

Geometry optimization in redundant internal coordinates

P. Pulay and G. Fogarasi^{a)}

Department of Chemistry and Biochemistry, The University of Arkansas, Fayetteville, Arkansas 72701

(Received 25 October 1991; accepted 1 November 1991)

The gradient geometry-optimization procedure is reformulated in terms of redundant internal coordinates. By replacing the matrix inverse with the generalized inverse, the usual Newton–Raphson-type algorithms can be formulated in exactly the same way for redundant and nonredundant coordinates. Optimization in redundant coordinates is particularly useful for bridged polycyclic compounds and cage structures where it is difficult to define physically reasonable redundancy-free internal coordinates. This procedure, already used for the geometry optimization of porphine, $C_{20}N_4H_{14}$, is illustrated here at the *ab initio* self-consistent-field level for the four-membered ring azetidine, for bicyclo[2.2.2]octane, and for the four-ring system $C_{16}O_2H_{22}$, the skeleton of taxol.

I. INTRODUCTION

One of the most important areas of applied quantum chemistry is the determination of equilibrium and transition-state structures (for a recent review, see Ref. 1). For larger systems, geometry optimizations is only practical using analytical gradient techniques as reviewed in, e.g., Ref. 2. Essentially all optimization methods are based on the Newton–Raphson technique, i.e., on a local quadratic approximation to the surface. The second derivatives (the *Hessian*) in this formalism are, however, only estimated in most cases, as their exact evaluation, especially in *ab initio* work, is usually not cost effective. With the recent dramatic expansion of computer power, much larger systems became tractable, and this, in turn, stimulated interest in optimization methods.

Three choices influence the efficiency of geometry optimization, measured by the number of energy and gradient evaluations needed to achieve convergence: (1) the coordinates used to describe the system, (2) the mathematical technique used for optimization, and (3) the quality of the quadratic approximation (the force constants or *Hessian*). At first sight, the importance of the coordinates may not be apparent. Two coordinate systems which are related by a *linear transformation* are equivalent in all gradient-optimization methods if the gradient vector and the *Hessian* are properly transformed. It has been claimed on this basis that Cartesians are equivalent,^{3–5} or even superior⁶ to valence-type internal coordinates. This reasoning^{3–5} overlooks, however, the importance of cubic and higher-order couplings. The latter are minimized in a properly chosen valence coordinate system, such as the natural internal coordinates.^{7,8} Only toward the end of the optimization procedure, when the steps are small, do Cartesians and internal coordinates become equivalent. Cartesians may be superior to the “*Z*-matrix”-type internal coordinates⁹ for ring systems.⁶ However, in our experience they are less efficient than the natural internal coordinates, particularly for flexible rings. Internal

coordinates, even the *Z*-matrix type, are invariably superior to Cartesians for open chains.⁸

One disadvantage of valence coordinates is that they are usually redundant, i.e., there are more coordinates than the $3N - 6$ required to characterize the relative positions of the nuclei. Arbitrarily omitting some valence coordinates, as in the *Z*-matrix method, introduces strong linear and nonlinear couplings between the coordinates. This is undesirable for geometry optimization. Our solution to this problem^{7,8} has been to use a carefully selected set of coordinates which we call natural coordinates.^{7,8} In this system, each bond is a stretching coordinate, and redundancy-free linear combinations of bond angles and torsions are used as deformational coordinates. The deformational coordinates are approximately orthogonal to local angle and ring redundancies. Similar coordinates are widely used in vibrational spectroscopy.¹⁰ Our natural internal coordinates are constructed on the basis of local pseudosymmetry. For instance, the four bonds around an sp^3 hybridized atom are considered equivalent for the purpose of defining the coordinates, and so are all torsional angles around a given bond. Ring deformation coordinates are constructed on the basis of idealized D_{nh} symmetry. In our opinion, natural internal coordinates reduce harmonic and anharmonic couplings between coordinates as much as possible on a purely geometrical basis, and are thus the preferred choice for organic molecules. Their main disadvantage, the substantial effort required to set them up, has been solved with a computer program which generates them automatically.⁸

Most redundancies can be eliminated by using natural internal coordinates. However, in molecules with complex topology, in particular, in cage-type polycyclic organic compounds, it is difficult to find physically reasonable nonredundant coordinates. Similarly, large, irregular macrocycles may be too far from the ideal D_{nh} symmetry, and thus the ring coordinates may not work best. In these cases, it is difficult to find a good alternative to using redundant valence-type coordinates. In this paper, we describe how the usual gradient-optimization techniques can be generalized to handle redundant coordinates directly.

^{a)} Permanent address: Laboratory for Theoretical Chemistry, Eötvös L. University, Budapest, Hungary.

The first method known to us which could use redundant coordinates in geometry relaxation was the normal coordinate relaxation method of Sellers *et al.*¹¹ This method used a force field, usually expressed in redundant internal coordinates, to determine the approximate normal coordinates. Geometry relaxation was carried out in normal coordinates which automatically eliminated redundancies. This is, however, a disadvantage of their method as the optimization is less efficient in rectilinear coordinates than in proper curvilinear ones. The first contribution which strongly emphasized the utility of redundant coordinates in geometry optimization is due to Schlegel,¹² although Sellers *et al.* also stress this point in their second paper. Schlegel realized that the strong-coupling terms in the potential function, which arise in ring systems using Z-matrix-type coordinates, can be estimated by setting up an approximate quadratic potential function in redundant valence coordinates, and transforming this to a nonredundant (e.g., Z matrix) set. A similar transformation program (FCT, force-constant transformation) has been in use in our group since the mid-1970s; it was briefly described earlier.^{13,14} The purpose of FCT was to transform vibrational force fields between different internal or Cartesian coordinate systems, including the redundant Urey-Bradley type. A disadvantage of the transformation method¹² is the need to specify two sets of coordinates, and the fact that the transformation itself should be repeated, in principle, in every step if the geometry changes significantly.

We describe here a method which is fully automatic, avoids the explicit use of intermediate coordinate systems, and is fully analogous to the methods widely used in optimizing geometries in nonredundant coordinates. This method has been included in our *ab initio* suite of programs TX90.¹⁵ However, it should be useful both for semiempirical quantum chemical methods, and also for empirical force fields. The latter are usually defined originally in terms of a highly redundant set of variables. We assume that most of these methods carry out the optimization itself in Cartesians. Optimizing directly in redundant parameters will probably improve convergence enough to offset overhead from the larger size of the matrices.

II. THEORY

The first difficulty in working with redundant coordinates is the suitable definition of the force (or gradient) vector. The usual definition, in terms of partial derivatives of the energy, is obviously not applicable. Let us consider a set of redundant internal displacement coordinates $\mathbf{q} = (q_1, q_2, \dots, q_{n+r})^T$. They are related to the Cartesian displacement coordinates \mathbf{X} through the familiar \mathbf{B} matrix:^{10,16}

$$\mathbf{q} = \mathbf{B}\mathbf{X}. \quad (1)$$

In the case of redundancy, the rows of \mathbf{B} are linearly dependent. The redundancy condition can be easily determined by diagonalizing the matrix \mathbf{G} , defined as

$$\mathbf{G} = \mathbf{B}\mathbf{u}\mathbf{B}^T. \quad (2)$$

In Eq. (2), \mathbf{u} is an arbitrary nonsingular matrix, usually the unit matrix. If \mathbf{u} is a diagonal matrix, consisting of triplets of inverse nuclear masses, then this is identical to the spectro-

scopic \mathbf{G} matrix.¹⁰ The eigenvalue equation of \mathbf{G} ,

$$\mathbf{G}(\mathbf{K} \ \mathbf{L}) = (\mathbf{K} \ \mathbf{L}) \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (3)$$

has n nonzero eigenvalues $\mathbf{\Lambda}$ and r (numerically) zero eigenvalues corresponding to the redundancies. In the general case, $n = 3N - 6$, where N is the number of nuclei. \mathbf{K} consists of the first n eigenvectors of \mathbf{G} , corresponding to nonzero eigenvalues; \mathbf{K} is an $(n+r) \times n$ matrix. Similarly, \mathbf{L} consists of the r redundant eigenvectors of \mathbf{G} which correspond to the zero eigenvalues. A set of n nonredundant coordinates can be defined as

$$\mathbf{Q} = \mathbf{K}^T \mathbf{q}, \quad (4)$$

along with the redundant coordinates \mathbf{R} ,

$$\mathbf{R} = \mathbf{L}^T \mathbf{q}. \quad (5)$$

It is straightforward to transform¹⁷ the Cartesian force vector \mathbf{f} to forces Φ expressed in the nonredundant coordinates \mathbf{Q} , using the fact that the transformation matrix from Cartesians to \mathbf{Q} is $\mathbf{K}^T \mathbf{B}$:

$$\Phi = (\mathbf{K}^T \mathbf{B} \mathbf{u} \mathbf{B}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{B} \mathbf{u} \mathbf{f}. \quad (6)$$

We can formally define the forces (gradients) in the redundant coordinate system $(\mathbf{Q}^T \ \mathbf{R}^T)^T$ by simply assigning zero forces to the redundant coordinates. The extended gradient vector, $(\Phi^T \ \mathbf{0}^T)^T$, can now be transformed back to original coordinates \mathbf{q} , using the nonsingular transformation matrix $(\mathbf{K} \ \mathbf{L})$, yielding a force (gradient) vector in redundant coordinates which is very much analogous to the usual gradient vector. We do not go into the details of this transformation because the result can be formulated in a simpler way using the concept of *generalized inverse*. The latter will be denoted here by a minus sign superscript, to distinguish it from the ordinary inverse. The generalized inverse of a singular matrix \mathbf{G} can be expressed from the eigenvalue equation (3) as

$$\mathbf{G}^- = (\mathbf{K} \ \mathbf{L}) \begin{pmatrix} \mathbf{\Lambda}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{K}^T \\ \mathbf{L}^T \end{pmatrix}, \quad (7)$$

i.e., by transforming the matrix to diagonal form, inverting the *nonzero* diagonal elements, and transforming it back. For a nonsingular matrix, the generalized inverse is identical to the ordinary inverse. An alternative, and numerically more efficient method of calculating the generalized inverse, is based on the singular-value decomposition.¹⁸ We used Eq. (7) exclusively to calculate the generalized inverse and found it completely satisfactory. However, the singular-value decomposition method may be useful in optimizing very large molecules which occur in empirical or semiempirical calculations. In practice, the numerical eigenvalues of a singular matrix do not vanish exactly, and therefore the zero condition should be replaced by a check for smallness. In the usual units, a good threshold is 10^{-7} .

Using the generalized inverse, the gradient vector in redundant coordinates is given by

$$\phi = \mathbf{K}\Phi = \mathbf{G}^- \mathbf{B} \mathbf{u} \mathbf{f}, \quad (8)$$

which differs from the usual formula¹⁷ only in substituting the generalized inverse for the common inverse. In Eq. (8) ϕ denotes the force (negative gradient) vector in the original

redundant coordinates, Φ is the force vector in the intermediate nonredundant coordinate system Q , and f is the Cartesian force vector. Note that Eq. (8) does not require the explicit definition of the redundancy condition; the latter is implicitly included through the generalized inverse. The fact that the forces in Eq. (8) refer to the original coordinate system is an advantage; in contrast, the forces Φ are expressed in coordinates which may be complicated linear combinations of the original coordinates. The physical significance of the force (or gradient) vector in redundant coordinates is of some interest. It is easy to prove that the above definition is unique if we require the following two conditions: (1) the first-order change in the energy is given by $dE = -\phi dq$; and (2) at a stationary point, $\phi = 0$. Both of these conditions are satisfied by the usual gradient vector.

This line of thought can be followed to the next stages of geometry optimization: the updating of the internal coordinates, and the calculation of the Cartesians for the next gradient evaluation. We require that the updating in the redundant coordinate system q should give the same results as the updating in the nonredundant coordinates Q . It is useful to introduce the projector P to the nonredundant subspace of the coordinates:

$$P = GG^{-1} = G^{-1}G = KK^T. \quad (9)$$

The requirement formulated above leads to the following inverse Hessian in redundant coordinates:

$$H = P(PFP)^{-1}P, \quad (10)$$

where F is the original force-constant matrix (Hessian) in redundant coordinates. Note that it is essential to include the projector P both before and after the generalized inverse operation. Equation (10) is identical to F^{-1} if the coordinates system is nonredundant. The updating formula in the simplest case becomes

$$\Delta q = -H\phi. \quad (11)$$

The geometry direct inversion in the iterative subspace (DIIS) method¹⁹ can be directly applied to the redundant case. It appears that the usual force-constant updating methods, in particular, the widely used BFGS (Broyden–Fletcher–Goldfarb–Shanno²⁰) variable metric method, are also applicable directly in redundant coordinates.

The final step in geometry optimization using internal coordinates is the generation of the new Cartesian coordinates. Because of the nonlinear nature of this transformation, this is usually done iteratively,⁷ using the first-order formula

$$\Delta X = Aq \quad (12)$$

with

$$A = uB^T G^{-1}. \quad (13)$$

By simply substituting the generalized inverse for the inverse in Eq. (13), this formula becomes applicable to the redundant case.

Although the principle used in the above derivation is the same as Schlegel's suggestion,¹² working directly with redundant coordinates, instead of transforming the force-constant matrix to a nonredundant coordinate system, offers

both practical and theoretical advantages. There are two practical advantages. (1) There is no need to construct a nonredundant coordinate system; this procedure is cumbersome and error prone in larger molecules. (2) With use of the formalism developed above, the computer program is exactly the same for the redundant and nonredundant cases. The only disadvantage is that the present scheme involves some extra computational work in the transformation stage if the coordinate system is not redundant. However, in *ab initio* work, this is negligible.

The present method is also theoretically superior to the explicit transformation method because it performs the (implicit) transformation to nonredundant coordinates in every step. This is likely to produce a better Hessian than a single transformation at the initial geometry if geometry changes in the course of the iteration are significant.

III. EXAMPLES

The technique described above has been implemented in our TX90 program package.¹⁵ All three examples given below employ *ab initio* self-consistent-field (SCF) theory, using the 4-21G(*)^{7,8} or the 6-31G* (Ref. 9) basis sets, as noted in Tables I and II. The starting geometries were obtained by preoptimizing with the MMX force field included in PCMODEL.²¹ The Hessian was a simple diagonal force field, set automatically by the program.

Table I compares the history of the optimization of azetidine, $(CH_2)_3NH$ in redundant and nonredundant coordinate systems. The redundant coordinate system consists of all bond lengths, the four-ring angles, and the four-ring torsions, and our usual natural CH_2 and NH deformational coordinates;^{7,8} the latter include scissoring, rocking, wagging, and twisting^{7,8} for CH_2 , as well as rocking and out-of-plane bending for the NH group. This coordinate system has

TABLE I. Iteration history for the geometry optimization of azetidine in redundant and nonredundant coordinates.*

Cycle no.	Energy + 171.0	Maximum force	Maximum coord. change
Redundant coordinates			
1.	-0.719 694	0.2246	0.0778
2.	-0.731 841	0.0299	0.0293
3.	-0.732 256	0.0106	0.0115
4.	-0.732 351	0.0070	0.0134
5.	-0.732 399	0.0020	0.0062
6.	-0.732 410	0.0010	0.0040
7.	-0.732 413	0.0006	0.0045
8.	-0.732 415	0.0002	0.0021
Natural internal coordinates			
1.	-0.719 694	0.2246	0.2498
2.	-0.731 715	0.0392	0.1222
3.	-0.732 306	0.0116	0.0252
4.	-0.732 382	0.0075	0.0143
5.	-0.732 410	0.0033	0.0057
6.	-0.732 415	0.0013	0.0023
7.	-0.732 416	0.0002	0.0005

* Energies in atomic units, forces in aJ/Å or aJ/rad, coordinates in Å or rad. Basis: 4-21G(*) ; see Ref. 8.

33 coordinates and thus contains 6 redundancies. The non-redundant coordinate system replaces the four deformations and four torsions of the ring by one ring-bending and one ring-puckering coordinate; the latter are linear combinations of the four individual angles or torsions, adapted to D_{4h} pseudosymmetry. There is no important reason to use redundant coordinates for azetidine and we thus prefer to use the redundancy-free natural coordinates. However, as Table I shows, the convergence characteristics of the two methods are very similar.

Another example, bicyclo[2.2.2]octane, is given in Table II. In such a cagelike structure it is difficult to define redundancy-free coordinates in a physically reasonable way. The use of ring coordinates for more than one ring inevitably leads to redundancy due to shared angles and torsions. We used the coordinates provided by our automatic internal coordinate generator. Cagelike structures are not yet included in this program, and thus the program generates the usual ring coordinates⁸ for only one of the rings, and treats the bridge simply as a chain, using the regular internal coordinates for chains. This leads to 66 coordinates, i.e., six redundancies. As Table II shows, this coordinate system, although not symmetrical, works very well for this system: strict convergence is achieved in only four steps. The symmetry of the initial geometry was close to D_{3h} , and we have maintained a subgroup of this point group, C_{2v} , during optimization. The final geometry has D_{3h} symmetry and it is a minimum within this point group. However, the potential energy along a path which lowers the symmetry to D_3 is very shallow, and we have not investigated whether our structure is a true minimum or possibly a saddle point along this path. Any possible D_3 distortion must be small. The number of independent geometry parameters is 6 under the D_{3h} point group and increases to 11 under D_3 .

A more complex case is the skeleton of the taxol molecule (Fig. 1). Taxol²² is a promising anticancer agent; it differs from the $C_{16}O_2H_{22}$ molecule in Fig. 1 by substituting four methyl and six hydroxy or esterified hydroxy groups for hydrogens. As in the previous case, the redundancy is not easy to avoid because of the bridge between the eight-ring and one six-ring. As discussed in Ref. 8 and above, there is no natural redundancy-free coordinate system in bridged polycyclic compounds. The use of natural coordinates for each of

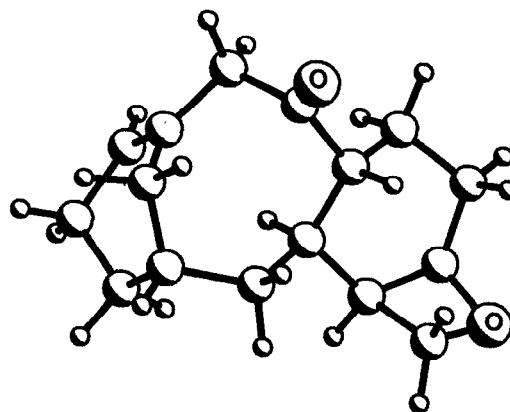


FIG. 1. Perspective view of the optimized taxol skeleton.

the four rings yields 115 internal coordinates, i.e., one redundancy, caused by the 1,3 ring fusion. Table III shows the optimization history for this molecule, using the same procedure as for azetidine. The only significant difference is that for taxol we use *offset forces*⁸ to correct for systematic errors in the limited basis-set SCF procedure. Considering the size of the system, and the very simple initial Hessian, convergence in 10–11 steps is satisfactory. We have not checked rigorously whether the resulting geometry is a true local minimum because of the cost involved. However, the optimization history strongly indicates that the final geometry, available from the first author upon request, is a minimum.

Two other examples, the geometry optimization of porphine, $C_{20}N_4H_{14}$, and of norbornane, C_7H_{12} , were given in Ref. 8(a) without a detailed discussion. The results in these cases confirm the present conclusions.

TABLE III. Iteration history for the optimization of the taxol skeleton.*

Cycle no.	Energy + 766.0	Corrected energy + 766.0	Maximum force	Maximum coord. change
1.	− 0.778 592	− 0.956 850	0.1450	0.1169
2.	− 0.795 162	− 0.967 739	0.0761	0.1573
3.	− 0.782 802	− 0.954 965	0.2312	0.1241
4.	− 0.797 868	− 0.969 809	0.0432	0.0531
5.	− 0.798 568	− 0.970 549	0.0344	0.0341
6.	− 0.798 808	− 0.970 729	0.0144	0.0123
7.	− 0.798 822	− 0.970 779	0.0046	0.0107
8.	− 0.798 864	− 0.970 810	0.0040	0.0119
9.	− 0.798 890	− 0.970 824	0.0019	0.0074
10.	− 0.798 902	− 0.970 827	0.0023	0.0055
11.	− 0.798 907	− 0.970 828	0.0007	0.0021

* $C_{16}H_{22}O_2$; see Fig. 1. For units see Table I. 4-21G(*) basis, 216 contracted basis functions. The corrected energy is the *ab initio* energy minus the sum of the bond distances times offset forces; see Ref. 8. The actual quantity minimized is the corrected energy.

TABLE II. Iteration history for geometry optimization in bicyclo[2.2.2]octane.*

Cycle no.	Energy + 311.0	Maximum force	Maximum coord. change
1.	− 0.095 099	0.1740	0.0316
2.	− 0.103 521	0.0105	0.0107
3.	− 0.103 585	0.0088	0.0044
4.	− 0.103 597	0.0012	0.0001

* For units see Table I. 6-31G* basis (Ref. 9). The starting geometry was obtained by the MMX force field (Ref. 21) and had almost perfect D_{3h} symmetry. C_{2v} symmetry was maintained during optimization. The resulting geometry has virtually perfect D_{3h} symmetry. See text.

ACKNOWLEDGMENTS

This research has been supported by the National Science Foundation under Grant No. CHE-8814143. We are grateful to IBM for the donation of an RS6000 workstation to our department.

¹ H. B. Schlegel, *Adv. Chem. Phys.* **67**, 249 (1987).

² P. Pulay, *Adv. Chem. Phys.* **69**, 241 (1987).

³ J. D. Head and M. C. Zerner, *Chem. Phys. Lett.* **122**, 264 (1985).

⁴ J. D. Head, B. Weiner, and M. C. Zerner, *Int. J. Quantum Chem.* **33**, 177 (1988).

⁵ J. D. Head, *J. Comput. Chem.* **11**, 67 (1990).

⁶ J. Baker and W. J. Hehre, *J. Comput. Chem.* **12**, 606 (1991).

⁷ P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs, *J. Am. Chem. Soc.* **101**, 2550 (1979).

⁸ (a) G. Fogarasi, X. Zhou, P. W. Taylor, and P. Pulay, *J. Am. Chem. Soc.* (submitted); (b) P. Pulay, G. Fogarasi, X. Zhou, and P. W. Taylor, *Vibr. Spectrosc.* **1**, 159 (1990).

⁹ See, e.g., W. J. Hehre, L. Radom, P. R. v. Schleyer, and J. A. Pople, *Ab initio Molecular Orbital Theory* (Wiley, New York, 1986).

¹⁰ E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).

¹¹ (a) H. L. Sellers, V. J. Klimkowski, and L. Schäfer, *Chem. Phys. Lett.* **58**, 541 (1978); (b) H. L. Sellers, J. F. Pinegar, and L. Schäfer, *ibid.* **61**, 499 (1979).

¹² H. B. Schlegel, *Theor. Chim. Acta* **66**, 333 (1984).

¹³ C. E. Blom and P. Pulay, *Mol. Phys.* **39**, 1537 (1980).

¹⁴ G. Fogarasi and P. Pulay, in *Vibrational Spectra and Structure*, edited by J. R. Durig (Elsevier, Amsterdam, 1985), Vol. **14**, p. 125.

¹⁵ P. Pulay and co-workers, TX90, Fayetteville, Arkansas 1990; P. Pulay, *Theor. Chim. Acta* **50**, 299 (1979).

¹⁶ M. Eliashevich, *C. r. Acad. Sci. U.R.S.S.* **28**, 605 (1940).

¹⁷ P. Pulay, *Mol. Phys.* **17**, 197 (1969).

¹⁸ See, e.g., W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University, Cambridge, 1989), p. 52.

¹⁹ P. Csaszar and P. Pulay, *J. Mol. Struct.* **114**, 31 (1984).

²⁰ See, e.g., R. Fletcher, *Practical Methods of Optimization* (Wiley, Chichester, 1980), Vol. 1.

²¹ PCMODEL, Serena Software, Box 3076, Bloomington, IN 47402 (1990).

²² M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon, and A. T. McPhail, *J. Am. Chem. Soc.* **93**, 2325 (1971).