A review of contact algorithms

N. G. Bourago and V. N. Kukudzhanov

The Institute for problems in mechanics of RAS, Moscow, burago@ipmnet.ru


translation into english
A review of contact algorithms

N. G. Bourago and V. N. Kukudzhano

The Institute for problems in mechanics of RAS, Moscow, burago@ipmnet.ru

Moving interfaces between media play an important role in technological and natural processes. The development of methods for solving problems with moving interfaces is one of the major aims of continuum mechanics. This review concentrates on parts of the numerical methods of continuum mechanics named contact algorithms that serve to track and calculate moving interfaces such as contact, phase change, and moving free boundaries.

1. Types of contact algorithms under consideration

Contact algorithms can be classified according to the concept utilized for the description of motion of a continuous medium. In Lagrangian contact algorithms, the nodes move with the velocity of the material medium. In non-Lagrangian contact algorithms, the nodes either are fixed (Eulerian algorithms) or move independently of the material medium (Arbitrary Lagrangian–Eulerian (ALE) algorithms). A characteristic feature of non-Lagrangian algorithms is the occurrence of convective terms in the evolution equations due to a difference in the velocities of the grid (coordinate system) and the medium.

In both cases (Eulerian and Lagrangian), the moving interfaces can be treated explicitly by tracking algorithms as a set of surface nodes (or markers) and cells or be defined implicitly by capturing algorithms. The capturing algorithms are based on continuous marker functions that take on a certain constant values at the moving boundaries.

This known classification underlies the systematization of the publications to be reviewed. In the present paper, we consider all types of contact algorithms, irrespective of the types of contacting media, which is in agreement with the modern tendency to unify the methods of solid mechanics and hydrodynamics. This tendency is accounted for by the requirement for construction of unified computational models for technological and natural processes.

2. Types of boundaries under consideration

In addition to the classical initial-boundary value problems of continuum mechanics, the contact problems involve specific boundary conditions (kinematic and dynamic constraints, phase change laws) that govern the motion of the interfaces and possible boundary singularities.

For the classical contact problems, such conditions express the impenetrability constraint, the action-reaction law (Newton’s third law), and the surface friction law. The normal contact constraints prevent mutual penetration of immiscible media, while the tangent contact constraints represent friction between the contacting media. The extended physical-chemical formulation involves also boundary conditions for heat transfer, electromagnetic interaction, diffusion, chemical reactions, etc.

Additional relevant cases of boundary conditions determining the behavior of free and phase change boundaries are also considered. The moving free boundaries are Lagrangian surfaces between a condensed (liquid or solid) medium and a rarefied medium. The boundary conditions at the free boundaries describe the influence of the rarefied medium characterized by the forces of external pressure and friction, as well as the surface tension forces, which depend on the orientation and the curvature of the boundary (besides the dependence on state variables).

Unlike the classical contact and free boundaries, the phase change boundaries are non-Lagrangian, and the motion of such boundaries in a continuum is governed by the phase equilibrium conditions (such as, for example, Stephan’s law, Chapman-Jougeot detonation condition, yield condition, fracture condition, etc.) The phase change boundaries are weak discontinuity surfaces that move along the continuous medium tracking the phase change process characterized by sharp changes in the properties of the continuum. The temperature, velocities, displacements, and stresses are continuous at the phase change boundaries, whereas the coefficients of elasticity and plasticity, heat capacity, compressibility, and other media properties can undergo a jump discontinuity.

One more important special limiting case of contact is a contact of deformable solid and liquid media with rigid bodies. This type of contact is also considered in the present review.

---

3. Reviews of formulations of contact problems

Initial-boundary value problems can be formulated in the differential, integral or variational forms. The variational formulations, especially the formulations in terms of variational inequalities, play an important role in the analysis of the existence and uniqueness of solutions of initial-boundary value problems and their well-posedness.

Formulations of contact problems have been dealt with in [136, 156, 170, 193, 369-371, 410, 423, 501, 503, 544, 545, 556, 557], where one can find references to additional sources. Theoretical foundations for the statement of contact problems related to phase changes have been reviewed in [76, 199, 299, 351, 420, 421, 428, 524]. Bibliographic lists of publications devoted to the existence and uniqueness of solutions of contact problems can be found in book of collected reviews [587].

In many of the modern publications on contacts of media the initial-boundary value problems are formulated in the Galerkin variational form. In this case, the boundary conditions are treated as constraints to be taken into account in the variational equation by using Lagrange multipliers or penalty functions. For a review of methods to take into account constraints in general variational problems, see, for example [497].

A reviews of Arbitrary Lagrangian-Eulerian formulations of continuum mechanics problems are given in [63, 89, 165, 214, 282, 299, 361].

4. Reviews of methods for contact calculations.

A large number of recent reviews of numerical-analytical methods for contact calculations, initiated by Hertz's work [275], are given in the book of collected reviews [587].

In the present paper, we consider numerical-discrete methods for the calculation of contact boundaries. These methods comprise finite-difference, finite-volume, and finite-element algorithms, as well as Galerkin's "meshless" methods and contact algorithms of the boundary element method.

Large bibliographies on contact algorithms in solid mechanics have been presented in [6, 76, 99, 199, 232, 233, 289, 331, 374, 397, 625, 627] and in a number of other publications cited in what follows farther in connection with algorithms of specific types.

Reviews of algorithms for calculating the interfaces of immiscible media within the hydrodynamics framework are presented in [11, 12, 58-60, 76, 214, 315, 359, 360, 494, 509, 538].

5. Lagrangian algorithms for calculating contacts with rigid bodies.

An overwhelming number of publications on contact problems relates to contact of deformable solid and liquid media with rigid bodies (walls, punches, impactors or obstacles). In this case, the moving boundary of rigid bodies is regarded as a prescribed slip surface. This surface can be smooth or rough. The motion of the rigid bodies is either assumed to be predetermined or calculated by using the methods of theoretical mechanics taking into account the mass of the rigid bodies and the contact reaction forces.

Examples of calculations for contacts with rigid bodies, as well as additional references, can be found in [31-33, 198, 248, 344, 346-348, 374, 431, 485, 495, 616, 617]. Reviews of investigations in this direction are given in [150, 199, 550]. In Lagrangian contact algorithms, the velocities at the boundaries between rigid and deformable media are either prescribed or determined by the penetration of the nodes of the moving grid in deformable media into the prohibited spatial regions that represent rigid bodies. The velocity and displacement components normal to the surface of the rigid body are corrected to prevent the penetration. In many algorithms, such a correction is reduced to the equating of normal velocities of the deformable and rigid boundaries. The correction can be also performed by the elimination of the penetration by applying external normal loads, which represent the contact pressure. The calculation of friction forces does not have specific features in comparison with more general contact algorithms and is considered farther.

6. Lagrangian capturing contact algorithms.

In the Lagrangian capturing algorithms, the solution is continuous at the contact boundary and contact discontinuities are modeled by large gradients of the solution.

Models of ideal contact. Matched grids. A simple approach to the calculation of a contact in the case of small strains for a prescribed (possibly moving) Lagrangian contact boundary involves an approximate representation of the full adhesion conditions that characterize the ideal contact. The Lagrangian grids in the contact region are matched node to node, slip and rebound (coming unstuck) of the contacting bodies are prohibited. The solution at the contact boundary is continuous for the displacement and velocity (ideal contact). There are many publications in which such a scheme is used for contact calculations. Any algorithm for solving continuum mechanics problems by a grid based method can serve as an example. If
different properties of the material have been prescribed in different domains, such an algorithm implements the conditions of ideal contact automatically.

Typical examples of calculations for the ideal contact can be found, for instance, in old Russian papers on finite-element method published 20-30 years ago [382, 440, 491, 493, 514, 515, 542, 582, 586]. For references to international publications on the ideal contact, see, e.g., [632].

In the general case of variable contact region, where the slip and coming unstuck of the contacting bodies can occur, the ideal contact model, which ignores these phenomena, is physically inadequate and, therefore, is not utilized.

Ideal contact. Unmatched grids. A contact algorithm for matching the solutions to satisfy the ideal contact conditions in the case of unmatched grids in the contact region has been proposed in [47] for 2D case and developed in [49, 50] for 3D case. Apart from the contact calculations, this algorithm is utilized for the calculation of joints in composite structures. This algorithm does not require one to worry about the matching of the grids at the boundaries of subdomains to provide continuity of the solutions. This simplifies the construction of grids in combined 3D subdomains of complex shape. Recently these algorithms have been rediscovered [191, 477, 478].

Buffer layer algorithms are based on the introduction of a fictitious buffer contact layer (contact pseudomedium) between the contacting bodies. The buffer layer consists of the contact cells, the nodes of which belong to the contacting boundaries. The introduction of the buffer layer of cells reduces the contact problem for many bodies to a problem for a single composite inhomogeneous body. As a rule, only one cell is utilized across the buffer layer thickness. Depending on the prescribed properties of the contact, these cells can play the role of an elastic spring, a viscous damper, an adhesive contact, etc. The stresses acting in the buffer layer imitate the contact loads. The success or failure of this imitation depends on the properties prescribed for the material of the buffer layer. These properties should provide the appearance of compressive contact loads, prevent the appearance of tensile contact loads (to model coming unstuck), and model friction forces. The resulting mathematical model should be well-posed. In addition, for “cosmetic” and accuracy reasons, the buffer layer thickness is desired to be much smaller than the characteristic size of the contacting bodies.

The issues of the implementation and theoretical substantiation of the buffer layer algorithms have been considered, for example, in [155, 211, 362, 381, 430, 433, 447, 448, 480, 492, 516, 518, 519, 555, 588, 619–621].

Lagrangian united-equation-of-state algorithms serve for the calculation of the development of internal contact discontinuities such as macro-cracks. In these algorithms, cracks are modeled by narrow internal zones with weakened resistance of the material to deformation. These zones are formed if some damage criterion has been satisfied. For the description of such through calculation algorithms, see, e.g., [65, 66, 87, 88, 96–98, 100, 199, 253, 254, 300, 312, 349–351, 374–376, 404, 405, 416, 442, 457, 527, 576]. If the damage criterion (for example, formulated in terms of the limiting principal strains or stresses) has been satisfied in a Lagrangian cell, the elastic moduli and stresses relax to zero forcing strain localization and only the compression strength is preserved in the material. A comprehensive review and the description of this type of through calculation algorithms is given in [199]. It can be easily seen that the narrow damage zones in such algorithms are similar to the buffer contact layers that have been considered previously.

Multiple contact. Problems of multiple contact of deformable bodies appear in many applications, for example, in the numerical study of the properties of composite materials consisting of many contacting components, in the calculations associated with hitting a target with a case-shot or in the analysis of the interaction between rough surfaces. Publications in this direction have been reviewed in [199, 220, 235, 363, 599, 602, 606].

Algorithms for calculating multiple contact in accreting bodies have been considered in [16, 420, 421]. This type of problems covers the layer-by-layer formation of composite materials, loading of a soil during the construction work, growth of crystals in the process of solidification of metallic melts and polymer solutions, spraying, deposition, and freezing.

When solving multiple contact problems, helpful are the models and algorithms for through calculation of contact boundaries that have been utilized for studying the strength properties of composite materials by considering the deformation of an idealized small domain containing a fairly large number of dissimilar contacting subdomains (the matrix and inclusions).

As the number of the contacting elements increases, the direct numerical modeling becomes complicated and one has to involve continuum models for the multiple contact. For example, an effective method for solv-
ing multiple contact problems for sandwich and block structures is based on the asymptotic homogenization for media with periodic structure. A review of publications in this direction is given in [446], with particular cases of homogenization models and their numerical implementation being presented. The approach developed in the cited publication allows effective modeling of fracture in masonry structures, a brickwork or layered rock, with delamination and friction on contact surfaces being taken into account. The constitutive relations for these structures are similar to those for an elastoviscoplastic medium. The delamination is taken into account by specific functions.

The modeling of non-Lagrangian interfacial boundaries (e.g., liquid-vapor or liquid-solid for melting or crystallization) [531, 538] on the basis of united-equation-of-state algorithms will be considered below when describing Eulerian through calculation contact algorithms.

7. Contact detection algorithms.

For the numerical calculations on the basis of Lagrangian grid methods, the boundaries of bodies are represented by a set of boundary surface cells. In most cases, the contact region is unknown in advance and has to be detected during the calculation. It is identified either by the penetration of “alien” boundary nodes through “friendly” boundary cells in the preliminary calculation, in which the contact is not taken into account, or by the approach of the boundaries to a prescribed small distance which the fact is detected by the pairwise check of the mutual positions of boundary nodes and cells. The search for the contact region results in a list of contact pairs (for example, of the “alien node-friendly cell” type), which introduce the contact or buffer elements. Sometimes, the contact pairs can be represented by combinations of “node-fictitious node” or “cell-cell”, i.e., by pairs of discrete elements of the contacting boundaries.

The number of operations for searching for contact pairs is proportional to the square of the number of boundary cells or nodes. For problems with large number of nodes this leads to unacceptably great computational effort. Further, we consider available algorithms for accelerated search for the contact region.

Master-slave (node-to-segment) algorithm is one of the first contact searching (detection) algorithms. It was proposed in [238, 260]. To reduce the computational effort, the regions of possible contact are specified in advance. One of the contacting surfaces is considered to be the master surface and the other (slave surface) is subordinated to it. The master surface is represented by the boundary cells and the slave surface by the boundary nodes. The algorithm utilizes a priori information about the region of possible contact and identifies contact pairs by the penetration of the slave nodes through the master’s boundary cells. To detect the penetration, the algorithm checks the sign of the normal projection of the slave node on to the master cell and the fact that the normal dropped from this node onto the cell intersects this cell.

Note that in many problems, it is impossible to predict master and slave contact surfaces in advance. The necessity to describe the contact regions in the initial data for bodies of complicated geometry is rather burdensome. An additional argument against the utilization of a priori information about the contact region is the possibility of self-contact. By self-contact, the contact between different parts of the surface of the same body is understood. Self-contact can occur in the case of large deformation.

In the contact searching algorithms to be considered in what follows, the search process is divided into two or more levels to accelerate the process of detection of contact pairs. These levels are usually referred to as the global and local ones. On the global levels, the regions of possible contact are searched for among groups of neighboring nodes. The groups of nodes that lie far away from the region of possible contact and, therefore, are not involved in contact are rapidly discarded in accordance with an appropriate group criterion associated with the distance. On the last (local) level, contact pairs “node-boundary cell” are identified by the violation of the impenetrability constraint or by a sufficient proximity criterion. The contact searching algorithms differ from one another in the criteria for grouping the nodes, group characteristics, hierarchy, and methods for quick sorting.

Among the well-known global algorithms of searching for contact region, we mention the regular cell algorithm [530], the hierarchy-territory algorithm [626, 628], the linear position code algorithm [623, 624], the bucket sorting algorithms [64, 75], and the spherical sorting algorithm [472].

The hierarchy-territory algorithm and the linear position code algorithm are most widely utilized and efficient global contact searching algorithms.

The Hierarchy-Territory Algorithm (HTA) [626] is based on the grouping of the boundary elements lying close to one another and searching for possible contact regions by means of the analysis of the distances between the groups of elements. After having found possible contact groups, the local search is performed.
If the number of the boundary elements is very large, a hierarchy of the groups is constructed and the search for a possible contact occurs consecutively from higher-level to lower-level groups.

The Linear Position Code Algorithm (LPOCA) [455] artificially orders all boundary nodes of the finite-element grid. To that end, the parallelepiped bordering the solution domain is divided into a large number of small “bricks” with a structured $ijk$ numbering, i.e., an additional uniform regular $ijk$ grid is introduced. Each node of the unstructured finite-element grid is assigned an $ijk$ number, depending on the small “brick” entered by this node. Such an additional numbering contains information about the arrangement of the nodes, which enables one to use this numbering for forming clusters (groups of neighboring nodes) and reducing the number of checks, thereby accelerating the search for contact regions (see also [21, 273, 418, 419]).

The Space Filling Curve algorithm (SFC) [157] hierarchically divides the solution region into squares as is the case in the familiar problem of catching a lion in Sahara desert. Parts of the desert are consecutively bisected with the selection of the part in which the lion is located. The bisection continues until the current part becomes so small that the lion has no place to hide. In a similar way, the SFC algorithm divides the searching domain into four rectangular parts (in the 2D case) until only one node remains in the cell. This node is assigned a position code (an address) formed by a chain of binary codes (00 for the left lower part, 01 for the right lower part, 10 for the left upper part, and 11 for the right upper part) to indicate the path to the given node. To detect the neighboring nodes for the given one, the algorithms of quick comparison of position codes are utilized. These algorithms are based on XOR-operations combined with binary and exponential search.

The local inside-outside algorithm [590] effectively copes with the “deadzone” problem. The matter is that at the inflections of the boundary there appear internal dead zones and it is unclear where the alien boundary node penetrated the boundary and situated in the dead zone could be “pushed out”. The algorithm suggests a simple strategy, according to which this node should be pushed out backward by means of the reaction forces directed along the Lagrangian trajectory of the node (inside and outside in the same track).

The local gap function algorithm (GFA) [280], according to the concept of its authors, should enable all contacts for problems characterized by very complicated geometry and very high dimension to be considered in a unified way. The algorithm is based on the scalar gap (material depth) function defined in the domain of the solution. For each node, this function is calculated one time for the initial position of the moving boundary and is equal to the initial distance from this node to the nearest boundary. Outside the solution domain this function takes on zero value. Penetrated alien node lies inside some cell of the spatial grid and, hence, is characterized by two values of the gap function. One of these values (the proper value for this node) is equal to zero, while the other one corresponds to the position of the node inside the cell and is nonzero. The direction of push back contact forces is defined by the negative of the gap function gradient; the penetration depth and reaction force magnitude are proportional to the gap function value. The efficiency of the gap function algorithm depends on the efficiency of the interpolation of the gap function.

The gap function approach is criticized in [470]. According to this study, preferred are conventional contact pair algorithms, which fit better to the general case of nonsmooth boundaries with interactions and the case of multiple contact.

The local pinball algorithm [68, 69] has been proposed to identify complex cases of contact of dissimilar elements (rods, shells, an spatial members). This method considers a spherical neighborhood of each element irrespective of its nature (3D, 2D shell, or 1D rod/beam) and contact detection is simplified to an interference check between these bounding spheres. This procedure is simple, since contact takes place if current distance between centers of pinballs is less than the sum of their radii. Push back contact forces are applied to the centers of the overlapped pinballs. The magnitude of push back forces is proportional to the overlap measure. Then these forces are recalculated from pinball centers to the nodes of the elements to which these pinballs are related. Since the pinball algorithm is based on simple checks, this algorithm is very quick. It spends about 15% of time required for the calculation of one time step, while other algorithms sometimes spend more than 65–70% [76].

The hierarchical pinball algorithm has been developed to remove the drawbacks of the pinball method. These drawbacks were indicated in the paper [69], in which the pinball method was utilized for thin shells. It turned out that the pinball method failed if the contacting elements were very thin. To overcome this difficulty, two modifications of the pinball algorithm have been proposed—the quick pinball algorithm and the splitting pinball algorithm [70]. As is the case for the conventional pinball algorithms, in the splitting
algorithm, a pinball is associated with each element, although in the modified method the radius of the pinball is always chosen to be large enough to surround the element. This large pinball is referred to as the parent one. The overlap of the parent pinballs indicates only the possibility of the penetration. If the overlap of the parent pinballs has been identified, smaller lower-level pinballs are arranged. The radii of these pinballs coincide in order of magnitude with the thickness of the shell. The smaller pinballs, distanced by this radius from one another, cover the surface of a shell element or the length of a beam element. The overlap of the smaller pinballs with alien pinballs indicates the penetration. In this case, the desired contact forces are calculated in accordance with the overlap of the small pinballs and then are recalculated for the nodes of the corresponding surface grid elements.

In a frame of meshfree Galerkin methods an effective meshfree contact algorithm based on the principle of moment of meshfree interpolation is proposed in [459] (see also [57, 458]).

8. Lagrangian contact algorithms. Calculation of contact forces and velocities.

Let the contact region has been identified and let the node-cell contact pairs have been constructed. Consider algorithms for calculation of the contact velocities and forces.

Slip contact algorithms have been proposed in pioneering works by M. Wilkins [511, 593–598] on the contact between deformable bodies in the case of variable contact region. The works by M. Wilkins have greatly influenced the development of the contact algorithms for the case of large deformation.

In the slip algorithms, at each time instant, one of the contacting surfaces (alternately) is regarded as the slip surface (a surface of a rigid body), while the other surface follows the motion of the first one. Although the contact conditions in this case are satisfied approximately, on the average, this method provides plausible results. The contact forces and velocities in the numerical solution obtained by means of this method oscillate, which reduces the accuracy of the computation.

For the extension of Wilkins method to 3D case, see, for example [48, 248, 310, 311, 313, 352, 353].

Fictitious node contact algorithms have been introduced in various forms in a number of publications. In the paper [47], which has already been mentioned in connection with ideal contact algorithms, a 2D algorithm utilizing the impenetrability constraint along the normal and the free slip condition along the tangent has been proposed. In the case of mismatch of the grids in the contact region, this algorithm uses an auxiliary matched grid that consists of the boundary nodes of the primary grid in one of the bodies and the corresponding (having the same position) fictitious boundary nodes in the other body. The values of the solution at the fictitious nodes on the old time layer (iteration) should be determined by means of interpolation. On the new time layer, each pair of adjacent (primary-fictitious) boundary nodes is subjected to the kinematic constraint (equality of the corresponding normal velocity components) and the static condition (equality in magnitude of the normal force components). These relations are utilized to determine the contact pressure magnitudes and correct the boundary nodal velocities at each boundary node of the primary grid. This algorithm has been extended to 3D contact problems in [50].

Fictitious nodes have been introduced also in many other contact algorithms, for example, in the characteristic algorithms [617] to be considered below and in the hierarchical pinball algorithm [69] that has been considered previously.

A logical completion of the line of fictitious-node algorithms is provided by the adaptive contact algorithms. These algorithms at each time step reconstruct the grids locally in the contacting bodies in the neighborhood of the contact region to provide the node-to-node match of these grids [434].

Characteristic contact algorithms. The relations on the characteristics for hyperbolic systems of equations of mechanics of elastoplastic media have been utilized in [347, 348, 485, 617] for the calculation of contact boundaries in 2D problems. The systems of characteristic relations have been written out for each boundary node in one body and the corresponding fictitious node in the other body. The solutions of these systems of equations have been used to determine the contact velocities and forces. The characteristic algorithms can be applied only to unsteady hyperbolic problems. For an update review of the characteristic algorithms, see, for example, [378].

Contact algorithms based on the Riemann problem and Godunov scheme. The Godunov scheme [223] utilizes the solution of the Riemann problem for the decay of an arbitrary discontinuity. It has been applied by Godunov with co-authors [225] to the calculation of contact interactions in the explosion welding processes. In [225], one can find the description of the contact algorithm and references to other publications of the authors. This line of algorithms has been developed in [1, 9, 10, 520–524]. The family of Godunov algorithms
has been described in the review of [76] and monographs [378, 578].

Inelastic impact contact algorithms have been applied in various modifications in [19, 243, 244, 296, 297, 354, 525, 526, 533]. In these algorithms, the velocities of the Lagrangian nodes in the impact contact region are corrected on the basis of the solution of the problem of an inelastic impact for the nodal masses that form node-surface cell or node-fictitious node contact pairs.

Contact algorithms based on Lagrange multipliers take into account the desired normal contact loads in the virtual work principle or in discrete equations of the contact nodes motion taking into account the impenetrability constraint. The normal contact loads in this formulation are the Lagrange multipliers for the impenetrability constraints. The difference between various modifications of this approach can consist, for example, in the form (continuous or discrete) in which the impenetrability constraints are taken into account, in the methods (direct or iterative) utilized for solving the system of algebraic equations for the nodal contact pressures, in the interpretation (physical or mathematical) of the algorithms, and in the methods of approximation of the solution in the contact region. However, despite the apparent differences, all these modifications are various versions of the realization of the same concept.

Note that most of the contact algorithms applied in explicit computational schemes can be regarded as versions of the Lagrange multiplier method.

For small deformations, the algorithms based on Lagrange multipliers have been constructed in [204, 256, 286] - [262], [17], [39-41, 245, 399]. For large deformations such algorithms have been developed in [50, 90, 91, 93, 95, 109, 113, 228, 230, 251, 355, 357, 473, 507, 529, 546, 547, 551, 623, 624, 628].

Penalty function algorithms provide another version of dynamic contact algorithms applied most frequently in implicit schemes to solve contact problems in quasi-static and dynamic formulations. In these algorithms, the normal load magnitude is assumed to be proportional to the residual of the impenetrability constraints with large coefficients of proportionality (penalty coefficients).

Contact penalty function algorithms for implicit schemes have been developed and described in [18, 41, 93, 95, 111, 119, 202, 245, 258, 291, 325, 329, 331, 356, 377, 409, 450, 451, 474, 488, 546, 606, 609, 626].

The hybrid algorithms, which combine the penalty function and Lagrange multiplier methods, have been also developed; see, for example [43, 181, 182, 272, 388, 401, 548, 600, 608, 609, 618].

Lagrangian algorithms for explicit treatment of internal contact boundaries. An alternative to the through calculations algorithms for main cracks is an approach to the fracture modeling in which contact discontinuities corresponding to main cracks are determined explicitly [82, 129, 293, 440, 517]. In this approach the contact surface is introduced in advance and is defined by paired nodes. In the general case, additional nodes are introduced in the process of solution [244, 432]. Algorithms for reconstructing the grids in the neighborhood of contact discontinuities are reviewed and described in [214, 384, 385].

A method for explicit treatment of newly formed contact discontinuities, based on the local reconstruction of the grid by means of “collapse” of the destroyed cell (by means of shifting the nodes of this cell onto the fracture surface), has been proposed and described in [199, 250, 335]. This method does not require introducing new nodes.

In a number of algorithms, crack-type discontinuities are modeled on the level of elements without reconstruction of the grid [549, 559], by using the moving ALE grid technique [506], as well as by introducing additional degrees of freedom in the elements containing a contact discontinuity [149, 162, 163, 169, 194, 196, 437]. These additional degrees of freedom are associated with specific shape functions.

9. Eulerian algorithms for calculating contact boundaries.

Let us imagine the case where two or more domains occupied by one phase of the material coalesce (e.g., merging liquid drops). It is difficult to model such a process by means of Lagrangian grid methods with explicit treatment of interfaces. This is especially difficult in the case of 3D problems, since the combination of nodes into Lagrangian boundary cells is defined by lists, and these lists would have had to be continuously updated. Additional difficulties would have arisen because of unacceptable Courant’s restrictions for the time step in the case of too close approach of the Lagrangian nodes. Similar difficulties are characteristic of the Lagrangian approach when applied to the modeling of fragmentation processes (for example, separation of a liquid drop).

These difficulties in tracking the interfacial boundaries can be overcome by utilizing Eulerian and Eulerian-Lagrangian front-tracking algorithms. In the broad interpretation of this review, these algorithms are treated as contact algorithms.
Reviews of Eulerian front-tracking algorithms. Eulerian methods for through calculation of contact discontinuity form a specific rich world of algorithms and deserve a separate review. In the present review, these methods are described briefly. Other bibliographies of publications on the Eulerian contact algorithms can be found, for example, in [59, 76, 151, 152, 189, 214, 292, 358, 460, 462, 463, 494, 536, 537, 564, 584].

Characteristic of this group of contact algorithms is that the calculation is performed on an Eulerian (fixed) grid (often uniform and rectangular), which contains the contacting material bodies and media with margin, and that the contact boundaries (common boundaries of media and bodies, free boundaries, and phase change boundaries) are tracked by means of Lagrangian discrete or continuous markers. Sometimes, an Eulerian-Lagrangian arbitrary moving (dynamically adaptive) grid is utilized instead of an Eulerian grid.

The velocity field calculated on an Eulerian grid is utilized for calculating the motion of discrete or continuous Lagrangian markers on the basis of the transport equations in the Lagrangian (for discrete markers) or Eulerian (for continuous markers) form.

Discrete Lagrangian marker algorithms form a large family that involves the basis algorithms for the particle-in-cell method [266], the boundary marker method [449], and the marker-and-cell method [436, 443, 574, 575, 591]. In these algorithms, mass, momentum, and energy is transported by particles, while the markers serve for the identification of the interfaces and the motion of the phases.

Cited above marker-and-cell algorithms deal with fluid-fluid interactions and free boundary motion. Versions of the marker-and-cell method for contact of elastoplastic bodies are presented in [199, 322].

In the case of discrete particle and marker methods when dealing with complex boundary conditions (e.g., friction, surface tension, and phase change), one has to use boundary Lagrangian markers to determine the geometry of contact (interfacial and free) boundaries. The boundary Lagrangian cells enable one to calculate normals, tangents, and curvatures of the interfacial surface to formulate the boundary conditions. As was the case for purely Lagrangian grid algorithms, such a Lagrangian description of the boundaries encounters considerable difficulties when identifying disappearing or arising boundaries. Therefore, the particle and marker methods are good while simple boundary conditions (which do not require calculating the geometrical characteristics of the contact boundary) can be utilized.

Besides, the particle and marker methods strike with problems of correct description of the boundary markers motion, problems of maintenance of conservation laws near boundaries, problems of insufficient number of markers or particles in rarefaction regions and problems of the generation and removal of markers at open boundaries. These problems can be solved in principle, but complication of the algorithms can lead to an unacceptably large number of operations.

Continuous Lagrangian marker methods enable one to simplify the description of boundary conditions and physical phenomena at the contact boundaries and the identification of these boundaries. This is especially important in the cases of variable topology of the subdomains occupied by various phases, when the phases merge or separate. The type of the medium is identified by the values of functions of continuous Lagrangian markers that remain constant along Lagrangian trajectories. These functions obey Eulerian transport equations. The interfacial boundaries are defined as an equal level surfaces of the marker functions. The same method for describing boundaries has been adopted, for example, for describing coast lines in cartography.

Various versions of continuous Lagrangian marker algorithms based on the continuous marker concept are presented in a number of publications, some of which are cited below.

The large particle method has been applied for interfacial boundary tracking in [103, 104, 283]. In these publications, the boundaries between the heavier and lighter media are identified by the level surface of the density.

In the volume-of-fluid method [103, 104, 283, 424], pseudo-concentration method [572], and scalar equation method [318], interfacial boundaries are identified by the level surfaces of the volume concentrations or "color" functions of various media.

In level set methods [174, 380, 461, 463, 464, 536-539, 561, 562], the interfaces are tracked by the level function that indicates the distance to the interfacial boundary.

The idea of boundary tracking based on the concentration or domain function has been described also in publications on the through calculation of the boundaries [176, 456, 531], fictitious domain method [108], and R-function method [492, 519].

The method of continuous markers that take on certain constant values for each of the contacting media
(i.e., have a shape of the Heaviside step function) suffers from interface blurring due to numerical diffusion. The zone of jump change of the continuous marker function from one value to another at the interface gradually increases, which is accounted for by errors in the numerical solution of the Eulerian transport equation for the continuous marker.

This difficulty is not encountered in the level set method applied to a slowly changing marker function based on the distance to the interfacial boundary, since the accuracy of the solution of the Eulerian transport equations for this slowly changing function is much higher.

Nevertheless, conservation laws are violated near the interfacial boundary for all marker (discrete and continuous) methods and, therefore, some additional techniques to control and correct the solution are necessary [219].

As compared with particle methods and discrete marker methods, continuous marker algorithms enable one to calculate normals, tangents, and curvatures of the contact surface in terms of derivatives of level function in Eulerian cells containing the boundary surface, thereby simplifying the computational procedure. The corresponding mathematical expressions for the geometrical characteristics can be utilized in boundary conditions on contact interfaces. These boundary conditions can express, for example, Stephan’s law, or force conditions for pressure, friction, and surface tension.

In many cases, marker methods solve problems that have failed to be solved by methods implying explicit description of contact boundaries. This is the case, for example, for multiple contact or phenomena with variable topology of material subdomains (e.g., breaking waves, fountains, fragmentation of bodies during fracture, drop separation or merging, filling reservoirs, bubbles, and cavitations).

Eulerian methods for through calculation of a contact are especially efficient for transient processes, in which case the errors that violate conservation laws and blur boundaries do not have enough time to become noticeable. This is the case, for example, for impact phenomena with velocities close to that of the sound velocity, detonation, explosion, cumulation, explosion welding, and transient motions of liquids with free surfaces. Note that because of the accumulation of errors in the boundary conditions, the solution accuracy provided by through calculation Eulerian algorithms is frequently insufficient for acceptable modeling of contact phenomena in materials with clearly pronounced hyperelastic properties that characterize the resistance of the material subjected to long-term, slow, low loading. In the case of modeling long-term processes, a loss of accuracy can be observed.

Shock capturing algorithms have been developed for through calculations of discontinuous solutions in Eulerian algorithms of hydromechanics. These algorithms can be applied together with Lagrangian marker methods to increase the accuracy of the numerical integration of the transport equation when tracking interfacial boundaries and contact discontinuities. To familiarize oneself with shock capturing methods, one can turn to reviews and outlines presented in many modern publications on the numerical methods for calculation of compressible flows, e.g., [76, 378, 610]. The further refinement of the solution near the interfacial boundaries can be achieved by applying adaptive grids.

Adaptive grid algorithms reduce the approximation errors of grid methods in high-gradient regions by means of local reduction of the grid step, optimization of the cell shape, and adaptation of grids to subdomain boundaries and solutions, in particular, to discontinuities and boundary layers. Such an approach has been suggested by the analysis of grid approximation errors. This analysis shows that these errors are proportional to the norm of the derivatives and a certain power of the characteristic step of the grid and increase in case of the appearance of salient points or distortions in the grid cell shape [558].

The development of the moving adaptive grid method is dealt with in [89, 121, 153, 214, 224, 244, 281, 294, 295, 305, 361, 407, 411, 412, 563, 577]. In these publications, the reader can find more comprehensive analysis and review of investigations related to the method. Characteristic of such algorithms are the preservation of the number of grid nodes during the solution and increase in the accuracy due to an optimal arrangement of the nodes (adaptation).

An approach to the description of contact boundaries based on local refinement and reconstruction of grids (adaptive mesh refinement method) is presented and discussed in [79–81, 102, 214, 415, 573, 589]. In this method, the number of the grid nodes and cells is variable.

10. From mesh-based algorithms to meshless ones

A tendency to avoid complications associated with the construction of grids and to design a more economic and simple technique for the numerical solution has led to two extensive fields in the development
A simplification of the derivation of equations for the discrete problem is based on the collocation method, in which friction contact loads depend on the jump in the tangential velocity [452]. A detailed analysis of friction laws for calculating large strains is given in [130].


Publications [38, 45, 46, 120, 175, 183, 184, 320, 321, 452, 467, 468] are devoted to the modeling of contact problems. One of these fields deals with boundary element (boundary integral equation) methods, and the other with a large family of meshless free point algorithms.

The boundary element algorithms have been considered in [6, 7, 78, 144, 171, 190, 226, 227, 367, 368]. The boundary element methods are based on the reduction of the classical linear equations of continuum mechanics to boundary integral equations by representing solutions of these equations as the superposition of fundamental solutions corresponding to unit excitations or by using Green's influence functions. The discretization of these equations does not require a volume grid; it suffices to introduce a grid of boundary elements, which reduces the dimension of problems to be solved by one. In a number of algorithms, the stage of the derivation of the boundary integral equations is omitted and the discretized equations are derived straightforwardly on the basis of the fundamental solutions for unit excitations [61, 141, 494].

In nonlinear problems, the solution is constructed by means of external iterations with respect to nonlinear terms. At each iteration, a linearized problem with the classical linear operator is solved. The right-hand sides of these equations involve the nonlinear terms calculated on the basis of the previous iteration. Difficulties arising in this approach are associated with the necessity to calculate volume integrals on the right-hand sides, which requires a volume grid to be introduced, as well as with catastrophic deterioration of convergence of the external iterations as the influence of nonlinear terms increases.

Free Lagrangian discrete algorithms are an attempt to avoid difficulties associated with the distortions of the cells of moving grids when tracking the interfacial boundaries. To that end, the grid is reconstructed step by step with variable neighborhood of grid nodes [2, 12, 159, 160, 214, 218, 438, 481, 583]. The conservative interpolation of the solution on moving grids with variable topology is a nontrivial operation. An algorithm for this operation has been proposed in [2].

Meshless contact algorithms are the further step to the rejection of grids for free Lagrangian nodes (markers/particles). These algorithms are based on the Galerkin-Petrov method with specific basis functions. These functions are not compactly supported but decay rapidly. The basis functions are constructed with the arrangement of the Lagrangian nodes (particles) being taken into account. (The unification of the nodes into cells is not required.) The development of these algorithms can be traced by [71, 72, 125, 127, 168, 215, 408, 413, 414, 429, 438, 458, 459].

In the first studies, Gaussian numerical integration procedures were utilized to derive the system of equations for the discrete problem. This caused certain computational difficulties. Improved numerical integration algorithms (stabilized conforming nodal integration (SCNI) algorithms) are presented in [127]. A simplification of the derivation of equations for the discrete problem is based on the collocation method, which corresponds to the utilization of projection basis of delta-functions.

At the present time, meshless methods have started successfully competing with the traditional grid methods when utilized for the numerical modeling of contact in the case of large deformation and complex rheology [255, 319, 630] and in the case of complex fluid-structure interactions [458, 459].
contact processes with rolling friction. The general Eulerian-Lagrangian approach to rolling is presented in [403].

12. Reviews of problem-oriented contact algorithms.

In this section, we indicate reviews and separate papers that deal with problem-oriented contact algorithms adapted to specific applications.

**Animation contact algorithms** utilize simple non-iterative explicit methods which do not imply the laws of mechanics to be observed and satisfy the impenetrability constraints by means of heuristic approaches to provide a minimum plausibility for the result. The animation algorithms are aimed at the creation of a cartoon (of dinosaur film type) to visualize contact phenomena. A distinctive feature of the animation algorithms, as compared with continuum mechanics algorithms, is that the former algorithms should allow the animator to change the scenario of motion of the contacting bodies and media when creating the cartoon. As an example of algorithms of that type, one could mention the free shape deformation algorithm [534] and the algorithm for imitating the skin/skeleton deformation [406]. For more detail see the review of [54].

**Contact algorithms for applications in medicine and biology.** A more proximate approach to reality is necessary in medicine for modeling surgical operations, designing training programs, and predicting the results of plastic surgery. These goals can be achieved by models that partly satisfy the laws of solid mechanics and take into account the strength properties of materials [166, 239, 343].

These simplified models should be replaced by full-scale contact algorithms (developed within the framework of continuum mechanics) as the desired degree of agreement between reality and the model increases. Continuous increase in the productivity of computers favors the intensification of research on the development of algorithms for modeling the motion of living bodies, with the complex internal structure of these bodies being taken into account. However, there have been very few such works so far and they utilize rather simple mechanical models. Mechanical properties of living bodies have been considered in [203]. The animation of contact interactions of elastic bodies for applications in medicine has been attempted in [34, 435, 570, 631]. Reviews of contact algorithms that can be applied in medicine are given in [280, 342].

For a review of contact algorithms for geomechanics see [439], for flows with free moving boundaries [25, 197, 213, 305, 427, 543, 552, 571, 605, 622], for phase change boundaries [299, 567, 584], for thermal effects [540], for cavitation [592], and for modeling of the behavior of cloths [35]. The publications of [536, 538], on the contact tracking on the basis of level set function methods deserve special attention in view of astonishing variety of applications.

13. Contact interaction optimization algorithms.

In the general case, the distribution of the contact loads is nonuniform and depends on the shape of the contact surface. The contact loads can have undesirable peaks which make worse the characteristics of technological processes and reduce the life time of engineering products. One of the first attempts to make the load distribution more uniform by means of optimizing the shape of the contact surface was done in [139]. The optimization was performed on the basis of linear programming. This theme was developed in [271] where the linear programming was utilized in combination with the finite element method. In [56, 132, 135, 475], the optimization was performed on the basis of nonlinear programming. Objective functions for the optimization of contact surfaces were proposed in [179, 180].

A simplified iterative procedure for smoothing peaks of the contact loads for constant volume of the contacting bodies was proposed in [565]. The further simplification was provided by the evolutionary structural optimization (ESO) algorithm. This algorithm was proposed in [606, 607] and can be applied to a wide class of contact optimization problems.

All optimization problems solved in the cited publications have been considered for the cases of rather simple rheology (linear isotropic elasticity) and geometry (2D static problems for two contacting bodies) and have illustrative character. This direction is at the development stage yet [535].

14. Vectorization and parallelizing of contact algorithms.

Discrete models for the analysis of contact interactions in complex structures involve a very large number of nodes. To obtain solutions for such models in a reasonable time, the utilization of vector and multiprocessor computers can be of help. For that reason, many studies related to contact problems have been devoted to vectorization and parallelizing of the available contact algorithms. Issues of the vectorization of contact algorithms were considered in [92, 93, 95, 217, 256, 259, 261].

Parallel computers of various types were tested in the 1980-90s. At the present time, the most suitable
The domain decomposition method (DDM) turned out to be most convenient. For the fundamentals of this method as applied to contact problems, see [114, 185–187, 417, 453].

The fictitious domains formed in accordance with the domain decomposition method have overlapped boundaries on which the solution must be continuous. This continuity is provided by the exchange of the boundary data between the processors during the iterative process. The conjugate gradient method turned out to be one of the most effective and convenient for parallelizing and vectorization. The application of this method to contact problems has been dealt with in [92, 93, 95, 422, 425, 453, 611, 612].

Considered above contact searching algorithms (such as global LPOCA, HTA and local pinball, master-slave, etc.) primarily are not intended for use in parallel computing and their implementation may sufficiently dismiss the profit in performance expected. In parallel codes the contact surface is represented as a set of boundary subdomains, which are processed by using additional processors, which are different from those used for interior subdomains (see [285, 453]). Parallelized contact algorithms are presented also in [20, 21, 216, 418, 419, 496].

The highest performance of computations achieved for 2001 by means of parallelizing of high-speed impact contact problems was due to a state research organization Sandia National Laboratories (SNL) in the USA. This record has been established by running the parallelized software package PRONTO on the Intel Teraflopp Computer (3600 processors). A speed of 1/10 second per time step was achieved for models with the number of 3D 8-node finite elements exceeding 10–15 million. A number of test problems have been solved, from simple tests to an important practically problem of an aircraft crash, with the deformation in the aircraft structure, fuel hydrodynamics (particle method), and the ground deformation being taken into account. The last problem has been hopeless to be solved numerically until recently. In this problem, all contact phenomena that have been discussed previously (including self contact and multiple contact) and through calculation method have occurred. This record computation is presented in the SNL report [22].

Taking into account an intensive development of parallel computers, changes in their architecture and software, and possible expansion into the PC world, one could anticipate the issue of the adaptation of contact algorithms to parallel computers to become rather topical in the nearest time.

On the other hand, one should not overestimate the value of this direction. The point is that the share of parallel computers among the available computer park has been rather small so far and that such computers have been utilized in a remote access mode. This sharply increases the time of waiting for the solution and creates a paradoxical situation. Formally, the time of computation is sharply reduced but in fact for particular user this time may be even increased because of small speed of data transfer through a network and because of possible competition between users, since parallel computers are multiple-user shared computers. As compared with the traditional scalar programming, the parallelizing of codes requires much more effort and cardinal reprogramming and revision of common scalar algorithms. Therefore, to be a success the parallelizing requires high-performance hardware and a powerful financial support, which is feasible only for large research centers working under large-scale state contracts.

When planning work on the parallelizing of contact algorithms, one should also take into account temporal characteristics, specifically, the life time of the parallel computer and its software, the time required for the development of the parallel version of the code, and the time during which the parallel computer can be regarded as a supercomputer. The last remark is due to the fact that scalar computers (including PCs) are also rapidly developing and very often can compete in all aspects with many types of parallel computers that until recently had been acknowledged to be supercomputers. For example, a comparison of a modern PC Pentium/4 with a ten-years-old parallel "8-head" Parsytec will hardly be in favor of the latter.

15. Accuracy analysis and comparison of contact algorithms.

Virtually all contact algorithms provide only approximate solutions. An important aspect of research related to contact algorithms involves the investigation of the accuracy of such algorithms and a posteriori analysis of errors of the numerical solution. It should be noted that this direction has not been thoroughly developed yet. For the evaluation of the accuracy of Lagrangian algorithms, see, for example, [23, 24, 25, 400, 401, 403, 541, 580, 585, 633]. A review of publications on the evaluation of the accuracy of the Eulerian continuous marker algorithm is given in [115].
A review of contact algorithms

Comparative analysis of various approaches is complicated by the fact that the success or failure of a contact algorithms is influenced by the quality of the code and specific features of the algorithms that have not been documented. Unfortunately, there have been only few publications so far in which the same authors compare various approaches. Comparisons that have been performed at the same “kitchen” provide clearer and more definite results. The point is that a minor feature of the algorithm that seems to be unimportant at first sight and has not been mentioned in papers and reports can frequently play a decisive role in the success or failure of the entire algorithm. This is not due to malicious intent of the authors to “keep secret” but most frequently because of a plenty of components of the algorithm and ambiguity of possible formulations of it.

Paper [509] can serve as an example of classic comparative analysis of contact algorithms. In this paper, the quality of numerical modeling of contact discontinuities by means of various through calculation methods is assessed on the basis of the solution of four test transport problems for a specific distributed scalar substance in prescribed steady velocity fields. These fields correspond to simple translation, rigid-body rotation, an isolated vortex, and a complex strain. Four methods have been tested—the most recent version of the marker-and-cell algorithm, the fluid-in-cell method, the level set method, and the shock capturing methods such as TVD and ENO. These methods are arranged from better to worse as they have been listed above. The continuous marker method has been improved in [174] by means of the new Hybrid Particle Level Set (HPLS) method which combines the discrete and continuous Lagrangian marker techniques in a single algorithm.

Among the publications related to critical analysis of contact algorithms there is a very interesting lecture [315] which presents an ironic collection of typical samples of author partialities in self-estimations of merits of developed codes for continuum mechanics: “it will solve your problem without modifications”; “the manual has everything you need to run the code”; “standardized graphics output, compatible with third party post-processors”; “minimal learning curve”; “executeable on all machines with no modifications”; “robust and accurate”; “all physics are compatible”; “user friendly”; “there are no more bugs in the code, only undocumented features”; “you can run the code without the manual”; “the technique was first developed here”. It is hardly possible to contest this criticism. The very fact of simultaneous existence of numerous contact algorithms indicates that these algorithms are not perfect. The author assessments of algorithms and results that appear in papers should be considered very carefully.

16. Conclusion.

There are many hundred works on numerical methods of analysis of contact interactions that have been published in the world during the last 3 or 4 decades. We confined our review to publications on contact algorithms. If we had included, in addition, studies on the physics of contact and publications on solving particular problems, the list of bibliography would have consisted of several thousand items and the length of the review would have exceeded all reasonable limits.

This review can be used as a guide in contact algorithms. It should facilitate the choice of an appropriate algorithm and help one to assess the novelty of newly designed algorithms and to select publications for more detailed study and citing.

Although a great number of contact algorithms have been designed, the basic concepts of these algorithms can be comprehended and classified. One possible classification has been presented in the given review. This classification is not optimal but has been suggested by the material to be sorted.

Acknowledgments

This review was carried out within the framework of the Programs of the Department of Power Engineering, Mechanical Engineering, Mechanics, and Control Processes of the Russian Academy of Sciences “Structural mechanics of materials and structural elements. Interaction of nano-, micro-, meso-, and macroscale phenomena in deformation and fracture” and “Cumulative damage, fracture, and structural changes in materials due to mechanical and thermal loads” and under financial support of the Russian Foundation for Basic Research (Projects Nos. 01-01-00659 and 05-01-245a).
References


5. V. V. Alexandrov and S. M. Mkhitaryan, Contact Problems for Bodies with Thin Coatings and Interlayers [in Russian], Nauka, Moscow, 1983.


A review of contact algorithms

A review of contact algorithms

Internet: http://www.ipmnet.ru/burago/papers/grid.htm
A review of contact algorithms


A review of contact algorithms


A review of contact algorithms


207. L. A. Galin, Contact Problems of Elasticity [in Russian], Gostekhizdat, Moscow, 1953.


A review of contact algorithms


29
A review of contact algorithms

Internet: http://granity.lanl.gov/history/CFD_paper_5.24.96.pdf


N. G. Bourago and V. N. Kukudzhanov

A review of contact algorithms


A review of contact algorithms


A review of contact algorithms


446. I. S. Nikitin, Dynamics of Laminated Beam Media with Slip and Friction [in Russian], Preprint No. 366, Institute for Problems in Mechanics of the USSR Academy of Sciences, Moscow, 1989.


A review of contact algorithms


39
A review of contact algorithms

N. G. Bourago and V. N. Kukudzhanov


169. M. V. Zernin, “Problems and prospects of construction of an effective finite-element model of the oil flow in liquid-friction bearing clearances, with nonuniform distribution of temperature and surface
strains being taken into account," Pt. 1, Problemy Tribologii, No. 1, pp. 73-78, 1997; Pt. 1, Problemy Tribologii, No. 2, pp. 57-64, 1997.


