^i = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^i, i = 0, 1, 2, \ldots, n$. The number of iterations until tolerance of the $\mathcal{L}^2$-norm of the residual vector $\boldsymbol{r}$ should be not larger than about 100.

Since the 1950s, Newton's method and quasi-Newton methods (for sparse systems) as well as Jacobi and Gauss-Seidel methods with the decomposition $\boldsymbol{A} = \boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U}$ are applied.

Very large systems are predominantly solved with conjugate gradient (CG) methods, see e.g., Hackbusch, [41].

The gradient (steepest descent) method with line search for minimization of a quadratic function $f(\boldsymbol{x}) = 1/2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$ needs the negative gradient $\boldsymbol{d}^i = -\nabla f(\boldsymbol{x}^i)$. With the line search $\boldsymbol{x}^{i+1} = \boldsymbol{x}^i + \alpha^i\boldsymbol{d}^i$, a sequence $f(\boldsymbol{x}^0) > f(\boldsymbol{x}^1) > \cdots > f(\boldsymbol{x}^i)$ can be constructed where equality holds if the gradient vanishes. For a quadratic function $f(\boldsymbol{x})$ one gets $\boldsymbol{d}^i = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^i$ with the line search

$$\alpha^i = \frac{\boldsymbol{d}^{i\mathrm{T}}\boldsymbol{d}^i}{\boldsymbol{d}^{i\mathrm{T}}\boldsymbol{A}\boldsymbol{d}^i}.$$

The gradient method for solving $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ uses the distance measure $\|\boldsymbol{x}\|_A \coloneqq \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}}$. If $\boldsymbol{x}^*$ is a solution of the equation system, then the function $f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|_A^2$ yields the sequence $\|\boldsymbol{x}^i - \boldsymbol{x}^*\|_A \le \left(\frac{\kappa - 1}{\kappa + 1}\right)^i \|\boldsymbol{x}^0 - \boldsymbol{x}^i\|_A$, where κ is the spectral condition number, see, e.g. Braess, [18].

The conjugate gradient method, developed by Hesteness and Stiefel in 1952, becomes more effective by preconditioning due to Reid in 1971. The idea is to construct conjugate directions $\boldsymbol{d}^i$ and $\boldsymbol{d}^{i+1}$ in iterative steps fulfilling the generalized orthogonality condition $\boldsymbol{d}^{i+1\mathrm{T}}\boldsymbol{A}\boldsymbol{d}^i = 0$. Then, the wanted solution $\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$ has the expansion $\boldsymbol{x}^* = \sum_{i=0}^{n-1} \alpha^i\boldsymbol{d}^i$, with the line search $\alpha^i = \frac{\boldsymbol{d}^{i\mathrm{T}}\boldsymbol{b}}{\boldsymbol{d}^{i\mathrm{T}}\boldsymbol{A}\boldsymbol{d}^i}$.

Furthermore, **Multi-Grid Methods (MGM)** and **Domain-Decomposition Methods (DDM)** play an important role for solving large systems since the 1980s and, moreover, **parallel computing** with distributed processor techniques (with more than $100,000$ processor units) are used for solving the big coupled system simulating multiphysics coupled with chemistry and other disciplines, e.g., for the prediction of the earth climate and expectable earthquakes on our planet, today requiring 1000-times solutions of non-linear equation systems of the dimension 10^8.

13. DEVELOPMENT OF GENERAL PURPOSE FINITE ELEMENT PROGRAM SYSTEMS

Early finite element programs for static and dynamic analysis with primal FEM (with some exceptions) were developed in the 1960s and 1970s mostly using FØRTRAN program language and some were written in ALGOL. Those are, see [25]:

1. AMSA 9 and AMSA 20, FIAT, Divisione Aviazioni – SCV, Turin, Italy

2. ASAS, Atkins Research and Development, Woodcote Grove, England

3. ASKA, Institut für Statik und Dynamik der Luft- und Raumfahrtkonstruktionen, Universität Stuttgart, Germany and IKOSS – Software Service Stuttgart

4. BERSAFE, Central Electricity Generating Board, Berkeley Nuclear Laboratories, Berkeley, Gloucestershire, England

5. EASE (FIDES), FIDES Rechenzentrum, Dr. D. Pfaffinger, Zürich, Switzerland

6. MARC 2, Marcal, Div. of Eng., Brown University, Providence, RI, USA and Control Data GmbH, Stuttgart, Germany

7. NASTRAN, NASTRAN Systems, Langley Research Center, Hampton, VA, USA

8. PRAKSI-RIB, Rechen-Institut fuer EDV im Bauwesen, Stuttgart, Germany

9. SAP, Prof. E. L. Wilson, Dept. of Civil Eng., University of California at Berkeley, CA, USA

10. SESAM 69, A/S Computas, Økernvein 145, Oslo 5, Norway

Internationally available and competitive general purpose finite element program systems for linear and non-linear thermomechanical (and today also for multi-physical) mathematical models were developed since the 1980s. They are mostly written in FØRTRAN 77 and FØRTRAN 90; those are, e.g., NASTRAN, ANSYS, Abaqus and ADINA.

Different from these commercial program systems, the general FE program FEAP by Prof. R.L. Taylor, University of California at Berkeley, has a command language and is written in FØRTRAN 77. The complete source code is available for the users mostly working in academia.

Remark

It has to be emphasized that various object-oriented finite element programs with distributed vector and matrix data structures have been developed especially at university institutes since the 1990s, usually with the C++ program language. Within this concept, strict separation of topology, geometry, mathematical modeling, numerical methods, equation solvers, data evaluation and visualization was realized, e.g., in Niekamp and Stein, [65]. This provides a lot of fundamental benefits, but is not realized so far in the above mentioned commercial program systems. This has crucial disadvantages for the users; some examples are:

(i) optimization problems are hard to realize because the geometrical data should be based on topological and parametric representations, and

(ii) mesh adaptivity with residual a posteriori error estimators requires the access to neighboured elements in order to calculate the difference of interface tractions, but none of the mentioned program systems admits this. Only access to element data within element domains is provided.

Therefore, it is due time that the relevant companies – why not joint ventures with capable industries – develop new object-oriented program systems for the next technological generation with virtual product development and multi-coupled problems. These new developments are necessary to keep computational methods economical and payable.

REFERENCES

[1] M. Ainsworth, J.T. Oden. *A posteriori error estimators in finite element analysis*. John Wiley & Sons, New York, 2000.
[2] J.H. Argyris. Energy theorems and structural analysis. *Aircraft Engineering*, **26:** 347–356, 383–387, 394, 1954.
[3] J.H. Argyris. Energy theorems and structural analysis. *Aircraft Engineering*, **27:** 42–58, 80–94, 125–134, 154–158, 1955.
[4] J.H. Argyris, K.E. Buck, D.W. Scharpf, H.M. Hilber, G. Mareczek. Some new elements for the matrix displacement method. In *Proc. 2nd Conf. on Matrix methods in Structural mechanics*. Wright-Patterson Air Force Base, Ohio, 1968.
[5] J.H. Argyris, S. Kelsey, H. Kamel. *Matrix methods of structural Analysis*. AGARDograph 72. Pergamon Press, London, 1963.
[6] D.N. Arnold, F. Brezzi, J. Douglas Jr. PEERS: A new mixed finite element for plane elasticity. *Japan. J. Appl. Math.*, **1:** 347–367, 1984.
[7] D.N. Arnold, J. Douglas Jr., C.P. Gupta. A family of higher order finite element methods for plane elasticity. *Numer. Math.*, **45:** 1–22, 1984.
[8] J.P. Aubin. Behaviour of the error of the approximate solution of boundary value problems for linear elliptic operators by Galerkin's and finite difference methods. *Ann. Scuola Norm. Sup. Pisa*, **21:** 599–637, 1967.
[9] I. Babuška. The finite element method with Lagrangian multipliers. *Numer. Math.*, **20:** 179–192, 1973.
[10] I. Babuška, A. Miller. A feedback finite element method with a posteriori error estimation: Part i. the finite element method and some basic properties of the a posteriori error estimator. *Comput. Methods Appl. Mech. Engng.*, **61:** 1–40, 1987.
[11] I. Babuška, W.C. Rheinboldt. A-posteriori error estimates for the finite element method. *Int. J. Num. Meth. Engng.*, **12:** 1597–1615, 1978.
[12] I. Babuška, B. Szabó, I.N. Katz. The p-version of the finite element method. *SIAM J. Numer. Anal.*, **18:** 515–545, 1981.
[13] R.E. Bank, A. Weiser. Some a posteriori error estimators for elliptic partial differential equations. *Math. Comp.*, **44:** 283–301, 1985.
[14] T. Belytschko, T. Black. Elastic growth in finite elements with minimal remeshing. *Int. J. Num. Meth. Eng.*, **45:** 601–620, 1999.
[15] E. Betti. Teorema generale intorno alle deformazioni che fanno equilibrio a forze che agiscono alla superficie. *Il Nuovo Cimento, series 2*, **7−8:** 87–97, 1872.
[16] M. Bischoff, E. Ramm, D. Braess. A class of equivalent enhanced assumed strain and hybrid stress finite elements. *Comp. Mech.*, **22:** 443–449, 1999.
[17] H. Borouchaki, P.L. George, F. Hecht, P. Laug, E. Saltel. Delaunay mesh generation governed by metric specifications. Part I. Algorithms. *Finite Elem. Anal. Des.*, **25:** 61–83, 1997.
[18] D. Braess. *Finite elements*. Cambridge University Press, 2nd edition, 2001, (1st edition in German language, 1991).
[19] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers. *Rev. Fr. Automat. Inf. Rech. Opérat. Sér. Rouge*, **8:** 129–151, 1974.
[20] F. Brezzi, M. Fortin. *Mixed and hybrid finite element methods*. Springer New York, 1991.

[21] F. Brezzi, J. Douglas Jr., M. Fortin, L.D. Marini. Efficient rectangular mixed finite elements in two and three space variables. *RAIRO M$_2$AN*, **21:** 581–604, 1987.

[22] F. Brezzi, J. Douglas Jr., L.D. Marini. Two families of mixed finite elements for second-order elliptic problems. *Numer. Math.*, **47:** 217–235, 1985.

[23] I. Bubnov. *Structural Mechanics of Ships (translated from Russian)*. 1914.

[24] K.E. Buck, D.W. Scharpf. Einführung in die Matrizen-Verschiebungsmethode. In K.E. Buck, D. W. Scharpf, E. Stein, W. Wunderlich, editors, *Finite Elemente in der Statik*, pages 1–70. Wilhelm Ernst & Sohn, Berlin, 1973.

[25] K.E. Buck, D.W. Scharpf, E. Schrem, E. Stein. Einige allgemeine Programmsysteme für finite Elemente. In K.E. Buck, D.W. Scharpf, E. Stein, W. Wunderlich, editors, *Finite Elemente in der Statik*, pages 399–454. Wilhelm Ernst & Sohn, Berlin, 1973.

[26] K.E. Buck, D.W. Scharpf, E. Stein, W. Wunderlich, editors. *Finite Elemente in der Statik*. Wilhelm Ernst & Sohn, Berlin, 1973.

[27] J. Céa. Approximation variationnelle des problčmes aux limites (Ph.D. Thesis). *Annales de l'institut Fourier 14*, **2:** 345–444, 1964.

[28] C. Carstensen, S.A. Funken. Averaging technique for FE-a posteriori error control in elasticity. *Comput. Methods Appl. Mech. Engng.*, **190:** 2483–2498, 4663–4675; 191: 861–877, 2001.

[29] C.A. Castigliano. *Théorie de l'équilibre des systčmes élastiques et ses applications*. Nero, Turin, 1879.

[30] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc.*, **49:** 1–23, 1943.

[31] M. Crouzeix, P.A. Raviart. Conforming and non-conforming finite element methods for solving the stationary Stokes equations. *RAIRO Anal. Numér.*, **R3:** 33–76, 1973.

[32] P.A. Cundall, O.D.L. Strack. Discrete numerical model for granular assemblies. *Geotechnique*, **29:** 47–65, 1979.

[33] P. de Fermat. *Analysis ad refractiones*. Oeuvres, I. 1661.

[34] B.F. de Veubeke, editor. *Matrix methods of structural analysis*. Pergamon Press, Oxford, 1964.

[35] R. Descartes. *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*. Leiden (French) (1637), Amsterdam (Latin), 1656.

[36] I.S. Duff, J.K. Reid. A set of FŘRTRAN-subroutines for sparse symmetric sets of linear equations. *Report R-10533, Computer Science and Systems Devisions, AERE Harwell, HMSO, London*, 1982.

[37] L. Euler. *Methodus inveniendi lineas curvas maximi minimive proprietate gaudens sive solutio problematis isoperimetrici latissimo sensu accepti*, volume 25 of *Opera omnia, Series I*. Lausanne and Genevae, 1744.

[38] B.G. Galerkin. Beams and plates, series for some problems of elastic equilibrium of beams and plates. *Wjestnik Ingenerow*, **10:** 897–908, 1915.

[39] G. Galilei. *Discorsi e dimonstrazioni matematiche, intorno a due nuove scienze*. Leiden, 1638.

[40] R.H. Gallagher, I. Rattinger, J.S. Archer. *A correlation study of methods of matrix structural analysis*. AGAR-Dograph 69. Pergamon Press, Oxford, 1964.

[41] W. Hackbusch. *Iterative solvers for large sparse systems of equations (English translation from German language)*. Springer Verlag, Berlin, 1994.

[42] E. Hellinger. Die allgemeinen Ansätze der Mechanik der Kontinua. In F. Klein and C. Müller, editors, *Enzyklopädie der Mathematischen Wissenschaften 4 (Teil 4)*, pages 601–694. Teubner, Leipzig, 1914.

[43] C. Hermite. Sur un nouveau developpment en série de fonctions. *C. R. Acad. Sci., Paris*, **58:** 93–100, 1864.

[44] L.R. Herrmann. Finite-element bending analysis for plates. *J. Engg. Mech. EM 5*, **93:** 13–26, 1967.

[45] D. Hilbert. *Die Grundlagen der Mathematik. Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität, Band VI*. 1928.

[46] H.H. Hu, D.D. Joseph, M.J. Crochet. Direct simulation of fluid particle motions. *Theor. Comp. Fluid Dyn.*, **3:** 285–306, 1992.

[47] Th.J.R. Hughes, J.A. Cottrell, Y. Bazilevs. Isogeometric analysis: Cad, finite elements, nurbs, exact geometry and mesh refinement. *Comput. Methods Appl. Mech. Engng.*, **194:** 4135–4195, 2005.

[48] Ch. Huygens. *Horologicum oscillatorium sive de motu pendularium*. Apud F. Muguet, Paris, 1673.

[49] B.M. Irons, O.C. Zienkiewicz. The isoparametric finite element system – a new concept in finite element analysis. In *Proc. Conf.: Recent advances in stress analysis*, Royal Aeronautical Society, 1968.

[50] J.W. Strutt, Baron Rayleigh, M.A., F.R.S. *The theory of sound*, volume I. Macmillan and Co., London, 1877.

[51] I. Kant. *Gedanken von der wahren Schätzung der lebendigen Kräfte*. Königsberg, 1746.

[52] K. Klöppel, D. Reuschling. Zur Anwendung der Theorie der Graphen bei der Matrizenformulierung statischer Probleme. *Der Stahlbau*, **35:** 236–245, 1966.

[53] J.S. Koenig. De universali principio aequilibrii et motus in vi viva reperto, deque nexu inter vim vivam et actionem, utriusque minimo dissertatio. *Nova Acta Eruditorum*, March 1751.

[54] P. Ladevčze, J.-P. Pelle. *Mastering calculations in linear and nonlinear mechanics*. Springer Science+Business Media, Inc., 2005.

[55] R. Lagrange. *Polynomes et fonctions de Legendre*. Gauthier-Villars, Paris, 1939.

[56] A.M. Legendre. Sur l'attraction des sphéroides. *Mém. Math. et Phys. présentés ŕ l'Ac. r. des. sc. par divers savants*, 10, 1785.

[57] G.W. Leibniz. Brevis demonstratio erroris memorabilis Cartesii... *Acta Eruditorum*, pages 161–163, 1686.

[58] G.W. Leibniz. Communicatio suae pariter, duarumque alienarum ad adendum sibi primum a Dn. Jo. Bernoullio, deinde a Dn. Marchione Hospitalio communicatarum solutionum problematis curvae celerrimi descensus a Dn. Jo. Bernoullio geometris publice proposti, una cum solutione sua problematis alterius ab eodem postea propositi. *Acta Eruditorum*, pages 201–206, May 1697.

[59] J.C. Maxwell. *Phil. Mag.*, **27:** 294, 1864.

[60] R.J. Melosh. Basis for derivation of matrices for the direct stiffness method. *AIAA J.*, **1:** 1631–1637, 1963.

[61] F.L. Menabrea. *Sul principio di elasticitŕ, delucidazioni di L.F.M.* 1870.

[62] K.-H. Möller, C.-H. Wagemann. Die Formulierungen der Einheitsverformungs- und der Einheitsbelastungszustände in Matrizenschreibweise mit Hilfe der Graphen. *Der Stahlbau*, **35:** 257–269, 1966.

[63] H. Müller-Breslau. *Hütte – Des Ingenieurs Taschenbuch*, chapter Elastizität und Festigkeit und Baumechanik. Ernst & Korn, Berlin, 1877.

[64] G.E. Moore. Cramming more components onto integrated circuits. *Electronics*, **38:** 4 pages, 1965.

[65] R. Niekamp, E. Stein. An object-oriented approach for parallel two- and three-dimensional adaptive finite element computations. *Computers & Structures*, **80:** 317–328, 2002.

[66] J.A. Nitsche. l_∞-convergence of finite element approximations. In *Mathematical aspects of finite element methods*, volume 606 of *Lecture Notes in Mathematics*, pages 261–274. Springer, New York, 1977.

[67] A.S. Ostenfeld. *Teknisk Statik I.* 3rd edition, 1920.

[68] E.C. Pestel, F.A. Leckie. *Matrix methods in elastomechanics.* McGraw-Hill, New York, 1963.

[69] T.H.H. Pian. Derivation of element stiffness matrices by assumed stress distributions. *AIAA J.*, **2:** 1533–1536, 1964.

[70] T.H.H. Pian, P. Tong. Basis of finite element methods for solid continua. *Int. J. Num. Meth. Eng.*, **1:** 3–28, 1964.

[71] W. Prager. Variational principles for elastic plates with relaxed continuity requirements. *Int. J. Solids & Structures*, **4:** 837–844, 1968.

[72] G. Prange. *Die Variations- und Minimalprinzipe der Statik der Baukonstruktion.* Habilitationsschrift. Technische Hochschule Hannover, 1916.

[73] E. Reissner. On a variational theorem in elasticity. *J. Math. Phys.*, **29:** 90–95, 1950.

[74] W. Ritz. Über eine neue Methode zur Lösung gewisser Probleme der mathematischen Physik. *Journal für die reine und angewandte Mathematik*, **135:** 1–61, 1909.

[75] M.C. Rivara. Mesh refinement processes based on the generalized bisection of simplices. *SIAM Journal on Numerical Analysis*, **21:** 604–613, 1984.

[76] J. Robinson. *Structural Matrix Analysis for the Engineer.* John Wiley & Sons, New York, 1966.

[77] R. Rodríguez. Some remarks on Zienkiewicz-Zhu estimator. *Numer. Methods Partial Diff. Equations*, **10:** 625–635, 1994.

[78] K.H. Schellbach. Probleme der Variationsrechnung. *Crelle's Journal für die reine und angewandte Mathematik*, **41:** 293–363 + 1 table, 1851.

[79] J.C. Simo, M.S. Rifai. A class of mixed assumed strain methods and the method of incompatible modes. *Int. J. Num. Meth. Eng.*, **29:** 1595–1638, 1990.

[80] S.L. Sobolev. *Applications of Functional Analysis in Mathematical Physics*, volume 7 of *Mathematical Monographs*. AMS, Providence (RI), 1963.

[81] E. Stein. Gottfried Wilhelm Leibniz, seiner Zeit weit voraus als Philosoph, Mathematiker, Physiker, Techniker... – ein Extrakt der gleichnamigen Ausstellungen. *Abhandlungen der Braunschweigischen Wissenschaftlichen Gesellschaft*, **54:** 131–171, 2005.

[82] E. Stein. Theoria cum praxi: Leibniz als technischer Erfinder. In Thomas A.C. Reyden, H. Heit, P. Hoyningen-Huene, editors, *Der universale Leibniz*, pages 155–183. Franz Steiner Verlag, Stuttgart, 2009.

[83] E. Stein, R. de Borst, T.J.R. Hughes, editors. *Encyclopedia of Computational Mechanics*, vol. 1: Fundamentals, vol 2: Solids and Structures, vol 3: Fluids. John Wiley & Sons, Chichester, 2004, (2nd edition in Internet, 2007).

[84] E. Stein, M. Rüter. Finite element methods for elasticity with error-controlled discretization and model adaptivity. In E. Stein, R. de Borst, T.J.R. Hughes, editors, *Encyclopedia of Computational Mechanics*, volume 2: Solids and Structures, pages 5–58. John Wiley & Sons, Chichester, 2004.

[85] E. Stein, M. Rüter, S. Ohnimus. Implicit upper bound error estimates for combined expansive model and discretization adaptivity. *Comput. Methods Appl. Mech. Engrg.*, **200:** 2626–2638, 2011.

[86] E. Stein, W. Wunderlich. Finite-Element-Methoden als direkte Variationsverfahren in der Elastostatik. In K.E. Buck, D.W. Scharpf, E. Stein, W. Wunderlich, editors, *Finite Elemente in der Statik*, pages 71–125. Wilhelm Ernst & Sohn, Berlin, 1973.

[87] R. Stenberg. Analysis of mixed finite element methods for the Stokes problem: A unified approach. *Math. Comp.*, **42:** 9–23, 1984.

[88] H. Stolarski, T. Belytschko. Limitation principles for mixed finite elements based on the Hu-Washizu variational formulation. *Comput. Methods Appl. Mech. Engng.*, **60:** 195–216, 1987.

[89] B. Szabó, I. Babuška. *Finite Element Analysis.* John Wiley & Sons, New York, 1991.

[90] B.A. Szabó, A.K. Mehta. p-convergent finite element approximations in fracture mechanics. *Int. J. Num. Meth. Engng.*, **12:** 551–560, 1978.

[91] I. Szabó. *Geschichte der mechanischen Prinzipien.* Birkhäuser, Basel, 3rd edition, 1987.

[92] R.L. Taylor, O.C. Zienkiewicz, J.C. Simo, A.H.C. Chan. The patch test – a condition for assessing f.e.m. convergence. *Int. J. Num. Meth. Eng.*, **22:** 39–62, 1986.

[93] E. Torricelli. *De Motu gravium*, volume 2 of *Opere.* Faenca, 1919 (reprint).

[94] M.J. Turner, R.W. Clough, H.C. Martin, L.J. Topp. Stiffness and deflection analysis of complex strucutres. *Journal of the Aeronautical Sciences*, **23:** 805–823, 1956.

[95] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh refinement technis.* Wiley-Teubner, Chichester, 1996.

[96] K. Washizu. *Variational Methods in Elasticity and Plasticity.* Pergamon Press, 1968.

[97] N. Willems, W.M. Lucas Jr. *Matrix analysis for structural engineers.* Prentice-Hall, INC. / Englewood Cliffs, N.J., 1968.

[98] G. Zavarise, P. Wriggers. *Trends in Computational Contact Mechanics.* LNACM, vol. 58, Springer, Berlin.

[99] O.C. Zienkiewicz, S. Qu, R.L. Taylor, S. Nakazawa. The patch test for mixed formulations. *Int. J. Num. Meth. Eng.*, **23:** 1873–83, 1986.

[100] O.C. Zienkiewicz, J.Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Num. Meth. Eng.*, **24:** 337–357, 1987.

[101] O.C. Zienkiewicz, J.Z. Zhu. The superconvergent patch recovery (SPR) and adaptive finite element refinements. *Comput. Methods Appl. Mech. Engng.*, **101:** 207–224, 1992.