DOI: 10.46299/ISG.2024.MONO.TECH.1.2.5

2.5 Numerical methods in soil mechanics and geomechanics

When writing this monograph, an attempt was made to represent in a concentrated form the currently accepted classification of numerical methods that are used in solving problems of soil mechanics and geomechanics in general. At the same time, the goal was also to highlight the advantages and disadvantages of each of the considered numerical methods.

First, let’s answer the question of what "numerical methods" are.

Since, as a result of solving a particular problem, we obtain either one number (for example, the stability coefficient of a slope or inclination) or a certain set of numbers (for example, when calculating contact stresses at the "foundation-base" contact, stress distribution fields, pressure in the pore fluid of the base, etc.) all methods for solving engineering problems are numerical.

Therefore, the currently accepted classification of foundation calculation methods is somewhat arbitrary. Up to now, the following practice has developed for categorizing methods for calculating soil foundations:

1. **Analytical methods** are typically used to solve simple problems, for example, when calculating stresses in the foundation from the natural weight of the soil, concentrated force applied to the upper boundary of the foundation, stability of a vertical slope, etc.

The advantages of analytical methods are their simplicity and accessibility, and the disadvantages are a very limited range of engineering problems that can be solved using them [119, 120]).

2. **Graphic-analytical** (for example, Kuhlman’s method [121]).

The advantages of graphic-analytical methods are their simplicity, clarity, and accessibility, and the disadvantages are a very limited range of engineering problems solved with their use, as well as low calculation accuracy (it is determined by the angles, scales, and thickness of lines on the diagrams available when performing graphical constructions).

3. **Tabular** (these include, for example, the methods of Sokolovsky, Goldstein, and many others [121, 122, 123]).

The advantages of tabular methods are their simplicity and accessibility, and the disadvantages are a very limited range of engineering problems that can be solved using them.

4. **Engineering empirical and semi-empirical methods** (these include most of those set out in SCN, as well as procedures, for example, the method of layer-by-layer summation when calculating the settlement of foundations, the method of a circular-cylindrical sliding surface when determining the stability coefficient of slopes and inclinations, etc.) [121, 124, 125, 126, 127].

The advantages of engineering empirical and semi-empirical methods are:

4.1. Simplicity and accessibility.

4.2. Relationship with experimental data.

4.3. Clear boundaries for changing the parameters introduced to empirical and semi-empirical formulas, within which engineering empirical and semi-empirical methods should be applied.

4.4. Good agreement between the calculation results and experimental and field data.

The disadvantages of engineering empirical and semi-empirical methods are:

4.5. A narrow range of problems that can be solved using the indicated methods.

4.6. The detail of calculation schemes is insufficient to solve many practical problems.

5. **Numerical methods based on the principles of continuum mechanics** [128, 129, 130, 131, 132, 133, 134].

Recently, these methods have become quite widespread. Therefore, let us dwell on their features, advantages, and disadvantages in more detail.

Depending on the technique of performing calculations, these methods are divided into:

5.1. **Finite difference method** (when implemented, the derivatives of the sought functions included in the equilibrium and state equations within the computational domain are replaced by finite differences) [128, 129].

Typically, this version of numerical methods is used to solve the following problems in soil mechanics and foundation engineering:

– thermal conductivity;

– filtration of pore fluid;

– filtration consolidation;

– determination of the stressed-strained state, strength, and stability of soil foundations.

The advantages of the finite difference method are:

5.1.1. Ease of implementation: the finite difference method is based on approximating derivatives by difference ratios, making it relatively simple to understand and implement.

5.1.2. Versatility: the finite difference method can be used to solve a wide class of differential equations, including ordinary differential equations and partial differential equations. This makes it a versatile tool for modeling various physical and engineering problems.

5.1.3. Flexibility: the finite difference method makes it possible to choose different grids and discretization steps, which makes it possible to adapt the method to a specific problem and achieve the required solution accuracy.

5.1.4. Efficiency: the finite difference method can be effectively applied to solve problems with nonlinear and inhomogeneous equations, as well as to model complex geometric structures.

The disadvantages of the finite difference method are:

5.1.5. Approximation error: the finite difference method is based on approximation of derivatives, which can lead to errors in the solution. This is especially noticeable when using coarse meshes or large sampling steps.

5.1.6. Geometry constraints: the finite difference method requires discretization of space, which can be difficult for complex geometric structures. For example, when modeling objects with curved boundaries, one may need to use unstructured meshes.

5.1.7. Computational complexity: solving problems using the finite difference method can be computationally complex, especially at large grid sizes or when there are nonlinear equations. This may require the use of powerful computing resources.

5.1.8. Sensitivity to boundary conditions: the finite difference method can be sensitive to the choice of boundary conditions, especially in the presence of nonlinear equations. Setting boundary conditions incorrectly can lead to incorrect results or an unstable solution.

5.1.9. The impossibility of joint calculation of systems of the type "soil foundation – foundations – above-foundation structure" (Fig. 1).

Currently, the finite difference method is used to solve problems of geomechanics listed in paragraph 5.1.

5.2. **The collocation method** is a projection method for solving integral and differential equations, in which an approximate solution is determined from the condition of satisfying the equation at certain specified points [129, 130].

In simple words: if, when solving a problem using the finite difference method, we replace finite differences in one or more coordinates with ordinary or partial derivatives (while we must leave finite differences in the remaining coordinates), **then we obtain the collocation method**.

The advantages of the difference collocation method are:

5.2.1. All the inherent advantages of the finite difference method.

5.2.2. The ability to reduce the dimensionality of a problem by one or several orders of magnitude.

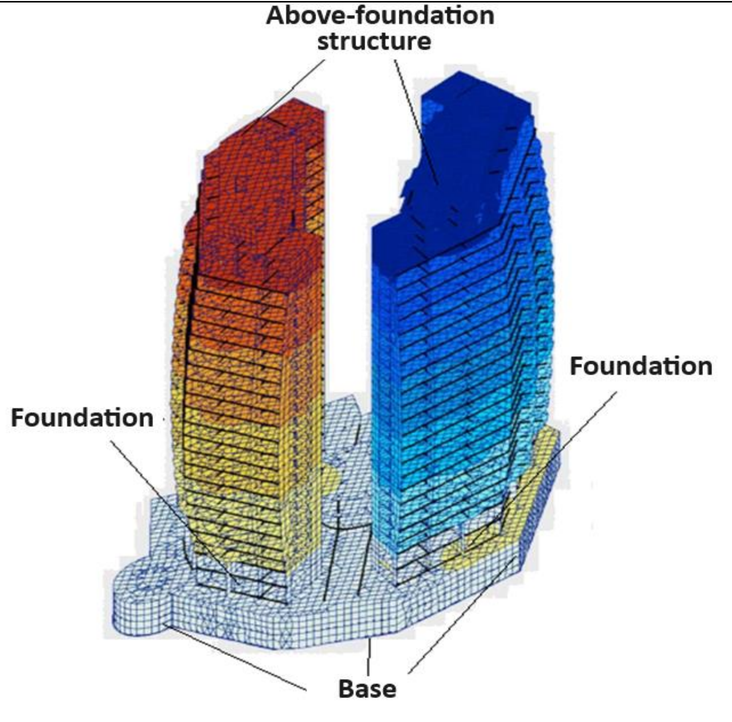


Fig. 1. The calculation scheme "base – foundations – above-foundation structures", split into volumetric, flat, and linear finite elements

The disadvantages of the collocation method are:

5.2.3. All the inherent disadvantages of the finite difference method.

5.2.4. Problems arising when satisfying boundary conditions and assessing the accuracy, convergence, and stability of solutions obtained using the finite difference method.

Currently, the collocation method has limited use in solving geomechanics problems.

5.3. **Boundary element method** [129, 131, 132]. When solving practical problems, it is necessary to take into account rather complex boundary conditions. Therefore, it is impossible to obtain analytical solutions to such problems. At the same time, it is quite easy to obtain analytical solutions for the simplest boundary conditions.

Therefore, it is proposed to find a solution to complex problems by dividing the boundary of the considered area into a number of sections, the loads within each of which correspond to the loads adopted when constructing analytical solutions.

As a result, a system of linear algebraic equations is built, the unknowns of which are either the loads within each of the boundary elements (**force method**) or the displacement of their centers (**displacement method**).

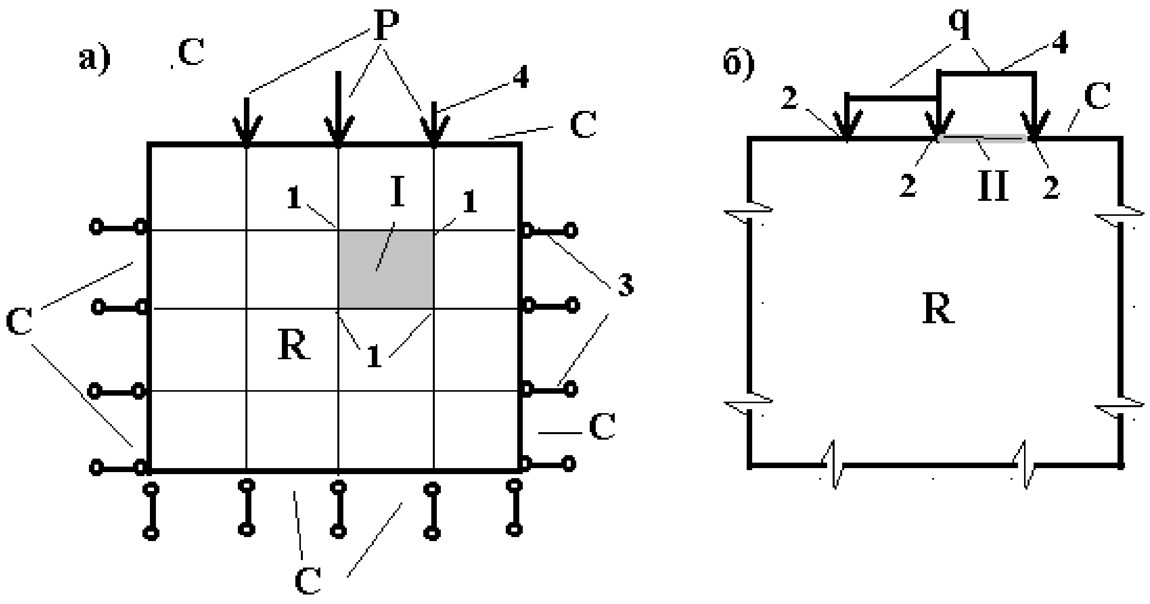


Fig. 2. Finite-element (*a*) and boundary-element (*b*) approximations of the calculation domain R.

I – finite element; II – boundary element

1 – nodes of finite elements; 2 – nodes of boundary elements; 3 – connections; 4 – nodal loads, distributed loads

The solution of the problem with complex boundary conditions is obtained by superposition of the simplest solutions.

This approach to solving boundary value problems was termed the **boundary element method** (BEM, Fig. 2*b*). Usually, it is used to solve linear problems.

There are **direct**, **semidirect**, and **indirect** boundary element methods.

The essence of the **direct boundary element method** is that, in this case, the functions to be defined in the integral equations have a specific physical meaning.

The essence of the **indirect method of boundary elements** is that, first, singular solutions of the problem that do not have a specific physical meaning and satisfy the boundary conditions are found, using which the remaining required boundary parameters are calculated.

There is also a distinction between **internal and external problems** of the boundary element method. The **internal problem of BEM** is understood as the variant when the calculation domain R is bounded by the closed circuit C.

The **external problem of BEM** is understood as such a variant when the calculation domain R is unbounded.

The advantages of the boundary element method are:

5.3.1. The possibility of constructing solutions for areas that are unlimited in plan and in depth (this is fundamentally impossible to do using the finite element method).

5.3.2. From a computational point of view, the boundary element method leads to a system of equations of a lower order than the finite element method when solving the same problem.

5.3.3. There is no need to interpolate the results of the solution within the calculation area (this drawback is inherent in FEM method).

5.3.4. Exact satisfaction of the initial differential equation within the calculation domain.

The disadvantages of the boundary element method are:

5.3.5. The traditional setting of the method considers boundary conditions of one type, either Dirichlet or Neumann. At the same time, the mixed problem is not considered.

5.3.6. The border of the calculation area must be smooth.

5.3.7. The fundamental impossibility of solving problems in which it is impossible to use the principle of superposition.

5.4. **Finite element method** [129, 133, 134]. In some cases (geometrical and physical nonlinearity, different equations of state in different subregions of the calculation domain, etc.), there is a need to divide the entire calculation domain into sections. This approach to solving the boundary value problem was termed the **finite element method** (FEM, Fig. 2a).

In the course of solving the problem using the finite element method, it is necessary to find the solution of the problem in the nodes of the grid. Values of the desired function outside the grid nodes should be determined by interpolation.

The essence of the method is to divide the studied area into separate parts, called finite elements, which connect to each other at a certain number of points, termed nodes. Each element is described by an approximate function, the values of which at the nodes of the element are its solution and are unknown in advance. Outside its element, the function is zero. Separate equations are assembled into a general system of equations describing the entire domain. The number of equations is equal to the number of unknown values at the nodes and is limited only by the computational capabilities of the computer. Since the element is connected to a small number of neighboring elements, the matrix of equations has a sparce form, which greatly simplifies the solution.

Depending on the unknowns sought, three main forms of FEM are distinguished:

– FEM in the form of the displacement method;

– FEM in the form of force method;

– FEM in the form of a mixed method.

In the first case, the unknowns are displacements, in the second, forces, and in the third, both forces and displacements.

The most widely used version of FEM at present is the displacement method.

Advantages of the finite element method.

At present, the finite element method (FEM) firmly occupies a leading position in the practice of engineering calculations of construction structures, buildings and facilities. It is a powerful tool for scientific research.

The following advantages contributed to the enormous popularity of the FEM method:

5.4.1. The generality of the approach when solving various problems of calculating structures (including complex ones consisting of various structural elements of different dimensions).

5.4.2. The relative simplicity of accounting for the interaction of the structure with the environment (mechanical, temperature, corrosion effects, boundary conditions, etc.).

5.4.3. A high degree of adaptability to the automation of all calculation stages.

5.4.4. Natural mechanical interpretation and the possibility of building models based on a physical rather than a mathematical approach, 5.4.5. Possibilities of consideration of inhomogeneous bodies of arbitrary configuration, etc.

5.4.6. The presence of a large number of universal calculation software packages (at present, for example, ANSYS, NASTRAN, ABAQUS, SCAD, Robot Millenium, MicroFE, "LIRA", and many others, characterized by an increasing degree of automation of the generation of finite element networks, the formation and solution of a huge number of algebraic equations, as well as effective and efficient means of specifying initial data and visualizing results (preprocessor and postprocessor).

Disadvantages of the finite element method.

5.4.7. The most important drawback of the method is the need to discretize the entire calculation area, which inevitably leads to a large number of finite elements, and therefore unknown problems, especially for bodies with remote boundaries. In this case, there is a need to solve huge systems of algebraic equations that either take a lot of time or are impossible at all.

5.4.8. FEM is fundamentally unsuitable for solving problems with infinite boundaries (it is impossible to divide an infinite domain into a finite number of elements with finite dimensions).

5.4.9. The use of FEM in some cases leads to discontinuities in the values of the studied quantities since the procedure of the method imposes continuity conditions, usually only at the nodes.

At the same time, if we are talking about the prediction of cracks and destruction, this disadvantage becomes an advantage.

5.4.10. FEM is fundamentally inapplicable for solving problems of the stability of slopes, inclinations, and the like, if the stability coefficient is less than or equal to one.

At the same time, if we are talking about the prediction of cracks and destruction, this disadvantage becomes an advantage.

Next, let's focus on the most frequent mistakes that engineers make when using the finite element method while solving soil mechanics and geomechanics problems (Tables 1–4).

In order for our comments to be understandable and transparent, we considered the problem of determining the stress in a square plate with dimensions of 10x10 meters and a thickness of 1.0 meter with the following initial data:

1. The modulus of elasticity of the plate material is E=2000 t/sq.m.

2. The Poisson's ratio of the plate is ν =0.3.

3. A vertical force equal to 100 tons is applied to the upper face of the plate.

Five calculations were performed.

1. In the first case, ties were imposed on the sole and two sides of the calculation area, prohibiting movement in the vertical (axis 0z) and horizontal (0x) directions. At the same time, the calculation area was divided into square finite elements with dimensions of 1.0x1.0 m.

The results of these calculations are given in the first columns of Tables 1, 2, 3, and 4.

2. In the second case, ties were imposed on the sole and two sides of the calculation area, prohibiting movement in the vertical (0z axis) and horizontal (0x) directions. At the same time, the calculation domain was divided into finite elements in the form of right triangles with legs equal to 1.0 m and a diagonal equal to $\sqrt{2}$. Moreover, the finite elements were oriented in the opposite direction – symmetrically relative to the vertical passing through the point of application of the force.

The results of these calculations are given in the second column of Table 1.

3. In the third case, ties were imposed on the sole and two sides of the calculation area, prohibiting movement in the vertical (axis 0z) and horizontal (0x) directions. At the same time, the calculation area was divided into square finite elements with side dimensions of 0.1x1.0 m and 1.0x1.9 m. The location of these finite elements and the results of the calculations are given in the second column of Table 2.

4. In the fourth case, ties were imposed on the sole and two sides of the calculation area, prohibiting movement in the vertical (axis 0z) and horizontal (0x) directions. At the same time, the calculation area was divided into finite elements in the form of rectangles with dimensions of 0.4x1 m, and the total height of the calculation area was reduced from 10 to 4 meters (i.e., 2.5 times).

The results of these calculations are given in the second column of Table 3.

5. In the fifth case, ties were imposed on the sole of the calculation area, prohibiting movements in the vertical (axis 0z) and horizontal (0x) direction. There are no connections on both sides. At the same time, the calculation area was divided into square finite elements with dimensions of 1.0x1.0 m.

The results of these calculations are given in the first columns of Table 4.

The data presented in Tables 1–4 allowed us to draw the following conclusions:

1. As a result of the replacement of rectangular finite elements with triangular ones, a significant distortion of the stress isofields occurred. At the same time, the stresses given in column 2 of Table 1 differ from the reference stresses (column 1 of Table 1) by 1.1–2.6 times.

2. As a result of replacing equilateral rectangular finite elements with non-equilateral finite elements, a significant distortion of tension isofields occurred. At the same time, the stresses given in column 2 of Table 2 differ from the reference stresses (column 1 of Table 2) by 1.06–4.35 times.

3. As a result of the decrease in the height of the calculation area, there was a significant distortion of the stress isofields. At the same time, the stresses given in column 2 of Table 3 differ from the reference stresses (column 1 of Table 3) by 1.05–3.1 times.

4. As a result of the removal of connections on the sides of the calculation area, a significant distortion of the stress isofields occurred. At the same time, the stresses

Table 1.

Dependence of stress distribution on the shape and location of the finite elements

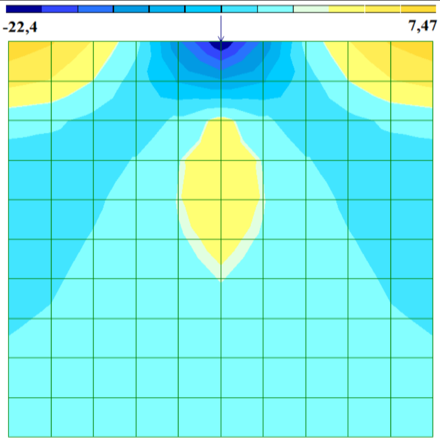
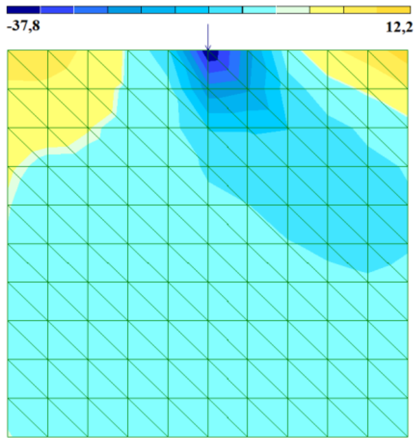
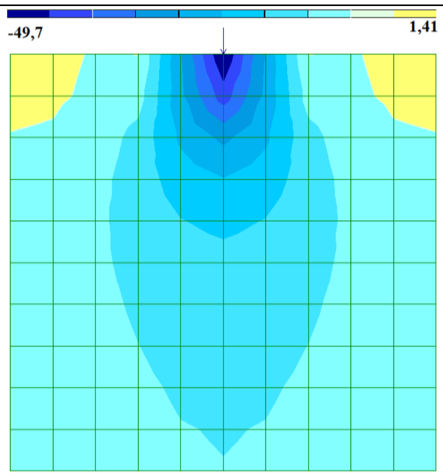
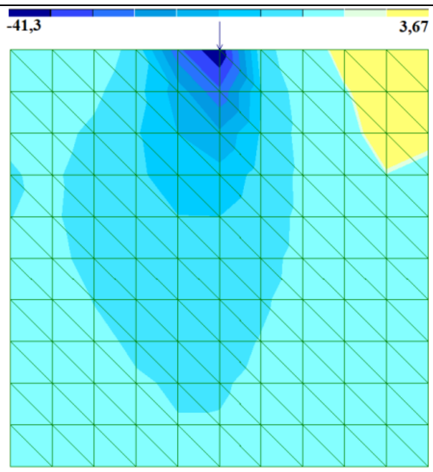
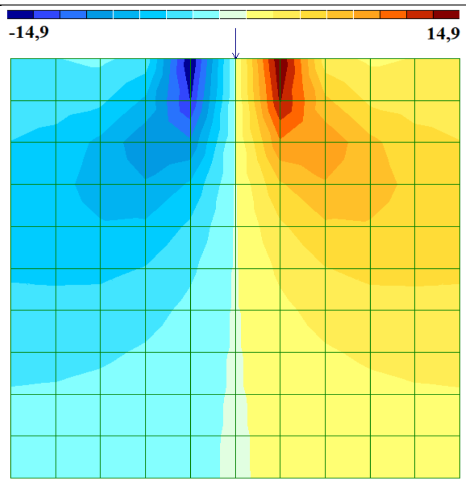
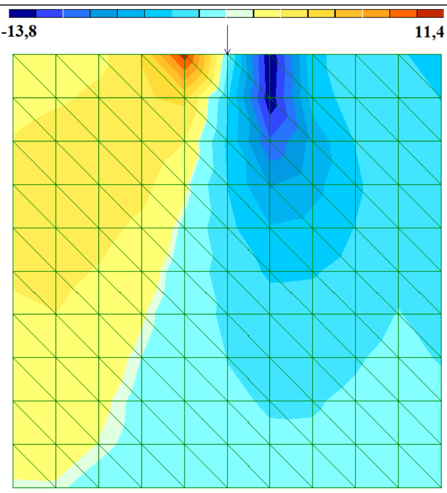
Splitting into rectangular elements (reference)	Splitting into triangular elements	Notes
1	2	3
		Horizontal stresses σ_x. The results in columns 1 and 2 differ by a factor of 1.63–1.69
		Vertical stresses σ_z. The results in columns 1 and 2 differ by 1.2–2.6 times
		Tangential stresses τ_{xz}.The results in columns 1 and 2 differ by a factor of 1.1–1.3

Table 2.

Dependence of stress distribution on the ratio between the sides of the finite elements

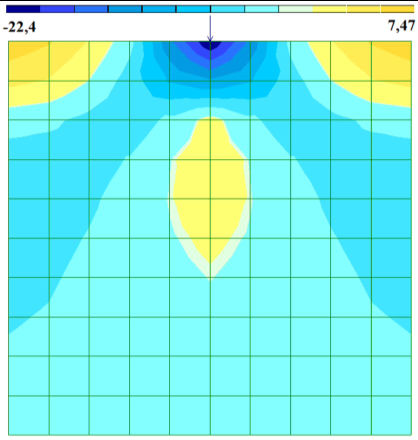
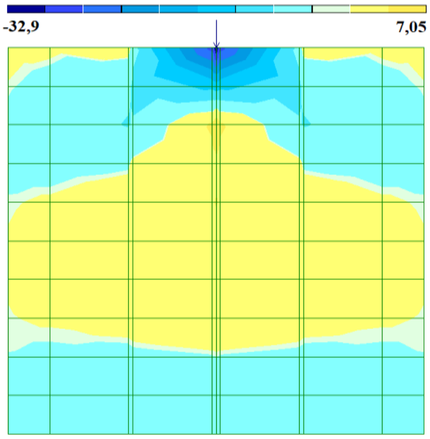
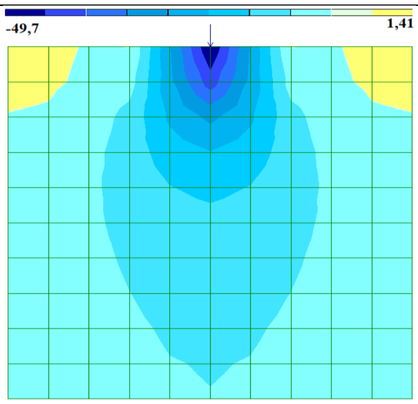
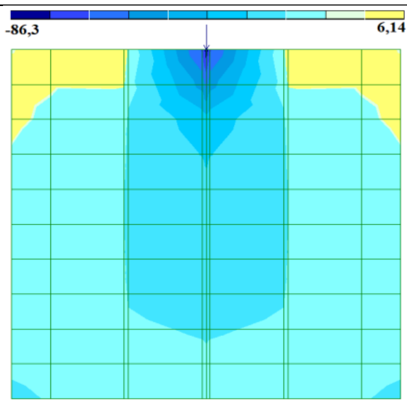
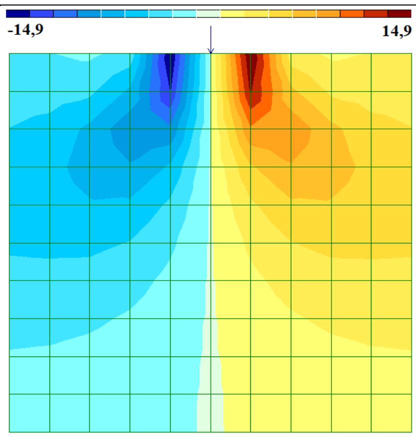
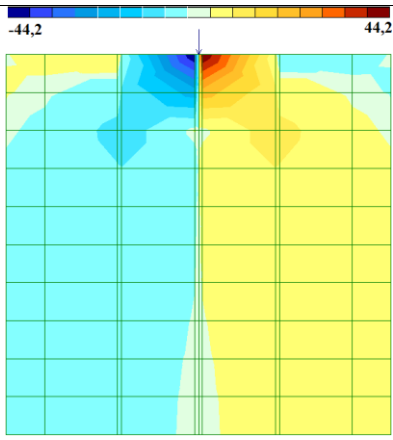
Splitting into equilateral rectangular elements (reference)	Splitting into rectangular elements with different aspect ratios (1:19 horizontal)	Notes
1	2	3
		Horizontal stresses σ_x. The results in columns 1 and 2 differ by a factor of 1.06–1.47
		Vertical stresses σ_z. The results in columns 1 and 2 differ by 1.74–4.35 times
		Tangential stresses τ_{xz}. The results in columns 1 and 2 differ by a factor of 2.99

Table 3.

Dependence of stress distribution on the dimensions of calculation area

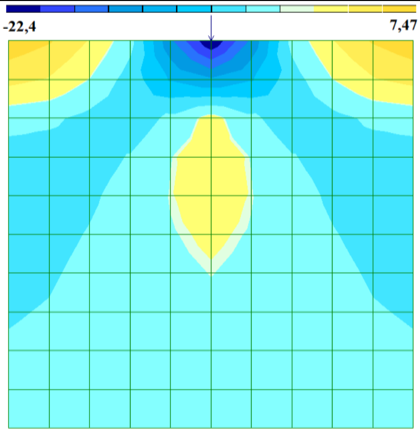
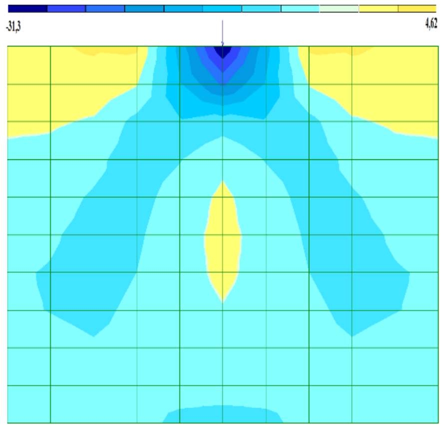
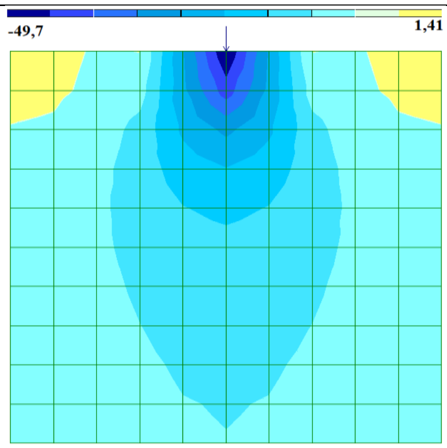
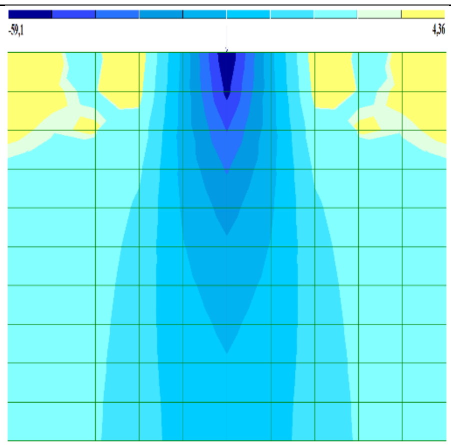
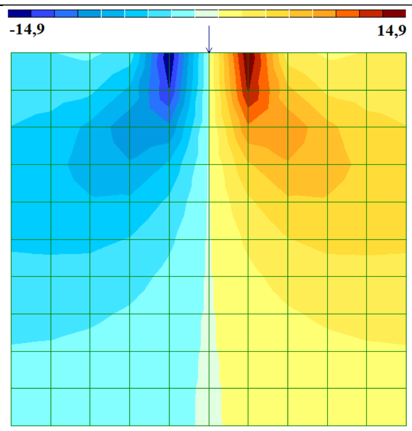
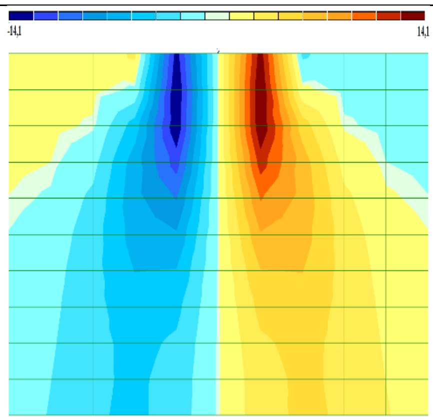
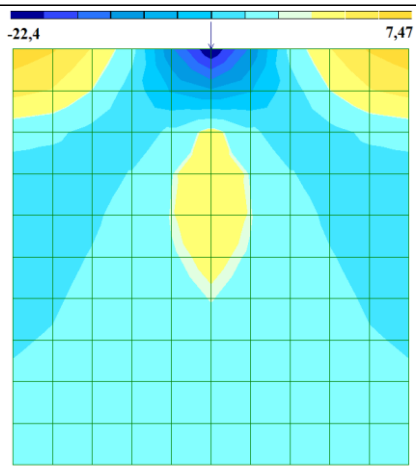
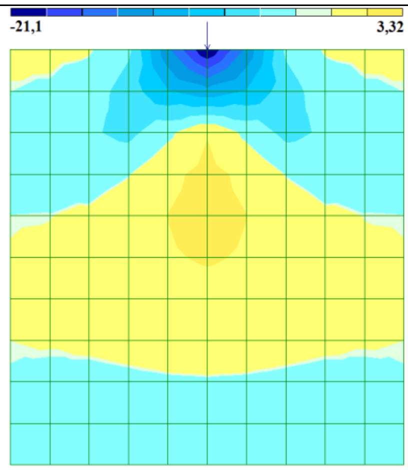
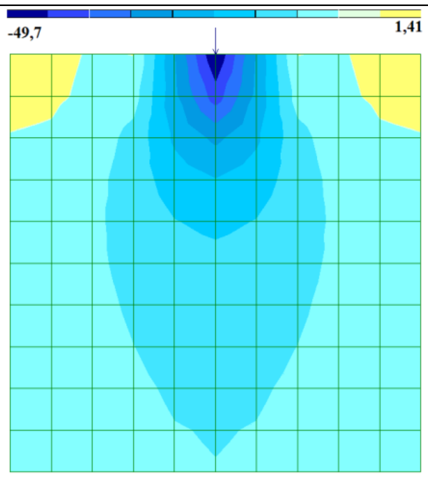
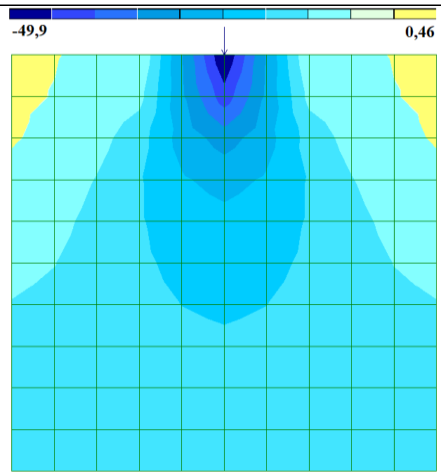
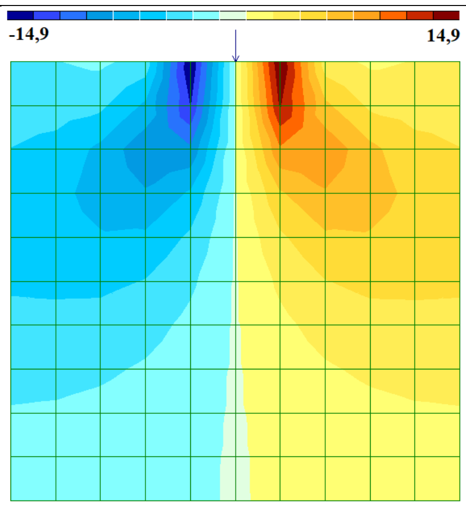
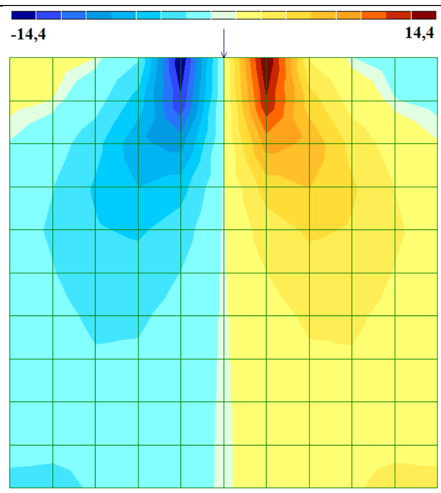
Splitting into rectangular elements (reference)	Splitting into rectangular elements. The height of the calculated area has been reduced by 2.5 times	Notes
1	2	3
		Horizontal normal stresses σ_x. The results in columns 1 and 2 differ by a factor of 1.40–1.62
		Vertical normal stresses σ_z. The results in columns 1 and 2 differ by a factor of 1.20–3.10
		Tangential stresses τ_{xz}. The results in columns 1 and 2 differ by a factor of 1.05

Table 4.

Dependence of stress distribution on anchoring conditions at the border of the calculation area

Splitting into rectangular elements (reference)	There are no fasteners on the sides of the calculation area	Notes
1	2	3
		Horizontal normal stresses σ_x. The results in columns 1 and 2 differ by a factor of 1.06–2.25
		Vertical normal stresses σ_z. The results in columns 1 and 2 differ by a factor of 1.10–3.05
		Tangential stresses τ_{xz}. The results in columns 1 and 2 differ by a factor of 1.05

given in column 2 of Table 1 differ from the reference stresses (column 1 of Table 1) by 1.05–3.05 times.

These data are an illustration of the following basic rules that must be observed when performing calculations using the finite element method technique:

1. The dimensions, shape, aspect ratio of the finite elements, conditions at the boundary of the computational domain and its dimensions should be controlled.

2. To obtain adequate results, the results obtained using the finite element technique should be compared with exact data or with data obtained using the semi-empirical methods set out in SCN.

3. If the requirements set out in paragraph 2 are not feasible, it is necessary to perform several parallel calculations of the same problem using different software packages, sizes, and configurations of finite elements, etc. (according to the requirements of the Ukrainian State Construction Standards, these requirements must be met when designing high-rise buildings).

4. When using one or another computer package intended for calculations applying the finite element method technique, one should definitely familiarize oneself with the requirements set out in them for the size, shape, aspect ratio of the finite elements, conditions at the boundary of the computational domain and its dimensions.

The above fully applies to other numerical methods of soil mechanics and geomechanics in general.