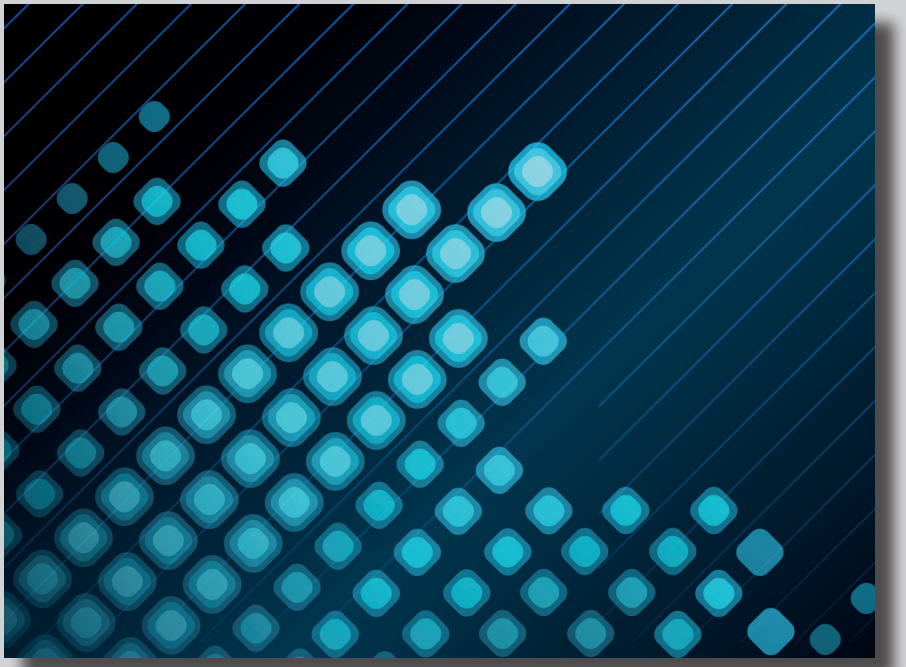
POLITECHNIKA LUBELSKA

PODRĘCZNIKI

Jerzy Warmiński, Jarosław Latalski, Rafał Rusinek
Andrzej Mitura, Marek Borowiec

Numerical modelling of mechanical systems



Lublin 2015

Numerical modelling of mechanical systems

Monografie – Politechnika Lubelska

HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Publication co-financed by the European Union
under the European Social Fund

Jerzy Warmiński, Jarosław Latalski, Rafał Rusinek
Andrzej Mitura, Marek Borowiec

Numerical modelling of mechanical systems

Politechnika Lubelska
Lublin 2015

Adviser by:
Dr. hab. inż. Paweł Droździel, prof. Politechniki Lubelskiej
Prof. Ing. Milan Saga, PhD

Technical editor: Jarosław Latalski

Translation at the sole request of Lublin University of Technology

Free of charge publication.

The publication was prepared and published as a part of the project *Engineer with a warranty of quality – tailoring the course offer of the Lublin University of Technology to the requirements of the European labour market* (agreement number: UDA-POKL.04.01.01-00-041/13-00), co-financed by the European Social Fund, Human Capital Operational Programme, Submeasure 4.1.1.

Publication approved by the Rector of Lublin University of Technology

ISBN: 978-83-7947-130-0

Publisher: Lublin University of Technology
ul. Nadbystrzycka 38D, 20-618 Lublin, Poland
Realization: Lublin University of Technology Library
ul. Nadbystrzycka 36A, 20-618 Lublin, Poland
tel. (81) 538-46-59, email: wydawca@pollub.pl
www.biblioteka.pollub.pl
Printed by : TOP Agencja Reklamowa Agnieszka Łuczak
www.agencjatop.pl

The digital version is available at the Digital Library of Lublin University of Technology: www.bc.pollub.pl
Circulation: 200 copies

Table of contents

1. Modelling of dynamical systems – introduction and basic definitions

Modelling of any dynamic system is a process allowing learning, understanding and explaining the main features of a real object. Often, the models give the possibility to detect new phenomena which is sometimes impossible or difficult to get by physical experiments, due to the limited accuracy of research equipment or a limited number of samples. High costs of experiment are usually associated with expensive equipment, preparation of appropriate number of samples, which after the conducted tests may not be suitable for further use. Therefore, the numerical models of the real objects are a very important in the process of designing of machines and structures.

Generally, it is difficult to represent precisely the real systems by the model. Fundamental problems occurring during the modelling process result from high complexity of the real systems and a large number of physical phenomena occurring during their operation. Thus, by the *physical*, *mathematical* or *numerical model* we understand the approximate representation of the real system dynamics. The modelling process can be represented in the form of an algorithm, shown schematically in figure 1.1.

1.1. Physical model

Knowing the real object, in the first step we must try to build a *physical model*, which will allow determining the selected characteristics of the real system. In relation to this, we have to make simplifying assumptions [4]. This step is mainly based on the knowledge and intuition of the researcher (engineer, designer). Simplifications are attractive because they lead to a simple physical model, but unfortunately they may also lead to erroneous results. Therefore, this step of modelling requires intuition based on the engineer experience in order not to omit the behaviour and physical phenomena occurring in a real object.

While creating physical model, we usually make simplifications regarding geometry and material. We assume constancy of the selected parameters, skip less significant internal and external interactions, neglect deformations of the selected elements treating them as rigid bodies, skip the masses of ”very light” elements. In addition, random excitations we replace by harmonic or polyharmonic forces.

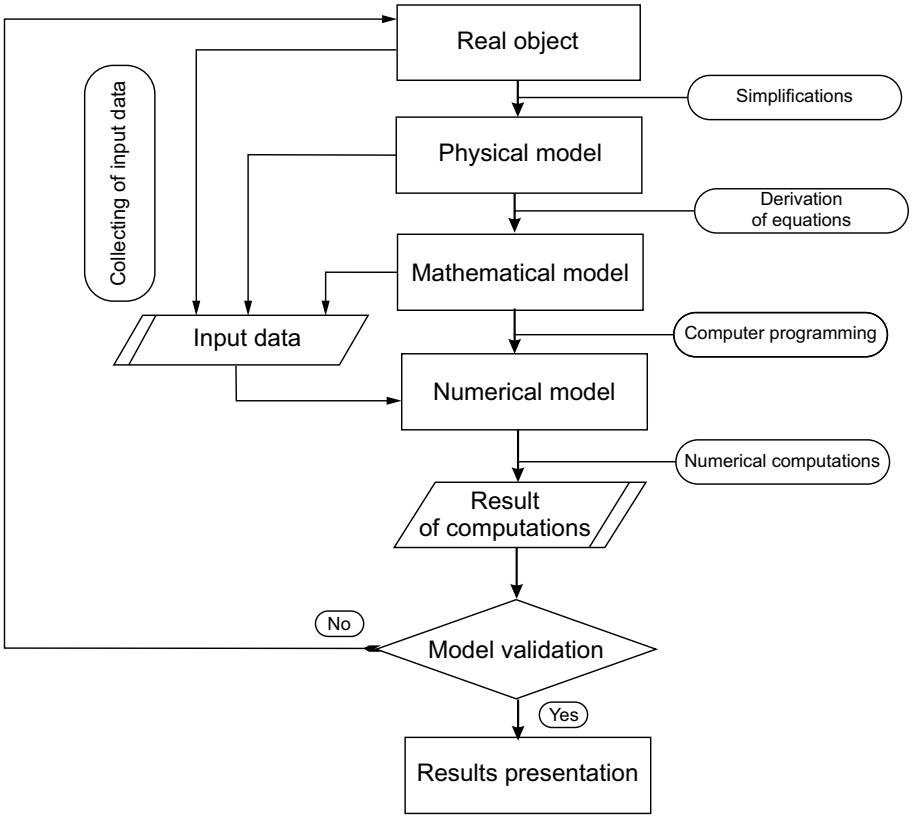


Figure 1.1. *Block diagram of modelling process*

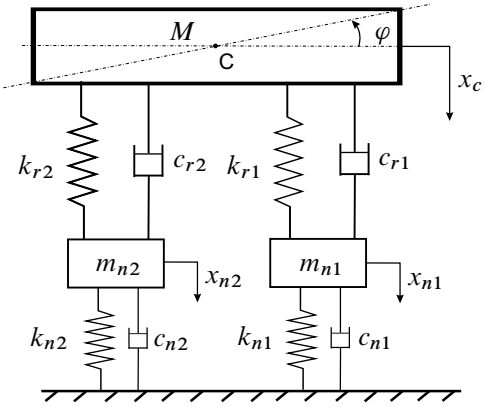


Figure 1.2. *Physical model of a car with four degrees of freedom*

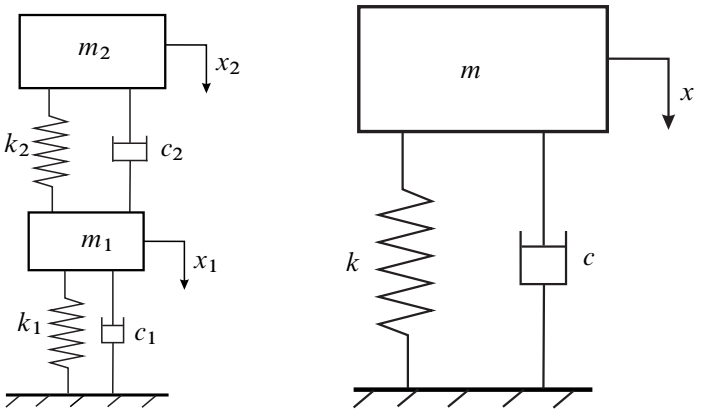


Figure 1.3. *Reduced model of a car with two (left side) and one degree of freedom (right side)*

Motion of the system elements can be described in various reference systems, using different types of coordinates. In mechanics, the most commonly, so called *generalized coordinates*, are used, which are usually marked with the letter q, and defined as independent physical quantities, determining the position of the considered system in space. They may be rectilinear coordinates, so called translational coordinates or angular coordinates, so called rotational coordinates.

The number of independent coordinates, necessary for the unique description of the system motion in the space is called *a number of degrees of freedom*. For an unique description of free motion of a rigid body i.e. motion without constraints, a six generalized coordinates are needed: three translational coordinates describing and three rotational coordinates (e.g. Euler angles). Therefore, the free motion of the rigid body is a motion with six degrees of freedom.

In the case when the object is modelled with the use of many rigid bodies and when the motion *constraints* occur, there's a necessity to select coordinates in such a way that the motion of the whole system is uniquely determined. For example, a car model presented in figure 1.2 is a system with 4 degrees of freedom $\{x_{n1}, x_{n2}, x_C, \varphi\}$. The bodies m_{n1} i m_{n2} are treated as lumped the masses, M is a rigid body performing plain motion, considered as a combination of translational and rotary motion.

Then, the car model can be simplified (reduced) to the lower dimension e.g. to the system with 2 degrees of freedom, presented in the form of sprung mass m_2 and unsprung mass m_1. Such a model can represent vibrations of e.g. front suspensions of the car. Obviously, for the model to be adequate, it's necessary to correctly determine its equivalent parameters i.e. masses, stiffness and damping.

The most simplified is the model with 1 degree of freedom, representing motion of the whole car by a single substituted mass, damping and stiffness. Despite its simplicity, such models are also used in the mechanical engineering.

Motion of the system may be limited by the imposed constraints. In the modelling, the constraints are replaced by reactions (passive forces). In mechanics, the constraints have been divided into specific classes, which depend on the adopted criteria. Taking into account *friction* the constraints are divided into:

- *ideal constraints* – without friction, with the reactions forces normal to a contact surface of the bodies,
- *nonideal constraints* – with friction, with the reaction forces deviated from normal due to the occurrence of contact friction force.

Depending on the method of motion limitations, we distinguish:

- *one sided constraints* – in which the reaction is directed one way, and the body can detach from the surface constraining its motion. These constraints are described with inequalities:

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s, t) > 0,$$

- *two sided constraints* – in which the reaction can change the sign and the motion is constrained in both sides preventing the detachment. These equations generally have the following form:

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s, t) = 0.$$

where s is a number of degrees of freedom and $k = 1, 2, \ldots, r$ is a number of constraints equations.

In analytical mechanics, the constraints have been also classified depending on the time or derivatives of generalized coordinates (generalized velocities) occurring in the equations that describe the motion limitations. Therefore, depending on time occurrence, we distinguish:

- *scleronomic constraints*, which are described by equations independent of time

$$f_k(q_1, q_2, \ldots, q_s) = 0$$

or

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s) = 0,$$

- *reonomic constraints*, described by equations depend on time

$$f_k(q_1, q_2, \ldots, q_s, t) = 0$$

or

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s, t) = 0.$$

Due to the occurrence of the derivative (generalized velocity), the constraints are divided into:

– *geometric constraints* – independent of generalized velocities

$$f_k(q_1, q_2, \ldots, q_s) = 0$$

or

$$f_k(q_1, q_2, \ldots, q_s, t) = 0,$$

– *kinematic constraints* – dependent on the generalized velocities

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s) = 0$$

or

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s, t) = 0.$$

The kinematic constraints

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s) = 0$$

or

$$f_k(q_1, q_2, \ldots, q_s, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s, t) = 0$$

are divided into two classes:

– *holonomic* – these are so called integrable constraints which after integrating can be reduced to geometric ones,
– *nonholonomic* – these are dependent on the generalized velocities, which can't be integrated and reduced to geometric ones.

The separation into *holonomic* and *nonholonomic* is very important in modelling of the mechanical systems.[1] Systems with nonholonomic constraints result in much more difficulties already at the stage of physical model construction, and even more difficulties during attempts to determine the solutions. This is due to the fact that no generalized coordinates exist for them, perturbation of which does not disturb the constraints' equations. It's worth noting that the existence of one non-integrable equation among the constraints' equations does not mean that the whole system is nonholonomic [6].

[1] The authors of the work [6] mention Lagrange mistake who claimed that for every mechanical system, we can select independent coordinates which have independent variations. However, after the analysis of the systems rolling without sliding on the horizontal plane or surfaces with complex shapes, it has been demonstrated that so called nonholonomic constraints exist. The division into holonomic and nonholonomic constraints has been introduced by Hertz in 1894. [2]

A common case of systems with nonholonomic constraints are such systems, in which the constraints are described by linear equations in regard to generalized velocity

$$\begin{aligned} A_{11}\dot{q}_1 + A_{12}\dot{q}_2 + \ldots + A_{1s}\dot{q}_s + B_1 &= 0 \\ A_{21}\dot{q}_1 + A_{22}\dot{q}_2 + \ldots + A_{2s}\dot{q}_s + B_2 &= 0 \\ &\vdots \\ A_{r1}\dot{q}_1 + A_{r2}\dot{q}_2 + \ldots + A_{rs}\dot{q}_s + B_r &= 0 \end{aligned} \tag{1.1}$$

where r is the number of constraints' equations, s is a number of degrees of freedom, and the coefficients $A_{ij}, B_i, j = 1, 2 \ldots, s, i = 1, 2, \ldots, r$ are the functions of generalized coordinates i.e.

$$A_{ij} = A_{ij}(q_1, q_2, \ldots, q_s, t), B_i = B_i(q_1, q_2, \ldots, q_s, t)$$

Constraints described by equations (1.1) are called kinematic, linear and nonintegrable. Moreover, if the coefficients A_{ij}, B_i do not involve time directly, we call them independent of time, and furthermore if all coefficients $B_i = 0$, then we call them homogeneous.

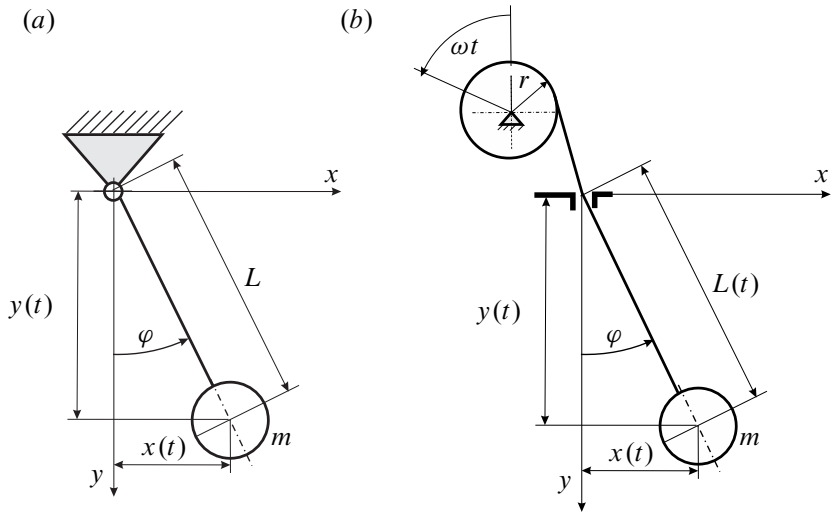


Figure 1.4. *System with scleronomic (a) and reonomic (b) constraints*

The mathematical pendulum presented in figure 1.4(a) is an example of a system with scleronomic constraints. Motion of the mass m is restricted by an inextensible line with constant length L which guarantees that at any moment of time the following equation is fulfilled

$$x(t)^2 + y(t)^2 - L^2 = 0 \tag{1.2}$$

In the next example (figure 1.4b) the length of the line is varied due to rotation of the cylinder with a radius r, rotating with angular speed ω. In this case the length of the pendulum is determined by equation $L(t) = L_0 - \omega r t$, where L_0 is the initial length of the line. In this case the equation of the constraints is:

$$x(t)^2 + y(t)^2 - (L_0 - \omega r t)^2 = 0 \tag{1.3}$$

In this case the constraints are reonomic, they directly dependent on time.

Let us now consider the disk which rolls on the horizontal plane (figure 1.5). Let us assume that the plane is perfectly rough, thus there is no sliding between the disk and the substrate. We are dealing with a plain motion of the disk which, in this case, due to the constraints applied has two degrees of freedom described by coordinates x and φ.

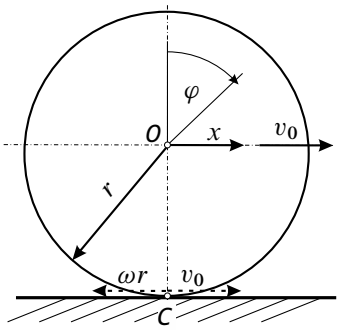


Figure 1.5. *System with kinematic constraints*

Velocity of disk centre O equates to v_0. Because there is no slip, velocity of point C is equal to zero, $v_C = 0$. Thus, for this point we may write the following equation

$$\dot{x} - \dot{\varphi} r = 0\,, \tag{1.4}$$

where $\dot{\varphi} = \omega$, and $\dot{x} = v_0$. The equation (1.4) describes the kinematic constraints. In fact, after integration of the equation (1.4)) we obtain geometric constraints

$$x - \varphi r = 0\,. \tag{1.5}$$

Thus, the equation (1.4) describes the holonomic constraints, that is kinematic integrable constraints. Taking into consideration the time, these are also the sclereonomic constraints.

In the subsequent part of this work we will refer mainly to the models with holonomic constraints.

The physical models of real systems may be created by lumped masses or non-deformable bodies connected by massless springs and damping elements. These

are the so called *discreet models* – with a finite number of degrees of freedom, described by ordinary differential equations.

We can also propose another type of model assuming deformability of bodies and considering a continuous mass distribution. These are the so called *continuous models* with infinite number degrees of freedom, described by partial differential equations.

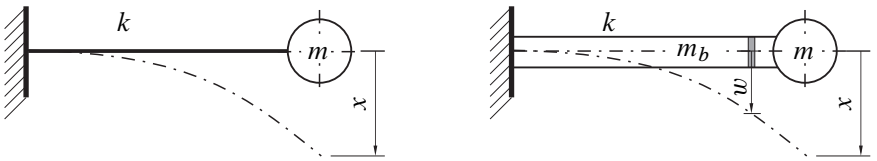


Figure 1.6. *Discreet model (a) and continuous model (b) of a beam with lumped mass*

Figure 1.6 presents a beam with point mass at the end. If the mass of the beam m_b is substantially smaller than the point mass m then we may disregard it, treating the entire system as an oscillator which consists of a spring of k stiffness and point mass m. In such a case the continuous model has been reduced to a discrete model with the omission of beam's mass. As it may be noticed the first bending form of the beam was assumed. The frequency of natural vibrations of the model equates to $\omega_0 = \sqrt{k/m}$. With an increase of the beam's mass its participation becomes more and more important. At that stage we have to describe not only the lumped mass motions m but also the motion of every element of the beam, treated as a continuous system represented by – coordinate w. This requires introduction of differential equations of the motion of a hybrid system -discreet-continuous – having an infinite number of degrees of freedom and an infinite number of frequencies and modes of vibrations [10].

Let us assume that we want to take into account the mass of the beam but we want to reduce the analysis to only i.e. the first vibration mode. In such a case the analysed discreet-continuous system may be replaced by a discreet one, taking into account the influence of the beam's mass. In order to do that we may use e.g. the Rayleigh method [5], according to which the first natural frequency of the equivalent one degree of freedom discrete system is equal to $\omega_{01} = \sqrt{k/(m+m_b/3)}$, where m_b is the mass of a uniform beam.

Within the process of modelling we need to consider the fact that in reality all the characteristics of i.e. elasticity or damping forces are nonlinear. Very often the nonlinearities are approximated by linear characteristics. In many cases such approximation is allowed and the results are in a good agreement with the experiment. The linear models are attractive because in many cases they are easier to solve and it is often possible to determine strict analytical solutions. Furthermore, numerical calculations are less time-consuming and the results are more predictable.

Unfortunately in many cases nonlinearities are significant and often, in order to explain the phenomena which occur it is necessary to build a more precise nonlinear model.

Nonlinearities may be divided as follows:

- *geometric nonlinearities* (structural),
- *material nonlinearities* (physical),
- *nonlinearities which result from nonlinear external excitations*.

An example of the system where geometric nonlinearity occurs is ie. a mathematical pendulum presented on figure 1.4. The equation of free vibrations of the pendulum has the form of

$$\ddot{\varphi} + \omega_0^2 \sin\varphi = 0\,.$$

Nonlinearity arises from the $\sin\varphi$ function present in the above equation. Geometric nonlinearity may result from natural features of the system or it may be purposefully introduced by the designer. The example may be the spring presented on figure 1.7(a), the stiffness of which changes with the increase of deflection.

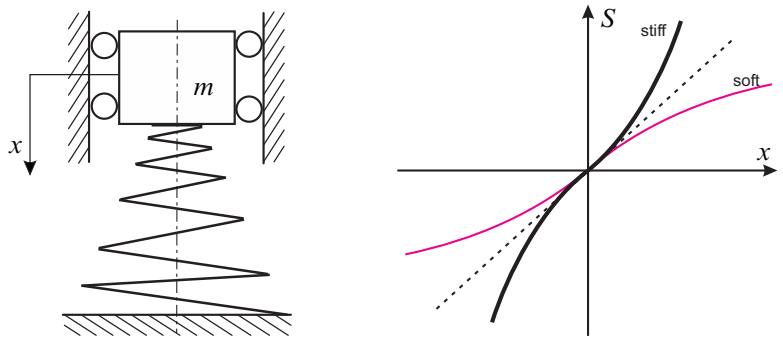


Figure 1.7. *The model of the system with geometric nonlinearity (a) and nonlinear characteristics of spring's force (b)*

Physical nonlinearities are caused by deviation from the Hook's law. Fig 1.7(b) presents the characteristics of elasticity force against the deflection. For small deflections both the stiff and the soft characteristics are close to the straight line (dotted line). However, with an increase of deflection the actual characteristics (soft or stiff) start to differ significantly from the linear resulting from Hook's law.

The reason of nonlinearity may be the external factors caused by e.g. non-linear magnetic field, nonlinear forces coming from an outflow of fluids leading to nonlinear aerodynamic forces.

The division into *linear models* and *nonlinear models* is very significant. The non-linear models introduce the possibility of qualitatively different behaviours, different to their linear counterparts. In case of nonlinear systems, apart from regular motions, we may also expect the occurrence of *chaotic motions* [8].

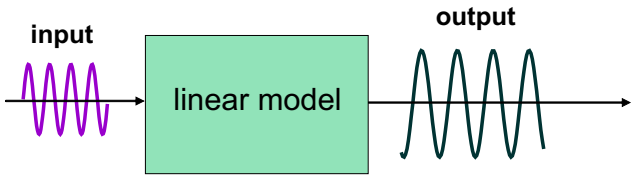


Figure 1.8. *Possible responses of the deterministic linear model*

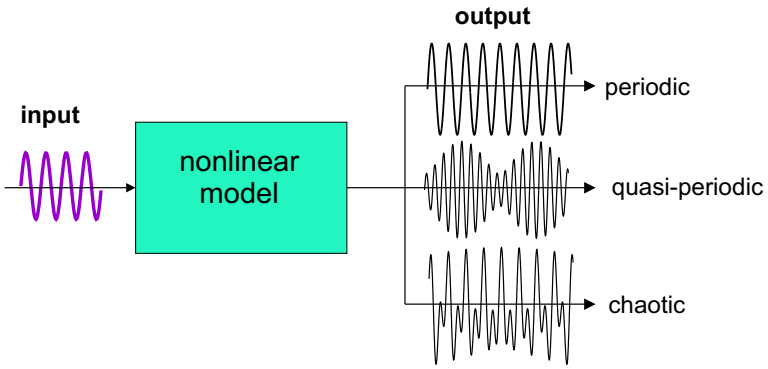


Figure 1.9. *Possible responses of deterministic nonlinear model*

The possible responses of the deterministic linear model (dissipative) are presented schematically on figure 1.8. If the system is forced by the harmonic force, its response is also harmonic of the same frequency as the force with a difference only in the amplitude and phase. The response of the nonlinear model may differ not only in terms of the amplitude and phase but also in terms of the frequency which might be a multiple of the force frequency or its fraction. We may also obtain either a quasi-periodic or a chaotic responses (figure 1.9).

Nonlinearity causes the appearance of *bifurcations*, which normally proceed the appearance of chaotic motion [3]. The chaotic vibrations occur only in nonlinear systems.[2]

In mechanics, we distinguish three main mechanisms of vibration excitation:

- excited vibrations – by force, excited kinematically, inertia excited,
- parametric vibrations – caused by periodically variable coefficients,
- self-excited vibrations – caused by nonlinear properties of the system which, cause the appearance of vibrations by the constant energy supply [9].

[2]It is also worth remembering that the modes of natural vibrations of the linear model are straight lines and their number is equal to the number of degrees of freedom. In case of a nonlinear model the modes are nonlinear, they may bifurcate, and their number may be greater than the number of degrees of freedom. [10].

The model of the system should also contain the influence of the energy source which generates the vibrations [9]. If the interactions of the source of vibrations and the vibrating system is omitted, it is a so called an *ideal system* or a system with an ideal source of energy. These are the so called systems with unlimited power. If we assume the interactions of a vibrating object and the source of vibrations in the model (by adding the model of e.g DC motor), then such system is called *nonideal* with limited power.

1.2. Mathematical model

Mathematical model of the system are the differential equations of the motion obtained from the physical model discussed in chapter 1.1. The type of differential equations depends on the physical model and the accepted simplifying assumptions. In case of continuous systems the mathematical model is represented by partial differential equations (PDE). Partial differential equations require elaborating solutions which fulfil the required boundary and initial conditions. The solutions are the functions of space and time. There are several methods of determining solutions, for example by applying Galerkin's method and reducing the partial equations to ordinary ones, by using finite element method, or by the application of the finite difference method [8]. The basis of the finite element method and finite difference method will be discussed in subsequent chapters of this book.

When the physical model is a discreet system with a finite degrees of freedom, the mathematical model is represented by ordinary differential equations (ODE). In such a case the solution is indicated for the given initial condition (so called initial problem). Within the hereby chapter we will focus on the discreet systems described by the ordinary differential equations.

The differential equations of motion (ODE) can be obtained through the direct use of the *Newton's principles*. In such an approach, it is necessary to analyse forces and moments acting on the bodies which are the elements of the entire model. When applying the second Newton's principle we obtain differential equations of motion as the second order ODEs. In case of the motion of a single solid body the ODEs have the form of:

$$m\frac{d^2\mathbf{r}_c}{dt^2} = \sum_{i=1}^{N}\mathbf{F}_i \qquad \text{and} \qquad \frac{d\mathbf{K}_c}{dt} = \sum_{i=1}^{N}\mathbf{r}_i \times \mathbf{F}_i \tag{1.6}$$

where $\mathbf{F}_i$ – i^{th} force which acts on the body, $\mathbf{r}_i$ – is the radius vector determining the force placement, $\mathbf{r}_c$ – radius vector which defines the location of mass centre C, $\mathbf{K}_c$ – angular momentum with respect to the centre of mass, N – total number of acting forces.

The second approach consists in the energy use. In mechanics, the equations of Lagrange of the second type are widely used. In this case it is possible to derive

differential equations motion from the energy of a studied system. The equation (1.7) presents one of the possible forms of Lagrange's equations of the second type for the systems with holonomic constraints

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_j}\right) - \frac{\partial T}{\partial q_j} + \frac{\partial V}{\partial q_j} + \frac{\partial R}{\partial \dot{q}_j} = Q_j\,, \quad j = 1, 2, \ldots, s \tag{1.7}$$

where q_j, $\dot{q}_j$ – denote the generalized coordinate and velocity, T i V are, accordingly kinetic and potential energies of the entire system, R is Rayleigh's function of dissipation describing damping, Q_j represents other generalized forces which are not counted neither in potential V nor dissipation function R, $j = 1, 2, \ldots, s$.

Generalized force occurring on the right side of the equation (1.7) is defined as

$$Q_j = \sum_{i=1}^{N} \mathbf{F}_i \frac{\partial \mathbf{r}_i}{\partial q_j}\,,$$

where $\mathbf{F}_i$ – force acting on i^{th} material point, $\mathbf{r}_i$ radius vector defining the location of the point, N – number of material points. It is worth noting that the generalized force may have a dimension of the force given in N or a moment of force in Nm. More detailed information on the Lagrange's equations may be found in the handbooks on analytical mechanics. We refer the reader to e.g. item [5].

1.3. Numerical model

Once the mathematical model is derived, the next step is to establish the solutions of differential equations. This can be done by analytical methods, by determining strict solutions or by approximate analytical methods, accepting some deviations. Unfortunately, determination of strict solutions is possible only for a limited class of equations. Whilst, analytical approximate methods are normally cumbersome and an analysis is valid only for a limited range of parameters. The development of computer techniques and numerical methods gave a brand new quality of dynamic systems analysis. Numerical methods allow establishing solutions on the basis of mathematical model with a high precision. Despite the fact that these are also the approximate solutions, due to their precision, they can often be treated equally to strict solutions.

In order for the mathematical model to be "understood" by computer it needs to be written by means of a computer eligible language. This requires writing a subsidy of adequately encoded commands called *computer programs*. For the engineering topics the basic program applied is the language called *Fortran*. This is one of the first languages developed until now which has a rich and well tested numerical libraries. This language allows for the conduct of advanced numerical calculations on large digital machines, using many processes at the same time, conducting the

so-called parallel calculations. The information on the structure of the Fortran language may be found in work [1]. There are commercial versions of compiler of Fortran language,[3] as well as open free versions.[4]

Apart from the historically developed Fortran language there is also a more modern programming language C or C++.[5] The principles of programming and the structure of the C language are presented in the work [7]. In order to conduct calculations it is possible to use the library with programs which allow for the use of standard calculations, e.g. inverting matrices, solving ordinary differential equations, solving a standard eigenvalue problem etc. On the websites of the software suppliers we may find paid and free libraries dedicated to a given compiler and operating system. Furthermore, software producers continue to work on the upgrades of compilers in such a way so that combining the modules written in different languages into one final code is possible.

Recently many modern higher level computational packages have been launched which enable introducing differential equations of motion in the form similar to the classical mathematical notation of writing equations, without the necessity of programming "line by line". Furthermore, such systems have many functions which facilitate the performance of numerical calculations. Sometimes the mathematical model is built on the basis of a block scheme where each unit performs more or less complex operations. Among many available systems in research in the field of technical sciences the most popular is the Matlab[6] package as well as Mathematica.[7] The packages are also equipped in their own programming languages which enable the creation of one's own codes.

Usually numerical simulations are cheaper than experimental tests. They allow a comprehensive analysis of influence of individual parameters and enable making a choice of the optimal solution. Numerical solutions are close to strict ones. The disadvantage of the solution obtained by numerical methods is a lack of analytical dependencies between the parameters and the response of the system, contrary to analytical solutions where dependencies are defined.

1.4. Model validation

The numerical model is created on the basis of the physical and mathematical models. The quality of a physical and subsequently – mathematical and numerical models – is determined by a comparison with the real object. As shown on the block

[3] Information on current software for Windows is available on the supplier website https://software.intel.com/en-us/intel-visual-fortran-compiler-for-windows

[4] Information on free software for the Linux system is available on the website https://gcc.gnu.org/wiki/GFortran

[5] Free software for Linux is available on the website https://gcc.gnu.org/gcc-4.9/

[6] http://www.mathworks.com/products/matlab/

[7] http://www.wolfram.com/mathematica/

scheme in figure 1.1, once we have the results of calculations we need to validate them. This is most often conducted by performing a physical experiment in characteristic points and then-comparing them to numerical results. If the differences are within the accepted range then the numerical model adequately represents the real object. If the results of the experiment are unavailable or difficult to be obtained then the results of numerical simulation are compared to the results obtained by other methods, e.g. analytical solutions can be established for specific values of parameters (for which this is possible) and it may be compared with the results of the numerical simulation.

The quality of the numerical model is determined by simplifications accepted at the stage of building the physical model. But apart from the simplifications the values of coefficients needed for the numerical model, called as input data, play a key role. These data are gathered on the basis of observations of the real object and formulation of the physical and mathematical models (figure 1.1). The more precise the model requires a larger number of coefficients necessary to perform the numerical simulations. Very often, the measurement of some values may be hard, and then we must approximate them, sometimes intuitively. On the other hand, too large simplifications may lead to omitting important physical phenomena. Therefore, one of the more important stages in the process of modelling is validation of the numerical model. If the validation is correct then we may conduct a series of numerical simulations in order to analyse comprehensively system dynamics.

1.5. Calculation results and their presentation

The results of numerical calculations are collected in the form of files and saved on devices on which they can be stored even after switching off the computer. Normally, the results are presented in columns representing a given physical value, e.g. time t, generalized coordinate q_j, generalized velocity $\dot{q}_j$ etc. Often, some additional values are calculated on the basis of given solutions, i.e. elasticity force, kinetic energy, potential energy etc. and they also can be stored in the files. The results are presented in the form of graphs such as time courses of selected values, phase planes, Poincaré's maps and others. Methods of analysis of the results have been discussed in detail in the subsequent chapter. It is possible to present the results of calculations by direct animations which show the behaviour of the studied system. The preparation of data needed for computation is performed by specific software. This stage is called *pre-processing*. The process of the results presentation after the main computation is called *post-processing*.

Bibliography

[1] Etzel M., Dickinson K. (1999): *Digital Visual Fortran Programmer's Guide*. Digital Press, Boston.
[2] Hertz H. (1894): *Die Prinzipien der Mechanik*. Gesammettte Werke t.3, Leipzig (Lipsk).
[3] Kapitaniak T., Wojewoda J. (1994): *Bifurkacje i chaos*. Wydawnictwo Politechniki Łódzkiej, Łódź.
[4] Kruszewski J., (red.) (1984): *Metoda elementów skończonych w dynamice konstrukcji*. Arkady, Warszawa.
[5] Meirovitch L. (2001): *Fundamentals of vibrations*. McGraw-Hill International Edition, New York.
[6] Nejmark J.I., Fufajew N.A. (1971): *Dynamika układów nieholonomicznych*. PWN, Warszawa.
[7] Sokół R. (2013): *Microsoft Visual Studio 2012. Programowanie w C i C++*. Wydawnictwo Helion, Gliwice.
[8] Thomsen J.J. (1997): *Vibrations and Stability. Order and Chaos*. McGraw-Hill, London.
[9] Warmiński J. (2001): *Drgania regularne i chaotyczne układów parametryczno-samowzbudnych z idealnymi i nieidealnymi źródłami energii*. Wydawnictwo Uczalniane Politechniki Lubelskiej, Lublin.
[10] Warmiński J. (2011): *Nieliniowe postacie drgań*. PWN, Warszawa.

2. Ordinary differential equations – methods of analysis

Modelling of dynamic systems is a complex process which consists of a number of key stages. First, it is necessary to create a physical model, then, we have to do simplifying assumptions in order to get the mathematical model, and in the final stage we create the numerical model which is the basis for numerical simulations. The model should be validated and in a case of occurrence of significant errors, has to be corrected. The detailed description of particular stages of medelling was presented in chapter 1.

In this chapter we will discuss the methods of analysis of discreet systems, described by ordinary differential equations [3], [4]. Let us consider the mechanical system described by n differential equations of the first order

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2, \ldots, x_n) \\ \frac{dx_2}{dt} &= f_2(x_1, x_2, \ldots, x_n) \\ &\vdots \\ \frac{dx_n}{dt} &= f_n(x_1, x_2, \ldots, x_n) \end{aligned} \tag{2.1}$$

with initial condition

$$x_i(t_0) = x_{i0}\,, \tag{2.2}$$

where $i = 1, 2, \ldots, n$.

A set of equations (2.1) may have solutions [1] which depend on initial conditions. Imposing the certain condition for the sought solution guarantees its *uniqueness*. Most often it is initial conditions assigned by the equation (2.2).[1] The equation (2.1) with initial condition (2.2) is called *initial problem* or Cauchy's problem. If time does not appear in (2.1) in a direct form in the right sides of the equations then the system is called *autonomous*. In the opposite case, the system is called *non-autonomous* and then, the equations have the following form

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \ldots, x_n, t) \tag{2.3}$$

[1] It is possible to set boundary conditions at the range $[a, b]$, $g_i(x_i(a), x_i(b)) = 0$ where g_i is a function of variables x_i, $i = 1, 2, \ldots, n$. The equation (2.1) with the boundary condition is called *boundary problem*.

In mechanics, very often functions $f_i\,(x_1, x_2, \ldots, x_n, t)$ are periodic, so

$$f_i\,(x_1, x_2, \ldots, x_n, t) = f_i\,(x_1, x_2, \ldots, x_n, t + T)\,,$$

where T is the period. This system is called n-dimensional non-autonomous periodic. Formally, time t may be treated as additional coordinate used to describe the motion. Then the dimension of the problem is increased to $n + 1$. By applying the notation $x_{n+1} = t$, n-dimensional non-autonomous system is transformed to $n+1$ dimensional autonomous system of the form

$$\begin{aligned}
\frac{dx_1}{dt} &= f_1\,(x_1, x_2, \ldots, x_n, x_{n+1}) \\
\frac{dx_2}{dt} &= f_2\,(x_1, x_2, \ldots, x_n, x_{n+1}) \\
&\vdots \\
\frac{dx_n}{dt} &= f_n\,(x_1, x_2, \ldots, x_n, x_{n+1}) \\
\frac{dx_{n+1}}{dt} &= 1\,.
\end{aligned} \tag{2.4}$$

This means that time increased the problem by one coordinate x_{n+1}. However, due to significantly different behaviours, in mechanics the autonomous and non-autonomous systems are studied separately. Non-autonomous periodical systems describe for instance the parametric vibrations in which characteristic zones of parametric resonances occur.

2.1. Basic concepts in dynamics of mechanical systems

We will introduce the most important concepts related to the dynamics of mechanical systems [5]:

- **phase space** – n dimensional space, where n is a number of differential equations given in Cauchy's form (2.1),
- **phase plane** – phase space of $n = 2$ dimension,
- **phase point** – point of coordinates $(x_1, x_2, \ldots, x_n)$, also called the representative or regular point,
- **critical pint or singular point** – point having coordinates $(x_{10}, x_{20}, \ldots, x_{n0})$ for which right hand sides of equations (2.1) are equal to zero, $f_i(x_{10}, x_{20}, \ldots, x_{n0}) = 0$,
- **phase trajectory or orbit** – integral curve of the system of equations (2.1) obtained through subsequent locations of the phase point,
- **Poincaré plane** – stroboscopic map of trajectory on the phase plane, also called Poincaré's map.

The motion of point takes place in n dimensional phase space according to the phase trajectory which are the integral curves of the system of equations (2.1). The feature of phase trajectories is the fact that they cannot cross one another. In order to justify the above statement let us hypothetically assume that the trajectories do cross one another, as presented on 2.1. Trajectories commence in the initial points, first

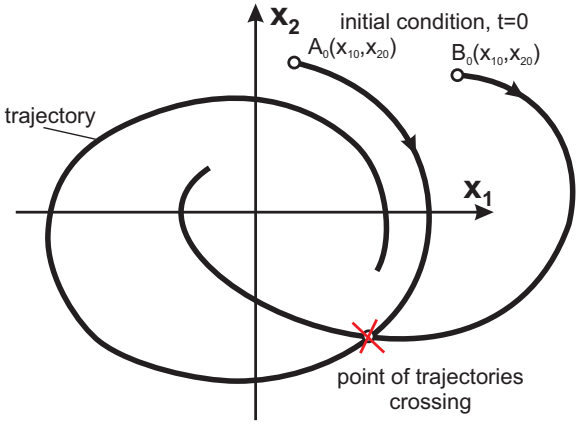


Figure 2.1. *Phase plane with two different trajectories obtained for different initial conditions*

of them in point $A_0(x_{10}, x_{20})$, second $B_0(x_{10}, x_{20})$. Because the initial conditions may be freely selected, we can change them in such a way so that the initial points A_0 and B_0 are found on the point where the trajectories cross. This would mean that when starting from the place of crossing it would be possible to obtain two solutions starting from the same initial condition which contradicts with the statement of uniqueness of the solutions [1].

In the phase space there are singular points (critical) which correspond to the steady state of the dynamic system. The singularity of these points will be shown by transforming the system of equations (2.1). Let us divide each of the equations by the first equation. Then we obtain:

$$\begin{aligned}
\frac{dx_2}{dx_1} &= \frac{f_2(x_1, x_2, \ldots, x_n)}{f_1(x_1, x_2, \ldots, x_n)} \\
\frac{dx_3}{dx_1} &= \frac{f_3(x_1, x_2, \ldots, x_n)}{f_1(x_1, x_2, \ldots, x_n)} \\
&\vdots \\
\frac{dx_n}{dx_1} &= \frac{f_n(x_1, x_2, \ldots, x_n)}{f_1(x_1, x_2, \ldots, x_n)}
\end{aligned} \tag{2.5}$$

In this way we have eliminated time from the system. If there is a point having coordinates $(x_{10}, x_{20}, \ldots, x_{n0})$, for which

$$f_i(x_{10}, x_{20}, \ldots, x_{n0}) = 0$$

that is the numerator and denominator in equations (2.5) is equal to zero, then in fact we obtain a singularity

$$\frac{dx_2}{dx_1} = \frac{0}{0}, \quad \frac{dx_3}{dx_1} = \frac{0}{0}, \quad \ldots, \quad \frac{dx_n}{dx_1} = \frac{0}{0}.$$

Singular points can "attract" or "repel" the phase trajectory, which means that they are stable or unstable. A phase trajectory may get close to the stable singular point at an infinitely small distance, but it will never reach it, or more precisely, it will reach at $t \to \infty$.

As an example let us consider the viscous damped oscillator. Its motion is defined by the ordinary differential equation of the second order

$$\ddot{x} + 2\xi\dot{x} + \omega_0^2 x = 0\,, \tag{2.6}$$

where ξ is a damping coefficient and ω_0 is the natural frequency. This is a system with one degree of freedom with variable x as a generalized coordinate. Once the equation (2.6) is transformed to the form of Cauchy (2.1) we obtain the system of two differential equations of the first order

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -2\xi v - \omega_0^2 x\,, \end{aligned} \tag{2.7}$$

where two phase coordinates (state variables) x and v occur. These are displacement and velocity of the oscillator. As result from the above example dynamics of the system with s degrees of freedom may be transformed into a set of differential equations of the first order with dimensions $n = 2s$. As noted earlier, the space $\mathbb{R}^n$ with dimension of n is called *phase space*. The phase trajectory for viscous damped oscillator (2.7) is drawn in the phase plane in fig 2.2. Phase trajectory starts in the initial point A_0, and then, when time t goes to infinity, the trajectory goes to the singular point which, as it is easy to establish, is placed in the origin of reference system. This singular point is stable and the trajectory reaches it at $t \to \infty$.

Apart from the classical phase plane the *stroboscopic* mapping of a solution (trajectory) on the phase plane is applied for the analysis of non-linear dynamic systems. This method of observation has been introduced by Ueda in order to study non-linear vibrations of oscillators excited externally [9]. This method is commonly used in the analysis of dynamics, in particular in case of systems with periodic excitations e.g. external or parametric. This kind of projection is called *Poincaré map*, stroboscopic map or stroboscopic portrait.

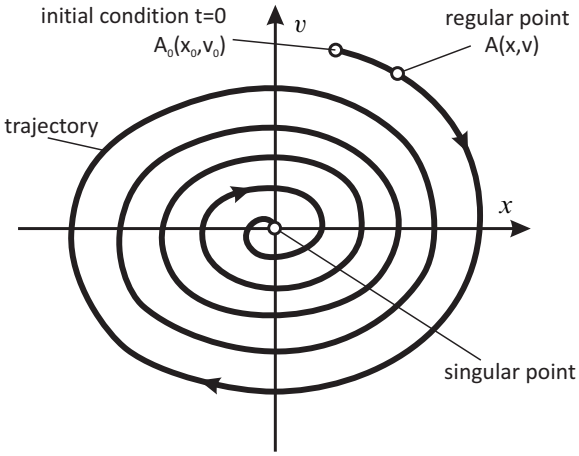


Figure 2.2. *Trajectory, regular and singular points on the phase plane*

Let us consider that the solution of the system is defined by the harmonic function in the form of $x(t) = A\sin(\omega t + \varphi)$. In such case instead of observing the entire phase trajectory we can observe its stroboscopic portrait. We register the solution at the periods corresponding to the frequency of ω. These moments have been marked by points on time courses of displacement and velocity (fig. 2.3(a) i (b)). Projecting it on the phase plane we obtain stroboscopic mapping in the form of one point M with coordinates $M(x_s, v_s)$. The location of the point M depends on the amplitude of vibrations A and phase φ. Considering that $\dot{x}(t) = A\omega\cos(\omega t + \varphi)$ we obtain

$$x_s{}^2 + \frac{v_s{}^2}{\omega^2} = A^2, \tag{2.8}$$

and

$$\mathrm{tg}\,\varphi = \frac{x_s}{v_s}. \tag{2.9}$$

Of course the solution may have a more complex nature than the periodical function presented on figure 2.3. Then, there may appear more points on the stroboscopic map, or a closed line or more complex structures, such as *strange chaotic attractors* [5, 7, 9, 10, 2].

2.2. Stability of singular points

Let us consider the system of differential equations presented in the form (2.1). The singular point [8] of the system has the coordinates marked by superscript 0. We

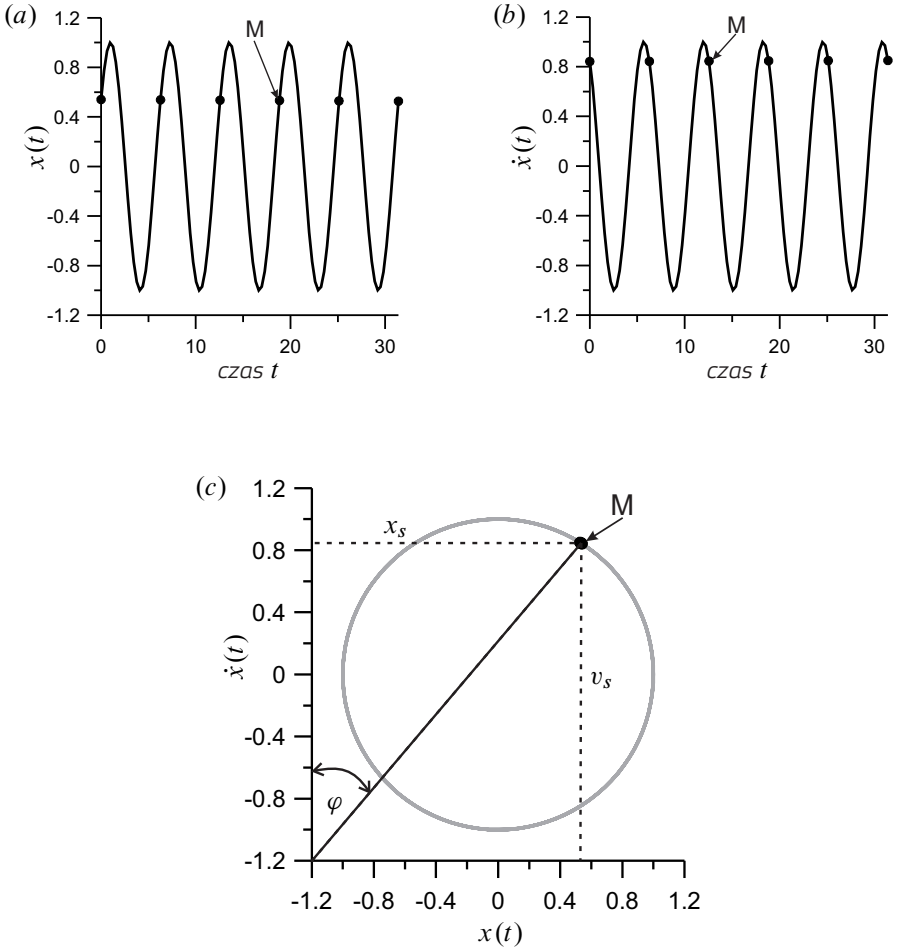


Figure 2.3. *Time histories of generalized coordinate $x(t)$ (a) and generalized velocity $\dot{x}(t)$ (b) and the method of Poincaré map creation (stroboscopic portrait) (c)*

remember from the chapter (2.1), that values of the function which are located on the right side of the equations (2.1), in the singular point are equal to zero

$$\begin{aligned} f_1\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right) &= 0, \\ f_2\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right) &= 0, \\ &\vdots \\ f_n\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right) &= 0. \end{aligned} \tag{2.10}$$

We disturb the solutions in the vicinity of the singular point by introducing the variation δ_i to each coordinate

$$\begin{aligned} \tilde{x}_1 &= x_1 + \delta_1 \\ \tilde{x}_2 &= x_2 + \delta_2 \\ &\vdots \\ \tilde{x}_n &= x_n + \delta_n, \end{aligned} \tag{2.11}$$

where $\tilde{x_i}$ stands for the disturbed and x_i undisturbed solution. Equations (2.1), after substituting in (2.11) and subtracting from undisturbed equations take the form

$$\begin{aligned} \frac{d\delta_1}{dt} &= f_1(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) - f_1\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right), \\ \frac{d\delta_2}{dt} &= f_2(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) - f_2\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right), \\ &\vdots \\ \frac{d\delta_n}{dt} &= f_n(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) - f_n\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right). \end{aligned} \tag{2.12}$$

Expanding the disturbed functions $f_1(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ in Taylor series in the vicinity of the singular point and considering the first order terms we obtain

$$\begin{aligned} \tilde{f_1}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) = & f_1\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right) + \left(\frac{\partial f_1}{\partial x_1}\right)_0 \delta_1 \\ & + \left(\frac{\partial f_1}{\partial x_2}\right)_0 \delta_2 + \ldots + \left(\frac{\partial f_1}{\partial x_n}\right)_0 \delta_n, \\ \tilde{f_2}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) = & f_2\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right) + \left(\frac{\partial f_2}{\partial x_1}\right)_0 \delta_1 \\ & + \left(\frac{\partial f_2}{\partial x_2}\right)_0 \delta_2 + \ldots + \left(\frac{\partial f_2}{\partial x_n}\right)_0 \delta_n, \\ & \vdots \\ \tilde{f_n}(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) = & f_n\left({x_1}^0, {x_2}^0, \ldots, {x_n}^0\right) + \left(\frac{\partial f_n}{\partial x_1}\right)_0 \delta_1 \\ & + \left(\frac{\partial f_n}{\partial x_2}\right)_0 \delta_2 + \ldots + \left(\frac{\partial f_n}{\partial x_n}\right)_0 \delta_n. \end{aligned} \tag{2.13}$$

Then, substituting (2.13) into equation (2.12) we obtain the system of differential equations in variations

$$\begin{aligned}
\frac{d\delta_1}{dt} &= \left(\frac{\partial f_1}{\partial x_1}\right)_0 \delta_1 + \left(\frac{\partial f_1}{\partial x_2}\right)_0 \delta_2 + \ldots + \left(\frac{\partial f_1}{\partial x_n}\right)_0 \delta_n \\
\frac{d\delta_2}{dt} &= \left(\frac{\partial f_2}{\partial x_1}\right)_0 \delta_1 + \left(\frac{\partial f_2}{\partial x_2}\right)_0 \delta_2 + \ldots + \left(\frac{\partial f_2}{\partial x_n}\right)_0 \delta_n \\
&\vdots \\
\frac{d\delta_n}{dt} &= \left(\frac{\partial f_n}{\partial x_1}\right)_0 \delta_1 + \left(\frac{\partial f_n}{\partial x_2}\right)_0 \delta_2 + \ldots + \left(\frac{\partial f_n}{\partial x_n}\right)_0 \delta_n .
\end{aligned} \tag{2.14}$$

We get a set of linear, homogeneous differential equations of first order the solutions of which are sought in the form

$$\delta_i = C_i e^{\lambda t} , \tag{2.15}$$

where: C_i is an amplitude, and λ an unknown parameter, $i = 1, 2, \ldots, n$. Introducing the solution (2.15) to equations (2.14) we obtain the set of linear algebraic equations, homogeneous with respect to amplitudes C_i

$$\begin{aligned}
\left[\left(\frac{\partial f_1}{\partial x_1}\right)_0 - \lambda\right] C_1 + \left(\frac{\partial f_1}{\partial x_2}\right)_0 C_2 + \ldots + \left(\frac{\partial f_1}{\partial x_n}\right)_0 C_n &= 0, \\
\left(\frac{\partial f_2}{\partial x_1}\right)_0 C_1 + \left[\left(\frac{\partial f_2}{\partial x_2}\right)_0 - \lambda\right] C_2 + \ldots + \left(\frac{\partial f_2}{\partial x_n}\right)_0 C_n &= 0, \\
&\vdots \\
\left(\frac{\partial f_n}{\partial x_1}\right)_0 C_1 + \left(\frac{\partial f_n}{\partial x_2}\right)_0 C_2 + \ldots + \left[\left(\frac{\partial f_n}{\partial x_n}\right)_0 - \lambda\right] C_n &= 0.
\end{aligned} \tag{2.16}$$

In order to get non-trivial solutions, $C_i \neq 0$, the main determinant of (2.16) must be equal to zero

$$\begin{vmatrix}
\left(\frac{\partial f_1}{\partial x_1}\right)_0 - \lambda & \left(\frac{\partial f_1}{\partial x_2}\right)_0 & \cdots & \left(\frac{\partial f_1}{\partial x_n}\right)_0 \\
\left(\frac{\partial f_2}{\partial x_1}\right)_0 & \left(\frac{\partial f_2}{\partial x_2}\right)_0 - \lambda & \cdots & \left(\frac{\partial f_2}{\partial x_n}\right)_0 \\
\vdots & \vdots & \ddots & \vdots \\
\left(\frac{\partial f_n}{\partial x_1}\right)_0 & \left(\frac{\partial f_n}{\partial x_2}\right)_0 & \cdots & \left(\frac{\partial f_n}{\partial x_n}\right)_0 - \lambda
\end{vmatrix} = 0. \tag{2.17}$$

Expanding determinant (2.17) we obtain the characteristic equation as n-degree polynomial of λ parameter. Stability and a type of a singular point depends on the

values of roots λ_i of the equation (2.17). If all roots λ_i are real negative numbers or complex numbers with real negative parts, then the singular point is stable. Bearing in mind, that perturbation was assumed in form (2.15) we see that in fact that the perturbation δ_i will go to zero. The perturbed solution tends to the singular point.

Equation (2.17) may be saved in the matrix form

$$\frac{d\boldsymbol{\delta}}{dt} = \mathbf{J}\left(\mathbf{x}^0\right)\boldsymbol{\delta}\,, \tag{2.18}$$

where $\boldsymbol{\delta}$ is a column matrix of perturbations, whilst

$$\mathbf{J}\left(\mathbf{x}^0\right) \equiv \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)_0 \tag{2.19}$$

is Jacobian - a square matrix of derivatives determined in the singular point.

2.3. Singular points classification – phase plane

The classification of singular points is based on analysis of stability presented in sub-chapter 2.2. The classification is limited to the phase plane (phase space $\mathbb{R}^2$). The perturbation of the solution in the vicinity of the singular point is defined by (2.15). In case of a dimension $n = 2$ we have

$$\delta = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}. \tag{2.20}$$

Therefore, depending on the value of parameters λ_1 and λ_2 we get the different behaviours in the vicinity of the singular point. The type of singular point depends on the value of two roots of characteristic equation (2.17), or in other words on the eigenvalues of Jacobian (2.19).

2.3.1. Node

If roots λ_1 i λ_2 are *real negative numbers* then the disrupted trajectory gets close to the singular point having the shape presented on figure 2.4(a). In such a case the singular point is called *stable node*. When both roots are *real positive numbers*, the disrupted trajectory goes away from the singular point, and then the singular point is a *unstable node* (figure 2.4(b).

2.3.2. Focus

The solution gets quite different shape in a case when roots λ_1 and λ_2 are *complex, conjugate numbers*. When the real part of both roots is negative, we obtain the singular point called *stable focus* (2.5(a)), or if the real part is positive we obtain *unstable focus* (figure 2.5(b)).

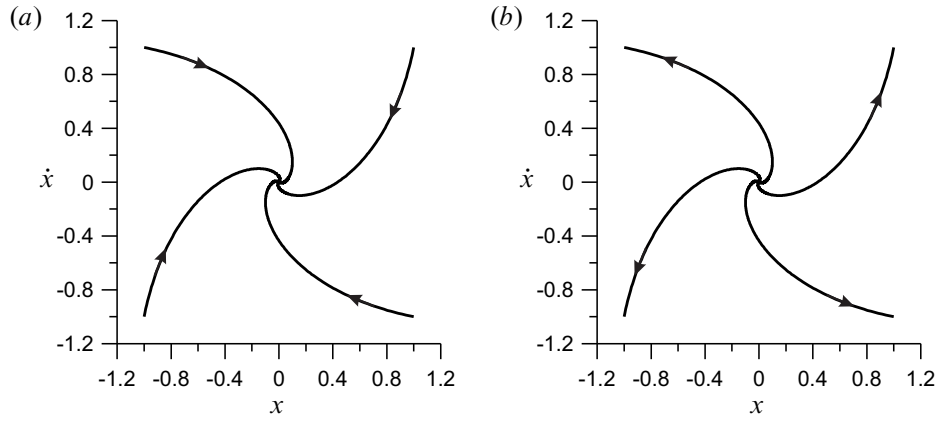


Figure 2.4. *Trajectories around the stable (a) and unstable (b) node*

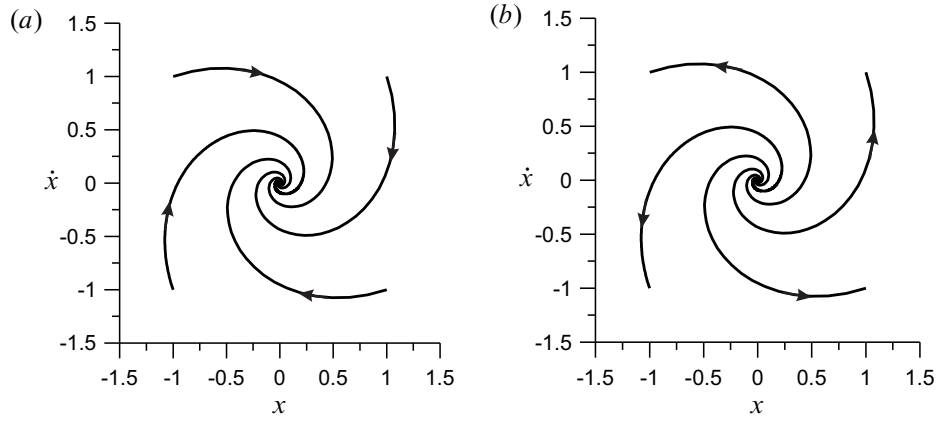


Figure 2.5. *Trajectories around the stable (a) and unstable (b) focus*

2.3.3. Centre

In a particular case the roots λ_1 and λ_2 may be complex conjugate numbers with zero values of a real part, i.e. they are *imaginary numbers*. This solution is located on the border and it is neither stable nor unstable. The trajectory neither gets closer to the singular point nor does it go away from it (figure 2.6). This is a neutral point called *centre*.

2.3.4. Saddle

When both roots λ_1 and λ_2 are real numbers with opposite signs (one of them is positive while the other one is negative) then we obtain a singular point of sad-

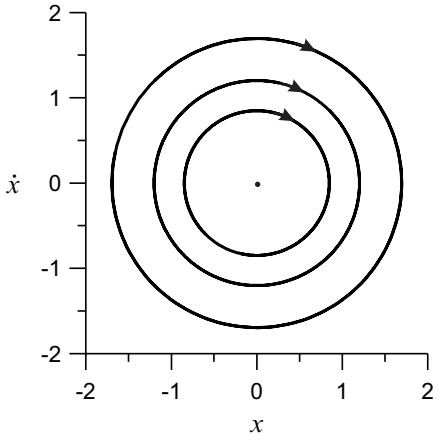


Figure 2.6. *Trajectories around a centre point*

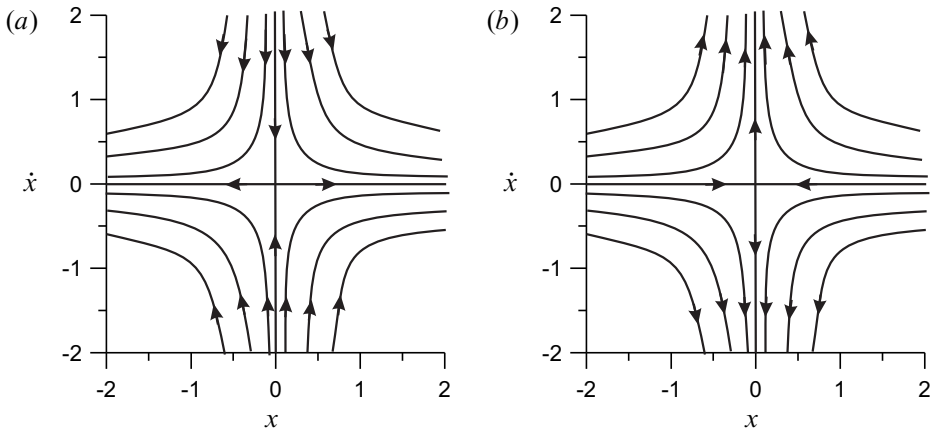


Figure 2.7. *Trajectories around the saddle point*

dle type (figure 2.7), which is always unstable, regardless of which of the roots is positive and which is negative.

2.4. Singular points in case of mathematical pendulum

As an example we present the singular points of a mathematical pendulum, and we determine their stability. Viscous damped vibrations of the pendulum are presented in figure 2.8 and they are defined by the equation:

$$J_0\ddot{\varphi} + c\dot{\varphi} + mgL\sin\varphi = 0. \tag{2.21}$$

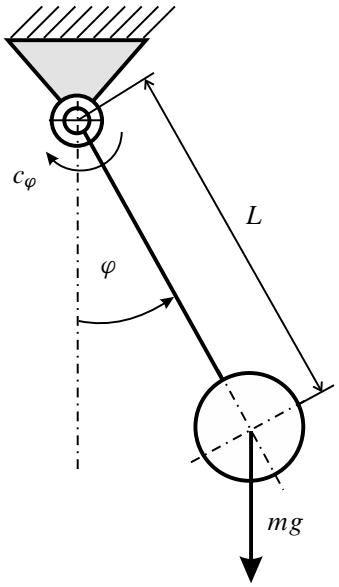


Figure 2.8. *Mathematical pendulum with damping*

Dividing both sides of the equation by J_0 and introducing the notation $2\zeta = c/J_0$, $\omega_0 = \sqrt{mgL/J_0}$, we obtain the differential equation of motion in a standard form

$$\ddot{\varphi} + 2\zeta\dot{\varphi} + {\omega_0}^2 \sin\varphi = 0. \tag{2.22}$$

Substituting $\Omega = \dot{\varphi}$, the equation (2.22) is written in the form of the set of two differential equations of the first order

$$\begin{aligned} \dot{\varphi} &= \Omega \\ \dot{\Omega} &= -2\zeta\Omega - {\omega_0}^2 \sin\varphi \end{aligned} \tag{2.23}$$

Singular point is determined by equaling right sides of the equations (2.23) to zero. Thus, we obtain $\Omega = 0$ and $\sin\varphi = 0$, which allows determining the coordinates of the singular points φ^0, Ω^0

$$\begin{Bmatrix} \varphi^0 \\ \Omega^0 \end{Bmatrix} = \begin{Bmatrix} k\pi \\ 0 \end{Bmatrix}, \qquad k = \ldots, -1, 0, 1, \ldots \tag{2.24}$$

Characteristic equation (2.17) takes the form

$$\begin{vmatrix} \left(\frac{\partial f_1}{\partial\varphi}\right)_0 - \lambda & \left(\frac{\partial f_1}{\partial\Omega}\right)_0 \\ \left(\frac{\partial f_2}{\partial\varphi}\right)_0 & \left(\frac{\partial f_2}{\partial\Omega}\right)_0 - \lambda \end{vmatrix} = \begin{vmatrix} -\lambda & 1 \\ -{\omega_0}^2(-1)^k & -2\zeta - \lambda \end{vmatrix} = 0\,, \tag{2.25}$$

and its roots are calculated as

$$\lambda_{1,2} = -\zeta \pm \sqrt{\zeta^2 - (-\omega_0{}^2)^k} \tag{2.26}$$

In case of underdamping , i.e. if $0 < \zeta < \omega_0$ we may write (2.26) in the form

$$\begin{aligned} \lambda_{1,2} &= -\zeta \pm i\ \sqrt{\omega_0{}^2 - \zeta^2} \qquad \text{for } k \text{even}, \\ \lambda_{1,2} &= -\zeta \pm \sqrt{\omega_0{}^2 + \zeta^2} \qquad \text{for } k \text{ odd}, \end{aligned} \tag{2.27}$$

where $i = \sqrt{-1}$. Because we accepted $n < \omega_0$, then expressions under the root are larger than zero. Thus, when k is an even number then roots are $\lambda_{1,2}$ complex conjugate with negative real parts. This means that singular points with coordinates $\ldots(-2\pi, 0), (0, 0), (2\pi, 0), \ldots$, are stable focus types. Whilst, when k is an odd number the roots are real numbers of with different signs which means that the singular points $\ldots(-3\pi, 0), (-\pi, 0), (\pi, 0), (3\pi, 0), \ldots$, are saddles (unstable). The course of the example phase trajectories has been presented in figure 2.9. The letter F stands for the stable focus while letter S is a saddle point.

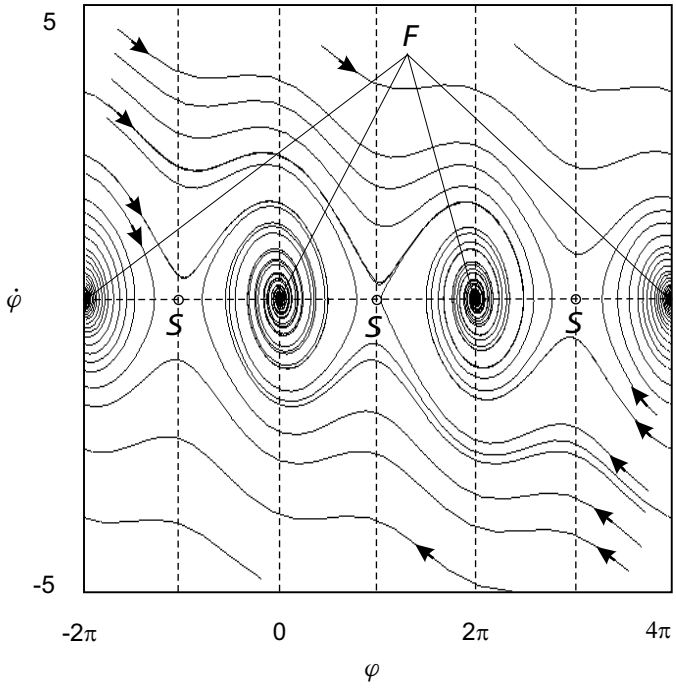


Figure 2.9. *Trajectories on the phase plane - underdamped mathematical pendulum, F-focus, S-saddle point*

In case of critical damping and overdamping $\zeta \geqslant \omega_0$, for k even the roots are real negative numbers while for odd they are real of opposite signs. Singular points are respectively stable nodes and saddles (unstable). A particular case is obtained when assuming that the motion of a pendulum is not damped ($\zeta = 0$). At that point the roots have the values $\lambda_{1,2} = \pm i\ \omega_0$ for k even, and $\lambda_{1,2} = \pm\omega_0$ for k odd. This means that singular points correspond to the centre and saddle.

2.5. Numerical methods

In many cases establishing strict solutions of the system of equations (2.1) is impossible. This concerns mainly non-linear equations of which the analytical solutions are established most often by approximate methods [6]. Currently it is possible to perform symbolic calculations in such packages as e.g. Mathematica. The advantage of such an approach is obtaining the solution in the form of analytical dependencies which gives the possibility of an easy and comprehensive parametric analysis. The disadvantage of symbolic calculations is their cumbersome complex formulas and rather easy possibility of making an error. Numerical calculations allows the direct integrating of differential equations of motion avoiding these obstacles. Nevertheless we must remember that numerical calculations also give approximate results which are burdened with an numerical error.

2.5.1. Representation of numbers, conditioning of the numerical problem and stability of algorithms

Numbers in computer systems are represented by a finite number of digits coded in the appropriate arithmetics. In numerical calculations *binary arithmetics* is commonly used. Although an equally important role is played by an octal or hexadecimal system. In computer systems we distinguish two representations of numbers *fixed point representation* or *floating point representation*. Fixed point notation is used for natural numbers (Integer type). In such case the result of basic actions, ie. adding, subtracting, multiplying is strict. Floating point notation is applied for real numbers (Real type). In general, in such case the result is an approximate number. Floating point representation is about presenting the number in the form of *mantissa* and *basis of representation* given in the power which is a natural number. The representation of floating point number has the form

$$x = S \cdot M \cdot B^E \tag{2.28}$$

- S – **sign** defined as $S = (-1)^z$ where z is the exponent defining the sign, when $z = 0$ the number is positive, when, $z = 1$ the number is negative,
- M – **mantissa**, $0 < M < 1$,
- B – **representation base**, B is the base of the power or the representation base. For the decimal system $B = 10$, for the binary system $B = 2$,

– E – **exponent**, exponent E of power part is called *feature of a number*.

For example the number $x = 123.45$ is written in the decimal system

$$(123.45)_{10} = 0.12345 \times 10^3 \tag{2.29}$$

$$123.45 = 1 \times 10^2 + 2 \times 10^1 + 3 \times 10^0 + 4 \times 10^{-1} + 5 \times 10^{-2}$$

In computer systems the numbers are written with the base 2. For example number $\frac{1}{10} = 0.1$ in binary system is written in the form

$$(0.1)_{10} = (0.00011\,(0011))_2 \tag{2.30}$$

As we can see the above number does not have a strict representation in the binary system. The floating point numbers may be written in a *Single* or *Double Precision*. In order to write the number in a double precision we need 64 bits (64 bit computer word). Figure 2.10 presents a scheme of coding of floating point number in a binary system.

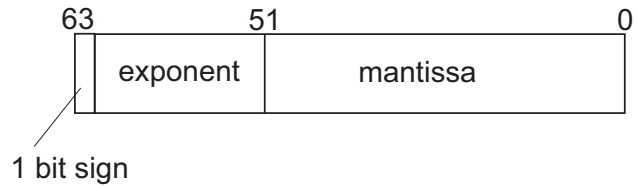


Figure 2.10. *The representation of a double precision number in a floating point binary system*

It must be pointed out that floating point arithmetic is not joint, thus

$$(x + y) + z \neq x + (y + z) \tag{2.31}$$

as well as it is not separable

$$x \cdot (y + z) \neq (x \cdot y) + (x \cdot z) \tag{2.32}$$

The order of performing operations has an impact on the final result.

We have to remember that while performing numerical calculations the following *numerical errors* may occur in the calculations:

– input errors – they occur when we introduce data to digital machine (to the memory or registers) which numerical representation is different than their strict values,

- truncation errors – errors which occur due to truncated number of actions e.g. computing finite sum instead of infinite one,
- error of rounding of numbers – errors which occur during calculations. These errors may be minimized by establishing a proper method and order of computing actions.

In numerical calculations we need to pay attention to two basic concepts: *conditioning of the problem*, *stability of the numerical algorithm*.

Conditioning of a problem means a property of *a mathematical problem*, that is a problem which consists of establishing a vector of results based on vector data, for example finding solutions of a set of ordinary differential equations. Problem conditioning is thus not related to the applied numerical algorithm but to the formulated task. We may distinguish *well-conditioned* and *ill-conditioned* problems.

As an example of an ill-conditioned problem let us determine solutions of a set of linear algebraic equations [11]

$$\begin{aligned}(2.5410 + \delta_{11})\, x_1 + (2.1120 + \delta_{12})\, x_2 &= 4.6530\\ (1.8720 + \delta_{21})\, x_1 + (1.5560 + \delta_{22})\, x_2 &= 3.4280\end{aligned} \tag{2.33}$$

We are looking for solutions x_1 i x_2. Let us assume that coefficients of the set (2.33) are entered with a certain "small" error. Perturbations of coefficients are denoted by δ_{ij}, $i, j = 1, 2$. When the perturbations are equal to zero $\delta_{ij} = 0$, then we obtain the strict solution which is:

$$x_1 = 1.0000\,, \quad x_2 = 1.0000.$$

Now we introduce into the system some inaccuracies of coefficients. Let us assume small perturbations:

$$\begin{aligned}\delta_{11} = 0.0010 \quad \delta_{12} = 0.0010\\ \delta_{11} = -0.0010 \quad \delta_{12} = -0.0020\end{aligned} \tag{2.34}$$

Then solutions of the system (2.33) are equal (2.34) equate to

$$x_1 = 3.9943\,, \quad x_2 = -2.6032\,.$$

We see that the task is ill-conditioned. Small inaccuracies in value of coefficients caused very large changes in the results.

Due to numerical errors for given data vector c and numerical precision $\varepsilon > 0$ of the numerical machine, the numerical result S_{num} will differ from the strict result S_c

$$S_{num}(c, \varepsilon) \neq S(c).$$

The conditioning or the problem (well or ill-conditioned) depends on variation of the difference between the results $S(c)$ and $S_{num}(c + \delta c)$.

Numerical stability is a property of calculation algorithms. If the result of calculations varies from the strict result and calculation errors cumulate then the numerical process is unstable [3]. We say that the algorithm is numerically stable if for the arbitrary selected data a_0 exists a precision of calculations ε_0 that for $\varepsilon < \varepsilon_0$

$$\lim_{\varepsilon \to 0} S_{num}(a_0, \varepsilon) = S(a_0)$$

Algorithm is numerically stable when increasing accuracy the of calculations it is possible to find the solution of the problem with arbitrary small error.

2.5.2. Numerical methods for ordinary differential equations – initial problem

The numerical solution of the system of ordinary differential equations (2.1) with a given initial condition (2.2) is a certain function which satisfies a given initial problem in discrete time domain. Because numerical calculations are performed with an assumed time step, the solution is a series of points obtained for selected moments of time within the interval $\langle t_0, t_k \rangle$. Starting from the initial condition t_0 we establish values in the moment $t_1 = t_0 + \Delta t$, $t_2 = t_1 + \Delta t$ etc. with the end at the t_k moment. This series is an approximation of the strict solution of the problem. The continuous solution is substituted by the discrete one in the subsequent moments of time t. Thus, this method is called the *difference method*. In order to obtain the solution in it is necessary to determine the approximate values of left sides of the equations (2.1) i.e. derivatives $\frac{dx_i}{dt}$ as well as right sides of functions $f_i\,(x_1, x_2, \ldots, x_n)$. The derivatives may be determine by various more or less complex numerical methods.

The numerical methods used for solving ordinary differential equations can be divided into:

– **single step methods** in which the solution $x(t)$ is constructed with the formula

$$\begin{aligned} x_{k+1} &= x_k + h\Phi_f\,(t_k, x_k, h) \\ x_0 &= x(t_0) \end{aligned} \tag{2.35}$$

where Φ_f may be the linear or non-linear function, while k is an integration step. In case of one step methods in order to establish a subsequent approximation x_k the solution from the previous x_{k-1} is sufficient. Starting from the initial point x_0 we establish further solutions x_i, $i = 1, 2 \ldots, N$, where N is a number of integration steps,

– **multi-step methods** are defined by dependencies

$$\begin{aligned} \alpha_m x_{k+m} + \ldots + \alpha_1 x_{k+1} + \alpha_0 x_k &= h\,(\beta_m f_{k+m} + \ldots + \beta_1 f_{k+1} + \beta_0 f_k) \\ x_j &= x_j(h) \end{aligned} \tag{2.36}$$

where $f_j = f(t_j, x_j)$; $k = 0, 1, 2, \ldots, N-m$; $j = 0, 1, \ldots, m-1$. The method defined by a dependency (2.35) is a peculiar case of a multi-step method. The method defined by the formula (2.36) is called m-steps method. In the multi--step method a certain number of previous solutions are applied. In order to start calculations it is necessary to know the solutions in the initial moment and in $k-1$ moments of time which are unknown. Thus, in order to "kick off" with the calculations it is necessary to establish solutions for several initial steps. These solutions may be established by means of a one-step method and then the multi-step method can be activated. Therefore, multi-step methods are not in the group of the so called *self-staring methods*, contrary to the single-step methods.

While establishing a numerical solution it is key to select the step of integration Δt. The length of this step is selected in such a way so that the error occurring throughout the calculations is minimal. Due to the fact that the strict solution is unknown therefore the error is also estimated by approximate methods, applying e.g. the *Runge approach* discussed in detail while presenting the *Runge-Kutta method*.

It is worth noting that in order to solve equations with elements of a very fast and a very slow dynamic actions which are *numerically stiff* we apply methods dedicated to them such as Adams or Gear method.

2.5.3. Euler method

One of the simplest methods of solving differential equations is Euler method, also called the *method of tangents*. This method consists in finding a solution in the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h\mathbf{f}(t_k, \mathbf{x}_k) \quad k = 1, 2, \ldots, N \tag{2.37}$$

where $\mathbf{x}, \mathbf{f}$ are column vectors which correspond respectively to coordinates x_i and right sides f_i of the system of equations (2.1). The Euler method will be explained for a single differential equation in the form

$$\frac{dx}{dt} = f(x, t) \tag{2.38}$$

The derivative $\frac{dx}{dt}$ is approximated by difference quotient determined in nodes t and $t + h$, h is an integration step. Applying the Taylor expansion we obtain

$$\frac{dx}{dt} = \frac{x(t+h) - x(t)}{h} + \frac{h}{2}\frac{d^2x}{dt^2} \tag{2.39}$$

with the dependency in nodes

$$x(t_{k+1}) = x(t_k) + hf(t_k, x(t_k)) + g_k \tag{2.40}$$

where g_k are expansions of a higher order. Omitting in the equation (2.40) the unknown functions g_k we obtain Euler method

$$\begin{aligned} x_{k+1} &= x_k + hf(t_k, x_k) \\ x_0 &= x(t_0) \end{aligned} \tag{2.41}$$

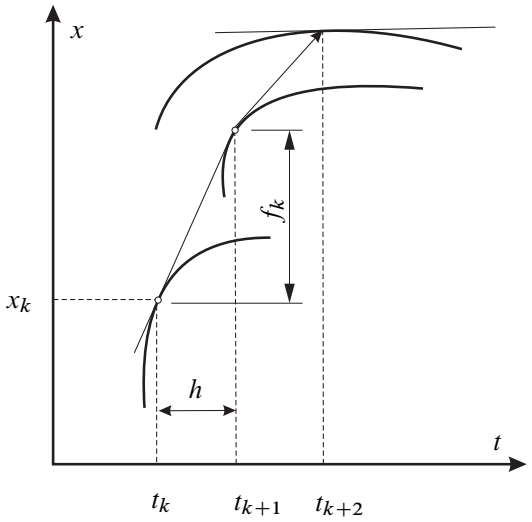


Figure 2.11. *Graphic interpretation of Euler's method*

A graphic interpretation of Euler method is presented in figure 2.11. In order to establish the solution in point t_{k+1}, we apply the equation of a tangent to curve in point t_k.

The Euler method is the easiest difference method used in practice only in specific cases. Its advantage is its simplicity. In order to obtain the required accuracy it is necessary to apply a small step Δt because we must determine the value of function $f(x(t), t)$ on the basis of just one time step.

A more precise result may be obtained by construing more advanced methods of calculations of a derivative through the use of information from several previous steps, that is by applying multi-step methods or considering the values of the functions $f(x(t), t)$ in a larger number of points located between the nodes [3, 11].

2.5.4. Runge-Kutta method

Runge-Kutta (RK) method is included in many standard software packages or libraries of Fortran or C languages. The Runge-Kutta method belongs to a one-step method so it is an self-starting method. An accuracy in this method is increased due to a larger number of test steps between nodes in which the right sides of equations

(2.1) are determined. The values of a vector of functions $\mathbf{f}(\mathbf{x}, t)$ are calculated also in intermediate points other than nodes defined by time step. This method is suitable for solving the majority of ordinary differential equations. It is also recommended for discontinuous systems. The most popular method is the fourth order RK method (RK4).

The general form of the explicit Runge-Kutta method is presented by the following dependencies:

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \sum_{i=1}^{s} w_i \mathbf{K}_i \tag{2.42}$$

$$\mathbf{K}_1 = \Delta t \mathbf{F}\left(\mathbf{x}(t), t\right) \tag{2.43}$$

$$\mathbf{K}_i = \Delta t \mathbf{F}\Big(\mathbf{x}(t) + \sum_{j=1}^{i-1} b_{ij} \mathbf{K}_j, t + a_i \Delta t\Big), \quad i > 1 \tag{2.44}$$

where w_i, a_i, b_{ij} coefficients – real numbers.

The equation (2.42) is used for calculation the values of solution in the subsequent node as the sum of solutions in the previous node and the weighted average of the values of the solution between the nodes, marked as $\mathbf{K}_i$ (2.44) with w_i weights assigned to this average. $\mathbf{K}_1$ (2.43) is the value of the solution computed as in Euler method. The remaining increments $\mathbf{K}_2, \mathbf{K}_3, \ldots, \mathbf{K}_s$, are calculated on the basis of recurrence scheme, that is each subsequent one is calculated on the basis of the previous one.

These values are calculated analogously as in the Euler method, although the values of function $\mathbf{F}$ are considered for the moment of time later than t and for the approximated values of solutions $\mathbf{x}$. In this case the trial steps are calculated in order to obtain greater accuracy of the solution. The later moment of time is calculated applying the coefficient a_i, and computing $t_{a_i} = t + a_i \Delta t$, and then the approximate solution for this moment as

$$x(t_{a_i}) = x(t) + \sum_{j=1}^{i-1} b_{ij} \mathbf{K}_j.$$

The value of a solution in the moment t_{a_i} is found by adding to the current solution a weighted average from the calculated previously $\mathbf{K}_j$. Iterative formula of the Runge-Kutta method of the fourth order allowing establishing the solution in the step $k + 1$ has the form of:

$$\begin{aligned} x_{k+1} &= x_k + \Delta x_k \\ \Delta x_k &= 16\,(K_1 + 2K_2 + 2K_3 + K_4) \\ K_1 &= hf(t_k, x_k) \\ K_2 &= hf\left(t_k + \frac{h}{2},\ x_k + \frac{h}{2}K_1\right) \\ K_3 &= hf\left(t_k + \frac{h}{2},\ x_k + \frac{h}{2}K_2\right) \\ K_4 &= hf\,(t_k + h, x_k + hK_3) \end{aligned} \tag{2.45}$$

where $h = \Delta t = t_{k+1} - t_k$ is a time step. For comparison we present the explicit RK2 method, called midpoint method and defined by the dependency

$$x_{k+1} = x_k + K_2 \tag{2.46}$$

where K_2 is the same as in the formula (2.45).

Coefficients within the Runge-Kutta method are established through the selection of the order of the method ensuring convergence of the solutions.

As noted in section 2.5.3 the error made while establishing numerical solution (which is an approximate one) strongly depends on the selected method and the choice of integration step. Within the Runge-Kutta methods to estimate the value of the error the so called *Runge method* is applied. This method consists in establishing the error in point x_{k+1} performing the following calculations [3]:

- we establish an approximate solution x_{k+1} going from x_k to x_{k+1} with a step size h,
- we establish approximate solution from x_k to $\tilde{x}_k$ taking a step $\frac{h}{2}$, and then from $\tilde{x}_k$ to x_{k+1} also with a step $\frac{h}{2}$.

The error of solution with the passing of the h step is defined as

$$x_{k+1} - x_{k+1}^{(1)} \approx \gamma_p\,(h)^{p+1} \tag{2.47}$$

while with $\frac{h}{2}$ step as

$$x_{k+1} - x_{k+1}^{(2)} \approx 2\gamma_p \left(\frac{h}{2}\right)^{p+1} \tag{2.48}$$

where p stands for the order of the method, γ_p a certain constant, h integration step, x_{k+1} is a strict solution, $x_{k+1}^{(1)}$ approximate solution obtained with step h, while $x_{k+1}^{(2)}$ is an approximate solution obtained by solving bisection step $\frac{h}{2}$. When subtracting the equation (2.48) from (2.47) we obtain

$$\delta = \frac{1}{2^p - 1}\left(x_{k+1}^{(1)} - x_{k+1}^{(2)}\right) \tag{2.49}$$

where $\delta = x_{k+1}^{(1)} - x_{k+1}^{(2)}$, is an estimated value of method error. For the Runge-Kutta method of 4-th order we obtain the measure of the error

$$\delta = \frac{1}{15} \left| x_{k+1}^{(1)} - x_{k+1}^{(2)} \right| . \tag{2.50}$$

The selection of the integration step is a compromise between accuracy and time of calculations. If we assume that the calculations ought to be conducted with the accuracy ε, then there may occur two cases:

- $\delta < \varepsilon$, the solution is sufficiently accurate. In addition, we may assume that if $\delta > \varepsilon/50$ a step h can be doubled (this will allow to shorten time of calculations) in other case, we move to the next point,
- $\delta > \varepsilon$, the solution is not sufficiently accurate, the step h is divided on half and the calculations are repeated.

In numerical libraries for the majority of computational systems we may find many procedures devoted to solving ordinary differential equations (initial problem). The Runge-Kutta method is available as a standard and recommended method for the most of typical problems. For example in the Matlab package the procedures of RK of 2^{nd} and 3^{rd} order `ode23` as well as of 4^{th} and 5^{th} order `ode45` are offered as default.

As an example, we may establish the solution of a non-linear Duffing equation

$$\ddot{x} + 2\zeta\dot{x} + x + \gamma x^3 = f_0 \sin \omega t. \tag{2.51}$$

The equation coefficients indicate: ζ – damping coefficient, γ – non-linear stiffness, f_0, ω – amplitude and excitation frequency.

On the example of Duffing's equation it is possible to verify the impact of the given parameters on obtained solutions, as well as to check the impact of the integrating step, the method order on the accuracy of the obtained results. In order to compare the results we take two extremely different integrating steps: "large" and "small" step. Fig 2.12 presents the solutions of natural vibrations for $\zeta = 0.05$, $\gamma = 0.25$, $f_0 = 0$. The solutions were determined by RK4 method with an integration step $h = 1$ (figure 2.12(a)) and $h = 0.01$ (figure 2.12(b)). The differences in the obtained solutions are clearly visible on the presented time histories.

2.6. Self-training problems

Problem 1

Derive ordinary differential equations of motion of the car model with four degrees of freedom, presented on figure 1.2. Differential equations to be written in Cauchy's form (2.1). Write the equations in a selected programming language (i.e. Fortran, C or in Matlab or Mathematica package), and then find the solutions describing the

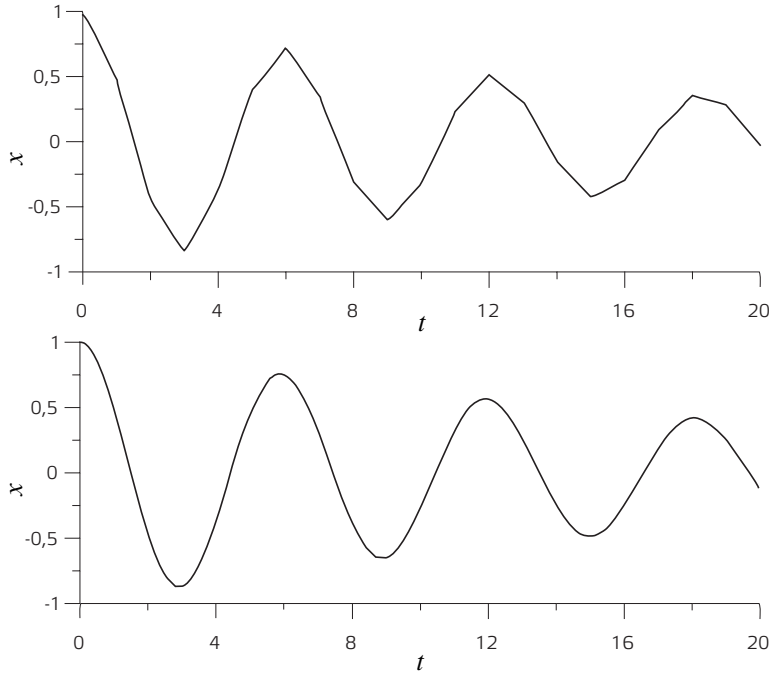


Figure 2.12. *Numerical solutions of Duffing equation, natural vibrations for* $\zeta = 0.05$, $\gamma = 0.25$, $f_0 = 0$ *and (a) integration step* $h = 1$, *(b) integration step* $h = 0.01$

motion of the given masses. Perform the simulations for the various time steps and initial conditions.

Problem 2

Reduce the model of a car with four degrees of freedom (figure 1.2) to the model with two degrees of freedom (1.3(a)) describing the motion of the front suspension. The differential equations must be written in Cauchy's form (2.1), and then in the selected programming language (e.g. Fortran, C or in Matlab package or Mathematica). Indicate the solutions describing motion of sprung and unsprung masses. Perform simulations for the various time steps and initial conditions. Compare results with the model of four degrees of freedom.

Bibliography

[1] Arnold W.I. (1975): *Równania różniczkowe zwyczajne*. PWN, Warszawa.

[2] Awrejcewicz J. (1996): *Drgania deterministyczne układów dyskretnych*. WNT, Warszawa.

[3] Fortuna Z., Macukow B., Wąsowski J. (1982): *Metody numeryczne*. Wydawnictwa Naukowo-Techniczne, Warszawa.
[4] Jankowska J., Jankowski M. (1988): *Przegląd metod i algorytmów numerycznych. Cz eść 1*. WNT, Warszawa.
[5] Kapitaniak T., Wojewoda J. (1994): *Bifurkacje i chaos*. Wydawnictwo Politechniki Łódzkiej, Łódź.
[6] Nayfeh A.H. (2000): *Nonlinear Interactions. Analytical, Computational and Experimental Methods*. Wiley Series in Nonlinear Science, New York.
[7] Szemplinska-Stupnicka W. (2002): *Chaos, bifurkacje i fraktale wokół nas. Najkrótsze wprowadzenie*. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
[8] Thomsen J.J. (2003): *Vibrations and Stability: Advanced Theory, Analysis, and Tools*. Springer, Berlin.
[9] Ueda Y. (2001): *The Road to Chaos - II*. Aerial Press, Inc., Santa Cruz, CA.
[10] Warmiński J. (2001): *Drgania regularne i chaotyczne układów parametryczno-samowzbudnych z idealnymi i nieidealnymi źródłami energii*. Wydawnictwo Uczalniane Politechniki Lubelskiej, Lublin.
[11] Zalewski A., Cegiela R. (1997): *Obliczenia numeryczne i ich zastosowania*. Wydawnictwo Nakom, Poznań.

3. Partial differential equations. Finite-difference method

3.1. Partial differential equations

The general form of partial differential equations (PDEs) which contains the derivatives of an unknown function $z(x, y, \dots)$ of two or more independent variables can be written in the following way

$$f\left(x, y, ..., z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, ..., \frac{\partial^2 z}{\partial x^2}, \frac{\partial^2 z}{\partial y^2}, \frac{\partial^2 z}{\partial x \partial y}, \dots\right) = 0 \tag{3.1}$$

In structural mechanics the equations of this type are used in many applications, inter alia, in research on elements with continuous mass distribution (bars, beams, plates etc.) In this work the scope of the considerations is limited to the analysis of second order linear PDE described in the form

$$a_1 \frac{\partial^2 z}{\partial x^2} + a_2 \frac{\partial^2 z}{\partial x \partial y} + a_3 \frac{\partial^2 z}{\partial y^2} + a_4 \frac{\partial z}{\partial x} + a_5 \frac{\partial z}{\partial y} + a_6 z = f(x, y) \tag{3.2}$$

A detailed classification of second order linear partial differential equations is performed based on the value of the determinant Δ which is defined in the following way $\Delta = a_2^2 - 4a_1a_3$. Thus, the classification was made using only first three terms of the equation (3.2). We distinguish the following type of the equations [3]:

– elliptic ($\Delta < 0$), e.g.:

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} + F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0 \tag{3.3}$$

– parabolic ($\Delta = 0$), e.g.:

$$\frac{\partial^2 z}{\partial x^2} + F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0 \tag{3.4}$$

or

$$\frac{\partial^2 z}{\partial y^2} + F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0 \tag{3.5}$$

– hyperbolic ($\Delta > 0$), e.g.:

$$\frac{\partial^2 z}{\partial x^2} - \frac{\partial^2 z}{\partial y^2} + F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0 \tag{3.6}$$

or

$$\frac{\partial^2 z}{\partial x \partial y 2} + F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0 \tag{3.7}$$

The presented classification is not only a formal and indicates the scope of possible applications in technical issues. Hyperbolic and parabolic equations can be used to describe dynamic processes (e.g. vibrations, thermal conduction etc.) where usually one of the variables is time. While the basic application of elliptic equations is the analysis of equilibrium conditions where function z depends only on spatial variables.

3.2. Finite-difference method

The solutions of second order linear partial differential equations can be determined by theoretical methods: analytical (strict solution) and numerical (approximate solution). One of the basic numerical tool is the finite difference method FDM. The application of this algorithm requires to replace derivatives by appropriate difference quotients. The first derivative of the function of one variable $z = f(x)$ in point $x = x_i$ can be approximated by several methods. Their geometric interpretations are presented in fig. 3.1:

– forward difference quotient

$$\left.\frac{dz}{dx}\right|_{x_i} \approx \left.\frac{\Delta z}{\Delta x}\right|_{x_i}^{+} = \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} = \frac{z_{i+1} - z_i}{\Delta x} \tag{3.8}$$

– backward difference quotient

$$\left.\frac{dz}{dx}\right|_{x_i} \approx \left.\frac{\Delta z}{\Delta x}\right|_{x_i}^{-} = \frac{f(x_i) - f(x_i - \Delta x)}{\Delta x} = \frac{z_i - z_{i-1}}{\Delta x} \tag{3.9}$$

– central difference quotient

$$\left.\frac{dz}{dx}\right|_{x_i} \approx \left.\frac{\Delta z}{\Delta x}\right|_{x_i} = \frac{f(x_i + \Delta x) - f(x_i - \Delta x)}{2\Delta x} = \frac{z_{i+1} - z_{i-1}}{2\Delta x} \tag{3.10}$$

Consider the continuous and differentiable function, for example: $z = x^2$. The exact value of the first derivative in point $x = 2$ can be defined as $\frac{dz}{dx} = 2x$, $\left.\frac{dz}{dx}\right|_{x=2} = 4$.

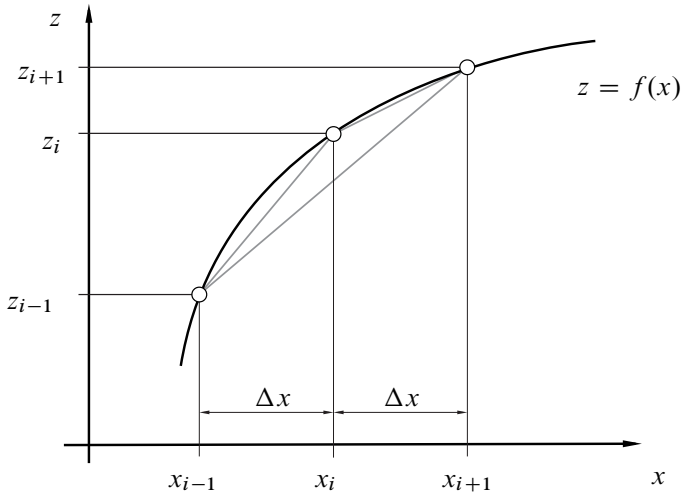


Figure 3.1. *Geometric interpretation of the difference quotients*

The approximate value of the first derivative of the function z was calculated using points $x_{i-1} = 1.9,\ x_i = 2,\ x_{i+1} = 2.1$ ($\Delta x = 0.1$). The obtained results are: $\frac{dz}{dx}\big|_{x=2} \approx \frac{\Delta z}{\Delta x}\big|^{+}_{x=2} = \frac{2.1^2-2^2}{0.1} = 4.1$, $\frac{dz}{dx}\big|_{x=2} \approx \frac{\Delta z}{\Delta x}\big|^{-}_{x=2} = \frac{2^2-1.9^2}{0.1} = 3.9$, $\frac{dz}{dx}\big|_{x=2} \approx \frac{\Delta z}{\Delta x}\big|_{x=2} = \frac{2.1^2-1.9^2}{0.2} = 4$.

Analysing the presented solutions we can see that the best approximation was obtained using central difference quotient to calculation. The accuracy of the finite difference method is assessed by Taylor's series expansion of searched function $z(x)$ about the point x_i:

$$z_{i+1} = z_i + \Delta x \frac{dz}{dx}\bigg|_{x_i} + \frac{\Delta x^2}{2!}\frac{d^2z}{dx^2}\bigg|_{x_i} + \frac{\Delta x^3}{3!}\frac{d^3z}{dx^3}\bigg|_{x_i} + \frac{\Delta x^4}{4!}\frac{d^4z}{d4^2}\bigg|_{x_i} + \ldots \quad (3.11)$$

$$z_{i-1} = z_i - \Delta x \frac{dz}{dx}\bigg|_{x_i} + \frac{\Delta x^2}{2!}\frac{d^2z}{dx^2}\bigg|_{x_i} - \frac{\Delta x^3}{3!}\frac{d^3z}{dx^3}\bigg|_{x_i} + \frac{\Delta x^4}{4!}\frac{d^4z}{d4^2}\bigg|_{x_i} - \ldots \quad (3.12)$$

The equations (3.11) and (3.12) have been transformed as follows:

– forward difference quotient

$$\frac{dz}{dx}\bigg|^{+}_{x_i} = \frac{z_{i+1} - z_i}{\Delta x} + 0(\Delta x) \quad (3.13)$$

– backward difference quotient

$$\frac{dz}{dx}\bigg|^{-}_{x_i} = \frac{z_i - z_{i-1}}{\Delta x} + 0(\Delta x) \quad (3.14)$$

Subtracting equation (3.12) from (3.11) leads to obtaining the expansion for central difference quotient.

$$\left.\frac{dz}{dx}\right|_{x_i} = \frac{z_{i+1} - z_{i-1}}{2\Delta x} + 0(\Delta x^2) \tag{3.15}$$

In equations (3.13), (3.14), (3.15) the term $0(\ldots)$ signifies the remainder of the Taylor's series expansion. In brackets are information about the lowest order of the part which has not been included within the finite difference method. For higher order of this term we can obtain better accuracy of calculations.

The simplest method to find the approximation of the second derivative of the function $z(x)$ in point x_i is applied the dependence for the first derivative in points $x_{i+1/2}$ i $x_{i-1/2}$

$$\left.\frac{d^2z}{dx^2}\right|_{x_i} \approx \frac{\left.\frac{dz}{dx}\right|_{x_{i+1/2}} - \left.\frac{dz}{dx}\right|_{x_{i-1/2}}}{\Delta x} \approx \frac{\frac{z_{i+1}-z_i}{\Delta x} - \frac{z_i-z_{i-1}}{\Delta x}}{\Delta x} = \frac{z_{i+1} - 2z_i + z_{i-1}}{\Delta x^2} \tag{3.16}$$

The above considerations showed method to replace the first and second derivative of the function of one variable $z(x)$ by difference quotients. The dependence (3.2) presents the equation which contains the partial derivatives of the function $z(x, y)$, where x, y are independent variables. Acting analogously as in one dimesional space case, we can written adequate derivatives in point x_i, y_j by the following difference equations:

$$\left.\frac{\partial z}{\partial x}\right|_{x_i,y_j} = \frac{z_{i+1,j} - z_{i-1,j}}{2\Delta x} \tag{3.17}$$

$$\left.\frac{\partial z}{\partial y}\right|_{x_i,y_j} = \frac{z_{i,j+1} - z_{i,j-1}}{2\Delta y} \tag{3.18}$$

$$\left.\frac{\partial^2 z}{\partial x^2}\right|_{x_i,y_j} = \frac{z_{i+1,j} - 2z_{i,j} + z_{i-1,j}}{\Delta x^2} \tag{3.19}$$

$$\left.\frac{\partial^2 z}{\partial y^2}\right|_{x_i,y_j} = \frac{z_{i,j+1} - 2z_{i,j} + z_{i,j-1}}{\Delta y^2} \tag{3.20}$$

$$\left.\frac{\partial^2 z}{\partial x \partial y}\right|_{x_i,y_j} = \frac{z_{i+1,j+1} - z_{i+1,j-1} - z_{i-1,j+1} + z_{i-1,j-1}}{4\Delta x \Delta y} \tag{3.21}$$

Figure 3.2 presents the plane of independent variables where points after discretization were placed. They create a rectangular grid which is characterized by two increments Δx and Δy. The equations from (3.17) to (3.21) precisely define

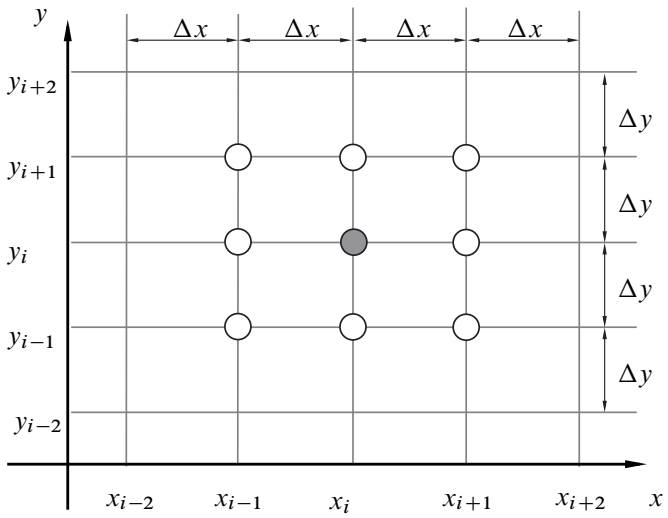


Figure 3.2. *The rectangular grid with general difference scheme for second order partial differential equations.*

a general difference scheme for the point (x_i, y_j) where solution of second order linear partial differential equations by finite difference method can be found.

Independent variables can be related with space, eg. coordinates of rectangular Cartesian reference system x, y, z or also with time t. In the finite difference method for the differential equations considered in time domain exist are two basic calculation algorithms: explicit and implicit. If the first derivative of certain function equates to $\frac{dz}{dt} = f(t)$ then we can find the value of the function z_{i+1} at the subsequent moment $t + \Delta t$ using the option:

– explicit

$$\left.\frac{dz}{dt}\right|_{t_i} \approx \frac{z_{i+1} - z_i}{\Delta t} \tag{3.22}$$

$$z_{i+1} = z_i + f(t_i)\Delta t \tag{3.23}$$

– implicit

$$\left.\frac{dz}{dt}\right|_{t_{i+1}} \approx \frac{z_{i+1} - z_i}{\Delta t} \tag{3.24}$$

$$z_{i+1} = z_i + f(t_{i+1})\Delta t \tag{3.25}$$

Comparing the equations (3.22) and (3.24) we can easily notice a fundamental difference in the definition of both algorithms. In explicit method the time derivative is defined at the moment t_i, whereas for implicit at the time t_{i+1}. The value of functions in point z_{i+1} can be calculated from so called explicit scheme (3.23). Searched value in this method is only at the left side of the equation. An alternative

scheme is implicit (3.24), because z_{i+1} is also inside the functions $f(t_{i+1})$. In this work the examples of explicit and implicit schemes will be presented for a function of two variables $z(x,t)$, where its derivatives create a differential equation:

$$\frac{\partial z}{\partial t} = c\frac{\partial^2 z}{\partial x^2} \tag{3.26}$$

Using the explicit method the first derivative with respect to the time variable t was replaced by forward difference quotient

$$\left.\frac{\partial z}{\partial t}\right|_{t_i,x_j} = \frac{z_{i+1,j} - z_{i,j}}{\Delta t} \tag{3.27}$$

whereas the second derivative with respect to the spatial variable x by central difference quotient at the moment t_i

$$\left.\frac{\partial^2 z}{\partial x^2}\right|_{t_i,x_j} = \frac{z_{i,j+1} - 2z_{i,j} + z_{i,j-1}}{\Delta x^2} \tag{3.28}$$

Substituting the equations (3.27) and (3.28) to the equation (3.26) was obtained

$$z_{i+1,j} = z_{i,j} + \frac{c\Delta t}{\Delta x^2}(z_{i,j+1} - 2z_{i,j} + z_{i,j-1}) \tag{3.29}$$

A difference scheme for the explicit method allowing to define z at time t_{i+1} is presented in the figure 3.3. The basic advantage of the algorithm is the use of simple calculation procedures. Unfortunately, the stability of the method requires the imposition of a limitation on the time increment length $\Delta t \leqslant \Delta t_{max}$, where limit value Δt_{max} is calculated from [2]:

$$\frac{c\Delta t_{max}}{\Delta x^2} = \delta \tag{3.30}$$

If $\delta \leqslant 1/2$ then the errors in a numerical procedures will not increase, but they can oscillate. In the event when $\delta \leqslant 1/4$ the oscillations disappear and for $\delta = 1/6$ errors in finite difference method are minimal for the considered equation (3.26).

Using the implicit method to find a solution of the differential equation (3.26) the first derivative with respect to the time variable t was replaced by backward difference quotient

$$\left.\frac{\partial z}{\partial t}\right|_{t_{i+1},x_j} = \frac{z_{i+1,j} - z_{i,j}}{\Delta t} \tag{3.31}$$

whereas the second derivative with respect to the spatial variable x by a central difference quotient at time t_{i+1}

$$\left.\frac{\partial^2 z}{\partial x^2}\right|_{t_{i+1},x_j} = \frac{z_{i+1,j+1} - 2z_{i+1,j} + z_{i+1,j-1}}{\Delta x^2} \tag{3.32}$$

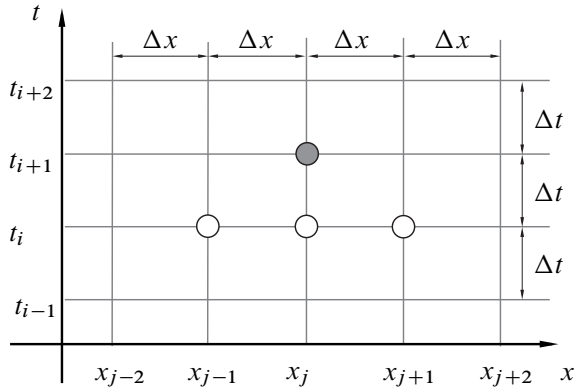


Figure 3.3. *Rectangular grid with difference scheme for the explicit method*

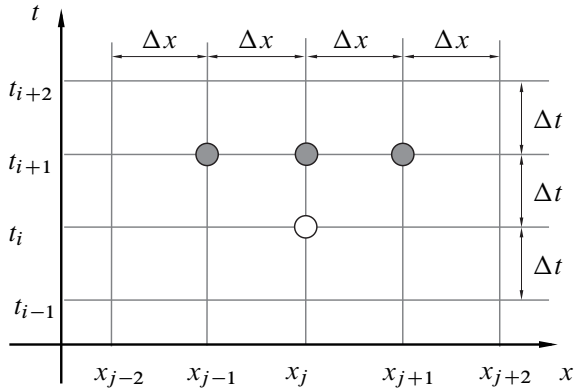


Figure 3.4. *Rectangular grid with difference scheme for the implicit method*

After substituting the equations (3.31) and (3.32) into the equation (3.26) was obtained

$$z_{i+1,j} = z_{i,j} + \frac{c\Delta t}{\Delta x^2}\left(z_{i+1,j+1} - 2z_{i+1,j} + z_{i+1,j-1}\right) \tag{3.33}$$

The presented scheme is implicit, because $z_{i+1,...}$ exist in both side of this equation (on the right and left) (3.33). Implicit algorithm requires the application of complicated calculation procedures, but is absolutely stable. A differential scheme for the implicit method allowing to define z at the moment t_{i+1} is presented in the figure 3.4.

In the FDM method the grid of points is limited, has a finite size. For the points located at the edges of the considered grid is required to include the limit conditions (boundary and/or initial). Consider the function z which depends on a spatial

variable x and time t. Boundary conditions will define additional dependencies in space domain. We can distinguish two main types of them [2]:

- type 1 or Dirichlet condition – boundary conditions imposed on the function z, e.g.:

$$z|_{x=0,\,t_i} = 0 \tag{3.34}$$

- type 2 or Neumann condition – the form of boundary condition includes the derivative function, e.g.

$$\left.\frac{\partial z}{\partial x}\right|_{x=0,\,t_i} = 0 \tag{3.35}$$

Initial conditions are defined in time domain at the moment $t = 0$. Analogously as for boundary conditions we can distinguish two basic types of initial conditions:

- initial condition imposed on the function z, e.g.:

$$z|_{x_i,\,t=0} = 0 \tag{3.36}$$

- initial condition, where its form includes the first derivative function, z, e.g.:

$$\left.\frac{\partial z}{\partial t}\right|_{x_i,\,t=0} = 0 \tag{3.37}$$

The application of the finite difference method to solve the second order linear differential partial equations requires consideration: boundary problem (elliptic equations) or initial - boundary problem (parabolic and hyperbolic equations).

3.3. Example – string vibrations

In the presented example, we consider the equation of transverse string vibrations in the form

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} \tag{3.38}$$

where: $c^2 = \frac{S}{\rho}$, S – Tension in the string, ρ – mass per unit length of string [6]. The scheme of the analysed system is presented in the figure 3.5.

In the finite difference method the explicit algorithm was used. The derivative with respect to the spatial variable x was replaced by central difference quotient:

$$\left.\frac{\partial^2 y}{\partial x^2}\right|_{x_j,\,t_i} = \frac{y_{j+1,\,i} - 2y_{j,\,i} + y_{j-1,\,i}}{\Delta x^2} \tag{3.39}$$

The same type of the difference quotient was used to replace the second time derivative of the function $y(x, t)$:

$$\left.\frac{\partial^2 y}{\partial t^2}\right|_{x_j,\,t_i} = \frac{y_{j,\,i+1} - 2y_{j,\,i} + y_{j,\,i-1}}{\Delta t^2} \tag{3.40}$$

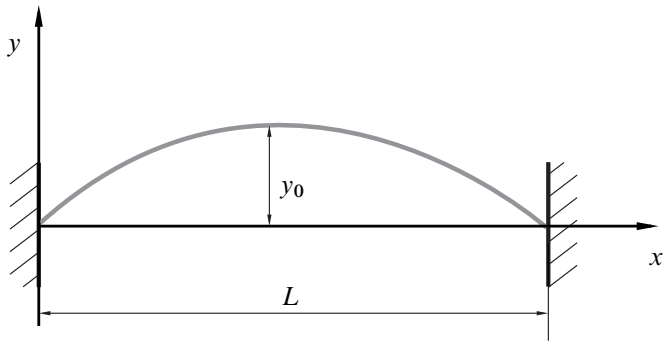


Figure 3.5. *Scheme of the string's initial deflection*

After substituting the equations (3.39) and (3.40) to the equation (3.38) was obtained

$$y_{j,i+1} = \delta^2(y_{j+1,i} - 2y_{j,i} + y_{j-1,i}) + 2y_{j,i} - y_{j,i-1} \tag{3.41}$$

where: $\delta = \frac{c\Delta t}{\Delta x}$. The condition of stability for the explicit method adopts the form $\delta \leqslant 1$ [1]. Using the equation (3.41) a difference scheme was made. It is presented in the figure 3.6. Such a diagram is valid for any moment of time, where $i \geqslant 0$, $(t \geqslant 0)$.

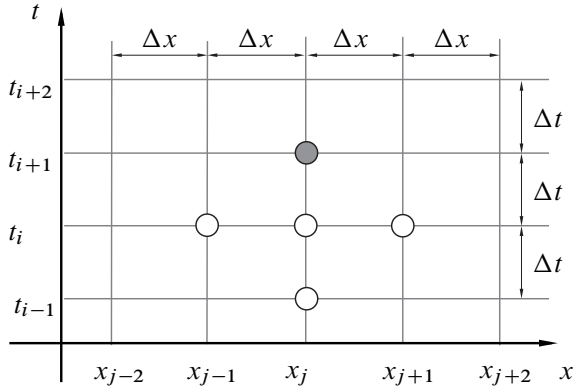


Figure 3.6. *Rectangular grid with the difference scheme for the considered example*

Finding a solution requires consideration of the boundary and initial conditions. The following conditions were adopted:

– boundary – resulting from clamped both ends of the string

$$z|_{x=0,\,t_i} = 0 \tag{3.42}$$

$$z|_{x=L,\,t_i} = 0 \tag{3.43}$$

– initial – dependent on the accepted initial deformation and velocity of the string

$$z|_{x_j,\,t=0} = y_0 \sin\left(\frac{\pi}{L} x_{j,\,i=0}\right) \tag{3.44}$$

$$\left.\frac{\partial z}{\partial t}\right|_{x_j,\,t=0} = \frac{y_{j,\,i=1} - y_{j,\,i=-1}}{2\Delta t} = 0 \tag{3.45}$$

On the basis of the equation (3.45) was defined a dependence, which is necessary to calculate the first iteration:

$$y_{j,i=-1} = y_{j,i=1} \tag{3.46}$$

The string was discretized into 6 elements with the length equal to $\Delta x = \frac{L}{6}$. Substituting the dependencies (3.42), (3.43), (3.44), (3.46) to the equation (3.41) was obtained the matrix form for the first iteration. It is used to calculated the values $y_{j,1}$:

$$\begin{Bmatrix} y_{1,1} \\ y_{2,1} \\ y_{3,1} \\ y_{4,1} \\ y_{5,1} \\ y_{6,1} \\ y_{7,1} \end{Bmatrix} = \begin{Bmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \\ y_{4,0} \\ y_{5,0} \\ y_{6,0} \\ y_{7,0} \end{Bmatrix} + \frac{\delta^2}{2} A \begin{Bmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \\ y_{4,0} \\ y_{5,0} \\ y_{6,0} \\ y_{7,0} \end{Bmatrix} \tag{3.47}$$

where:

$$\begin{Bmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \\ y_{4,0} \\ y_{5,0} \\ y_{6,0} \\ y_{7,0} \end{Bmatrix} = \begin{Bmatrix} 0 \\ y_0 \sin\left(\frac{\pi}{6}\right) \\ y_0 \sin\left(\frac{2\pi}{6}\right) \\ y_0 \sin\left(\frac{3\pi}{6}\right) \\ y_0 \sin\left(\frac{4\pi}{6}\right) \\ y_0 \sin\left(\frac{5\pi}{6}\right) \\ 0 \end{Bmatrix} \tag{3.48}$$

describes the initial deformation of the string. Whereas, matrix A contains coefficients defined during discretization in the space domain and including boundary conditions.

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.49}$$

For next iterations, at the moments where $i \geqslant 2$ the transverse vibrations of the string were calculated from the equation:

$$\begin{Bmatrix} y_{1,i} \\ y_{2,i} \\ y_{3,i} \\ y_{4,i} \\ y_{5,i} \\ y_{6,i} \\ y_{7,i} \end{Bmatrix} = 2 \begin{Bmatrix} y_{1,i-1} \\ y_{2,i-1} \\ y_{3,i-1} \\ y_{4,i-1} \\ y_{5,i-1} \\ y_{6,i-1} \\ y_{7,i-1} \end{Bmatrix} + \delta^2 A \begin{Bmatrix} y_{1,i-1} \\ y_{2,i-1} \\ y_{3,i-1} \\ y_{4,i-1} \\ y_{5,i-1} \\ y_{6,i-1} \\ y_{7,i-1} \end{Bmatrix} - \begin{Bmatrix} y_{1,i-2} \\ y_{2,i-2} \\ y_{3,i-2} \\ y_{4,i-2} \\ y_{5,i-2} \\ y_{6,i-2} \\ y_{7,i-2} \end{Bmatrix} \tag{3.50}$$

Calculations were perform in Matlab software [5] for the selected parameters: $\delta = 0.1$, $y_0 = 0.1$. In fig. 3.7 the obtained results have been presented. Approximated

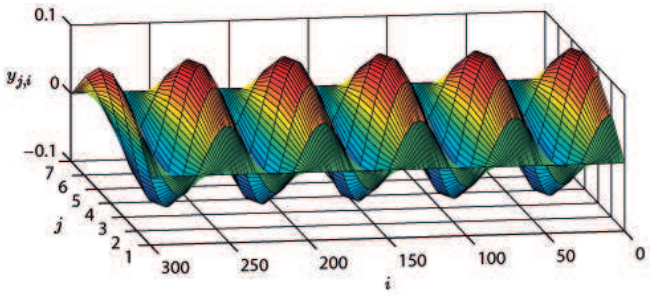


Figure 3.7. *Transverse vibrations of the string*

transverse vibrations of the string $y_{i,j}$ were presented on 3D graph which was

performed by use command mesh for subsequent points of FDM grid. The presented example the application of the finite difference method to finding solution of second order linear partial differential equations is not complicated. In scientific research FDM can be used to modelling many difficult problems, for example: research on damaged structures [4].

Bibliography

[1] Cichoń C. (2005): *Metody obliczeniowe: wybrane zagadnienia*. Wydawnictwo Politechniki Świętokrzyskiej.

[2] Cichoń C., Cecot W., Krok J., Plucński P. (2009): *Metody komputerowe w liniowej mechanice konstrukcji*. Wydawnictwo Politechniki Krakowskiej.

[3] Kącki E. (1989): *Równania różniczkowe cząstkowe w zagadnieniach fizyki i techniki*. Wydawnictwo Naukowo-Techniczne.

[4] Manoach E., Warmiński J., Mitura A., Samborski S. (2012): Dynamics of a composite Timoshenko beam with delamination. *Mechanics Research Communications* **46**: 47–53.

[5] Pratap R. (2009): *Matlab 7 dla naukowców i inżynierów*. Wydawnictwo Naukowe PWN.

[6] Szabelski K. (2002): *Zbiór zadań z drgań mechanicznych*. Wydawnictwo Politechniki Lubelskiej.

4. Analysis of non-linear signals

4.1. Introduction

The theory of signals is one of the fundamental areas of the technical knowledge. Its familiarity is necessary not only for designers of electronic devices, but also for automation specialists, IT scientists, electrical technicians and specialists for data communication and mechanics. The development of digital technique revolutionized the methods of processing signals, new methods of analysis appeared, but the basics of the mechanisms are invariable – till the of Fourier and Laplace's transforms are used and classical algorithms of modulation.

In a colloquial language a signal is a mark with some information text. In technical sciences a signal is defined as a function $f(t)$ dependent usually on time. In short words, a signal is a carrier of information.

By means of analysis of signals, we will understand in accordance with a definition taken from a dictionary of Polish language – thoughtful phenomenon, separation of features by definition, parts or components of the examined phenomenon or subject; examination of features of elements or structure of something and connections between them (...). The most convenient definition says that analysis of a signal is an examination the aim being identification of properties, features, measures of a signal as well as reproduction the information carried by the signal [12]. The most popular methods used in the analysis of the signal are:

- examination of statistical measures (moments),
- analysis of probability distribution,
- correlation analysis,
- spectrum analysis (spectral, fourier or frequency),
- wavelet analysis.

A frequently used definition is also processing the signal, i.e. change of property, form, features and measures of the signal, for easier analysis thereof, registration, storing. The signal, due to its nature, can be divided into:

- deterministic,
- stochastic (random),
- mixed.

A deterministic signal is foreseeable, whereas stochastic is a set of random information. A mixed signal is composed of at least two component signals one of which is deterministic and the other random. A deterministic signal in turn, is divided into periodical and non-periodical in accordance with a diagram of those presented in the figure 4.1.

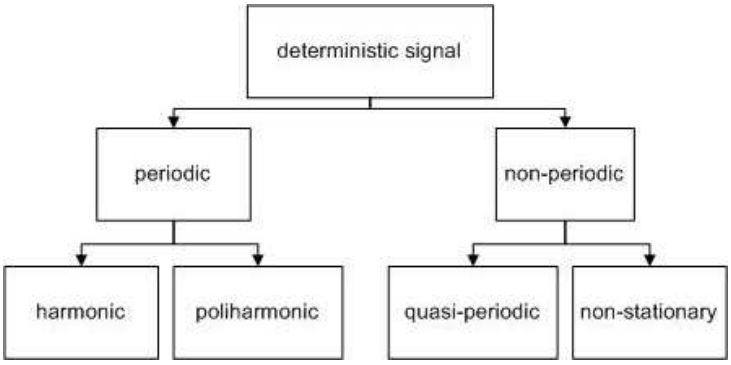


Figure 4.1. *Division of deterministic signals*

Through periodical signal we understand a time course which fulfils the dependence:

$$x(t) = x(t+T) = \ldots = x(t+kT), \tag{4.1}$$

where T is a period of signal, k an integral number. A typical example of the harmonic signal, namely also periodical one there is a sinusoidal course:

$$x(t) = A_o + A\sin(2\pi f t + \varphi), \tag{4.2}$$

where A_o is an average value, A – amplitude, f – frequency ($f = 1/T$) expressed in Hz, φ – phase shift.

A philharmonic signal is whereas a sum of harmonic signals described with the formula:

$$x(t) = A_o + \sum_{n=1}^{N} A_n \sin(2\pi n f t + \varphi_n) \tag{4.3}$$

f is basic frequency here, whereas n shall mean a number of harmonic component.

A quasi-periodical signal is similar in terms of mathematical formula, which is described with an equation:

$$x(t) = A_o + \sum_{n=1}^{N} A_n \sin(2\pi f_n t + \varphi_n). \tag{4.4}$$

However, the essence of the quasi-periodical signal involves that the relations of frequency f_i/f_k is an irrational number to at least one pair of components of the signal, so called harmonic.

In the chapter, the examinations of the signals will be conducted on the basis of a classic Fourier's method as well as by means of recurrence plots, based on the method of delayed coordinates and multi-scale entropy.

4.2. Fourier's Transform

Signals, most often obtained from measurement, are analysed by us in the time domain or in frequency domain. The tools which allow motion between these two areas are transform of Fourier's and reverse transform of Fourier's. The analysis in the frequency domain is called spectrum one. In a general case when we consider the problem of spectrum analysis of signals we consider four different methods of Fourier's analysis meaning: transformation of Fourier – frequency changes in the continuous manner and a series of Fourier – discrete frequency, respectively for continuous signals (continuous time) and discrete signals (discrete time). The result of the transformation conducted is transform [12]. However a definition of transformation and transforms are very often used interchangeably.

Transformation of Fourier's exchanges the function of real variable $f(x)$ into the function of complex variable $F(s)$. Transformation of Fourier's for continuous signals can be defined with the equation:

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ixs}dx. \tag{4.5}$$

Whereas reverse transformation with the equation:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(s)e^{2\pi ixs}ds, \tag{4.6}$$

where i shall mean imaginary unit. In practice often a variable x shall mean time (in seconds) and the argument of transforms s shall mean frequency expressed in Hz.

In practice, as the result of measurements we obtain data with discrete nature, and not continuous one, we will use a Discrete Fourier Transform (DFT) and Inverse Discrete Fourier Transform (IDFT). For the N-element course a discrete Fourier transform we define as follows:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi ink/N}, \ k \in \{0, 1, \ldots N-1\}, \tag{4.7}$$

where k is a number of harmonic, n number of the signal sample, x_n value of the signal sample (Fourier ratio), N number of samples.

Calculating DFT in required in the 60s of the 20th century so much calculation power that the machines from the period limited the usage of the algorithm. The

year 1965 brought a revolution, J. Cooley and J. Tuckey published a paper under the title „An Algorithm for the machine computation of complex Fourier series", in which they developed a faster algorithm of calculating a discrete Fourier transform commonly known as FFT — Fast Fourier Transfor). FFT is DFT with a reduced number of necessary arithmetical operations. The goal of FFT is to reduce a long calculation algorithm by its division into shorter and simpler calculations of DFT and shortening the time of calculation. The most popular version of FFT is FFT with the basis of 2. The Algorithm of FFT with the base of 2 is a very effective procedure for calculating DFT provided the dimension of DFT will be total power of two. A good approach is to add a required number of samples with zero values to the final part of the time series in order to adjust the number of its points to the next size of FFT on the basis of 2. Algorithms which calculate a fast Fourier transform are based on a method of "divide and win" on recurrence basis. It means that we divide a problem into sub-problems with smaller size and these recurrence ones we divide again into smaller ones, etc. Reaching satisfactory small problems we solve them. The solution of an initial problem is a sum of sub-problems.

In science and technique the measured values very often are of periodical nature, i.e. which causes the repetition of a given physician value with a defined period T. Such a periodical function may be presented in the form of infinite trigonometry series called also Fourier series:

$$f(x) = a_o + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{T} + b_n \sin \frac{n\pi x}{T} \right). \tag{4.8}$$

Due to the Fourier analysis we may learn which harmonic components (periodical) are present in a signal and in which relative amounts they appear in them. The ratios of the Fourier series, called in short Fourier ratios, a_o, a_n, b_n are interpreted as amplitudes of proper harmonic components. They are presented by means of formulas of Euler-Fourier:

$$\begin{aligned} a_o &= \frac{1}{2\pi} \int_{-T}^{T} f(x) dx, \\ a_n &= \frac{1}{\pi} \int_{-T}^{T} f(x) \cos \frac{n\pi x}{T} dx, \\ b_n &= \frac{1}{\pi} \int_{-T}^{T} f(x) \sin \frac{n\pi x}{T} dx. \end{aligned} \tag{4.9}$$

Each periodical signal $x(t)$ may be represented as a sum of series of harmonic functions if the Dirichlet's conditions are fulfilled:

– function $x(t)$ is absolutely of integral nature in any range with the period length T,

– in any limited range, $x(t)$ has a definite number of maximas and minimas with definite value,
– in any limited range, $x(t)$ has a definite number of non-continuance.

These conditions are fulfilled for a majority of signals encountered in reality.

The quality of the spectrum obtained from FFT is affected by so called frequency of Nyquist, namely maximum frequency of harmonic components of the harmonic signals being subject to the sampling process, which may be restored from a series of samples without deformations. Spectrum components with frequencies higher than frequencies of Nyquist are subject, during sampling, to overlapping on the components with other frequencies (aliasing phenomenon) which causes that they cannot be properly restored any more. In image the phenomenon was shown on the figure 4.2, where through the points representing the samples of signal one may conduct a few curves (signals) e.g. red and blue time series.

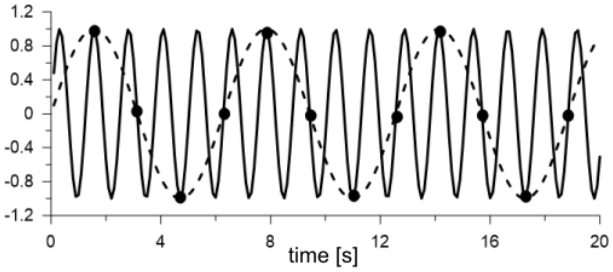


Figure 4.2. *Example of non-definiteness of a signal – aliasing*

In accordance with the theory of Kotielnikow-Shannon, with equal sampling with distance of sampling T_s (constant time step), a condition for a proper restoration of the signal is that its width of band B was strictly limited $B < 1/T_s$ or that the maximum frequency of a signal does not exceed the half of sampling frequency $f_{max} < f_s/2, or f_{max} < 1/2T_s$. In other words, the frequency of Nyquist is equal to half of sampling frequency $f_N = f_s/2$ or $f_N = 1/2T_s$.

As an example, for the sampling frequency 44.1 kHz used on CD records, the frequency of Nyquist is 22.05 kHz. If in the analogue signal the components are present with the frequency higher than frequency of Nyquist, it will cause the errors in sampling (aliasing). However, a human ear does not hear the frequencies higher than 22 kHz, therefore the components of the signal are cut out shortly before sampling by means of using a low-pass filter.

Although in theory, frequency of Nyquist indicates the limit of the band, which can be properly written in application of a defined sampling frequency, the limit is slightly lower than frequency of Nyquist.

In order to avoid aliasing, one should assure that the sampled signal was limited in bands to the Nyquist frequency namely the half of sampling frequency. The

phenomenon can be used by limiting the spectrum of a signal with the filter, called anti-aliasing filter. The filter should have the width of the band smaller than a half of sampling frequency. Usually filters with clearly smaller width of throttle band are used in order to include small damping, which takes place on the transition section of characteristics of the separating filter for the throttle band from the border one.

In practice, due to the fact that no signal with finite duration has limited band (which results from properties of the Fourier transform), and no filter dampens ideal in its border band aliasing appears always. In properly designed system using the sampling of the signal seeks to minimize the phenomenon so that the amplitude of alias components was small.

In nature, a majority of signals has a continuous nature as sound (changes in air pressure in time) or electroencephalogram (EEG, electric potential of the brain measured from the surface of the skull). Irrespective of the fact, modern analysis of signals refers in practice mainly to discrete signals the example being the value of shares in moments of closing next sessions of the stock exchange. The exchange of the continuous signal into discrete one has the name of discretization. The discretization process bears a danger of loss of information about the condition of the subject between samples and is strictly connected with the frequency of signal sampling. The example can be the record of an image being subject to improper discretization (figure 4.3). The image on the left side is original, whereas the one on the right side discretized.

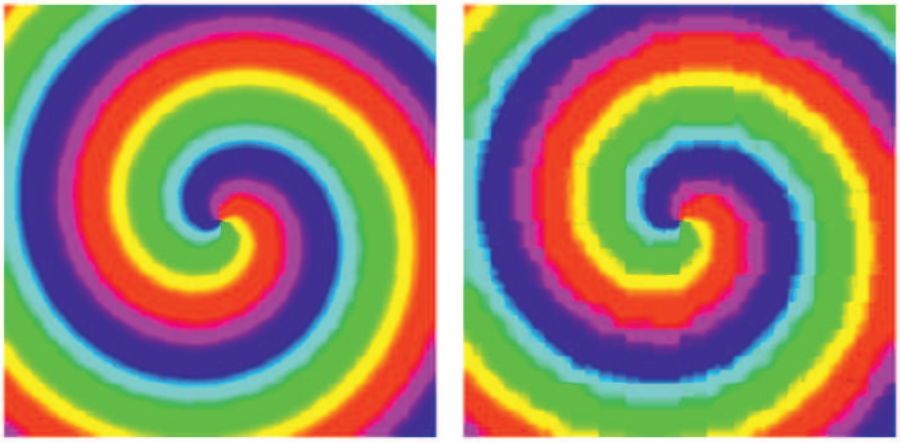


Figure 4.3. *Image before and after discretization*

In the next section, a proposal of an exercise conducted in the MATLAB program is presented. The exercise shows the effect of losing data as a result of discretization of a signal and allows to observe spectra of a few types of signals treated separately as well as after their summing in one course.

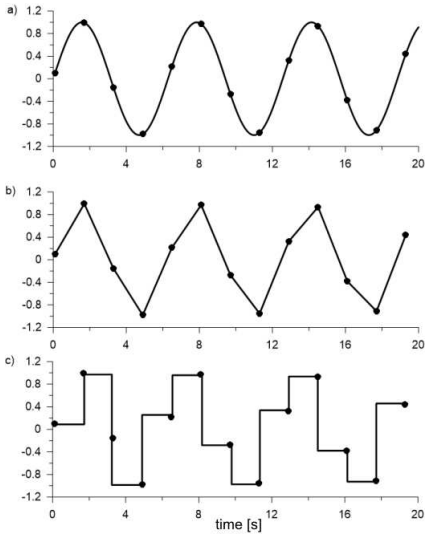


Figure 4.4. *Continuous signal (a) and discrete (b – d)*

Proposal of an exercise

If the continuous signal in nature (e.g. sound) we decide to analyse or store in digital form then this continuous function (e.g. air pressure) in time we have to replace with values measured in the finite (equal best) time intervals as it was shown in the figure 4.4.

Sampling exchanges the continuous signal (a) into points (b) with coordinates in moments of sampling and values of continuous signals corresponding to them. If we dispose only of a sampling signal (b), we can complete the values from among samples assuming that the signal between them is, e.g. linear (c) or constant from the previous point (d) — we see discrepancies with the original signal (a). Therefore a selection of sampling frequency is very significant.

We will conduct an example which will show possibilities to us of the MATLAB program and will explain an influence of the particular components of the signal on the Fourier spectrum. First, the following signals will be generated:

- harmonic $x_1 = 0.7\sin(2\pi 50t)$ with frequency f_1=50 Hz,
- harmonic $x_2 = 0.3\sin(2\pi 100t)$ with frequency f_2=100 Hz,
- quasi-periodical composed of two periodical signals 100 and $100\sqrt{3}$ Hz – i.e. $x_3 = 0.2\sin(2\pi 100t) + 0.3\sin(2\pi 100\sqrt{3}t)$,
- signal type "chirp": $x_4 = 0.4\sin(2\pi 130t^2)$,
- noise generated by the function "random": $x_5 = 0.1 * \mathrm{randn}(\mathrm{size}(1 * t))$.

It should be made with the commands line in MATLAB or use for this the (toolbox) Simulink package. The sampling frequency of these signals (time step)

should be fixed $f_{s1} = 1000$ Hz. Ready signals should look as those in the figure 4.5. Then we create a new course x, being the sum of above mentioned and we obtain the signal showed on the figure 4.6. In order to show the influence of sampling time of the signal on its exactness we change the number of samples in the signal x_1 so that the signal frequency will be reduced 8 times to adopt $f_{s2} = 125$ Hz. Then the time course x_1 with a changed number of samples (sampling frequency f_{s2}) takes now a completely different, deformed form drawn with a red colour on the figure

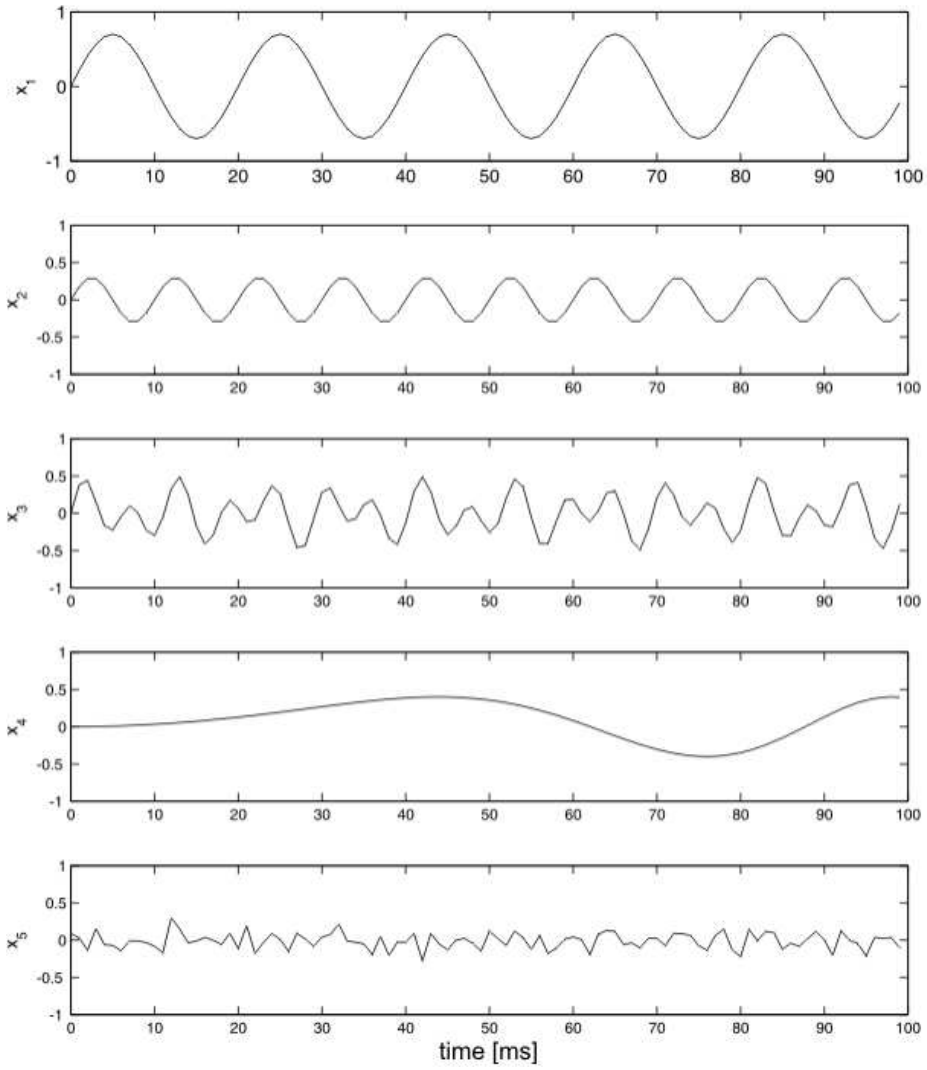


Figure 4.5. *Signals x_1–x_5 generated with sampling frequency $f_{s1} = 1000\,Hz$*

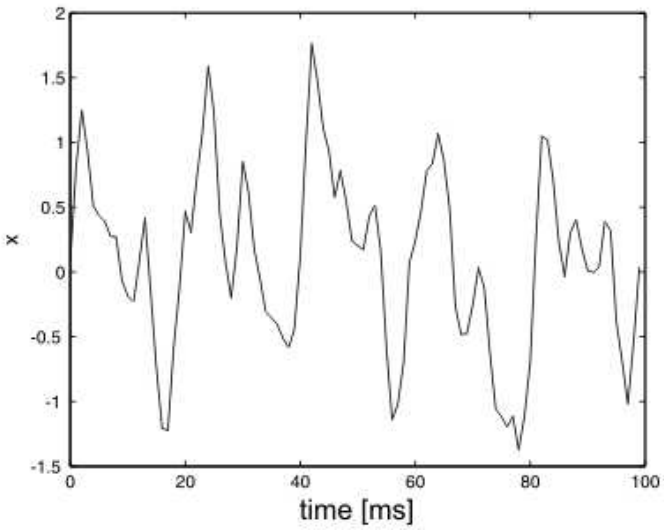


Figure 4.6. *Signal x being the sum of signals from x_1 to x_5*

4.7, in relation to the course of the original one (black colour). In order to change the number of samples in the signal one may use e,g. ”decimate” command, which is used just for changing the sampling frequency of the signals.

Then, we conduct the analysis of the frequency spectrum of the signal x performed by means of fast Fourier transform (FFT). There are many methods to perform FFT in the program MATLAB. Here, for this purpose, the script was used with the use of the command ”fft”, but one may use blocks with the module Simulink and create model showed in the figure 4.8. The model shall collect the signal from the work space of the program, perform the analysis and present the results in the form of frequency spectrum.

The results of the Fourier transforms for the signals x_1–x_5 and the signal x being the sum of courses from x_1 to x_5 were presented properly in the figures 4.9 and 4.10. Such spectra with characteristic peaks corresponding to given frequencies should be obtained in the result of performing the exercise although the methods of their calculation and outlining in the MATLAB are more. Here only two of them were presented as an example.

4.3. Method of delayed coordinates and recurrence diagrams

The analysis of experimentally measured data is very often complicated due to the complex nature of the process. Usually such courses are nor ordered or even may show the features of deterministic chaos. Then, the usage of a proper method to

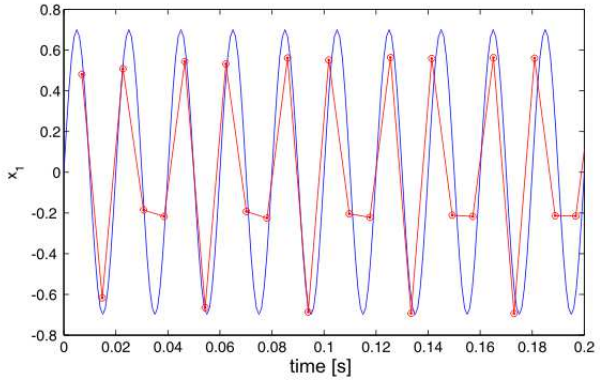


Figure 4.7. *Signal x_1 generated with the sampling frequency $f_{s1} = 1000Hz$ (blue) and $f_{s1} = 20Hz$ (red)*

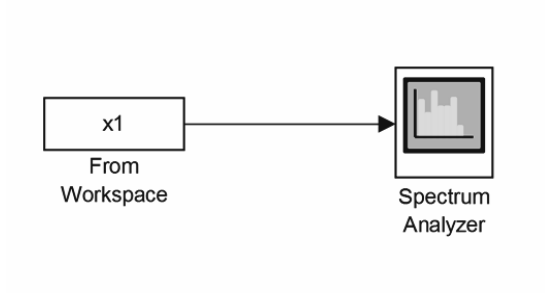


Figure 4.8. *Matlab-Simulink model for FFT analysis*

analyse them is required. One of the latest techniques used to examine non-linear time courses is the one involving reconstruction of the $x(t)$ vector in the phase space created for the new coordinates so called delays, thus the name method of delays. The vector obtained from the time series reconstructed in a news state phase has the form:

$$x = (x_i, x_{i+d}, x_{i+2d}, ..., x_{i+(m-1)d}) \tag{4.10}$$

where x_i mean coordinate (i the sample) in the time series $x(t)$, d – time of delay, m – embedding dimension.

Too small size of a dimension m makes that trajectories distance from each other in reality may seem to be close to each other. Whereas, too big embedding dimension complicates calculations and extends its time losing at the same time information about relations between particular points of time course.

In accordance with the Takens's theorem [8] the condition should be fulfilled:

$$m \geqslant 2D_2 + 1 \tag{4.11}$$

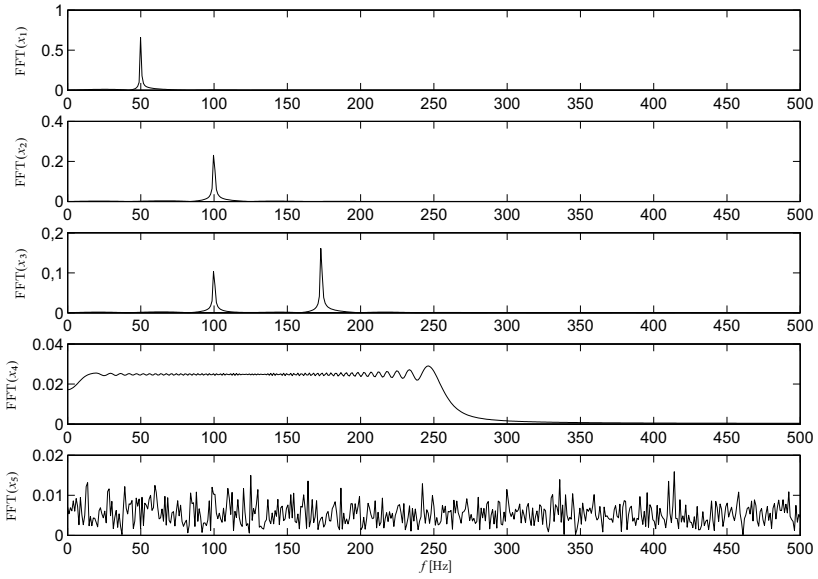


Figure 4.9. *FFT for signals* x_1–x_5

there is a diffeomorphism between original and reconstructed atractor which means that both attractors represent the same dynamic system in different systems of coordinates. D_2 means here attractor dimension [8].

The most popular method for calculating the dimension m is the False Nearest Neigbors – (FNN). It involves finding such a dimension m – in order to avoid false cutting of close trajectories. An exact description of the method can be found in the paper [9].

The time delay d is usually calculating with the Average Mutual Information method – (AMI). In this method, contrary to the function of auto-corelation, also non-linear corelations are included. The essence of AMI is to estimate how much information on average contained in one state may be the result of the prediction from the information contained in the previous one. The value of delay is adopted as the least level d, for which the function accepts local minimum. The function AMI is calculated from the formula:

$$\mathrm{AMI} = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j}, \tag{4.12}$$

where p_i is the probability of finding the value of time characteristics of the system and in this range, p_{ij} shall mean probability that the observation of a given moment of time t belongs to the range and the observation at a moment $(t + \tau)$ for the range. The exact description of the AMI method can be found e.g. in the paper [6].

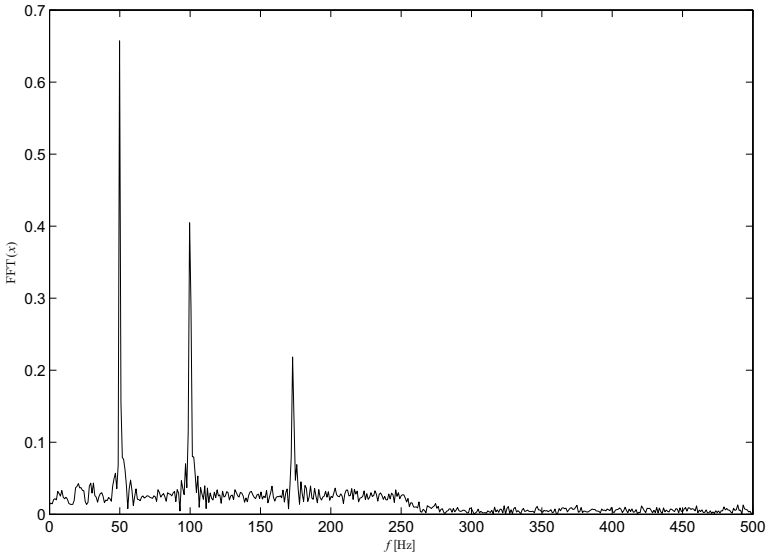


Figure 4.10. *FFT for the signal* $x = x_1 + x_2 + x_3 + x_4 + x_5$

On the basis of the method, of delayed coordinates, a recurrence plot which is a graphic interpretation of recurrence of conditions in the working space. The definition of recurrence conditions of the preventive systems was introduced in 1890 by Henri Poincare. Despite large interest in his discover, only after development of effective calculation systems, its practical application was possible in the numeric analysis of the dynamic systems.

The recurrence plot presents the recurrence of states of the process of the phenomenon (or the system). An important advantage of the method is a possibility to apply both for a large and a small set of data, including also non-stationary courses. A recurrence plot was introduced by Eckmann in 1987 in order to present the visualization of conditions of a certain variable x_i in the phase space [5]. In the subsequent years, a quantity method of analysis was developed called: Recurrence Quantification Analysis – (RQA) [22]. The recurrence plots is based on the dependence:

$$\mathrm{R}_{i,j} = \mathrm{H}\left(\varepsilon - \left\|\mathbf{x_i} - \mathbf{x_j}\right\|\right), \tag{4.13}$$

in which i, j is the number of state in space: $i, j = 1, \ldots, N$, where N is the number of considered points, x s a reconstructed vector of delay in the state space, H – Heaviside's function, $\|x\|$ – norm of the vector X in the space, most often Euclid's norm, ε – non-negative real number (so called radius of search o parameter of cut-off).

If in the phase space, the distance between points x_i and x_j is smaller than ε – $\mathrm{R}_{i,j} = 1$, in the contrary case $\mathrm{R}_{i,j} = 0$. The result $\mathrm{R}_{i,j} = 1$, shall mean the occurrence of recurrence which in the plot is marked with a point. If $\mathrm{R}_{i,j} = 0$, means the lack of recurrence – white point on the diagram (lack of a point). The diagram obtained must be symmetric towards so called line of identicality (main diagonal of the matrix). Depending on the nature and properties of the process in the recurrence plot, there can be continuous lines, interrupted, vertical, transverse or varied points (single or in concentrations).

The selection of proper value of a parameter ε is very important here, because if the parameter is too small, then the number of points on the recurrence plot is insufficient for analysis. On the other hand, however, if we take the value of the parameter ε which is too big, the number of recurrence points will be too high. As a consequence we will obtain so called artificial points which are not significant from the point of view of the process dynamics and even harmful as they obscure its image. Too big value of the parameter m and d may cause the occurrence of the calculation structures on the recurrence diagram not being the reflection of the actual dynamics of the system.

Due to the repeating formulas of the recurrence diagrams they are divided into:

- uniform: characteristic for the stationary courses, white noise (figure 4.11(a)),
- periodical and quasi-periodical: containing structures with periodical or quasi--periodical recurrence (figure 4.11(b)),
- drifting: obtained for the non-stationary systems (figure 4.11(c)),
- torn: characterized with white areas, repeated as a result of rapid changes in system dynamics (figure4.11(d)).

On the recurrence plots there can be single points, horizontal and vertical lines creating so called texture. Single isolated recurrence points correspond rarely to the states appearing, short-lasting in a single moment of time. The lines represent local relation between particular fragments of phase trajectory. Diagonal lines occur when a fragment of the phase trajectory runs parallel to another fragment of the trajectory, namely in case of the periodical motions.

Proposal of an exercise

The procedure of obtaining recurrence diagrams will be presented here on the example of signals generated in the previous sub-section and showed on the figure 4.5. A package TISEAN [7, 14, 13], will be used for that which consists of independent programs (commands) for the analysis of the signals with a delay methods. In the first stage, the function of mutual information will be calculated (AMI) by means of a program (commands) ”mutual”. The value of a delay for all signals, selected as the first minimum was established on the level 1. The diagrams of the function AMI for particular courses was showed on the figure 4.12.

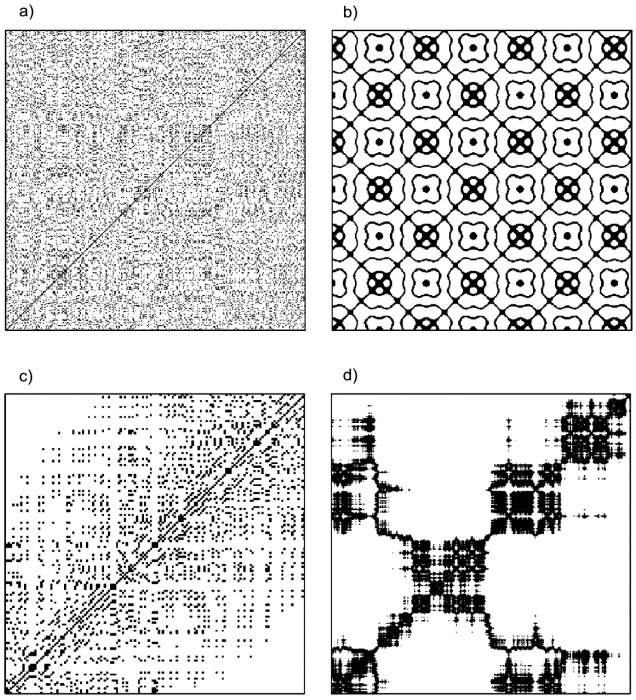


Figure 4.11. *Recurrence plots with the a) uniform (white noise),b) periodical, c) drifting, d) torn structure [13]*

In the next step, one should calculate the embedding dimension m. For this effect, the false nearest neighbours method should be used (FNN) and the command "false_nearest" from the TISEAN package was applied. The results of the calculations of the embedding dimension m for different time courses were showed on the figure 4.13. In case of periodical signals (x_1 and x_2) the embedding dimension m is 2, for quasi-periodical (x_3) $m = 3$. For other signals, which deviate in their nature from the periodical one, the dimension m is always higher, and the number of false neighbours never falls to zero, especially for data containing the noise. The last parameter which should be selected so as to outline the recurrence plots is the size of the surrounding ε. Optimum value of the parameter should assure proper degree of shadowing the recurrence plot.

Recurrence plots obtained for signals x_1–x_5 and x were presented on the figure 4.14. The diagrams (a) and (b) represent the periodical signal with different frequency and amplitude therefore long diagonal lines correspond to them in different distances from each other. Both these diagrams were outlined with the same parameters (m, d, ε) and therefore the distance between diagonal lines proves the

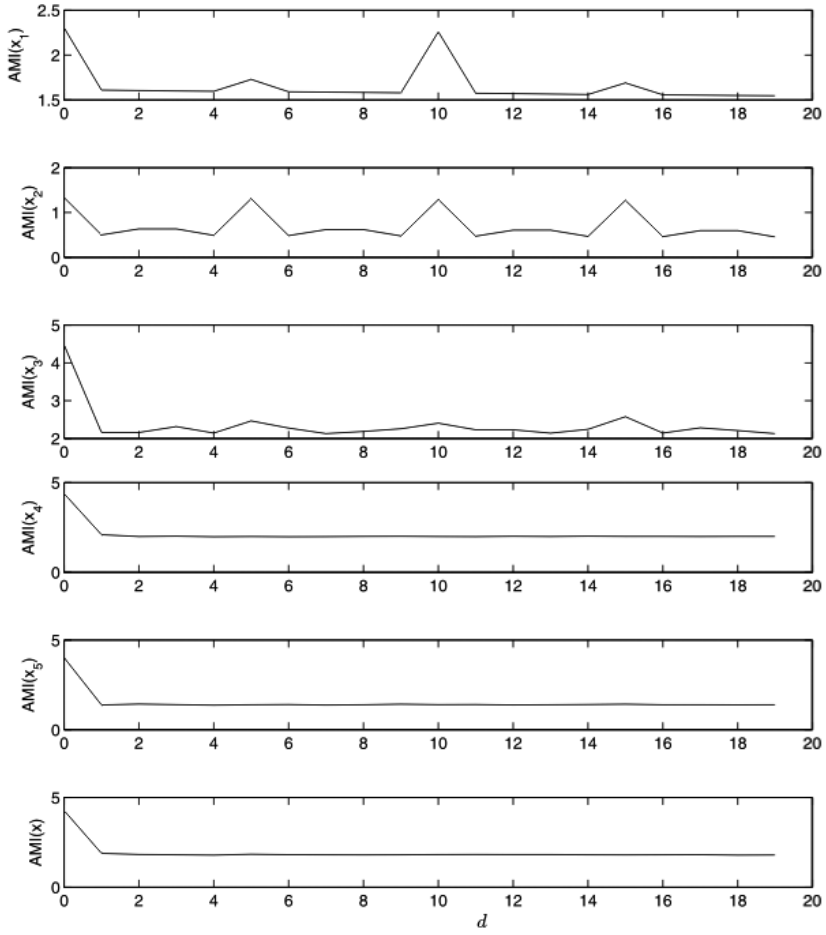


Figure 4.12. *Course of mutual function (AMI) for particular time courses* $x_1 - x_5$ *and* x

frequency of the signal. A larger degree of packing the diagonal lines (parallel to the diagonal one) corresponds to a larger frequency of the signal. In case of the quasi--periodical signal x_3 (figure 4.14(c)), except for diagonal lines also interrupted lines are showed with different component frequency which non-commensurate to the basic frequency (in case of 100Hz). Let us remember that the selection of surrounding size is of key importance here. The issue raises discussion among researchers of recurrence plots whether the analysis of comparable signals should be conducted with unique parameters m, d, ε, or not. Another plot (figure 4.14(d)) represents the signal with variable frequency so called "chirp" (x_4), which is reflected with completely different representation on the diagram. In such a case, the recurrence plot is characterized by densification of lines in a right corner, namely with the increase

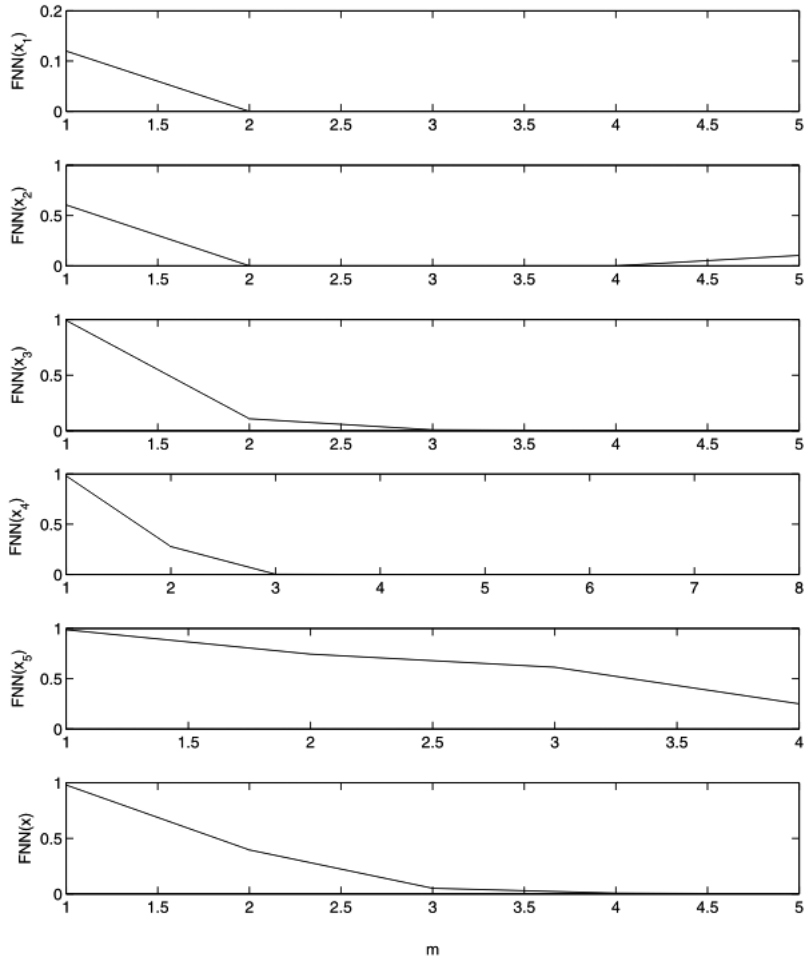


Figure 4.13. *Values of the false nearest neighbours function (FNN) depending on the embedding dimension m for particular time courses x_1–x_5 and x*

of frequency of the signal. Whereas, in case of a noise (x_5) the recurrence image is characterized with complete dispersion of recurrence points which in this case do not create the line (figure 4.14(e)). Making up all signals from x_1 to x_5 results in significant complexity of the recurrence image (figure 4.14(e)) and an increase of problem dimension in relations to the signals which are pure periodical. The noise (x_5), which was introduced as the representation of real measurement interruptions has a significant meaning here.

In the exercise, the calculation procedure can be repeated using another surrounding parameter in order to prove its influence on the appearance of the recurrence plots.

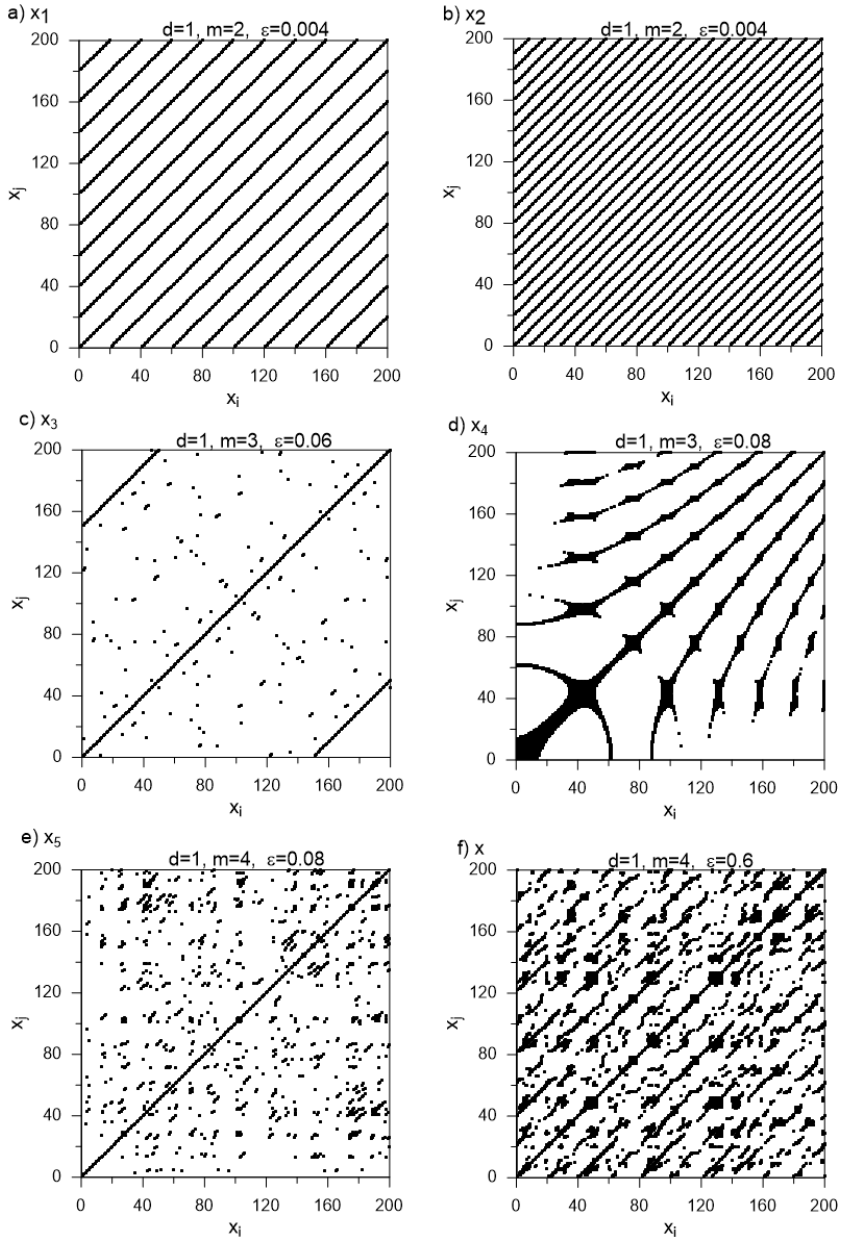


Figure 4.14. *Recurrence diagrams for certain time courses x_1–x_5 and x*

4.4. Method of multi-scale entropy MSE and CMSE

A definition of entropy can be frequently encountered in thermodynamics, where it is used for defining measurement level of non-ordering the particles. But entropy

turned out to be a useful tool to describe phenomena in many areas of science, i.a., in information theory, statistical mechanics or chaos theory. In literature one may find many definitions of entropy. In the chapter a few basic definitions were quoted on the basis of the paper [20].

In 1948 Shannon introduced a definition of entropy into the information theory. Since then, a few generalizations of the Shannon's entropy were created. In the subsequent years the definition appeared: Rényi's entropy (1965) and entropy of Havard-Charvát-Daróczy-Tsallisa (HCDT) (1970). Entropy of Rényi comes from a definition of an average mean value of Kołmogorow-Nagumo. Whereas, the HCDT entropy constitutes a certain function of Rényi's entropy. The entropies listed above in general are used for assessing the degree of non-commensurability of a discrete probability distribution. In applications, whereas, they constitute measurements to determine the degree of concentration of a discrete probability distribution.

To define the Shannon's entropies in the information theory one should require so that the searched function $\mathrm{H}_S(X) = \mathrm{H}_S\,(p(x_1), p(x_2), ..., p(x_n))$ fulfilled the following conditions:

- function H_S should be continuous towards all arguments $p(x_i)$, then small changes in probabilities correspond to small changes in entropy,
- function H_S grows monotonously along with a growth of n, if all n chance events of the X variable are equally probable ($p(x_1) = p(x_2) = ... = p(x_n) = \frac{1}{n}$),
- function H_S should be symmetric then the value of the function of entropy is a non-variable parameter of the permutation of probabilities $p(x_1)$, $p(x_2)$, $\ldots$, $p(x_n)$,
- fuction H_S should be coherent namely, when the realization of events takes place in two subsequently following stages after each other, then an initial entropy should be the weighed sum of entropies of these subsequent stages.

For a discrete random variable X with probability distribution $\{p(x_1), p(x_2), ..., p(x_n)\}$, in which i$-th$ probabilities $\{p(x_i)\}$ fulfil the conditions: $0 \leqslant p(x_i) \leqslant 1$ and $\sum\limits_{i=1}^{n} p(x_i) = 1$, exists with an exactness to the constant η function H_S, which fulfils the four postulates above. This is called Shannon's entropy:

$$\mathrm{H}_S(X) = \mathrm{H}_S(p(x_1), p(x_2), \ldots, p(x_n)) = \eta \sum_{i=1}^{n} p(x_i) \log \frac{1}{p(x_i)}. \tag{4.14}$$

The value η defines the unit of entropy [2] and for $\eta = \frac{1}{\log 2}$, the unit is bit, a dependency (4.14) accepts the form:

$$\mathrm{H}_S(X) = \sum_{i=1}^{n} p(x_i) \log_2 \frac{1}{p(x_i)}. \tag{4.15}$$

Shannon's entropy $\mathrm{H}_S(X)$ has the following properties:

- $\mathrm{H}_S(X)$ adopts 0 value, when $p(x_i) = 1$,
- $\mathrm{H}_S(X)$ is always a non-negative value,
- $\mathrm{H}_S(X)$ reaches a maximum value $H_S(X) = \log_2 \mathrm{n}$, when all $p(x_i)$ are equal to each other,
- $\mathrm{H}_S(X)$ is concave,
- $\mathrm{H}_S(X)$ fulfils property of additiveness for discrete independent random variables.

A dozen of years later, a Hungarian mathematician Alfred Rényi generalized a definition of Shannon's entropy and defined the Rényi's entropy in the form of:

$$\mathrm{H}_{R\alpha}(X) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{n} p(x_i)^{\alpha} \right), \tag{4.16}$$

where α defines the Rényi's entropy degree for $\alpha > 0$ and $\alpha \neq 1$. Entropies H_S and $\mathrm{H}_{R\alpha}$ fulfil the relations of $\mathrm{H}_{R\alpha_1} \geqslant \mathrm{H}_S \geqslant \mathrm{H}_{R\alpha_2}$, if $0 < \alpha_1 < 1$ and $\alpha_1 > 1$. Whereas when $\alpha \to 1$, then the Shannon's entropy H_S is the border of the Rényi's entropy $\mathrm{H}_{R\alpha}$.

The "α" entropy type was also independently proposed by Havrad, Charva´t and Daróczy, and then Tsallis. Entropy of Tsallis (or HCDT) was the first type of entropy in non-logarythmic form. While using properties of the logarithmic exponent functions by means of transformations which were omitted in the paper, the equation for the Tsallis's entropy was obtained (HCDT):

$$\mathrm{H}_{T\alpha}(X) = \frac{\sum\limits_{i=1}^{n} p(x_i)^{\alpha} - 1}{1-\alpha}. \tag{4.17}$$

The product of H_S and $\ln 2$ constitutes the border of Tsallis's entropy $\mathrm{H}_{T\alpha}$ for $\alpha \to 1$. One should add that entropy $\mathrm{H}_{T\alpha}$ fulfils the feature of sub-additiveness for independent random variables [11]. The sub-additiveness property differentiates the Tsallis's entropy from the entropy of Shannon and Rényi.

The quoted definitions of Shannon's, Rényi's and Tsallis's entropies prove well in mathematic linguistics as well as in the issues connected with a fractal theory. Presently, more and more often the so called multi-scale entropy is used, which is the support in the examination of dynamic processes in the mechanics and the medicine [10, 18, 1, 4, 3, 19]. A definition of multi-scale entropy constitutes an interesting measurement in the assessment of systems complexity, in analysing their behaviour to the external impulses. In the analysis of any signal, the entropy characterizes it by strengthening information and is the measure of non-order or uncertainty. In the examination of completed length of signals the "sampling entropy" is used (SampEn) [17].

For the signal of any time series $\mathbf{X}_i = \{x_1, x_2, ..., x_n\}$ with the length of N points, one may define so called m-measurement chains of vectors $\mathbf{v}(i) = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}$ and $\mathbf{v}(j) = \{x_j, x_{j+1}, \ldots, x_{j+m-1}\}$. Afterwards, one may define similarity between the vectors $\mathbf{v}(i)$ and $\mathbf{v}(j)$. The above vectors are similar to each other if the two conditions are fulfilled:

- $d(i, j) = \max\{|x(i + \kappa) - x(j + \kappa)| : 0 \leqslant \kappa \leqslant m - 1\}$ and
- $d(i, j) < r$, where r is a certain tolerance level [16].

The sampling entropy constitutes here the information of $\mathbf{v}$ vectors for one scale which is defined by the parameter $m \geqslant 2$. In order to estimate the complexity of the signal examined in a larger scale, the multi-scale entropy was introduced [3]. This entropy is not calculated by directly comparing $\mathbf{v}$, vectors, but through comparing newly created $\mathbf{y}^{(\tau)}$, vectors at so called the scale factor τ. Vectors are created from coarse-grained time series as follow (4.18)

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{i=j\tau} x_i, \quad 1 \leqslant j \leqslant N/\tau, \tag{4.18}$$

where $\tau = 1, 2, 3....$ In accordance with the above formula $y_j^{(\tau=1)} = x_i$. For the non-zero τ the analysed series $\mathbf{X}_i$ is a part of the average chain N/τ, where each one has the length of τ. The average value of calculated chains according to (4.18) constitutes now a new coarse-grained time series $\mathbf{y}^{(\tau)}$.

In the figure 4.15 a coarse-graining procedure is presented for $\tau = 2$ and $\tau = 3$. The averaging procedure introduces the smoothing of newly created $\mathbf{y}^{(\tau)}$, vectors, based on the original time series $\mathbf{X}_i$.

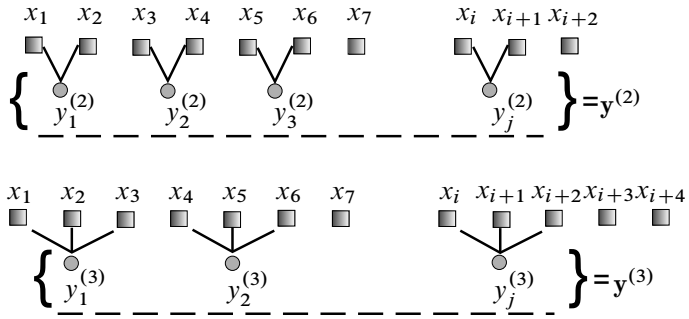


Figure 4.15. *Diagram of coarse-graining procedure for the scale factor* $\tau = 2$ *and* $\tau = 3$ *in MSE*

The multi-scale entropy for scales m and τ from the coarse-grained vector $\mathbf{y}^{(\tau)}$ is defined by equation (4.19)

$$\mathrm{MSE}(\mathbf{x}, \tau, m, r) = \mathrm{SampEn}\left(\mathbf{y}^{(\tau)}, m, r\right). \tag{4.19}$$

Whereas, $\mathrm{SampEn}\left(\mathbf{y}^{(\tau)}, m, r\right)$ in the equation (4.19) is defined by (4.20)

$$\mathrm{SampEn}\left(\mathbf{y}^{(\tau)}, m, r\right) = \ln\left(\frac{N_n}{N_d}\right). \tag{4.20}$$

values N_d and N_n are calculated from the previously prepared coarse-grained data $\mathbf{y}^{(\tau)}$ by the procedure (4.21)

$$\begin{aligned} N_d &= N_n = 1, \\ if \quad & |y^{(\tau)}(i) - y^{(\tau)}(j)| < r \ \ \& \ \ |y^{(\tau)}(i+1) - y^{(\tau)}(j+1)| < r \\ & N_n = N_n + 1, \\ if \quad & |y^{(\tau)}(i+2) - y^{(\tau)}(j+2)| < r \\ & N_d = N_d + 1. \end{aligned} \tag{4.21}$$

The result of the equation (4.19) is the probability of the occurrence in the next points of the time chain series with lengths m and $m+1$, which are similar towards each other within the tolerance r. In the literature the values of parameter m and r are provided [10] which are recommended to be used in calculations of multi-scale entropy. For the analysis of the time series, the $m = 2$, was accepted, whereas the tolerance of probability $r = 0.1\,\sigma_x$ where σ_x is a standard deviation of the original time series of the $\mathbf{X}_i$ vector. For introduced the whole scope of the scale factor parameter τ the level of tolerance r is established as constant [15]. Graphic presentation of chains similarity with different length at tolerance r (blue line) is presented in figure 4.16. While analysing the time series from the first section of two-point chains, i.e. $m = 2$, the similarity is noticeable between them (1-–2) and (22-–23). Whereas for $(m + 1) = 3$ similar chains consisting of three points are (1–2–3) and (22–23–24). In the first comparison, the number of two- and three--point sequences amounts to 2. Counting the similarities is repeated for the next two and three-component sequences (2–3) and (2–3–4) until the last ones occurring in the time series of the signal: ([N-2]–[N-1]) and ([N-2]–[N-1]–N). Obtained numbers of sequences similar to each other are summed up and the final result is the relations of a total number of the two-point sequences matching each other N_n with total number of the three-point sequences similar to each other N_d. Entropy SampEn is a natural logarithm from the product N_n/N_d in accordance with the definition (4.20).

If the next chains of the analysed signal are identical towards each other, then the result of entropy is zero and it means the lack of non-ordering the signal examined. For smaller values of tolerance r the level of entropy increases, because of then there is a smaller probability of occurring similar chains to each other.

Measurements of complexity of the signal by means of the multi-scale entropy MSE may be encumbered with some error. It depends on the accepted length of the index of scale factor τ [21]. Therefore, authors of the paper [21] introduced

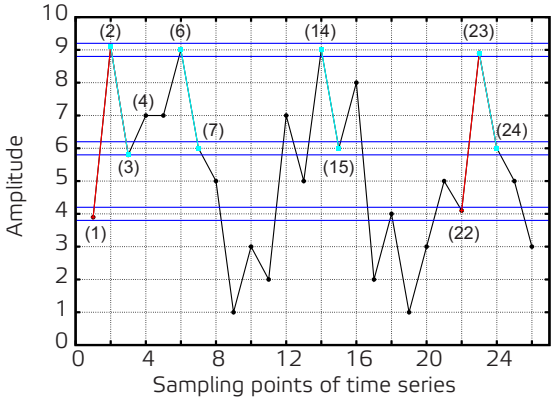


Figure 4.16. *Graphic illustration of estimating the similarities of next points of measurement*

a modified form of entropy, which eliminates the error, and they called it composited multi-scale entropy CMSE. If the calculations are conducted for the parameter $\tau \in (0–20)$, then the error is small and both results, the MSE and the CMSE are accepted as consistent. The discrepancies between the MSE and the CMSE appear at $\tau > 20$ which was presented on the view example in the figure 4.17(a) and magnified in the figure 4.17(b).

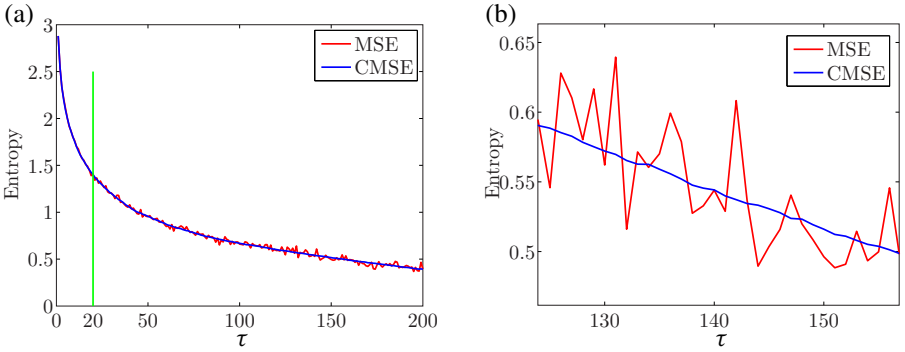


Figure 4.17. *View diagrams of multi-scale entropy MSE and composite multi--scale entropy CMSE (a) and illustration of differences at the scale factor* $\tau \gg 20$ *(b)*

In case of a composite entropy CMSE the coarse-graining procedure is presented in the figure 4.18. In comparison with the diagram presented in figure 4.15 and equation (4.18) describing the coarse-graining process only in the first grained series ($k = 1$) $y_1^{(\tau)}$, to mark CMSE the all coarse-grained time series are included

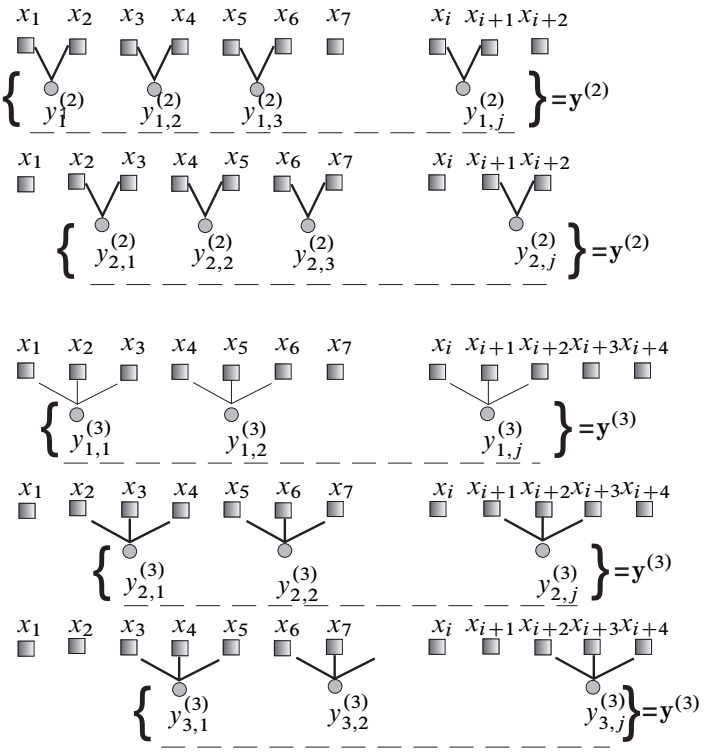


Figure 4.18. *Scheme of coarse-graining at the scale factor $\tau = 2$ and $\tau = 3$ in CMSE*

into equation (4.22)

$$y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{i=j\tau+k-1} x_i, \quad 1 \leqslant j \leqslant N/\tau, \quad 1 \leqslant k \leqslant \tau. \tag{4.22}$$

Then the formula which defines the composite multi-scale entropy CMSE takes the form of:

$$\mathrm{CMSE}(\mathbf{x}, \tau, m, r) = \frac{1}{\tau} \sum_{k=1}^{\tau} \mathrm{SampEn}\left(\mathbf{y}_k^{(\tau)}, m, r\right). \tag{4.23}$$

The calculation algorithms of both types entropies MSE and CMSE were presented in the figure 4.19 by means of block diagrams.

In order to illustrate the results of any signal analysis, calculations were made of composite multi-scale entropy CMSE for a few different signals (4.24):

$$\mathbf{X_1} = \sin(x_i) \quad \mathbf{X_2} = \sin(x_i + \Gamma_i) \quad \mathbf{X_3} = \Gamma_i \quad \mathbf{X_4} = 1 \tag{4.24}$$

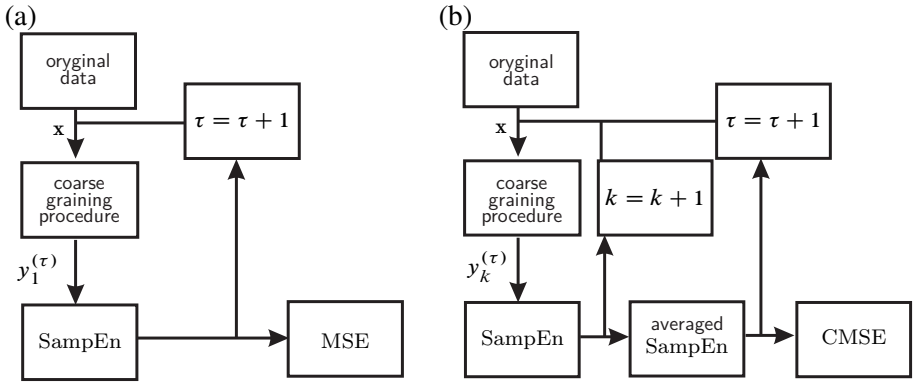


Figure 4.19. *The calculation algorithms of multi-scale entropy MSE (a) and composite multi-scale entropy CMSE (b)*

Appearing in vectors $\mathbf{X}_2$ and $\mathbf{X}_3$ the Γ function constitutes the Gaussian white noise which was generated in the Matlab environment using pseudo-random numbers with normal distribution.

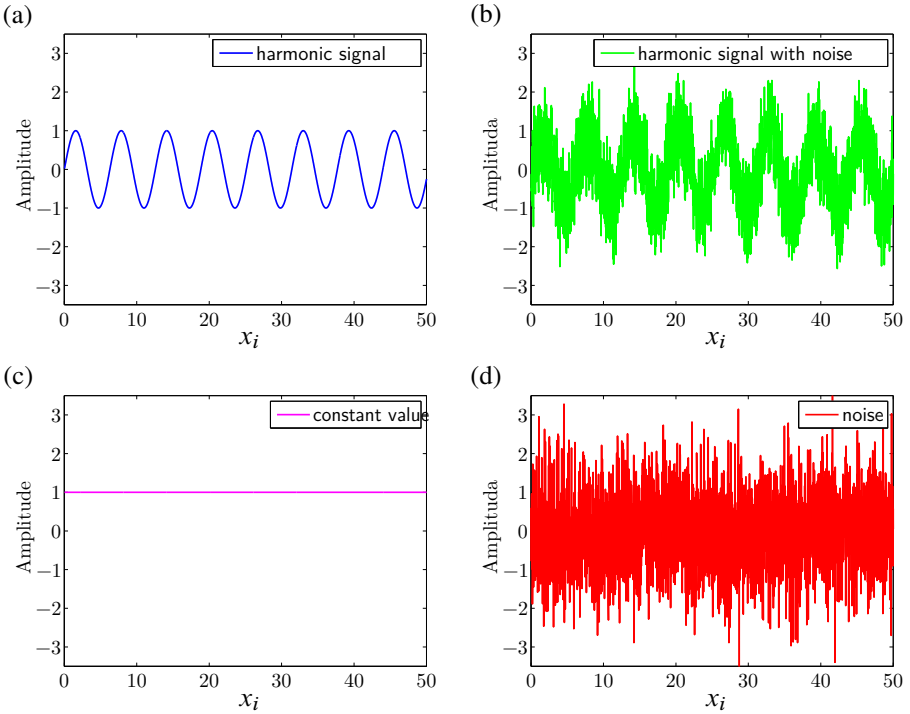


Figure 4.20. *Time series of the functions: sinus (a), sinus + noise (b), constant value (c) and Gaussian white noise (d)*

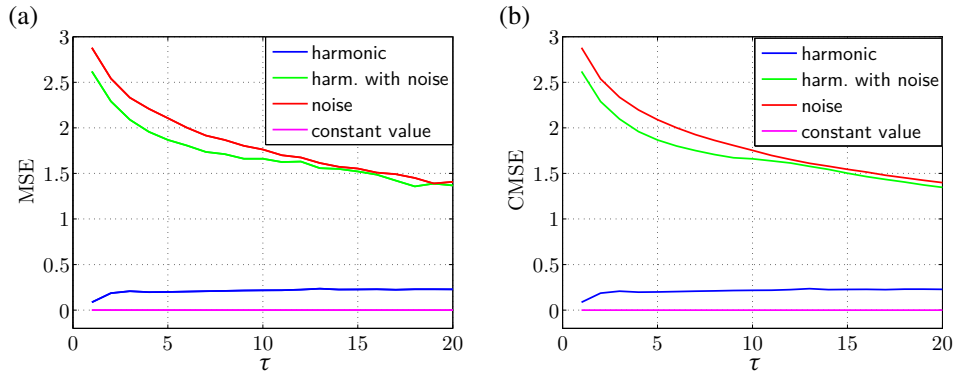


Figure 4.21. *Diagrams of multi-scale entropy MSE (a) and composite multi--scale entropy CMSE (b) via scale factor* τ*, calculated at parameters* $m = 2$*,* $r = 0.1\sigma_x$

The time series (4.24) presented in the figures 4.20 are specimen examples accepted for analysis. The goal is prove proper quantity identification of the non-ordering level by means of both types of multi-scale entropies MSE and CMSE.

In the figure 4.21(a) and 4.21(b) the diagrams present the entropy for all four vectors of signals (Eq. (4.24)). From the comparison of both entropy calculation approaches for the input data accepted, insignificant difference are visible only with reference to the signals containing noise at scale factor $\tau > 12$. But on the accepted τ level they do not appear any significant change in interpretation of the results obtained due to the level of complexity of the examined time series. Therefore, both from diagrams 4.21a and 4.21b one may obtain the same information in quantity sense with relative level of non-ordering.

Comparing the curves of the multi-scale entropy in the figure 4.21 for separate time series, one may easy observe maximum values of entropy calculated for the signal reflecting the white noise (red line). On a similar level the entropy is placed for the harmonic signal disturbed with the same type of the noise (green line). The changes observed of the entropy value at scale factor increasing are characteristic for the signals containing white noise (Gauss) [4]. Whereas the non-disturbed harmonic signal generated by function sinus, shows much larger order in comparison with the time series containing only noise. In this case, the entropy value is much smaller (blue line).

And finally, the minimum zero value of entropy from among the signals examined is the result for the time series of the constant equal value 1 (pink line). This is obvious because of such signal due to the lack of changes in time does not show any disorders.

Bibliography

[1] Borowiec M., Rysak A., Betts D., Bowen C., Kim H., Litak G. (2014): Complex response of a bistable laminated plate: Multiscale entropy analysis. *The European Physical Journal Plus* **129**: 211.

[2] Chakrabarti C., Chakrabarty I. (2005): Shannon entropy: axiomatic characterization and application. *International Journal of Mathematics and Mathematical Sciences* **17**: 2847–2854.

[3] Costa M., Goldberger A., Peng C. (2002): Multiscale entropy analysis of complex physiological time series. *Physical Review Letters* **89**: 068102.

[4] Costa M., Peng C.K., Goldberger A.L., Hausdorff J.M. (2003): Multiscale entropy analysis of human gaint dynamics. *Physica A* **330**: 53–60.

[5] Eckmann J., Kamphorst S.O., Ruelle D. (1987): Recurrence Plots of Dynamical Systems. *Europhysics Letters* **5**: 973–977.

[6] Fraser A., Swinney H. (1986): Independent coordinates for strange attractors from mutual information. *Physical Review A* **33**(2): 1134–1140.

[7] Hegger R., Kantz H., Schreiber T. (1999): Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos* **9**(2): 413–435.

[8] Iwaniec J. (2011): *Wybrane zagadnienia eksploatacyjnej identyfikacji układów nieliniowych. Rozprawy, Monografie Akademii Górniczo-Hutniczej im. Stanisława Staszica*, tom 231, Wydawnictwa AGH, Kraków.

[9] Kennel M., Brown R., Abarbanel H. (1992): Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* **45**(6): 3403–3411.

[10] Lake D., Richmann J., Griffin M., Moorman J. (2002): Sample entropy analysis of neonatal heart rate variability. *American Journal of Physiology. Heart and Circulatory Physiology* **283**: 789–797.

[11] Lavenda B. (2005): Mean entropies. *Open Systems and Information Dynamics* **12**: 289–302.

[12] Łuczko J. (2008): *Drgania regularne i chaotyczne w nieliniowych układach mechanicznych*. Monografie PK, Wydawnictwo Politechniki Krakowskiej, Kraków.

[13] Marwan N., Romano M.C., Thiel M., Kurths J. (2007): Recurrence plots for the analysis of complex systems. *Physics Reports* **438**: 237–329.

[14] Marwan N., Wessel N., Meyerfeldt U., Schirdewan A., Kurths J. (2002): Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E* **66**(2): 026702.

[15] Nikulin V., Brismar T. (2004): Comment on multiscale entropy analysis of complex physiological time series. *Physical Review Letters* **92**(8): 089803.

[16] Pan Y.H., Lin W.Y., Wang Y.H., Lee K.T. (2011): Computing multiscale entropy with orthogonal range search. *Journal of Marine Science and Technology* **19**(1): 107–113.

[17] Richmann J., Moorman J. (2000): Physiological time series analysis using approximate entropy and sample entropy. *American Journal of Physiology. Heart and Circulatory Physiology* **278**(6): 2039–2049.

[18] Starck J., Murthag F., Bijaoui A. (1998): *Image Processing and Data Analysis*. Cambridge University Press.

[19] Thuraisingham R., Gottwald G. (2006): *On multiscale entropy analysis for physiological data*. tom 366, Physica A 366.

[20] Wędrowska E. (2010): Wykorzystanie entropii Shannona i jej uogólnień do badania rozkładu prawdopodobieństwa zmiennej losowej dyskretnej. *Przegląd Statystyczny* **R. 57**(4): 39–53.

[21] Wu S.D., Wu C.W., Lin S.G., Wang C.C., Lee K.Y. (2013): Time series analysis using composite multiscale entropy. *Entropy* **15**: 1069–1084.

[22] Zbilut J.P., Webber C.L.J. (1992): Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A* **171**: 199–203.

5. Foundations of finite element method

5.1. Overview

Elementary course in mechanics of materials deals with the fundamental structural members and their simplest loading cases. Basic topics in continuum stress and strain analysis are discussed including bar tension/compression, torsion of a circular shaft and uniform beam bending. To solve the problem constitutive relations[1] relevant for the type of material and structure are to be formulated. Next, ordinary or partial differential equations of equilibrium formulated for infinitesimal structural elements are solved. Resulting mathematical expressions yield the values of desired variables e.g. stress and strain at any point within the body.

Following the above procedure leads to the exact solutions, but it's possible only for the simplest structural elements and loading cases. For problems involving complicated geometries, complex materials properties or combined loadings it is generally not possible to obtain analytical solutions.

Comparing to the mechanics of materials approach a wider class of problems can be resolved analytically by other methods. These include the use of elasticity theory and energy formulations. The theory of elasticity allows the solution of structural elements of general geometry under general loading conditions. However, closed form solutions are also limited to relatively simple cases, since the solution of elasticity problems also requires the solution of a system of partial differential equations. Use of energy formulation allows to solve structural problems, but it is best suited for specific types of structures like beams, trusses and frames, where loads or displacements at specific points may be found by means of e.g. Castigliano's theorem.

To analyse complex structures of arbitrary geometry and loading it is necessary to apply a more general and effective calculation approach. The available methods include, among others boundary element method (BEM), finite difference method (FDM) and finite element method (FEM). Nowadays, the most commonly used one is the finite element method. This is due to its versatility and simplicity to be implemented in the form of numerical algorithms and efficient computer codes.

[1] Formula defining stress-strain relation; for linear isotropic materials constitutive relation is commonly known as Hooke's law

Also hybrid methods are applied involving combination of the above techniques – in order to e.g. optimize the calculations time [13].

As it was mentioned above, the finite element method is a numerical calculation technique used for approximate solution of engineering problems. From the mathematical point of view, the task comes down to solving the boundary value problem, which is a mathematical problem of finding a certain function which describes a distribution of a dependent variable (e.g. displacements, temperature, etc.) within a given area. The function searched for should fulfill the governing differential equation everywhere within a domain of independent variables (i.e. within analysed structure) and satisfy the specific conditions on the boundary of this domain. Therefore, the boundary value problem is called sometimes a field problem, and the dependent variable searched for is the field variable. With respect to the type of physical problem being analysed, the field variables may include physical displacement, temperature, heat flux, and fluid velocity to name only a few. The domain of the solution of the boundary value problem is the analysed physical structure.

Finding the function which fulfils the governing differential equation in the whole area, often being a non-uniform one with complicated shapes etc. is an extremely difficult task. Therefore, a division of the structure for the finite number of sub-assemblies with small sizes and regular shapes is done. These sub-domains are mutually connected with each other by nodal points. Due to this approach individual fragments of the body are defined by a small number of parameters and relatively simple constitutive equations. Thus, it is possible to find local solutions to the governing equations for the structure under consideration.

The next stage of solution procedure is interpolation of the results obtained in nodal points of the domain over individual elements interior. In this way, approximate global solution is obtained, which describes the overall response of the structure. The division of the structure into a finite number of sub-domains (elements) and describing it by a finite number of state variables is called discretization of the system.

The solution of a general continuum problem by the finite element method always follows an orderly step-by-step process. With reference to the typical static structural analysis the procedure can be presented by the work-flow graph given in Figure 5.1. Individual stages are [11]:

– study of the physical problem and its mathematical model – this is an initial step of the analysis when identification of all distinctive features of the problem need to be done. This covers e.g. an acceptance of permitted simplifications, selection of possible axes and planes of symmetry which may facilitate the analysis, location of areas where stress concentrations are expected, possible material yielding due to high loads etc. An important issue is to establish additional criteria of project evaluation e.g. concerning requested accuracy of calculations and methods to review and verify the final numerical results,

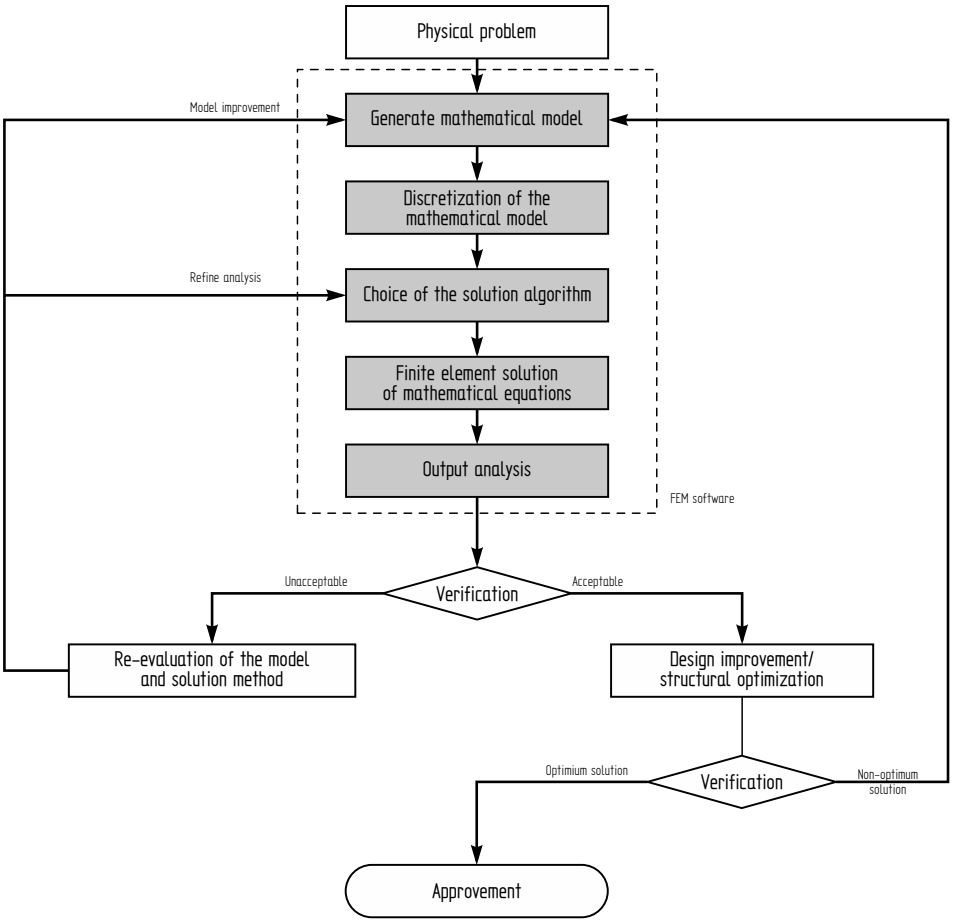


Figure 5.1. *Overview of consecutive steps of a finite element analysis*

– preparation of the numerical model of the system. This stage covers: the selection of finite elements types to be used, meshing the structure domain, definition of loads the structure is subjected to, imposing boundary conditions and possible interaction with other structures etc.,
– analysis – the software assembles the governing algebraic equations in a matrix form and computes the values of the unknown field variable(s). These values are then used to compute additional variables of request, such as reaction forces, element stresses etc. Appropriate solution techniques and numerical algorithms are to be chosen to reduce the data storage requirements and computation time,
– results assessment (postprocessing). This stage provides essential information required for the acceptance or rejection of the results. The standard treatment is data processing via graphical user interface e.g. outline of the structure in the deformed state, maps of strain and stress distribution, animations or time series

plots in case of dynamic problems. An important part of this stage is the physical interpretation of obtained results and – if possible – their comparison with experimental data,

– an optimization of the solution method. Before the final results approval additional simulations may be conducted to search for a better solution of the posed problem or to work-out more effective methods of its solving if following multiple repeated analyses are expected.

On the presented diagram, the stages run with the aid of the FEM code are indicated by a grey colour. Individual procedures (steps) may be performed by separate software packages or within one integrated environment. In view of the above discussed functions the FEM software packages may be divided into three basic groups: pre-processors i.e. the programs for data preparation, processors (solvers) and post-processors i.e. programs for outcomes evaluation.

One should point out that results of finite elements analysis are just the approximate ones, thus every time a rigorous review is necessary. Several methods of FEM output assessment are presented in the literature; readers interested in this topic are encourage to view references [4, 10, 11, 13].

In the next part of the chapter, the basic concepts related to structural modelling by the finite element approach are discussed. Mathematical formalism of the method is presented, but restricted to the necessary minimum. Two different methods for deriving the stiffness matrix for elementary finite elements are introduced. Next, we address a method of obtaining the global stiffness matrix via an element-by-element assembly procedure. Finally, the standard ways to solve the equilibrium equations are given.

The discussed topics are illustrated with simple numerical examples which facilitate the understanding of the core finite element method ideas. The given material is illustrated by two basic examples i.e. linear static analysis of a simple 2-D truss and linear static analysis of hyperstatic cantilever beam.

5.2. Linear static analysis of a truss structure

Truss structures are composed of straight elastic members subjected to axial forces only. Satisfaction of this restriction requires that all members of the truss need to be bar elements and that they are connected by pin joints such that each element is free to rotate about the joint. If the individual members are connected by rigid joints then the transverse forces and bending moments will arise. This types of structures are called frames and their proper modelling requires use of beam elements. Therefore in finite element method the bar is considered to be a one dimensional element subject to axial deformation. The analysis of rods is simpler, since the axial forces results only in longitudinal deformations of elements (tension or compression), whereas in case of beams it is necessary to include both transverse forces and bending moments as well.

The primary objective of any structural analysis is to determine internal loads and deformations at any point of the given structure as caused by external loads. Moreover, strains distribution maps, support reactions, frequencies of natural vibrations and system's stability data may be analysed. Numerous examples of different structural analyses are given in references [3, 5, 7, 9].

5.2.1. Assumptions and limitations of linear analysis

Before starting the analysis it is necessary to formulate assumptions concerning the mathematical model of the structure and further derivation steps. While limiting the considerations to the static linear case, one assumes the following:

- material of the structure is linearly elastic (i.e. the Hooke's law is fulfilled), so the displacements are expressed as linear functions of applied loads. Therefore, the superposition rule is valid and may be used to derive the overall equilibrium equations. Moreover, it is presumed that the modelled material is homogenous and fully isotropic,
- deformations of the system caused by the action of external loads are significantly smaller than the characteristic dimensions of individual elements of the system. That is to say displacements of the truss nodal points are smaller than dimensions of rods cross-sections, so the change in geometry of the structure as a result of its deformations does not affect the loading conditions,
- external loading forces are assumed to be imposed quasi-statically, so they do not cause any dynamic effects; thereby the structure remains in the static equilibrium state.

In addition to the above postulated assumptions concerning the linear nature of the given system, it is also presumed that the structure is manufactured perfectly – i.e. no dimensional imperfections of individual members are present. In case of statically indeterminate structures (hyperstatic ones) this assumption is especially important due to the possibility of initial assembly stress to occur. For the above reason, it is also assumed that the ambient temperature remains constant and does not impact the material behaviour.

The primary characteristics of any individual finite element, as well as the whole structure are embodied in the stiffness matrix. For a standard structural finite element, the stiffness matrix contains the geometric data and material behaviour information that indicates the resistance of the element to deformation when subjected to loading. The global stiffness matrix representing the whole structure is assembled from individual stiffness components. Next the system equilibrium equations are obtained. The solution of this set are displacements of nodal points.

In order to derive the expressions for the individual elements stiffness matrices the principle of virtual work or residual methods may be used [3]. Only in the simplest cases it is possible to formulate these relations by a direct stiffness approach. This method will be presented at first; for the comparison the requested relations

are derived by means of minimum potential energy method too. In the subsequent part of the chapter dealing with flexural (beam) elements energy approach is used.

5.2.2. Uniaxial bar element

Stiffness matrix for a rod element in local coordinate system

Let us consider a simple rod with a constant cross section, made of elastic uniform and isotropic material – see Figure 5.2. We define the nodes *1* and *2* at the end points of the bar, where the external axial forces F_1 and F_2 are applied. We assume these are the only loadings of the bar and the element stays in equilibrium. For convenience a coordinate system xy with its origin placed at the left end of the bar is introduced ($12 = Ox$). This is the element (local) coordinate system of reference.

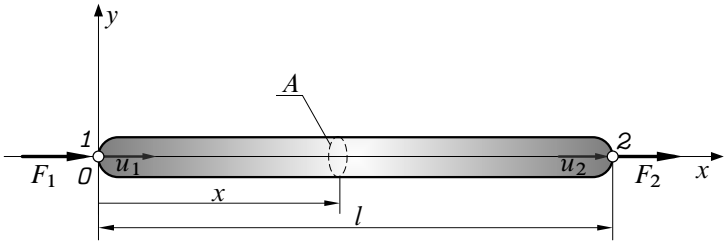


Figure 5.2. *Forces and knot shifts of the simple rod*

As a result of acting forces both nodes *1* and *2* are shifted along the rod axis by u_1 and u_2 respectively. If the rod is infinitely stiff, both displacements are equal $u_1 = u_2$, whereas in case of the deformable unit $u_1 \neq u_2$. Therefore, these two displacements uniquely define the position of the rod after deformation, so the bar element under discussion has two degrees of freedom (u_1 and u_2).

Static equation of equilibrium takes the form

$$\sum_{i=1}^{i=2} F_{ix} = F_1 + F_2 = 0 \tag{5.1}$$

and the net axial deformation of the element Δl is

$$\Delta l = u_2 - u_1, \tag{5.2}$$

According to the definition the axial strain in the element is given by

$$\varepsilon = \varepsilon_x = \frac{u_2 - u_1}{l}. \tag{5.3}$$

Following the accepted assumptions on linear nature of the considered system, the stress and strain variables are related by the Hooke's law

$$\sigma = E\varepsilon, \tag{5.4}$$

where E is Young's modulus of the material.

By definition stress resultants at nodal points `A` and `B` are:

$$\begin{aligned} &\text{node } 1\text{:} \quad \sigma = -\frac{F_1}{A} \qquad \leftarrow \text{compression (force towards the specimen)},\\ &\text{node } 2\text{:} \quad \sigma = \frac{F_2}{A} \qquad \leftarrow \text{tension} \end{aligned} \tag{5.5}$$

where A is the cross section of the rod (5.1).

Putting relations (5.2), (5.3) and (5.4) into (5.5) formula we obtain a set of equilibrium conditions

$$\begin{aligned} F_1 &= -\sigma A \ = -E\varepsilon A \ = -\frac{EA}{l}(u_2 - u_1) = \frac{EA}{l}u_1 - \frac{EA}{l}u_2,\\ F_2 &= \sigma A \ = E\varepsilon A \ = \frac{EA}{l}(u_2 - u_1) = -\frac{EA}{l}u_1 + \frac{EA}{l}u_2. \end{aligned} \tag{5.6}$$

Relations (5.6) constitute the equilibrium condition as a system of two linear equations with two unknown values (u_1 and u_2), which may be written in the matrix form

$$\begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix} = \frac{EA}{l}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \cdot \begin{Bmatrix} u_2 \\ u_1 \end{Bmatrix} \quad \Leftrightarrow \quad \mathbf{F} = \mathbf{K} \cdot \mathbf{u}, \tag{5.7}$$

where $\mathbf{K}$ is square, symmetric and positive definite stiffness matrix of the rod element in its local coordinate system. The symmetry property of the matrix results directly from the Maxwell-Betti's reciprocal work theorem.[2] Moreover, the matrix $\mathbf{K}$ is a singular one ($\det \mathbf{K} = 0$). Its size 2×2 corresponds to the fact that the element exhibits two nodal displacements (or degrees of freedom).

The presented above method of deriving the stiffness matrix by means of the Newton's law is effective only in the simplest cases, e.g. rod elements. As it was stated in the Overview section, much more powerful approach is to use calculus of variations and minimum potential energy theorem. In order to compare these two different methods derivation of matrix $\mathbf{K}$ for the truss element is repeated below.

Let's define the longitudinal displacement of an arbitrary cross-section of the rod (Figure 5.2) as $u = u(x)$. This displacement is a function of an independent variable x designating the position of the cross-section under consideration. Thus, we have the continuous field variable $u(x)$, which might be expressed approximately as an algebraic combination of displacements at nodal points `1` and `2`[3]

$$u(x) = N_1(x) \cdot u_1 + N_2(x) \cdot u_2. \tag{5.8}$$

[2] Each displacement is related to the other by the same physical phenomenon. In other words, if any load F is applied at node i resulting a deflection d at node j, then if same load F is applied at j, it will give the same deflection d at node i.

[3] Substitution of continuous field variable u by its values at specific points (nodes) is a part of discretization procedure

Functions $N_1(x)$ and $N_2(x)$ present in the above relation are referred to as shape functions (also known as blending or interpolating functions). One observes they are weighting factors for nodal displacements u_1 and u_2. In a general case, the expression (5.8) for $u(x)$ is an approximate one. However, for the specific case of uniform axial element discussed herein this formula is a strict one since within these specimens strain is distributed linearly.

Taking into account nodal points *1*, *2* and their displacements

$$u(x=0) = u_1, \qquad u(x=l) = u_2, \tag{5.9}$$

one gets the conditions to be identically satisfied by both unknown shape functions. Inserting (5.8) into the relations (5.9) one obtains

$$\begin{aligned} N_1(0) &= 1, \qquad N_2(0) = 0, \\ N_1(l) &= 0, \qquad N_2(l) = 1. \end{aligned} \tag{5.10}$$

It is required that the displacement expression, equation (5.8), satisfy the end conditions identically, since the nodes will be the connection points between neighbouring elements and the displacement continuity condition is enforced at those points. Since we have two conditions that must be satisfied by each of two one-dimensional functions, the simplest form for the interpolation functions are binomial expressions

$$N_1(x) = a_1 x + b_1 \qquad N_2(x) = a_2 x + b_2. \tag{5.11}$$

Bearing in mind relations (5.10) we obtain:

$$\begin{aligned} a_1 &= -\frac{1}{l} \qquad b_1 = 1, \\ a_2 &= \frac{1}{l} \qquad b_2 = 0. \end{aligned} \tag{5.12}$$

Thus, the final displacement field is described by the following formula

$$u(x) = \left(1 - \frac{x}{l}\right) \cdot u_1 + \frac{x}{l} \cdot u_2 \tag{5.13}$$

or in the contracted matrix form

$$u(x) = \begin{Bmatrix} N_1(x) & N_2(x) \end{Bmatrix} \cdot \begin{Bmatrix} u_A \\ u_B \end{Bmatrix} = \mathbf{N} \cdot \mathbf{u}. \tag{5.14}$$

According to the axial strain definition

$$\varepsilon = \frac{du(x)}{dx}, \tag{5.15}$$

and field variable approximation (5.13) the ε is equal to

$$\varepsilon = \frac{u_2 - u_1}{l}. \tag{5.16}$$

The above result is fully consistent with the previously derived expression (5.3).

Elastic potential energy of the rod in axial tension is

$$\mathcal{U} = \frac{1}{2}\int_l E(x)A(x)\varepsilon^2(x)dx, \tag{5.17}$$

thus after including the strain definition (5.16) and the fact that the considered element is uniform and constant cross-section the following expression is obtained

$$\mathcal{U} = \frac{1}{2}EA\,\frac{(u_2 - u_1)^2}{l}. \tag{5.18}$$

The first Castigliano's theorem states that for an elastic system in equilibrium, the partial derivative of total strain energy with respect to deflection at a certain point is equal to the force applied at that point projected on the direction of this deflection. Thus, applying the theorem to both nodes one obtains

$$\begin{aligned} \frac{\partial \mathcal{U}}{\partial u_1} &= F_1 = -\frac{AE}{l}\cdot(u_2 - u_1), \\ \frac{\partial \mathcal{U}}{\partial u_2} &= F_2 = \frac{AE}{l}\cdot(u_2 - u_1). \end{aligned} \tag{5.19}$$

The above may be written in the matrix form:

$$\frac{EA}{l}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\cdot\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix}. \tag{5.20}$$

The final result is the element equilibrium equation. It is exactly the same formula as (5.7) derived previously by direct stiffness approach.

Element transformation. Stiffness matrix in global system of reference

The methods presented above are capable of finding relations between external forces and nodal displacements – (5.7)/(5.20). These are expressed in the local system of reference, which conveniently represents an individual element. In case of constructions made of multiple elements that are located at different positions it is recommended to set a global coordinate system to represent the whole structure. Thus, the equilibrium equations for all the individual elements may be expressed in the same system of reference. The mathematical conversion from the local to the global reference system is done by means of a transformation matrix.

Before deriving the requested relations it is necessary to modify the uni-axial equilibrium condition (5.7). According to initially posed assumptions there are no transverse forces and displacements (i.e. along the Oy axis), so the stiffness matrix $\mathbf{K}$ may be extended by two additional zero rows and columns; similar zero rows are added to both column vectors. Finally, the equation (5.7) will take its new and more general two dimensional reference system form

$$\begin{Bmatrix} F_1 \\ V_1 \\ F_2 \\ V_2 \end{Bmatrix} = \frac{EA}{l} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix}, \tag{5.21}$$

where $\mathbf{F} = \{F_1, V_1 = 0, F_2, V_2 = 0\}^{\mathsf{T}}$ and $\mathbf{u} = \{u_1, v_1 = 0, u_2, v_2 = 0\}^{\mathsf{T}}$ are extended column vectors of nodal forces and nodal displacements respectively.

To develop the transformation formula that will subsequently be used to derive the global stiffness matrix let us consider the rod analysed in a previous paragraph now given in its new position. This orientation is defined by a directed angle θ measured from the $O\overline{x}$ axis of rectangular external system of coordinates to the local Ox axis – see Figure 5.3 (counterclockwise rotation is considered to be positive). The system of coordinates $\overline{x}O\overline{y}$ not directly related to the element is called the global system of reference.[4]

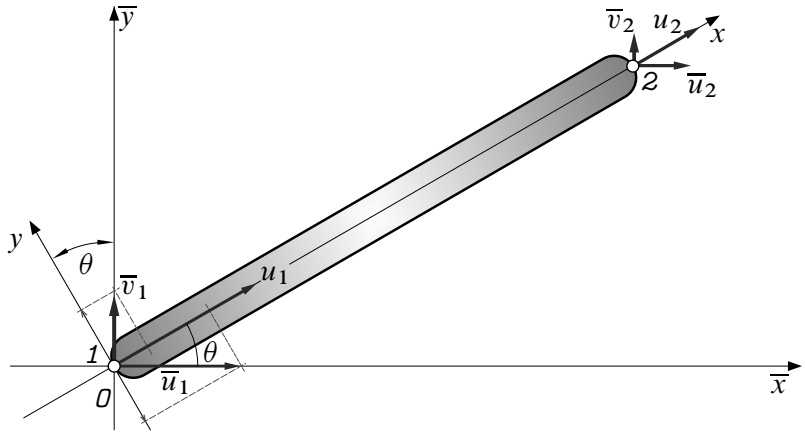


Figure 5.3. *Nodal displacements in local and global systems of reference*

[4] To distinguish local and global variables it is assumed, the quantities related to the global system of reference are denoted by an overbar symbol.

Let us consider the node *1* (Figure 5.3): from the sum of projections of displacements $\overline{u}_1$ and $\overline{v}_1$ (variable v denotes transversal displacement) onto the directions of local system of coordinates xOy we will obtain

$$\begin{aligned} u_1 &= \overline{u}_1 \cos\theta + \overline{v}_1 \sin\theta \\ v_1 &= -\overline{u}_1 \sin\theta + \overline{v}_1 \cos\theta, \end{aligned} \tag{5.22}$$

Which can be written in the matrix form as follows

$$\mathbf{u}_1 = \boldsymbol{\Theta} \cdot \overline{\mathbf{u}}_1. \tag{5.23}$$

The matrix present in the above relation

$$\boldsymbol{\Theta} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{5.24}$$

is the transformation (or rotation) matrix of any vector value from the global system of reference (the right hand side of the formula (5.23)) to the local system. So, the vector variable $\overline{\mathbf{u}}_1 = \{\overline{u}_1, \overline{v}_1\}^\mathsf{T}$ represents displacements of node *1* in the global system of coordinates.

Analysing the properties of the transformation matrix one may show that this is an orthogonal matrix – thus $\boldsymbol{\Theta}^{-1} = \boldsymbol{\Theta}^\mathsf{T}$, i.e. the inverse matrix is equal to its transpose.

Repeating the given above considerations for the second node, as well as for forces vectors acting in both nodes the following relations are obtained

$$\mathbf{u}_2 = \boldsymbol{\Theta} \cdot \overline{\mathbf{u}}_2 \tag{5.25}$$

and

$$\mathbf{F}_1 = \boldsymbol{\Theta} \cdot \overline{\mathbf{F}}_1, \qquad \mathbf{F}_2 = \boldsymbol{\Theta} \cdot \overline{\mathbf{F}}_2. \tag{5.26}$$

This is due to the fact the displacements as well as forces are vectorial variables, so they transform in the same manner.

Combining equations (5.23) and (5.25) we can obtain one common relation describing the transformation of displacements of the element. In order to do this the expansion of transformation matrix to the size of four is necessary, so that the order is consistent with the total number of global coordinates used

$$\mathbf{u} = \overline{\boldsymbol{\Theta}} \cdot \overline{\mathbf{u}}, \tag{5.27}$$

where the matrix $\overline{\boldsymbol{\Theta}}$ is created by juxtaposing two $\boldsymbol{\Theta}$ matrices

$$\overline{\boldsymbol{\Theta}} = \begin{bmatrix} \boldsymbol{\Theta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & \cos\theta & \sin\theta \\ 0 & 0 & -\sin\theta & \cos\theta \end{bmatrix} \tag{5.28}$$

The matrix $\overline{\boldsymbol{\Theta}}$ is transformation (rotation) matrix of the truss element from the global coordinate system to the local one. Relation similar to (5.27) is applicable for forces

$$\mathbf{F} = \overline{\boldsymbol{\Theta}} \cdot \overline{\mathbf{F}}. \tag{5.29}$$

Thus, the matrix written in (5.28) is a universal transformation matrix, which is capable of conversion of any vector variables between global and local coordinate systems.

Given the above derived rotation matrix the equation of equilibrium (5.21) formulated in the previous subsection may be converted from the local system of reference to the global one. For this effect, one should replace the vectors $\mathbf{u}$ and $\mathbf{F}$ by the appropriate formulas (5.27) and (5.29). As a result we will obtain

$$\overline{\boldsymbol{\Theta}} \cdot \overline{\mathbf{F}} = \mathbf{K} \cdot \overline{\boldsymbol{\Theta}} \cdot \overline{\mathbf{u}}. \tag{5.30}$$

Pre-multiplying the last relation by the inverse of transformation matrix we obtain

$$\overline{\boldsymbol{\Theta}}^{-1} \cdot \overline{\boldsymbol{\Theta}} \cdot \overline{\mathbf{F}} = \overline{\boldsymbol{\Theta}}^{-1} \cdot \mathbf{K} \cdot \overline{\boldsymbol{\Theta}} \cdot \overline{\mathbf{u}}, \tag{5.31}$$

which by virtue of the orthogonality condition (see page 97) is equal to

$$\overline{\mathbf{F}} = \overline{\boldsymbol{\Theta}}^{\mathsf{T}} \cdot \mathbf{K} \cdot \overline{\boldsymbol{\Theta}} \cdot \overline{\mathbf{u}}. \tag{5.32}$$

Introducing new definition

$$\overline{\mathbf{K}} = \overline{\boldsymbol{\Theta}}^{\mathsf{T}} \cdot \mathbf{K} \cdot \overline{\boldsymbol{\Theta}} \tag{5.33}$$

we obtain finally the equilibrium equation of the truss element in a global system of reference

$$\overline{\mathbf{F}} = \overline{\mathbf{K}} \cdot \overline{\mathbf{u}}. \tag{5.34}$$

Apparent in the above formula the $\overline{\mathbf{K}}$ matrix is the element stiffness matrix in the global system of reference.

The subsequent steps of derivation presented in this section need to be repeated for each element of the considered truss.

5.2.3. Global stiffness matrix of the structure

The stiffness matrix as introduced in the previous paragraph states the mathematical relation between element nodal displacements and imposed nodal forces as referred to global system of reference. Proceeding in the similar manner one may formulate the stiffness matrix of the whole structure to relate the overall displacements and overall loadings. To illustrate the procedure let us consider the three-element truss presented in the Figure 5.4.

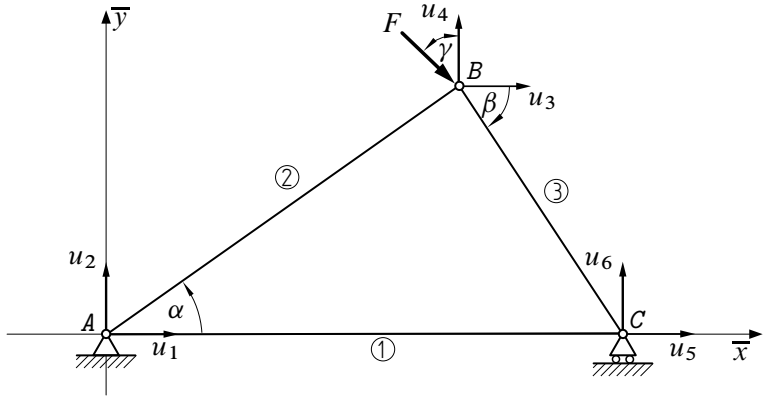


Figure 5.4. *Exemplary three-element truss structure and nodal displacements numbering scheme*

Following the given definitions (5.21) one may write down stiffness matrices of individual rods in their local reference frames:

$$\mathbf{K}_1 = \frac{EA_1}{l_1}\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{K}_2 = \frac{EA_2}{l_2}\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathbf{K}_3 = \frac{EA_3}{l_3}\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5.35}$$

Individual transformation matrices $\overline{\mathbf{\Theta}}_i$, $i = 1, 2, 3$ result from mutual orientations of reference frames defined by the angles present between the $O\overline{x}$ and Ox_i axes. Following the notation given in Figure 5.3 and eqn. (5.28) the appropriate angles are: $\theta_1 = 0°$, $\theta_2 = \alpha$ and $\theta_3 = -\beta$. Thus we obtain

$$\overline{\mathbf{\Theta}}_1 = \begin{matrix} \\ 1 \\ 2 \\ 5 \\ 6 \end{matrix}\!\!\begin{matrix} \begin{matrix} 1 & 2 & 5 & 6 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \qquad \overline{\mathbf{\Theta}}_2 = \begin{matrix} \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix}\!\!\begin{matrix} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} \cos\alpha & \sin\alpha & 0 & 0 \\ -\sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & \cos\alpha & \sin\alpha \\ 0 & 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \end{matrix} \tag{5.36}$$

$$\overline{\mathbf{\Theta}}_3 = \begin{bmatrix} \overset{3}{\cos\beta} & \overset{4}{-\sin\beta} & \overset{5}{0} & \overset{6}{0} \\ \sin\beta & \cos\beta & 0 & 0 \\ 0 & 0 & \cos\beta & -\sin\beta \\ 0 & 0 & \sin\beta & \cos\beta \end{bmatrix}_{\begin{matrix}3\\4\\5\\6\end{matrix}}.$$

In given above matrices additional denotation of the columns and rows has been introduced. The small digits correspond to the global degrees of freedom the considered element contributes to due to its deformation. Numbering order stays in accordance with the scheme given in the Figure 5.4.

Based on (5.34) the stiffnesses of all elements are set in one, common global system of coordinates as follows:

$$\overline{\mathbf{K}}_1 = \overline{\mathbf{\Theta}}_1^{\mathrm{T}} \cdot \mathbf{K}_1 \cdot \overline{\mathbf{\Theta}}_1 = \frac{EA_1}{l_1} \begin{bmatrix} \overset{1}{1} & \overset{2}{0} & \overset{5}{-1} & \overset{6}{0} \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{\begin{matrix}1\\2\\5\\6\end{matrix}} \tag{5.37}$$

$$\overline{\mathbf{K}}_2 = \overline{\mathbf{\Theta}}_2^{\mathrm{T}} \cdot \mathbf{K}_2 \cdot \overline{\mathbf{\Theta}}_2 =$$

$$= \frac{EA_2}{l_2} \begin{bmatrix} \overset{1}{\cos^2\alpha} & \overset{2}{\sin\alpha\cos\alpha} & \overset{3}{-\cos^2\alpha} & \overset{4}{-\sin\alpha\cos\alpha} \\ \sin\alpha\cos\alpha & \sin^2\alpha & -\sin\alpha\cos\alpha & -\sin^2\alpha \\ -\cos^2\alpha & -\sin\alpha\cos\alpha & \cos^2\alpha & \sin\alpha\cos\alpha \\ -\sin\alpha\cos\alpha & -\sin^2\alpha & \sin\alpha\cos\alpha & \sin^2\alpha \end{bmatrix}_{\begin{matrix}1\\2\\3\\4\end{matrix}}$$

$$\overline{\mathbf{K}}_3 = \overline{\mathbf{\Theta}}_3^{\mathrm{T}} \cdot \mathbf{K}_3 \cdot \overline{\mathbf{\Theta}}_3 =$$

$$= \frac{EA_3}{l_3} \begin{bmatrix} \overset{3}{\cos^2\beta} & \overset{4}{-\sin\beta\cos\beta} & \overset{5}{-\cos^2\beta} & \overset{6}{\sin\beta\cos\beta} \\ -\sin\beta\cos\beta & \sin^2\beta & \sin\beta\cos\beta & -\sin^2\beta \\ -\cos^2\beta & \sin\beta\cos\beta & \cos^2\beta & -\sin\beta\cos\beta \\ \sin\beta\cos\beta & -\sin^2\beta & -\sin\beta\cos\beta & \sin^2\beta \end{bmatrix}_{\begin{matrix}3\\4\\5\\6\end{matrix}}$$

whereas, additional numbering of global degrees of freedom has been repeated.

Given the notation in the Figure 5.4 one may state the equilibrium equations for each element as expressed in the global system of coordinates

$$\begin{aligned}
&\text{element 1:} \quad \{\overline{F}_1 \ \overline{F}_2 \ \overline{F}_5 \ \overline{F}_6\}^{\mathsf{T}} = \overline{\mathbf{K}}_1 \cdot \{\overline{u}_1 \ \overline{u}_2 \ \overline{u}_5 \ \overline{u}_6\}^{\mathsf{T}} \\
&\text{element 2:} \quad \{\overline{F}_1 \ \overline{F}_2 \ \overline{F}_3 \ \overline{F}_4\}^{\mathsf{T}} = \overline{\mathbf{K}}_2 \cdot \{\overline{u}_1 \ \overline{u}_2 \ \overline{u}_3 \ \overline{u}_4\}^{\mathsf{T}} \\
&\text{element 3:} \quad \{\overline{F}_3 \ \overline{F}_4 \ \overline{F}_5 \ \overline{F}_6\}^{\mathsf{T}} = \overline{\mathbf{K}}_3 \cdot \{\overline{u}_3 \ \overline{u}_4 \ \overline{u}_5 \ \overline{u}_6\}^{\mathsf{T}}.
\end{aligned} \tag{5.38}$$

Based on $\overline{\mathbf{K}}_i$ definitions given in (5.37) the relations written above may be combined in one equation of equilibrium valid for the whole structure – see (5.39) next page. The column vectors $\overline{\mathbf{F}}$ and $\overline{\mathbf{u}}$ contain all nodal forces and displacements as expressed in global coordinate frame. The square matrix appearing therein is called global stiffness matrix. It is obtained by putting individual components of stiffness matrices $\overline{\mathbf{K}}_i$ of each element $i = 1, 2, 3$ (see (5.37)) in proper cells. The place of input is defined row and column number proper to the corresponding degree of freedom. The approved numbering scheme of individual nodal displacements is described in the Figure 5.4. To some extent the (5.39) formula is similar to equation (5.34) where the last one is valid for individual element only.

Assembly of the systems stiffness matrix is called global matrix aggregation. Except for the method shown in the present paragraph matrix $\overline{\mathbf{K}}$ may also be derived by means of other methods – i.a. using the linkage matrices approach. However, the size of these matrices makes the operation quite troublesome, thus this procedure is not commonly used in numerical implementations of finite element method [3], [8], [13].

One should emphasize that global equilibrium equation as written in (5.39) has no unique solution. This results from the fact the global stiffness matrix, similar to the individual stiffness matrices of each element, is singular – $\det \overline{\mathbf{K}} = 0$. A unique solution can be obtained only taking into account the constraints imposed on the system displacements by the support conditions that preclude rigid body motion.

5.2.4. Boundary conditions and reduced global stiffness matrix of the structure

The form of the generalized stiffness matrix introduced in the previous paragraph may be used for the description of any truss structure having the geometry similar to the one presented in the Figure 5.4, irrespective of the support conditions. The form of this matrix strictly corresponding to the specific task under consideration is obtained after introducing the boundary conditions representing the support of the system.

$$\begin{Bmatrix} \overline{F}_1 \\ \overline{F}_2 \\ \overline{F}_3 \\ \overline{F}_4 \\ \overline{F}_5 \\ \overline{F}_6 \end{Bmatrix} = E \begin{bmatrix} \frac{A_1}{l_1} + \frac{A_2}{l_2}\mathrm{c}^2\alpha & \frac{A_2}{l_2}\mathrm{s}\alpha\,\mathrm{c}\alpha & -\frac{A_2}{l_2}\mathrm{c}^2\alpha & -\frac{A_2}{l_2}\mathrm{s}\alpha\,\mathrm{c}\alpha & -\frac{A_1}{l_1} & 0 \\ & \frac{A_2}{l_2}\mathrm{s}^2\alpha & -\frac{A_2}{l_2}\mathrm{s}\alpha\,\mathrm{c}\alpha & -\frac{A_2}{l_2}\mathrm{s}^2\alpha & 0 & 0 \\ & & \frac{A_2}{l_2}\mathrm{c}^2\alpha + \frac{A_3}{l_3}\mathrm{c}^2\beta & \frac{A_2}{l_2}\mathrm{s}\alpha\,\mathrm{c}\alpha - \frac{A_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta & -\frac{A_3}{l_3}\mathrm{c}^2\beta & -\frac{A_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta \\ & & & \frac{A_2}{l_2}\mathrm{s}^2\alpha + \frac{A_3}{l_3}\mathrm{s}^2\beta & \frac{A_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta & -\frac{A_3}{l_3}\mathrm{s}^2\beta \\ & & & & \frac{A_1}{l_1} + \frac{A_3}{l_3}\mathrm{c}^2\beta & -\frac{A_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta \\ & & \text{SYMMETRY} & & & \frac{A_3}{l_3}\mathrm{s}^2\beta \end{bmatrix} \cdot \begin{Bmatrix} \overline{u}_1 \\ \overline{u}_2 \\ \overline{u}_3 \\ \overline{u}_4 \\ \overline{u}_5 \\ \overline{u}_6 \end{Bmatrix} \tag{5.39}$$

To simplify the provision, additional marks were accepted $\mathrm{s}\alpha = \sin\alpha$, $\mathrm{c}\alpha = \cos\alpha$, $\mathrm{s}\beta = \sin\beta$ oraz $\mathrm{c}\beta = \cos\beta$.

Turning to the Figure 5.4, the structure is attached to the ground in points A and C by pin and roller supports respectively. Thus in point A all degrees of freedom are prohibited, while in point C only vertical, while horizontal displacement is allowed. In accordance with the given notation the excluded global degrees of freedom are $\overline{u}_1 = \overline{u}_2 = \overline{u}_6 = 0$. Therefore, one can remove rows and columns having numbers one, two and six from the global stiffness matrix $\overline{\mathbf{K}}$ (5.39). This new matrix $\overline{\mathbf{K}}_{ks}$ represents the behaviour of the specific system considered in this problem. The global stiffness matrix with the constraints applied as dictated by the boundary conditions is called reduced global stiffness matrix

$$\overline{\mathbf{K}}_{ks} = E \begin{bmatrix} \frac{A_2}{l_2}\mathrm{c}^2\alpha + \frac{A_3}{l_3}\mathrm{c}^2\beta & \frac{A_2}{l_2}\mathrm{s}\alpha\,\mathrm{c}\alpha - \frac{A_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta & -\frac{A_3}{l_3}\mathrm{c}^2\beta \\ & \frac{A_2}{l_2}\mathrm{s}^2\alpha + \frac{A_3}{l_3}\mathrm{s}^2\beta & \frac{A_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta \\ \text{SYMMETRY} & & \frac{A_1}{l_1} + \frac{A_3}{l_3}\mathrm{c}^2\beta \end{bmatrix}. \quad (5.40)$$

5.2.5. Truss equilibrium equation

As the reduced global stiffness matrix is defined one can set the equilibrium condition (5.39) of the truss structure in its final form

$$\begin{Bmatrix} \overline{F}_3 \\ \overline{F}_4 \\ \overline{F}_5 \end{Bmatrix} = E \begin{bmatrix} \frac{F_2}{l_2}\mathrm{c}^2\alpha + \frac{F_3}{l_3}\mathrm{c}^2\beta & \frac{F_2}{l_2}\mathrm{s}\alpha\,\mathrm{c}\alpha - \frac{F_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta & -\frac{F_3}{l_3}\mathrm{c}^2\beta \\ & \frac{F_2}{l_2}\mathrm{s}^2\alpha + \frac{F_3}{l_3}\mathrm{s}^2\beta & \frac{F_3}{l_3}\mathrm{s}\beta\,\mathrm{c}\beta \\ \text{SYMETRIA} & & \frac{F_1}{l_1} + \frac{F_3}{l_3}\mathrm{c}^2\beta \end{bmatrix} \cdot \begin{Bmatrix} \overline{u}_3 \\ \overline{u}_4 \\ \overline{u}_5 \end{Bmatrix} \quad (5.41)$$

The left-hand vector $\overline{\mathbf{F}} = \{\overline{F}_3, \overline{F}_4, \overline{F}_5\}^{\mathsf{T}}$ of external nodal loads is to be determined directly from the Figure 5.4. One immediately finds $\overline{F}_3 = F\sin\gamma$, $\overline{F}_4 = -F\cos\gamma$ and $\overline{F}_5 = 0$.

From the mathematical point of view, the above formula constitutes the system of $n = 3$ linear equations with $n = 3$ unknowns. As it was mentioned in the paragraph 5.2.2 the stiffness matrix is a square, positively definite one. Moreover due to imposed boundary conditions it is not singular any more ($\det \overline{\mathbf{K}}_{ks} \neq 0$), so the inverse of $\overline{\mathbf{K}}_{ks}$ exists. Thus, there exist unique and nonzero solution of (5.41).

In finite element method computer software the solution to the system of equilibrium is found by appropriate numerical procedures. Many algorithms are available and just a couple of them are mentioned below. Given the assumptions adopted, the final system of equations presented by (5.41) is a linear algebraic equations system. However, in a general case it may be a nonlinear one or it can be the set of differential equations. In these cases prior to the solution, the nonlinear terms need to be

approximated by linear ones or conversion from differential to the finite difference form has to be done. More information on this topic can be found in references e.g. [1], [3].

One of the fundamental methods of solving the system of linear equations is the Cramer's rule. The method comes down to calculating the value of the main determinant of the system (determinant of the stiffness matrix of the system) and n determinants formed by replacing every n-column in stiffness matrix by the left-hand side values (i.e. forces vector). Individual solutions to the system are expressed by quotients of subsequent determinants and main determinant. As the calculation of all $n+1$ determinants is very time-consuming, the Cramer's rule is computationally very inefficient for systems with many unknowns.

A more effective method is the Gauss's elimination rule [1]. It comes down to conversion of the square stiffness matrix to its triangular form. Next, the solution to the modified system is achieved by means of a recurrence formula.

To perform matrix trangularization, one uses a sequence of following elementary row operations to modify the matrix until the lower left-hand corner of the matrix is filled with zeros. At the first stage the procedure involves subtracting the first equation from all the remaining ones – i.e. from second to n-th one. In each of these operations the subtracted equation is multiplied by a factor, that eliminates the first unknown value from each of the resulting equations where subtracting is performed. Therefore, after this initial step, the first unknown is present in the first equation only. Then, the same procedure is repeated for the second equation and second variable – the second equation is subtracted from the third and all subsequent ones. So, after this step, the second unknown is present in the first two equations only. Repeating the above steps for the next variables leads to the form, where the last unknown is present in the last equation only. So the initial stiffness matrix is converted to its triangular form. The solution of this modified system is already the trivial task.

More often than not the described procedure is alternated by so called Crout modification, called partial selection of the basic element modification. In a general outline this modification involves the change of the equations sequence so that at a given step the expression with the largest possible modulus is eliminated. The advantage of the approach is to minimize the final calculation error.

Other methods of solving linear algebraic equations systems include e.g. Cholesky factorization, static condensation and frontal solution. Wider considerations on presented methods and other algorithms capable of solving linear algebraic equations systems are discussed in studies [1], [2] and [6].

5.2.6. Element strain and stress; axial force

Solution to the system of equations (5.41) gives displacements of nodal points of the structure expressed in the global reference frame. Therefore, one may reconstruct the full vector of global displacements $\overline{\mathbf{u}}$ (see (5.39)) by extending the obtained

solution with zero values resulting from the imposed boundary conditions. Thus, this is a reverse operation to eliminating appropriate rows and columns while converting from the generalized form to the reduced one as described on page 103. Finally, individual vectors of members displacements are expressed as follows:

$$\begin{aligned} &\text{element 1:} \quad \overline{\mathbf{u}}_1 = \begin{Bmatrix} 0 & 0 & \overline{u}_5 & 0 \end{Bmatrix}^{\mathsf{T}}, \\ &\text{element 2:} \quad \overline{\mathbf{u}}_2 = \begin{Bmatrix} 0 & 0 & \overline{u}_3 & \overline{u}_4 \end{Bmatrix}^{\mathsf{T}}, \\ &\text{element 3:} \quad \overline{\mathbf{u}}_3 = \begin{Bmatrix} \overline{u}_3 & \overline{u}_4 & \overline{u}_5 & 0 \end{Bmatrix}^{\mathsf{T}}. \end{aligned} \tag{5.42}$$

After using the rotation formulas (5.27) and rotation matrices definitions (5.35) above vectors may be converted to their local reference frames

$$\mathbf{u}_1 = \begin{Bmatrix} u_A \\ v_A \\ u_C \\ v_C \end{Bmatrix} = \overline{\mathbf{\Theta}}_1 \cdot \overline{\mathbf{u}}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{Bmatrix} 0 \\ 0 \\ \overline{u}_5 \\ 0 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ \overline{u}_5 \\ 0 \end{Bmatrix}, \tag{5.43}$$

$$\begin{aligned} \mathbf{u}_2 = \begin{Bmatrix} u_A \\ v_A \\ u_B \\ v_B \end{Bmatrix} = \overline{\mathbf{\Theta}}_2 \cdot \overline{\mathbf{u}}_2 = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 & 0 \\ -\sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & \cos\alpha & \sin\alpha \\ 0 & 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{Bmatrix} 0 \\ 0 \\ \overline{u}_3 \\ \overline{u}_4 \end{Bmatrix} = \\ = \begin{Bmatrix} 0 \\ 0 \\ \overline{u}_3 \cos\alpha + \overline{u}_4 \sin\alpha \\ -\overline{u}_3 \sin\alpha + \overline{u}_4 \cos\alpha \end{Bmatrix}, \end{aligned} \tag{5.44}$$

$$\begin{aligned} \mathbf{u}_3 = \begin{Bmatrix} u_B \\ v_B \\ u_C \\ v_C \end{Bmatrix} = \overline{\mathbf{\Theta}}_3 \cdot \overline{\mathbf{u}}_3 = \begin{bmatrix} \cos\beta & -\sin\beta & 0 & 0 \\ \sin\beta & \cos\beta & 0 & 0 \\ 0 & 0 & \cos\beta & -\sin\beta \\ 0 & 0 & \sin\beta & \cos\beta \end{bmatrix} \cdot \begin{Bmatrix} \overline{u}_3 \\ \overline{u}_4 \\ \overline{u}_5 \\ 0 \end{Bmatrix} = \\ = \begin{Bmatrix} \overline{u}_3 \cos\beta - \overline{u}_4 \sin\beta \\ \overline{u}_3 \sin\beta + \overline{u}_4 \cos\beta \\ \overline{u}_5 \cos\beta \\ \overline{u}_5 \sin\beta \end{Bmatrix}. \end{aligned} \tag{5.45}$$

Obtained expressions for $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ vectors allow to calculate elements elongations Δl_i $(i = 1, 2, 3)$ see (5.2) and their axial strains ε_i $(i = 1, 2, 3)$ – see also (5.3)

$$\Delta l_1 = u_C - u_A, \qquad \varepsilon_1 = \frac{\Delta l_1}{l_1} = \frac{\overline{u}_5}{l_1}, \tag{5.46}$$

$$\Delta l_2 = u_B - u_A, \qquad \varepsilon_2 = \frac{\Delta l_2}{l_2} = \frac{\overline{u}_3 \cos\alpha + \overline{u}_4 \sin\alpha}{l_2}, \tag{5.47}$$

$$\Delta l_3 = u_C - u_B, \qquad \varepsilon_3 = \frac{\Delta l_3}{l_3} = \frac{\overline{u}_5 \cos\beta - \overline{u}_3 \cos\beta + \overline{u}_4 \sin\beta}{l_3}. \tag{5.48}$$

Stresses are calculated according to the Hooke's law (see assumption of the linear material behaviour).

$$\sigma_1 = E\varepsilon_1 = \frac{\overline{u}_5}{l_1} E \tag{5.49}$$

$$\sigma_2 = E\varepsilon_2 = \frac{\overline{u}_3 \cos\alpha + \overline{u}_4 \sin\alpha}{l_2} E \tag{5.50}$$

$$\sigma_3 = E\varepsilon_3 = \frac{\overline{u}_5 \cos\beta - \overline{u}_3 \cos\beta + \overline{u}_4 \sin\beta}{l_3} E \tag{5.51}$$

whereas axial forces in rods are

$$\mathbf{N} = \begin{Bmatrix} A_1\sigma_1 \\ A_2\sigma_2 \\ A_3\sigma_3 \end{Bmatrix} = E \begin{Bmatrix} \dfrac{\overline{u}_5}{l_1} A_1 \\ \dfrac{\overline{u}_3 \cos\alpha + \overline{u}_4 \sin\alpha}{l_2} A_2 \\ \dfrac{\overline{u}_5 \cos\beta - \overline{u}_3 \cos\beta + \overline{u}_4 \sin\beta}{l_3} A_3 \end{Bmatrix}. \tag{5.52}$$

To complete the analysis of the structure, one may calculate the support forces (reaction of the base). Readers interested in this topic will find more information in references [5], [7].

5.3. Flexure elements. Linear analysis of beams

The rod (truss) elements discussed in the previous chapter are used to model parts and sub-assemblies loaded in axial direction only. Therefore they cannot be used to model structures where the transverse forces and bending moments occur. This corresponds to the case of e.g., welded or riveted structures like frames and grates,

since in these systems bending is a dominant type of internal loading. In the present sub-chapter, the flexure element will be discussed. This is the finite element of first choice for modelling slender members subjected to bending load. Similarly to the previously presented rods the degrees of freedom for flexure element will be discussed; next shape functions and the stiffness matrix will be derived. The relations will be formulated in accordance with the Euler-Bernoulli beam theory. In further part a simple example of a statically indeterminate beam analysis is performed.

5.3.1. Assumptions

Analogous to the case of axial elements discussed in sub-chapter 5.2 we will limit our considerations to the linear model. Apart from assumptions made in section 5.2.2 on page 92 we postulate the following:

- to simplify the analysis we restrict the discussion to a two dimensional problem, so the element and its loading are in the xy plane. Moreover the external loading acts only in end points of the beam,
- transverse deflections v of the element are small in relation to the characteristic dimension of its cross-section h, i.e. do not exceed $0{,}1h$,
- the cross-section of the beam is symmetric and symmetry axis stays in the bending plane xy,
- straight lines normal to the beam mid-surface remain straight and normal to the mid-surface after deformation (Kirchhoff assumptions),
- the thickness of the plate does not change during a deformation.

5.3.2. Shape functions of the beam element

In beam bending structural analysis the field variable of interest is the transverse displacement $v(x)$ of a representative point located on beam mid-surface. Similarly to the rod element problem, the value of field variable inside the finite element is expressed in terms of generalized displacements of nodal points. As shown in Figure 5.5, transverse deflection of the beam is not unequivocally described by displacement of its end points only. The end deflections can be identical, as illustrated, while the deflected shape of the two cases is completely different. Therefore, the flexure element formulation must take into account not only end-point displacements but the slope (rotation) of the beam as well. One may observe that introducing the rotation of a nodal cross-section as a degree of freedom automatically assures the consistency of cross-section rotations on the boundaries between neighbouring finite elements.

Therefore the linear beam finite element has four degrees of freedom, being the generalized displacements $v_1, \theta_1, v_2, \theta_2$. To this corresponds the generalized loading in the form of two transverse forces and two bending moments. The rule

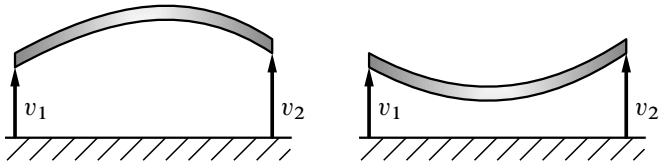


Figure 5.5. *Different forms of deformations with the same shifts of knots*

concerning the positive sign of generalized coordinates and forces has been presented in the Figure 5.6.

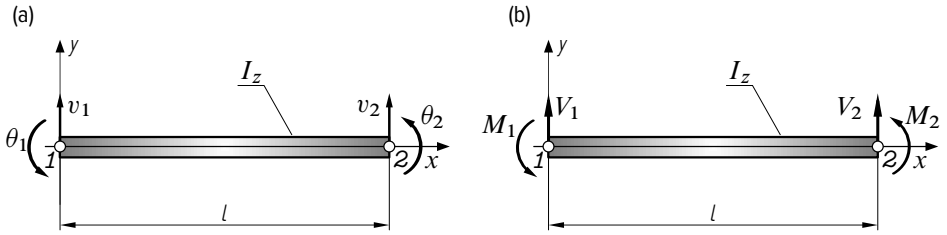


Figure 5.6. *Beam finite element: (a) degrees of freedom (b)generalized forces*

The yet unknown transverse displacement formula $v(x) = f(v_1, \theta_1, v_2, \theta_2, x)$ is the function of the independent variable x which defines the location of the considered point on the beam axis. As one observes in Figure 5.7(a) this function must fulfil the following boundary conditions:

$$\begin{aligned} v(x=0) &= v_1, \qquad \left.\frac{dv(x)}{dx}\right|_{x=0} = \theta_1, \\ v(x=l) &= v_2, \qquad \left.\frac{dv(x)}{dx}\right|_{x=l} = \theta_2. \end{aligned} \tag{5.53}$$

Any mathematical function of x to express the $v(x)$ variable may be assumed. However, polynomials are most frequently used due to simplicity of numerical computations. Considering four available boundary conditions and the one-dimensional nature of the problem in terms of the independent variable x, we assume the displacement function in the general form with maximum four unknown values. Let us adopt therefore polynomial of the third order:

$$v(x) = f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3. \tag{5.54}$$

In case of beams loaded by concentrated forces only, the choice of the third order polynomial is fully justified. This is due to the fact that in the discussed case the distribution of the bending moment $M_g(x)$ inside the element is linear. Since the

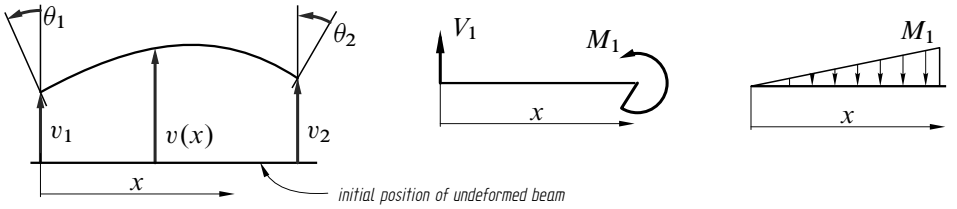


Figure 5.7. *Deformation field of the beam element and distribution of bending moments in the considered system*

bending moment is directly proportional to the second derivative of transverse displacement $\frac{d^2v(x)}{dx^2}$, the two-fold differentiation of third order polynomial results in the requested linear function.

Putting the general expression (5.54) into boundary conditions (5.53) the following equations are to be fulfilled:

$$\begin{cases} v(x=0) & = v_1 = a_0, \\ \left.\dfrac{dv(x)}{dx}\right|_{x=0} & = \theta_1 = a_1, \\ v(x=l) & = v_2 = a_0 + a_1 l + a_2 l^2 + a_3 l^3, \\ \left.\dfrac{dv(x)}{dx}\right|_{x=l} & = \theta_2 = a_1 + 2a_2 l + 3a_3 l^2, \end{cases}$$

Thus after transformations, individual terms of the polynomial (5.54) are:

$$a_0 = v_1 \qquad a_2 = \frac{3}{l^2}(v_2 - v_1) - \frac{1}{l}(2\theta_1 + \theta_2)$$

$$a_1 = \theta_1 \qquad a_3 = \frac{2}{l^3}(v_1 - v_2) + \frac{1}{l^2}(\theta_1 + \theta_2).$$

If put into the relation (5.54) and after re-ordering with respect to individual degrees of freedom $v_1, \theta_1, v_2, \theta_2$ the following function is obtained:

$$v(x) = \left(1 - \frac{3x^2}{l^2} + \frac{2x^3}{l^3}\right)v_1 + \left(x - \frac{2x^2}{l} + \frac{x^3}{l^2}\right)\theta_1 + \left(\frac{3x^2}{l^2} - \frac{2x^3}{l^3}\right)v_2 + \left(\frac{x^3}{l^2} - \frac{x^2}{l}\right)\theta_2. \tag{5.55}$$

While comparing the above expression with relations (5.8) and (5.13) one may notice that terms in brackets

$$\begin{aligned} N_1(x) &= 1 - \frac{3x^2}{l^2} + \frac{2x^3}{l^3}, & N_3(x) &= \frac{3x^2}{l^2} - \frac{2x^3}{l^3}, \\ N_2(x) &= x - \frac{2x^2}{l} + \frac{x^3}{l^2}, & N_4(x) &= \frac{x^3}{l^2} - \frac{x^2}{l} \end{aligned} \tag{5.56}$$

are the shape functions of the beam finite element

$$v(x) = N_1(x)v_1 + N_2(x)\theta_1 + N_3(x)v_2 + N_4(x)\theta_2. \tag{5.57}$$

5.3.3. Stiffness matrix of the beam element

Similarly to the approach presented in the previous sub-chapter, stiffness matrix will be derived on the basis of the Castigliano's theorem. To that end the total elastic energy of the system must be given. In accordance with the Euler's-Bernoulli theory the energy comes only from the bending effect:

$$\mathcal{U} = \mathcal{U}_g = \frac{1}{2}\int_V \sigma\varepsilon dV \tag{5.58}$$

where dV is the volume of the infinite small beam segment, whereas ε and σ are strain and stress acting on the considered section of the beam. The stress value may be found from the flexure formula

$$\sigma = \frac{M_g(x)}{I_z}y = E\frac{d^2v(x)}{dx^2}y, \tag{5.59}$$

where I_z is a moment of inertia of the cross-section about z axis perpendicular to the bending plane xy, whereas y is the coordinate defining the location of the point on the cross-section with respect to neutral plane. Inserting the transverse displacement definition (5.57) to the above formula the stress expression is obtained

$$\begin{aligned} \sigma &= Ey\left(\frac{d^2N_1}{dx^2}v_1 + \frac{d^2N_2}{dx^2}\theta_1 + \frac{d^2N_3}{dx^2}v_2 + \frac{d^2N_4}{dx^2}\theta_2\right) \\ &= Ey\left[\left(\frac{12x}{l^3} - \frac{6}{l^2}\right)v_1 + \left(\frac{6x}{l^2} - \frac{4}{l}\right)\theta_1 + \left(\frac{6}{l^2} - \frac{12x}{l^3}\right)v_2 + \left(\frac{6x}{l^2} - \frac{2}{l}\right)\theta_2\right]. \end{aligned} \tag{5.60}$$

Therefore, assuming strain and inertia definitions $\varepsilon = \frac{\sigma}{E}$ and $I_z = \int_A y^2 dA$ respectively the total potential elastic energy $\mathcal{U}$ of the element is

$$\mathcal{U} = \frac{EI_z}{2}\int_0^l \left(\frac{d^2N_1}{dx^2}v_1 + \frac{d^2N_2}{dx^2}\theta_1 + \frac{d^2N_3}{dx^2}v_2 + \frac{d^2N_4}{dx^2}\theta_2\right)^2 dx. \tag{5.61}$$

One should emphasize that in general case the above formula is only the approximate one. It results from the fact that the field variable $v(x)$ has been defined with an arbitrary assumed third order polynomial. As it was commented earlier in case of beams loaded in nodal points only this approximation gives strict results. However, for different loading scenarios this definition may give not fully accurate outcomes.

To derive the stiffness matrix **K** of the beam element the first Castigliano's rule is used. By virtue of the theorem the following equations are to be fulfilled:

$$\begin{aligned} \frac{\partial \mathcal{U}}{\partial v_1} &= V_1, & \frac{\partial \mathcal{U}}{\partial \theta_1} &= M_1, \\ \frac{\partial \mathcal{U}}{\partial v_2} &= V_2, & \frac{\partial \mathcal{U}}{\partial \theta_2} &= M_2. \end{aligned} \tag{5.62}$$

After inserting the energy definition, performing the differentiation and necessary manipulations we obtain the following relations at the first node

$$\begin{aligned} V_1 = EI_z \int_0^l \Bigg(& \frac{d^2N_1}{dx^2}\frac{d^2N_1}{dx^2}v_1 + \frac{d^2N_2}{dx^2}\frac{d^2N_1}{dx^2}\theta_1 \\ & + \frac{d^2N_3}{dx^2}\frac{d^2N_1}{dx^2}v_2 + \frac{d^2N_4}{dx^2}\frac{d^2N_1}{dx^2}\theta_2 \Bigg) dx \\ M_1 = EI_z \int_0^l \Bigg(& \frac{d^2N_1}{dx^2}\frac{d^2N_2}{dx^2}v_1 + \frac{d^2N_2}{dx^2}\frac{d^2N_2}{dx^2}\theta_1 \\ & + \frac{d^2N_3}{dx^2}\frac{d^2N_2}{dx^2}v_2 + \frac{d^2N_4}{dx^2}\frac{d^2N_2}{dx^2}\theta_2 \Bigg) dx \end{aligned} \tag{5.63}$$

and at the second one

$$\begin{aligned} V_2 = EI_z \int_0^l \Bigg(& \frac{d^2N_1}{dx^2}\frac{d^2N_3}{dx^2}v_1 + \frac{d^2N_2}{dx^2}\frac{d^2N_3}{dx^2}\theta_1 \\ & + \frac{d^2N_3}{dx^2}\frac{d^2N_3}{dx^2}v_2 + \frac{d^2N_4}{dx^2}\frac{d^2N_3}{dx^2}\theta_2 \Bigg) dx \\ M_2 = EI_z \int_0^l \Bigg(& \frac{d^2N_1}{dx^2}\frac{d^2N_4}{dx^2}v_1 + \frac{d^2N_2}{dx^2}\frac{d^2N_4}{dx^2}\theta_1 \\ & + \frac{d^2N_3}{dx^2}\frac{d^2N_4}{dx^2}v_2 + \frac{d^2N_4}{dx^2}\frac{d^2N_4}{dx^2}\theta_2 \Bigg) dx \end{aligned}$$

The above system of four linear equations can be written in the condensed matrix form

$$\begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{14} \\ K_{21} & K_{22} & K_{23} & K_{24} \\ K_{31} & K_{32} & K_{33} & K_{34} \\ K_{41} & K_{42} & K_{43} & K_{44} \end{bmatrix} \cdot \begin{Bmatrix} v_1 \\ \theta_1 \\ v_2 \\ \theta_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ M_1 \\ F_2 \\ M_2 \end{Bmatrix}$$

where each (m, n) term of the stiffness matrix is defined as

$$K_{mn} = EI_z \int_0^l \frac{d^2 N_m}{dx^2} \frac{d^2 N_n}{dx^2} dx \qquad m, n = 1, \ldots, 4$$

Inserting previously derived expressions (5.55) for shape functions $N_i(x)$ $(i = 1 \ldots 4)$ and performing the integration we obtain the final form of the stiffness matrix

$$\mathbf{K} = \frac{EI_z}{l^3} \begin{bmatrix} 12 & 6l & -12 & 6l \\ 6l & 4l^2 & -6l & 2l^2 \\ -12 & -6l & 12 & -6l \\ 6l & 2l^2 & -6l & 4l^2 \end{bmatrix}. \tag{5.64}$$

Thus, similarly to rod elements, symmetry of the linear flexure element stiffness matrix is apparent. Moreover, the given matrix is singular ($\det \mathbf{K} = 0$). Due to this last property it incorporates also rigid body behaviour if the element is not constrained anyhow. The size of the matrix $\mathbf{K}_{(4\times4)}$ results directly from the number of degrees of freedom – i.e. two transverse shifts and two angles of revolution – Figure 5.6.

5.3.4. Distributed load

In the present considerations the analysis has been made if external forces acting on the flexure element are imposed in nodal points only. However, the commonly encountered loading of beam elements is a distributed transverse force acting on the length of the element.

The usual approach is to replace this distributed load with substitute nodal forces and moments. The condition is that the mechanical work done by this nodal load system is equivalent to that done by the distributed load. If dynamic effects and energy dissipation are omitted, one may conclude the mechanical work done by continuous loading is equivalent to the elastic potential energy of the system $\mathcal{U}$ resulting from applied equivalent loads.

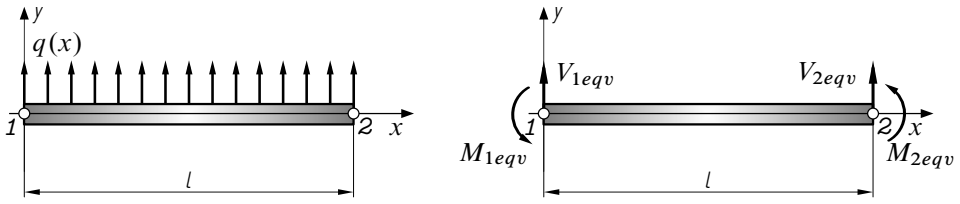


Figure 5.8. *Distributed load imposed on the beam element and equivalent nodal generalized forces*

In accordance with the definition, elementary work $d\mathcal{W}$ done by the continuous load $q(x)$ acting on infinitesimal section of the beam through the transverse displacement $v(x)$ is given by product $v(x)q(x)dx$. Therefore the total work done on the whole finite element is:

$$\begin{aligned}\mathcal{W}_c &= \int_0^l d\mathcal{W} = \int_0^l q(x)v(x)dx = \\ &= \int_0^l q(x)\left[N_1(x)v_1 + N_2(x)\theta_1 + N_3(x)v_2 + N_4(x)\theta_2\right]dx.\end{aligned} \tag{5.65}$$

Introducing equivalent generalized forces in beam nodes – see Figure 5.8 – one may write the work done by this load

$$\mathcal{W}_{\text{eqv}} = \mathbf{Q}_{\text{eqv}} \cdot \mathbf{v} = V_{1\text{eqv}}v_1 + M_{1\text{eqv}}\theta_1 + V_{2\text{eqv}}v_2 + M_{2\text{eqv}}\theta_2. \tag{5.66}$$

Multiplying in (5.65) the bracket terms by the $q(x)$ factor, separating into individual integrals and comparing to (5.66) we obtain

$$V_{1\text{eqv}} = \int_0^l q(x)N_1(x)dx \qquad M_{1\text{eqv}} = \int_0^l q(x)N_2(x)dx$$

$$V_{2\text{eqv}} = \int_0^l q(x)N_3(x)dx \qquad M_{2\text{eqv}} = \int_0^l q(x)N_4(x)dx.$$

The derived above expressions are general in nature so they enable calculating the equivalent generalized loads for any distribution of continuous loading $q(x)$.

While inserting shape function definitions (5.56) and presuming constant value $q(x) = q = \text{const}$ we get

$$V_{1\text{eqv}} = \int_0^l q\left(1 - \frac{3x^2}{l^2} + \frac{2x^3}{l^3}\right) dx = \frac{ql}{2} \qquad V_{2\text{eqv}} = \frac{ql}{2}$$

$$M_{1\text{eqv}} = \int_0^l q\left(x - \frac{2x^2}{l} + \frac{x^3}{l^2}\right) dx = \frac{ql^2}{12} \qquad M_{2\text{eqv}} = -\frac{ql^2}{12}$$

Example

To illustrate the application of the finite element method to solve beam structures let us consider a simple, statically indeterminate system presented in the Figure 5.9.

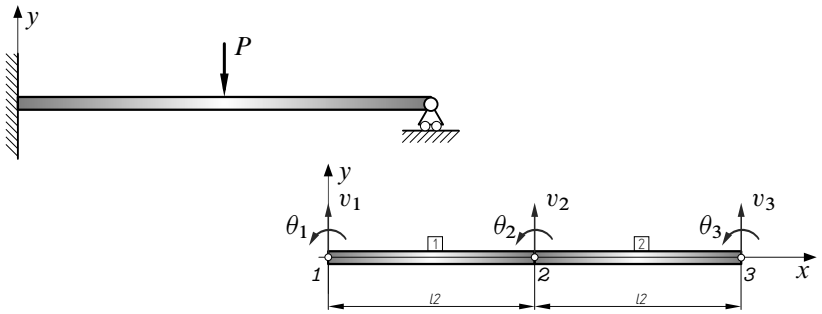


Figure 5.9. *Hyperstatic cantilever beam and its discretization into two finite elements*

Due to the force P acting in the middle of the beam the member must be divided into at least two finite elements.[5] The division point must coincide with the place where the force P is applied.

Given the division into two finite elements one may write down individual stiffness matrices. Based on previous derivations and taking into account the length of every finite element to be $\frac{l}{2}$ we obtain

$$\mathbf{K}_1 = \overline{\mathbf{K}}_1 = \mathbf{K}_2 = \overline{\mathbf{K}}_2 = \frac{8EI_z}{l^3}\begin{bmatrix} 12 & 3l & -12 & 3l \\ 3l & l^2 & -3l & l^2/2 \\ -12 & -3l & 12 & -3l \\ 3l & l^2/2 & -3l & l^2 \end{bmatrix} \tag{5.67}$$

The global stiffness matrix can now be assembled for the structure by using the direct stiffness method. When the global stiffness matrix is assembled, the external

[5] This condition results from the fact that the given formulas are derived for the case of transverse forces imposed only at the ends of the finite element.

nodal forces are directly related to the global nodal displacements. Through the superposition of individual stiffnesses the governing equations for the beam are thus given by

$$\frac{EI_z}{l^3}\begin{bmatrix} 96 & 24l & -96 & 24l & 0 & 0 \\ 24l & 8l^2 & -24l & 4l^2 & 0 & 0 \\ -96 & -24l & 192 & 0 & -96 & 24l \\ 24l & 4l^2 & 0 & 16l^2 & -24l & 4l^2 \\ 0 & 0 & -96 & -24l & 96 & -24l \\ 0 & 0 & 24l & 4l^2 & -24l & 8l^2 \end{bmatrix} \cdot \begin{Bmatrix} v_1 \\ \theta_1 \\ v_2 \\ \theta_2 \\ v_3 \\ \theta_3 \end{Bmatrix} = \begin{Bmatrix} P_1 \\ M_1 \\ P_2 \\ M_2 \\ P_3 \\ M_3 \end{Bmatrix}. \tag{5.68}$$

In the above expression, the fragments of the global stiffness matrix coming from the individual finite elements are highlighted in grey colour. Moreover, one should emphasize that due to the mutual position of individual finite elements the transformation of coordinates wasn't necessary.

Relation (5.68) is the general equilibrium condition of an arbitrary beam consisting of two equal spans. After setting the boundary conditions this equation will correspond to the discussed case.

Now considering the imposed constraints of clamping at the left end (node *1*) and pin support at node *3* the restrained degrees of freedom are

$$v_1 = 0 \qquad \theta_1 = 0 \qquad v_3 = 0. \tag{5.69}$$

To solve the set of equilibrium equations and find the unknown, non-zero generalized displacements one removes from the stiffness matrix $\overline{\mathbf{K}}$ (5.68) rows and columns with numbers one, two and five. This can be done since all entries in columns (1,2,5) of the global stiffness matrix are multiplied by zero nodal values. The reduced system of equilibrium equations is thus as follows

$$\frac{EI_z}{l^3}\begin{bmatrix} 192 & 0 & 24\,l \\ 0 & 16\,l^2 & 4\,l^2 \\ 24\,l & 4\,l^2 & 8\,l^2 \end{bmatrix} \cdot \begin{Bmatrix} v_2 \\ \theta_2 \\ \theta_3 \end{Bmatrix} = \begin{Bmatrix} -P \\ 0 \\ 0 \end{Bmatrix}$$

Solution to the system are:

$$v_2 = \frac{-7Pl^3}{768EI_z} \qquad \theta_2 = \frac{-Pl^2}{128EI_z} \qquad \theta_3 = \frac{Pl^2}{32EI_z}$$

Inserting these values into the full set of equilibrium equations (5.68) one may find the right-hand side vector of generalized forces in nodes `1`, `2` and `3`:

$$P_1 = \tfrac{11}{16}P, \qquad P_2 = -P, \qquad P_3 = \tfrac{5}{16}P,$$
$$M_1 = \tfrac{3}{16}Pl, \qquad M_2 = 0, \qquad M_3 = 0.$$

As expected, the value of the transverse force in node no. `2` corresponds to the external force P and the bending moment in pins `2` and `3` is equal to 0.

The shape of the beam deflection curve results directly from the the field variable $v(x)$ approximation and supposed shape functions – see (5.57) and (5.56)

$$\begin{aligned} N_1(x) &= 1 - \frac{12x^2}{l^2} + \frac{16x^3}{l^3}, & N_3(x) &= \frac{12x^2}{l^2} - \frac{16x^3}{l^3}, \\ N_2(x) &= x - \frac{4x^2}{l} + \frac{4x^3}{l^2}, & N_4(x) &= \frac{4x^3}{l^2} - \frac{2x^2}{l}. \end{aligned} \tag{5.70}$$

Since shape functions are the same for both finite elements the deflections for the left and right section of the beam are given by

$$\begin{aligned} v^{(1)}(x) &= N_1(x)v_1 + N_2(x)\theta_1 + N_3(x)v_2 + N_4(x)\theta_2, \\ v^{(2)}(x) &= N_1(x)v_2 + N_2(x)\theta_2 + N_3(x)v_3 + N_4(x)\theta_3. \end{aligned} \tag{5.71}$$

respectively. Thus finally

$$\begin{aligned} v^{(1)}(x) &= \frac{Px^2}{96EI_z}(-9l + 11x), \\ v^{(2)}(x) &= \frac{-P}{768EI_z}(7l^3 + 6l^2x - 60lx^2 + 40x^3) \end{aligned} \tag{5.72}$$

where in both cases $x \in \langle 0, \frac{1}{2}l \rangle$.

Distributions of bending moment and transverse force result directly from their definitions:

$$M(x) = EI_z\frac{d^2v}{dx^2}, \qquad V(x) = -\frac{dM_g(x)}{dx}. \tag{5.73}$$

After differentiation, for the individual sections the following results are obtained:

$$\begin{aligned} M_1(x) &= \frac{1}{16}(11Px - 3lP), \\ M_2(x) &= \frac{5}{32}P(l - 2x) \qquad\qquad x \in \left\langle 0, \frac{1}{2}l \right\rangle \end{aligned} \tag{5.74}$$

and

$$V_1(x) = -\frac{11}{16}P, \qquad V_2(x) = +\frac{5}{16}P. \tag{5.75}$$

The diagrams of the above variables are presented in the Figure 5.8. The obtained results of the reaction forces in knots and bending ate fully consistent with the results available in the literature e.g. [12].

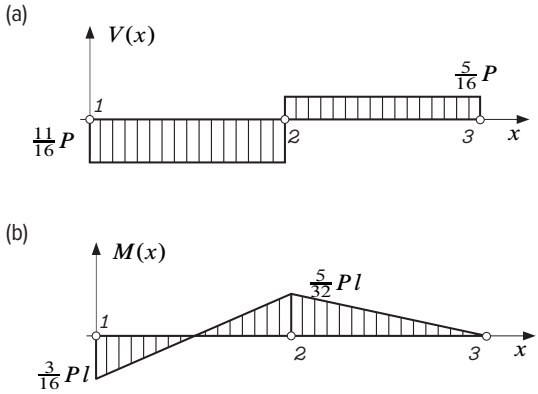


Figure 5.10. *Bending moment $M(x)$ (a), and shear force $V(x)$ (b) diagrams in the hyperstatic beam*

To sum up, one should pay attention to the fact that the solution of the statically determinate or indeterminate system by the finite element method do not differ in terms of work-flow and steps involved. In both problems the same system of equilibrium is solved, the only difference results from the number of imposed boundary conditions so the number of equations in reduced matrix (system) to be solved. Whereas, while solving the statically indeterminate problem with classic Newton's equations of equilibrium the additional geometrical conditions for deformed structure need to be added. In the discussed example the requested condition may be given as a virtual vertical force introduced in the right end and setting $f_3 = 0$ or the condition may result from the Menabrei's theorem.

Bibliography

[1] Bathe K.J. (2006): *Finite element procedures*. Klaus-Jurgen Bathe, Cambridge (MA), USA.

[2] Brandt S. (1998): *Analiza danych*. Wydawnictwo Naukowe PWN, Warszawa, pierwsze wyd.

[3] Chmielewski T., Nowak H. (1996): *Mechanika budowli*. Wspomaganie komputerowe CAD/CAM, Wydawnictwa Naukowo-Techniczne, Warszawa, drugie wyd.

[4] Cook R.D. (1995): *Finite Element Modeling for Stress Analysis*. Wiley, New York.
[5] Dąbrowski O. (1983): *Mechanika budowli*. Arkady, Warszawa.
[6] Fortuna Z., Macukow B., Wąsowski J. (1995): *Metody numeryczne*. Wydawnictwa Naukowo-Techniczne, Warszawa.
[7] Hutton D.V. (2013): *Fundamentals of Finite Element Analysis*. McGraw-Hill series in mechanical engineering, McGraw-Hill Higher Education, Boston.
[8] Kacprzyk Z., Rakowski G. (2005): *Metoda elementów skończonych w mechanice konstrukcji*. Oficyna Wydawnicza Politechniki Warszawskiej, Hoboken (NJ), USA.
[9] Kim N.H., Sankar B.V. (2009): *Introduction to Finite Element Analysis and Design*. John Wiley & Sons, New York.
[10] Logan D.L., Chaudhry K.K., Singh P. (2011): *A First Course in the Finite Element Method*. Cengage Learning, Stamford (CT), USA.
[11] Mac Donald B. (2011): *Practical Stress Analysis with Finite Elements*. Glasnevin Publishing, Dublin.
[12] Niezgodziński M., Niezgodziński T. (1996): *Wzory, wykresy i tablice wytrzymałościowe*. Wydawnictwo Naukowo-Techniczne, Warszawa.
[13] Rusiński E., Czmochowski J., Smolnicki T. (2000): *Zaawansowana metoda elementów skończonych w konstrukcjach nośnych*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.