Computational Mathematics and Mathematical Physics, Vol. 45, No. 8, 2005, pp. 1400–1415.
Translated from Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, Vol. 45, No. 8, 2005, pp. 1450–1465.

On the Convergence of a Gradient Method for the Minimization of Functionals in Finite Deformation Elasticity Theory and for the Minimization of Barrier Grid Functionals

V. A. Garanzha and I. E. Kaporin

Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia
e-mail: garan@ccas.ru, kaporin@ccas.ru

Received January 29, 2001

Abstract—Gradient descent methods are examined for the minimization of barrier-type polyconvex functionals arising in finite-deformation elasticity theory and grid optimization. The minimum of a functional is sought in the class of continuous piecewise affine deformations that preserve orientation. Sufficient conditions are found for a sequence of iterative approximations to belong to the feasible set and for the norm of the gradient of the functional to converge to zero on this set. As the functional, one can use a measure of the deformation of a grid, for instance, a grid formed of triangles or tetrahedra.

Keywords: nonlinear optimization, gradient method, finite deformation elasticity theory, polyconvex functionals, grid optimization.

1. INTRODUCTION

Consider the following variational problem that arises in elastostatics and in the mapping optimization in geometric modeling: find an invertible deformation (mapping) $x(\xi) : \Omega_\xi \longrightarrow \Omega_x$ as an extremal of the accumulated-energy functional

$$J(x) = \int_{\Omega_\xi} \varphi(\nabla_\xi x) d\xi. \tag{1}$$

Here, Ω_ξ and Ω_x are bounded connected Lipschitzian domains in $\mathbb{R}^d$ and the square d-by-d matrix $\nabla_\xi x$ with the entries $\partial x_i/\partial \xi_j$ is the Jacobian matrix of the mapping $x(\xi)$. Note that the assumption on the invertibility of $x(\xi)$ ensures that

$$\det(\nabla_\xi x) > 0.$$

In what follows, we usually assume that $d = 2$ or $d = 3$.

The integrand φ in functional (1), called the energy density of deformation, has the following properties (see [1]):

(1) polyconvexity, which means that $\varphi(A)$ can be written as a convex function of the minors in the matrix $A = \{a_{ij}\}_{1 \le i, j \le d}$;

(2) the barrier property, i.e.,

$$\varphi(A) \longrightarrow +\infty \quad \text{as} \quad \det A \longrightarrow +0; \tag{2}$$

(3) coercitivity, i.e., $\varphi(A) \ge 0$ and

$$\varphi(A) \longrightarrow +\infty \quad \text{as} \quad \|A\|_E \longrightarrow \infty, \tag{3}$$

where $\|A\|_E = (\mathrm{tr} A^\mathrm{T} A)^{1/2}$ is the Euclidean norm of A.

Polyconvexity implies the rank-one convexity (see [1]), i.e.,

$$\varphi(\lambda S_1 + (1-\lambda)S_2) \le \lambda\varphi(S_1) + (1-\lambda)\varphi(S_2),$$
$$\text{where} \quad \mathrm{rank}(S_1 - S_2) = 1, \quad 0 \le \lambda \le 1. \tag{4}$$

If, additionally, φ is a sufficiently smooth function (say, $\varphi \in C^2(\mathbb{R}^{d\times d})$), then the rank-one convexity is equivalent to the Legendre–Hadamard condition

$$\sum_{i,j,k,p=1}^{d} \frac{\partial^2 \varphi}{\partial a_{ij}\partial a_{kp}} t_i t_k s_j s_p \geq 0 \tag{5}$$

being fulfilled for all the nonzero vectors $t, s \in \mathbb{R}^d$. It is also called the ellipticity condition, because the Euler–Legendre equations for a functional satisfying this condition are elliptic equations.

Thus, unlike the case of scalar variational problems, the ellipticity condition for vector variational problems is broader than the convexity condition. An elliptic vector variational problem may have a nonunique solution or not have a solution at all (see [1, 2]). In a somewhat loose mathematical formulation, we can say that this paper examines iterative minimization methods for elliptic vector variational problems of barrier type.

It is assumed that Ω_ξ is a fixed domain. On $\partial\Omega_\xi$, a deformation (i.e., a continuous one-to-one mapping $\partial\Omega_\xi \longrightarrow \partial\Omega_x$) may be given. Another mathematically correct problem is that in which the boundary conditions are not given on the entire boundary or its part (see [1, 2]). This part of $\partial\Omega_x$ is found by minimizing functional (1).

To formulate the finite-dimensional optimization problem, we must first define a simplicial partition

$$\Omega_\xi = \bigcup_{1 \leq i \leq M} \mathcal{S}_i$$

of the domain Ω_ξ, i.e., the initial computational grid. We assume that this grid satisfies the conventional requirements for finite-element grids and that it is quasi-uniform.

To simplify the formulation of the problem, we assume that the boundary of Ω_ξ is piecewise linear; therefore, it is approximated exactly by the simplicial partition.

We seek the minimum of functional (1) in the class of continuous piecewise affine deformations that preserve orientation. It is assumed that the deformation φ is affine on each simplex $\mathcal{S}_i$. Thus, the Jacobian matrix is constant on $\mathcal{S}_i$ (i.e., $\nabla_\xi x = A_i$), and its determinant is positive.

In this paper, we do not examine the convergence of the solution to the discrete problem to the exact solution as the computational grid is refined. This is a difficult and not yet solved problem in approximation theory for large deformations.

The minimization of a discrete functional on a fixed grid, which is a simpler problem for the analysis, is the subject of this paper. It is required to determine the unknown vector R of a fixed dimension $n = Nd$, where N is the number of interior grid nodes; this vector consists of the coordinates (to be optimized) of the images of vertices in the simplicial partition. Accordingly, (1) is reduced to the discrete functional

$$\Phi(R) = \sum_{1 \leq i \leq M} \int_{\mathcal{S}_i} \varphi(\nabla_\xi x) d\xi = \sum_{1 \leq i \leq M} \mathrm{vol}(\mathcal{S}_i)\varphi(A_i). \tag{6}$$

The main theoretical result obtained by the authors is as follows. Under the additional condition that φ is a self-consistent function, it is proved that the preconditioned descent method

$$R^{k+1} = R^k - \alpha_k B_k \nabla\Phi(R^k),$$

which is analogous to the one used in [7, 8], converges to a stationary point of Φ for any feasible initial approximation (i.e., an approximation R_0 such that $\Phi(R^0) < \infty$). Moreover, the entire sequence $\{R^k\}_{k\geq 0}$ belongs to the feasible set; i.e., $\Phi(R^k) \leq \Phi(R^0)$. Here, α_k is a positive scalar and B_k is a specially chosen symmetric positive semidefinite n-by-n matrix (preconditioner). Actually, we find a fixed lower bound for α_k, which ensures that the norm of $\nabla\Phi(R^k)$ converges to zero as $k \longrightarrow \infty$.

The convergence is proved for the functionals corresponding to the models of neo-Hookean and Ogden materials, as well as the functional proposed in [3] for constructing quasi-isometric mappings.

1. THE DISTORTION MEASURE OF A CELL AND THE GRID FUNCTIONAL

Consider a grid formed of M cells, each of which is a triangle in the case of two dimensions (i.e., $d = 2$) and a tetrahedron if $d = 3$. It is assumed that the topology of a grid is fixed; however, the coordinates of most of its nodes are subject to changes in order to optimize the shape and size of the cells.

1.1. Local Characterization of Grid Cell Deformation

For each cell, we define the distortion measure as the value of the function

$$\varphi(A), \quad A \in \mathbb{R}^{d \times d}. \tag{7}$$

Here, A is the matrix

$$A = [r_2 - r_1 \ \ldots \ r_{d+1} - r_1]T, \tag{8}$$

where

$$r_i = \begin{bmatrix} x_i^{(1)} \\ \ldots \\ x_i^{(d)} \end{bmatrix}, \quad i = 1, 2, \ldots, d+1,$$

denotes a vertex (to be determined) of the simplex and T is a fixed d-by-d matrix (that is independent of r_i). The minimization of $\varphi(A)$ is related to the process in which A approaches a scalar multiple of an orthogonal matrix; i.e., $A^{\mathrm{T}}A \longrightarrow \mu^2 I_d$.

The construction just described covers two types of problems:

1. Improving the quality of cells in a prescribed grid, where T is the same for all the grid cells and A is the Jacobian matrix of the linear mapping that transforms the regular simplex of volume μ^d into the desired cell.

2. Deforming the original domain Ω_ξ, where the grid is prescribed, into the domain Ω_x with a (partially) prescribed boundary. Then, A is the Jacobian matrix of the linear mapping that transforms a cell of the prescribed grid into a cell of the desired grid. Here, we can set $\mu = 1$ for all the cells. In this case, we have (say, for $d = 2$)

$$A = [r_2 - r_1 \ \ r_3 - r_1][s_2 - s_1 \ \ s_3 - s_1]^{-1},$$

where s_1, s_2, and s_3 are the vertices in a triangular cell of the prescribed grid. Here, the matrix T changes from one cell to another.

In elasticity theory problems, it is natural to require that functional (1) attains its global minimum in the absence of deformations. This corresponds to the condition that $\varphi(A)$ attains its minimum at matrices with orthonormal columns ($\mu = 1$).

The formulation of the problem is different when computational grids are optimized. It is assumed that a *target cell*, which is a cell of the ideal shape and size, is given for each grid cell. The regular simplex of volume μ^d, where μ is the given normalizing constant, can serve as such a target cell. Then, A corresponds to the mapping of the regular simplex of volume μ^d to the desired cell. In this case, T is the same matrix for all the grid cells.

Thus, the following property is our basic requirement to $\varphi(A)$.

Property 1. The function $\varphi(A)$ is continuously differentiable and nonnegative on the set $\{A : \det A > 0\}$, and the global minimum of $\varphi(A)$ is attained if and only if $A^{\mathrm{T}}A = \mu^2 I_d$.

For instance, if $\mu = 1$ and φ approaches its global minimum, then the linear mapping under discussion approaches an isometry.

Throughout the paper, we assume that $\varphi(A)$ is a twice continuously differentiable function of A in the feasible domain.

1.2. Examples of Local Distortion Functions

As $\varphi(A)$, we can take the function

$$\varphi(A) = \frac{\theta}{2}\left(\frac{\det A}{\mu^d} + \frac{\mu^d}{\det A}\right) + (1-\theta)\frac{(d^{-1}\mathrm{tr}(A^{\mathrm{T}}A))^{d/2}}{\det A}, \tag{9}$$

proposed in [3]. Here, $\theta \in (0, 1)$ is a fixed scalar and $\min_A \varphi = 1$. An even simpler example is the function

$$\varphi(A) = \frac{1}{2\mu^2}\mathrm{tr}A^{\mathrm{T}}A - \log\det A, \tag{10}$$

where $\min_A \varphi = d/2 - d\log\mu$. This function can be considered as the simplest model of the density of accumulated energy in finite deformation elasticity problems (see [1]).

For compressible Neo-Hookean materials, the stored energy function is taken in the form $c_0\mathrm{tr}(A^{\mathrm{T}}A) + \Gamma(\det A)$, where $\Gamma(\cdot)$ is a convex function, for instance, $\Gamma(t) = c_1t^2 - c_2\log t$ (see [4]).

For compressible Mooney–Rivlin materials (see [5]), the stored energy function is taken in the form $c_0\mathrm{tr}(A^{\mathrm{T}}A) + c_3\mathrm{tr}(S^{\mathrm{T}}S) + \Gamma(\det A)$, where $S = \mathrm{cof}A$; hereafter, the symbol cof$A$ is used for the matrix of cofactors. If A is nonsingular, then

$$\mathrm{cof}A \equiv (\det A)A^{-\mathrm{T}}. \tag{11}$$

The coefficients c_1, c_2, and c_3 are appropriate constants independent of A.

The Ogden function (see [6]) has a similar form; however, it can use fractional powers rather than quadratic terms.

All the functions are set to $+\infty$ outside of the feasible set.

Consider the case $d = 3$. Define $B = A^{\mathrm{T}}A$ and $C = (\mathrm{cof}A)^{\mathrm{T}}\mathrm{cof}A$, and take into account the fact that $\det A = (\det B)^{1/2}$ in view of the inequality $\det A > 0$. Then, all the functions mentioned above lead to a local distortion function of the form

$$\varphi(A) = f(\mathrm{tr}B, \mathrm{tr}C, \sqrt{\det B}), \tag{12}$$

where $f(q_1, q_2, q_3)$ is a sufficiently smooth function on the positive orthant in $\mathcal{R}^3$. (Note that, for $d = 2$, this construction is simplified to the function $\varphi(A) = f(\mathrm{tr}B, \sqrt{\det B})$, because $\mathrm{tr}C = \mathrm{tr}B$.)

Define

$$f_1 = \partial f/\partial q_1, \quad f_2 = \partial f/\partial q_2. \tag{13}$$

It turns out that Condition 1 is fulfilled if

$$f_1 \ge 0, \quad f_2 \ge 0, \quad f_1 + f_2 > 0,$$

and the function

$$\psi(q) = f(dq^{1/d}, dq^{(d-1)/d}, \sqrt{q})$$

is nonnegative and has a unique stationary point (minimum) at $q = \mu^d$.

To prove this, we use the well-known inequality

$$\mathrm{tr}B \ge d(\det B)^{1/d}, \tag{14}$$

and the inequality

$$\mathrm{tr}C \ge d(\det B)^{(d-1)/d}. \tag{15}$$

The latter is valid in view of the identity

$$\mathrm{tr}C \equiv (\det B)\mathrm{tr}B^{-1}, \tag{16}$$

one only needs to apply (14) with B replaced by B^{-1} to the right-hand side of (16). Note that, both in (14) and (15), the equality is attained if and only if B is a scalar multiple of the identity matrix I_d.

Now, the condition $\varphi(A) = \psi(\mu^d)$ implies that

$$0 = f(\mathrm{tr}B, \mathrm{tr}C, \sqrt{\det B}) - f(d(\det B)^{1/2}, d(\det B)^{(d-1)/d}, \sqrt{\det B}) + \psi(\sqrt{\det B}) - \psi(\mu^d)$$

$$= (\mathrm{tr}B - d(\det B)^{1/d}) f_1(\tilde{q}_1, \tilde{q}_2, \sqrt{\det B}) + (\mathrm{tr}C - d(\det B)^{(d-1)/d}) f_2(\tilde{q}_1, \tilde{q}_2, \sqrt{\det B}) + (\psi(\sqrt{\det B}) - \psi(\mu^d)),$$

where $d(\det B)^{1/d} \le \tilde{q}_1 \le \mathrm{tr}B$ and $d(\det B)^{(d-1)/d} \le \tilde{q}_2 \le \mathrm{tr}C$ are the values appearing in the mean value theorem as applied to the first two variables in f. Since the sum of the first two terms can vanish only if $B = \mathrm{const} \times I_d$, while the third term can vanish only if $\det B = \mu^{2d}$, we obtain the desired equality $B = \mu^2 I_d$.

Thus, for $d = 3$, a tetrahedron preserves its shape under deformation and its volume is multiplied by μ^3 if and only if $f(A)$ attains its global minimum.

1.3. Assumptions on the Local Distortion Function

For a piecewise affine deformation to preserve its orientation, the following *barrier property* must hold:

Property 2. If $\det A \longrightarrow +0$, then $\varphi(A) \longrightarrow +\infty$.

Remark 1. The barrier property is inconsistent with the convexity of $\varphi(A)$ (see [1, 2]). Indeed, φ cannot be convex everywhere since its domain (given by the inequality $\det A > 0$) is not convex.

We call A a feasible matrix if the value $\varphi(A)$ is finite.

In [3], quasi-isometric grids are constructed with the help of a slightly more complicated function than (9), namely,

$$\varphi_1(t; A) = \frac{(1-t)\varphi(A)}{1 - t\varphi(A)}, \quad 0 < t < 1. \tag{17}$$

In this case, the barrier emerges on the boundary of a more narrow feasible set; however, the propositions given below on the convergence of iterative methods remain valid.

Finally, we demand that $\varphi(A)$ be a locally self-consistent function; namely, the gradient $\nabla\varphi$ of φ and its Hessian matrix $\nabla^2\varphi$ must have the following property:

Property 3. There exist nondecreasing positive functions $\omega_1(t)$ and $\omega(t)$ such that if $\det A > 0$ and C is an arbitrary perturbation matrix that is "small" in the sense of the inequality

$$\|C\|_E \le \frac{1}{\omega(\varphi(A))}, \tag{18}$$

then it holds that

$$\|\nabla\varphi(A + C)\| \le \omega_1(\varphi(A)), \tag{19}$$

and

$$\|\nabla^2\varphi(A + C)\| \le \omega(\varphi(A)). \tag{20}$$

Recall that $\nabla\varphi(A)$ and $\nabla^2\varphi(A)$ are a vector of dimension d^2 and a d^2-by-d^2 matrix, respectively, that are determined by the relation

$$\varphi(a + c) = \varphi(a) + c^{\mathrm{T}}\nabla\varphi(a) + \frac{1}{2}c^{\mathrm{T}}\nabla^2\varphi(a)c + O(\|c\|^3), \tag{21}$$

where a and c are the matrices A and C arranged as d^2-dimensional vectors. (If necessary, we use the identification $a = \mathbf{A}$.) Then, the symbol $\|\cdot\|$ in (20) stands for the conventional spectral norm of a matrix.

1.4. Sufficient Conditions for the Validity of Property 3

Hereafter, we use the notation

$$(U, V) = \sum_{i,j}(U)_{i,j}(V)_{i,j} \equiv \mathrm{tr}(V^{\mathrm{T}}U), \quad \|U\|_E = \sqrt{(U, U)}$$

for the scalar product and the norm of p-by-q matrices U and V. Indeed, since the argument is a d-by-d matrix rather than a vector, $\nabla\varphi$ is arranged as a matrix of the same size.

Assume that $d \le 3$. As before, we limit ourselves to examining functions φ of form (12), i.e.,

$$\varphi(A) = f(\|A\|_E^2, \|\mathrm{cof} A\|_E^2, \det A) \tag{22}$$

on the set of d-by-d matrices A such that $\det A > 0$.

Define

$$q_1 = \|A\|_E^2, \quad q_2 = \|\mathrm{cof} A\|_E^2, \quad q_3 = \det A; \tag{23}$$

then,

$$\varphi(A) = f(q_1(a), q_2(a), q_3(a)). \tag{24}$$

(Note that, for $d = 2$, we have $\|\mathrm{cof} A\|_E \equiv \|A\|_E$, which allows us to examine only the case where $\varphi = f(q_1, q_3)$.) Assume that f is a twice continuously differentiable function of q_k. Then, substituting

$$q_k(a+c) = q_k(a) + c^T \nabla q_k(a) + \frac{1}{2} c^T \nabla^2 q_k(a) c + O(\|c\|^3), \quad k = 1, 2, 3,$$

into (24) and using the similar formula

$$f(q+p) = f(q) + p^T \nabla f(q) + \frac{1}{2} p^T \nabla^2 f(q) p + O(\|p\|^3),$$

where $q = [q_1, q_2, q_3]^T$ and $p_k = q_k(a+c) - q_k(a)$, we conclude that the first-order terms in (21) have the form

$$c^T \nabla \varphi(a) = \sum_{k=1}^{3} c^T \nabla q_k(a) f_k, \tag{25}$$

while the second-order terms have the form

$$c^T \nabla^2 \varphi(a) c = \sum_{k=1}^{3} (c^T \nabla^2 q_k(a) c) f_k + \frac{1}{4} \sum_{k=1}^{3} \sum_{l=1}^{3} (c^T \nabla q_k(a))(c^T \nabla q_l(a)) f_{kl}. \tag{26}$$

Here, we use the same notation as in (13) and introduce the functions

$$f_{kl}(q_1, q_2, q_3) = \frac{\partial^2 f}{\partial q_k \partial q_l}.$$

Property 3 can easily be derived for a sufficiently broad class of functions f if the barrier property of φ and its coercitivity are specified as follows.

Property 2a (the barrier property). It holds that

$$\varphi(A) \ge \psi_1(1/\det A), \tag{27}$$

where $\psi_1(\cdot)$ is a positive monotonically increasing function.

Property 2b (coercitivity). It holds that

$$\varphi(A) \ge \psi_2(\|A\|_E), \tag{28}$$

where $\psi_2(\cdot)$ is a positive monotonically increasing function.

Corollary. It holds that

$$1/\det A \le \Psi_1(\varphi(A)) \tag{29}$$

and

$$\|A\|_E \le \Psi_2(\varphi(A)), \tag{30}$$

where $\Psi_1(\cdot)$ and $\Psi_2(\cdot)$ are positive monotonically increasing functions (that are the inverse functions of $\psi_1(\cdot)$ and $\psi_2(\cdot)$, respectively).

Remark 2. It is obvious that functions (9) and (10) satisfy the assumptions introduced above.

Lemma 1. *If* $\det A \neq 0$, *then it holds that*

$$\|\mathrm{cof} A\|_E \le \|A\|_E^{d-1}. \tag{31}$$

Proof. Let $B = A^{\mathrm{T}}A$; then, (31) is equivalent to the inequality

$$(\det B)(\operatorname{tr} B^{-1}) \le (\operatorname{tr} B)^{d-1}.$$

Denoting by $\lambda_i > 0$ ($i = 1, 2, \ldots, d$) the eigenvalues of B, we can rewrite this inequality as

$$\sum_{i=1}^{d}\left(\prod_{j \ne i}\lambda_i\right) \le \left(\sum_{i=1}^{d}\lambda_i\right)^{d-1}.$$

The desired relation follows from the inequality between the arithmetic and geometric means:

$$\sum_{i=1}^{d}\prod_{j \ne i}\lambda_i \le \sum_{i=1}^{d}\left(\frac{1}{d-1}\sum_{j \ne i}\lambda_i\right)^{d-1} \le \left(\sum_{i=1}^{d}\frac{1}{d-1}\sum_{j \ne i}\lambda_i\right)^{d-1} = \left(\sum_{i=1}^{d}\lambda_i\right)^{d-1}.$$

Lemma 2. *Matrix functions* (23) *satisfy the inequalities*

$$1/\Psi(\varphi(A)) \le q_k(A) \le \Psi(\varphi(A)), \quad k = 1, 2, 3, \tag{32}$$

where $\Psi(\cdot)$ is a positive monotonically increasing function.

Proof. Let $B = A^{\mathrm{T}}A$. In view of (30), $q_1(A)$ satisfies the inequality

$$q_1(A) = \|A\|_E^2 \le (\Psi_2(\varphi(A)))^2.$$

The lower bound in (32) follows from the relations

$$q_1(A) = \operatorname{tr} B \ge d(\det B)^{1/d} = d(\det A)^{2/d} \ge d(\Psi_1(\varphi(A)))^{-2/d},$$

where we used (29). From this, we also obtain the following bounds for $q_3(A)$:

$$q_3(A) = \det A \ge 1/\Psi_1(\varphi(A))$$

and

$$q_3(A) = (\det B)^{1/2} \le (d^{-1}\operatorname{tr} B)^{1/(2d)} = d^{-1/(2d)}(\|A\|_E)^{1/d} \le d^{-1/(2d)}(\Psi_2(\varphi(A)))^{1/d}.$$

Now, we bound $q_2(A)$ by applying Lemma 1:

$$q_2(A) = \|\operatorname{cof} A\|_E^2 \le (\|A\|_E)^{2d-2} \le (\Psi_2(\varphi(A)))^{2d-2}.$$

Using the bound $\|A^{-1}\|_E^2 \ge d(\det A)^{-2/d}$ (cf. (15)), which was already proved above, we have

$$q_2(A) = (\det A)^2\|A^{-1}\|_E^2 \ge d(\det A)^{2-2/d} \ge d(\Psi_1(\varphi(A)))^{-2+2/d}.$$

Thus, all the inequalities in (32) are valid if we set

$$\Psi(t) = \max\left((\Psi_2(t))^2, \frac{1}{d}(\Psi_1(t))^{2/d}, \Psi_1(t), \left(\frac{\Psi_2(t)}{d^{1/2}}\right)^{1/d}, (\Psi_2(t))^{2d-2}, \frac{1}{d}(\Psi_1(t))^{(2d-2)/d}\right).$$

Lemma 3. *Under the condition $\|C\|_E \le \|A\|_E$, the gradients and Hessian matrices of matrix functions* (23) *satisfy the inequalities*

$$\left\|\nabla^m q_1(a+c)\right\| \le \gamma_1(\Psi_2(\varphi(A)))^{2-m}, \tag{33}$$

$$\left\|\nabla^m q_2(a+c)\right\| \le \gamma_1(\Psi_2(\varphi(A)))^{2d-2-m}, \tag{34}$$

$$\left\|\nabla^m q_3(a+c)\right\| \le \gamma_1(\Psi_2(\varphi(A)))^{d-m}, \quad m = 1, 2, \tag{35}$$

where $\Psi_2(\cdot)$ is a positive monotonically increasing function defined in (30) *and γ_1 is a positive constant that depends only on d.*

Proof. The assertion of Lemma 3 is easily derived from the fact that $q_1(a+c)$, $q_3(a+c)$, and $q_3(a+c)$ are homogeneous polynomials in the components of the vector $a + c$. These polynomials have the degrees 2, $2d - 2$, and d, respectively. Each differentiation reduces their degrees by unity. It is obvious that, for each of the resulting Jacobians and Hessian matrices, the sum of the squares of the components can be bounded by a multiple of the appropriate power of the norm $\|A + C\|_E \leq 2\|A\|_E$. As for the quantity $\|A\|_E$, it can be bounded above with the help of (30).

Lemma 4. *If the norm of the perturbation matrix C is bounded by*

$$\|C\|_E \leq \frac{\varepsilon}{1+\varepsilon}(\Psi(\varphi(A)))^{-3/2}, \tag{36}$$

where $0 < \varepsilon < 1$, *then the corresponding values of matrix functions* (23) *satisfy the inequalities*

$$(1-\varepsilon)^4 \leq \frac{q_k(A+C)}{q_k(A)} \leq (1-\varepsilon)^{-4}. \tag{37}$$

Proof. For q_1, we have

$$q_1(A+C)/q_1(A) = (\|A+C\|_E/\|A\|_E)^2 \leq (1+\|C\|_E/\|A\|_E)^2.$$

Similarly,

$$q_1(A+C)/q_1(A) \geq (1-\|C\|_E/\|A\|_E)^2.$$

Thus, if, for a certain ε_1 $(0 < \varepsilon_1 < 1)$, it holds that

$$\|C\|_E \leq \varepsilon_1\|A\|_E, \tag{38}$$

then

$$(1-\varepsilon_1)^2 \leq q_1(A+C)/q_1(A) \leq (1-\varepsilon_1)^{-2}. \tag{39}$$

To bound the variation of q_3, we apply the well-known formula

$$\chi(a+c) = \chi(a) + c^{\mathrm{T}}\nabla\chi(a+\theta c), \quad 0 \leq \theta \leq 1,$$

to the function $\chi(a) = \log q_3(A) = \log\det A$, taking into account the relation $c^{\mathrm{T}}\nabla\chi(\tilde{a}) = \mathrm{tr}\tilde{A}^{-1}C$. Assuming that $\|A^{-1}\|_E\|C\|_E < 1$, we obtain

$$\left|\log\frac{q_3(A+C)}{q_3(A)}\right| = |\mathrm{tr}(A+\theta C)^{-1}C| = \left|\sum_{k=1}^{\infty}(-\theta^{k-1})\mathrm{tr}(A^{-1}C)^k\right| \leq \frac{\|A^{-1}\|_E\|C\|_E}{1-\|A^{-1}\|_E\|C\|_E}.$$

Thus, if, for a certain ε_2 $(0 < \varepsilon_2 < 1)$, it holds that

$$\|C\|_E < \varepsilon_2/\|A^{-1}\|_E, \tag{40}$$

then

$$\exp[-\varepsilon_2/(1-\varepsilon_2)] \leq q_3(A+C)/q_3(A) \leq \exp[\varepsilon_2/(1-\varepsilon_2)]. \tag{41}$$

Next, we bound the relative variation of the function $\mathrm{tr}A^{-1}$. We use the identity $(A + C)^{-1} = A^{-1} - A^{-1}C(I_d + A^{-1}C)^{-1}A^{-1}$, which, in view of (40), can be rewritten as

$$(A+C)^{-1} = A^{-1} + \sum_{k=1}^{\infty}(-A^{-1}C)^k A^{-1}.$$

We have

$$\|(A+C)^{-1}\|_E \leq \|A^{-1}\|_E + \left\|\sum_{k=1}^{\infty}(-A^{-1}C)^k A^{-1}\right\|_E,$$

$$\|(A+C)^{-1}\|_E \geq \|A^{-1}\|_E - \left\|\sum_{k=1}^{\infty}(-A^{-1}C)^k A^{-1}\right\|_E,$$

which implies that

$$\left|\|(A+C)^{-1}\|_E - \|A^{-1}\|_E\right| \le \sum_{k=1}^{\infty} \|A^{-1}\|_E^{k+1}\|C\|_E^k = \|A^{-1}\|_E^2\|C\|_E/(1-\|A^{-1}\|_E\|C\|_E)$$

or

$$1 - \frac{\|A^{-1}\|_E\|C\|_E}{1-\|A^{-1}\|_E\|C\|_E} \le \frac{\|(A+C)^{-1}\|_E}{\|A^{-1}\|} \le 1 + \frac{\|A^{-1}\|_E\|C\|_E}{1-\|A^{-1}\|_E\|C\|_E}.$$

Hence, under condition (40), we can derive from bounds (41) and the equality $q_2(A) = (q_3(A)\|A^{-1}\|_E)^2$ the following relations:

$$\left(1-\frac{\varepsilon_2}{1-\varepsilon_2}\right)^2 \exp\left(-\frac{2\varepsilon_2}{1-\varepsilon_2}\right) \le q_2(A+C)/q_2(A) \le \left(1-\frac{\varepsilon_2}{1-\varepsilon_2}\right)^{-2} \exp\left(\frac{2\varepsilon_2}{1-\varepsilon_2}\right). \quad (42)$$

It remains to observe that condition (38) with $\varepsilon_1 = \varepsilon_2/d$ follows from (40) in view of the elementary inequality $1/\|A^{-1}\|_E \le \|A\|_E/d$ (which is equivalent to the inequality $d^2 \le (\sum_{i=1}^d \lambda_i)(\sum_{i=1}^d \lambda_i^{-1})$, where $\lambda_i = \lambda_i(A^TA)$). Combining bounds (39), (42), and (41) and using the inequality $\exp(-2t) \ge (1-t)^2$, we conclude that the bounds

$$\left(1-\frac{\varepsilon_2}{1-\varepsilon_2}\right)^4 \le q_k(A+C)/q_k(A) \le \left(1-\frac{\varepsilon_2}{1-\varepsilon_2}\right)^{-4}, \quad k = 1, 2, 3 \quad (43)$$

are valid if (40) is fulfilled.

Finally, setting $\varepsilon_2 = \varepsilon/(1+\varepsilon)$, where $0 < \varepsilon < 1$, we obtain desired bounds (37). Moreover, condition (40) takes the form

$$\|C\|_E < \frac{\varepsilon}{1+\varepsilon}\frac{1}{\|A^{-1}\|_E}.$$

Since, in view of (32), we have

$$\frac{1}{\|A^{-1}\|_E} \equiv \frac{q_3(A)}{\sqrt{q_2(A)}} \ge (\Psi(\varphi(A)))^{-3/2},$$

condition (36) is obviously sufficient for assertion (37) in the lemma to hold.

Note that, to be specific, we can simply set $\varepsilon = 1/2$ in (36) and (37).

The following theorem makes it possible to indicate a simple sufficient condition for Property 3 to hold in the case under analysis (in particular, for hyperelasticity functionals).

Theorem 1. *Assume that the function* $\varphi(A)$ *defined in* (12) *has the barrier and coercitivity properties in the sense of* (27) *and* (28). *Moreover, assume that all the first and second derivatives of f satisfy the bound*

$$\max_{k,l}(|f_k(q_1, q_2, q_3)|, |f_{kl}(q_1, q_2, q_3)|) \le F(q_1, q_2, q_3), \quad (44)$$

where the function F is defined for all $q_k > 0$ *and is either monotone or convex with respect to each of its arguments. Then,* $\varphi(A)$ *satisfies Property* 3.

Proof. The maximum of a one-dimensional convex or monotone function on an interval is attained only at the endpoints of this interval. Hence, in view of definitions (22) and (23) and Lemmas 2 and 4, we have

$$|f_k(q_1(a+c), q_2(a+c), q_3(a+c))| \le F(\varphi(A)),$$

$$|f_{kl}(q_1(a+c), q_2(a+c), q_3(a+c))| \le F(\varphi(A)),$$

where

$$F(\varphi(A)) = \max_{\alpha_i = +1, -1} F(\gamma^{\alpha_1}, \gamma^{\alpha_2}, \gamma^{\alpha_3}), \quad \gamma = (1-\varepsilon)^{-4}\Psi(\varphi(A)).$$

Note that the right-hand side $F(\cdot)$ in these inequalities is a nondecreasing function of $\varphi(A)$. (Indeed, $\max_{\alpha^{-1} \le t \le \alpha} f(t)$ is a nondecreasing function of α.) Combining the bound just obtained with the result of Lemma 3, we can easily construct an appropriate upper bound for the norms of the gradient and Hessian matrix of φ, which are defined by formulas (25) and (26):

$$\|\nabla\varphi(a+c)\| \le \max_k |f_k| \sum_{k=1}^{3} \|\nabla q_k(a+c)\| \le F(\varphi(A))\beta(\varphi(A))\Psi_2(\varphi(A)) = \omega_1(\varphi(A)),$$

$$\|\nabla^2\varphi(a+c)\| \le \max_{k,l}(|f_k|, |f_{kl}|)\left[\sum_{k=1}^{3}\|\nabla^2 q_k(a+c)\| + \frac{1}{4}\left(\sum_{k=1}^{3}\|\nabla q_k(a+c)\|\right)^2\right]$$

$$\le F(\varphi(A))\left[\beta(\varphi(A)) + \left(\frac{1}{2}\beta(\varphi(A))\Psi_2(\varphi(A))\right)^2\right],$$

where

$$\beta(t) = \gamma_1[1 + (\Psi_2(t))^{d-2} + (\Psi_2(t))^{2d-4}].$$

It is obvious that the desired function $\omega(t)$ can be determined as the maximum of the right-hand side in the latest bounds and the inverse of the right-hand side in (18).

Theorem 2. *Let the function $\varphi_\Theta(A)$ be defined as*

$$\varphi_\Theta(A) = \Theta(\varphi(A)),$$

where $\varphi(A)$ satisfies all the conditions stated above, while $\Theta(z)$ is a monotonically increasing scalar function that defines a one-to-one mapping of the interval $[a_0, b_0]$, $a_0 \ge 0$, $b_0 \le +\infty$, onto the ray $[c_0, +\infty]$, $c_0 \ge 0$. Moreover, $\Theta(z) \in C^2[a_0, b_0)$, and $\Theta(z)$ can be set to $+\infty$ outside of this interval.

Also, assume that, for any $z \in (a_0, b_0)$, it holds that

$$\Theta'(z) \le \chi_1(\Theta(z)), \quad |\Theta''(z)| \le \chi_2(\Theta(z)),$$

where χ_1 and χ_2 are monotonically increasing functions, and, for each h such that

$$|h| \le 1/\chi_0(\Theta(z)), \tag{45}$$

we have $z + h \in (a_0, b_0)$ and

$$\Theta(z+h) \le \chi_0(\Theta(z)), \tag{46}$$

where χ_0 is a monotonically increasing function. Then, $\varphi_\Theta(A)$ satisfies Property 3.

Proof. We have

$$\nabla\varphi_\Theta(A) = \Theta'(\varphi(A))\nabla\varphi(A),$$

$$\nabla^2\varphi_\Theta(A) = \Theta'(\varphi(A))\nabla^2\varphi(A) + \Theta''(\varphi(A))(\nabla\varphi(A))(\nabla\varphi(A))^{\mathrm{T}}.$$

Assuming (18) and applying (19) and (20), we conclude from these relations that

$$\|\nabla\varphi_\Theta(A+C)\| = \Theta'(\varphi(A+C))\|\nabla\varphi(A+C)\| \le \chi_1(\Theta(\varphi(A+C)))\omega_1(\varphi(A))$$

and

$$\|\nabla^2\varphi_\Theta(A+C)\| \le \Theta'(\varphi(A+C))\|\nabla^2\varphi(A+C)\| + |\Theta''(\varphi(A))|\|\nabla\varphi(A+C)\|^2$$

$$\le \chi_1(\Theta(\varphi(A+C)))\omega(\varphi(A)) + \chi_2(\Theta(\varphi(A+C)))(\omega_1(\varphi(A)))^2.$$

Furthermore, taking into account (19), we find that, for a certain θ $(0 \le \theta \le 1)$, it holds that

$$|\varphi(A+C) - \varphi(A)| = |c^{\mathrm{T}}\nabla\varphi(A+\theta C)| \le \|C\|_E\|\nabla\varphi(A+\theta C)\| \le \|C\|_E\omega_1(\varphi(A)).$$

It follows that if we demand that

$$\|C\|_E\omega_1(\varphi(A)) \le \frac{1}{\chi_0(\Theta(\varphi(A)))},$$

in accordance with (45), then bound (46) with $z = \varphi(A)$ and $h = \varphi(A + C) - \varphi(A)$ takes the form

$$\Theta(\varphi(A + C)) \le \chi_0(\Theta(\varphi(A))) = \chi_0(\varphi_\Theta(A)).$$

Denote by Θ^{-1} the inverse function of Θ and note that Θ^{-1} is also an increasing function. We have $\varphi(A) = \Theta^{-1}(\varphi_\Theta(A))$. Thus, if

$$\|C\|_E \le \frac{1}{\omega_1(\Theta^{-1}(\varphi_\Theta(A)))\chi_0(\varphi_\Theta(A))}, \tag{47}$$

then $\varphi(A + C) \in (a_0, b_0)$ and it holds that

$$\|\nabla^2\varphi_\Theta(A + C)\| \le \chi_1(\chi_0(\varphi_\Theta(A)))\omega(\Theta^{-1}(\varphi_\Theta(A))) + \chi_2(\chi_0(\varphi_\Theta(A)))(\omega_1(\Theta^{-1}(\varphi_\Theta(A))))^2.$$

In a similar way, we derive a bound for the norm of the gradient $\|\nabla\varphi_\Theta(A + C)\|$. An explicit expression for ω_Θ can be obtained as the maximum of the right-hand side in the latest inequality, the inverse of the right-hand side in (47), and the function ω that meets self-consistency conditions (18) and (20) for $\varphi(A)$.

The validity of the conditions in Theorem 1 can easily be verified for function (10) and Ogden materials and also for function (9). For function (17), we must use Theorem 2.

Consider function (17),

$$\varphi_1 = \Theta(\varphi) = -\frac{1-t}{t} + \frac{1-t}{t^2}\frac{1}{t^{-1} - \varphi},$$

where φ is defined by (9). Here, we have $a_0 = 1$, $b_0 = 1/t$, and $c_0 = 1$; we also find that

$$\Theta' = \frac{t^2}{1-t}\left(\Theta + \frac{1-t}{t}\right)^2, \quad \Theta'' = \frac{2t^4}{(1-t)^2}\left(\Theta + \frac{1-t}{t}\right)^3,$$

i.e., functions χ_1 and χ_2 exist and satisfy the conditions in Theorem 2. The function

$$\chi_0(\Theta) = \frac{1-t+t^2}{t} + \frac{1-t+t^2}{1-t}\Theta$$

satisfies the property expressed by (45) and (46). Indeed, in this case, (45) takes the form

$$|h| \le \frac{t^2}{1-t+t^2}(t^{-1} - z);$$

then, $z - h \ge 1$, $z + h < t^{-1}$, and

$$\Theta(z + h) \le -\frac{1-t}{t} + \frac{1}{1 - t^2/(1-t+t^2)}\frac{1-t}{t^2}\frac{1}{t^{-1} - z} = \chi_0(\Theta(z)) - \frac{1-t}{t} \le \chi_0(\Theta(z)).$$

1.5. Global Characterization of Grid Deformation

To give a quantitative description of grid deformation as a whole, we introduce the distortion functional

$$\Phi(R) = \sum_{1 \le i \le M} \varphi(A_i)\mathrm{vol}(\mathcal{S}_i), \quad R = [r_1 \vdots \ldots \vdots r_N] \in \mathfrak{R}^{d \times N}, \tag{48}$$

where $A_i = A_i(R)$ is the Jacobian matrix corresponding to the ith cell. If, say, $d = 2$, then the vector $r_j = [x_j^{(1)}, x_j^{(2)}]^T$ determines the coordinates of the jth node, while the kth cell is determined by the nodes

$$[x_{j_1(k)}^{(1)}, x_{j_1(k)}^{(2)}]^T, \quad [x_{j_2(k)}^{(1)}, x_{j_2(k)}^{(2)}]^T, \quad [x_{j_3(k)}^{(1)}, x_{j_3(k)}^{(2)}]^T. \tag{49}$$

The symbol vol($\mathcal{S}_i$) denotes the volume of the ith simplex. Since ω is a monotone function, we have the obvious bound

$$\omega(\varphi(A_i(R))) \le \omega(\nu^{-1}\Phi(R)), \quad 1 \le i \le M, \tag{50}$$

where

$$\nu = \min_{1 \le i \le M} \mathrm{vol}(\mathcal{S}_i).$$

We will also need the following assertion.

Lemma 5. *For any n-by-N matrix S, it holds that*

$$\sum_{1 \le i \le M} \|A_i(S)\|_E^2 \le c_0^2 \|S\|_E^2,$$

where the constant c_0 *is proportional to the maximal number of edges that are incident to a grid node, while the matrix* $A_k(S)$ *is defined by* (8).

Theorem 3. *Let S be an arbitrary n-by-N matrix satisfying the inequality*

$$\|S\|_E \le \frac{1}{c_0\omega(\nu^{-1}\Phi(R))}. \tag{51}$$

Assume that Property 3 *is fulfilled. Then, it holds that*

$$\Phi(R+S) \le \Phi(R) + (S, \nabla\Phi(R)) + \frac{c_0^2\omega(\nu^{-1}\Phi(R))}{2}\|S\|_E^2. \tag{52}$$

Proof. We first note that, in view of (8), we have

$$A_i(R+S) = A_i(R) + A_i(S).$$

Also, the gradient of Φ, which determines the part of the increment $\Phi(R+S) - \Phi(R)$ that is linear in S, can be found from the relation

$$(S, \nabla\Phi(R)) + O(\|S\|^2) = \Phi(R+S) - \Phi(R) = \sum_{1 \le i \le M} [\varphi(A_i(R+S)) - \varphi(A_i(R))]$$

$$= \sum_{1 \le i \le M} [\varphi(A_i(R) + A_i(S)) - \varphi(A_i(R))] = \sum_{1 \le i \le M} (A_i(S), \nabla\varphi(A_i(R))) + O(\|S\|^2).$$

In accordance with the arrangement of components in the argument R and by analogy with the previous section, we consider $\nabla\Phi(R)$ as an n-by-N matrix.

Now, observe that, in view of Lemma 5 and relations (50), inequality (51) implies

$$\|A_i(S)\|_E^2 \le \sum_{1 \le k \le M} \|A_k(S)\|_E^2 \le c_0^2\|S\|_E^2 \le \frac{1}{\omega(\nu^{-1}\Phi(R))^2} \le \frac{1}{\omega(\varphi(A_i(R)))^2}.$$

Hence, Property 3 is fulfilled, which makes it possible to bound the nonlinear part of the increment by using the well-known formula

$$\varphi(a+c) = \varphi(c) + c^{\mathrm{T}}\nabla\varphi(a) + \frac{1}{2}c^{\mathrm{T}}\nabla 2\varphi(a+\theta c)c, \quad 0 \le \theta \le 1,$$

for $A = A_i(R)$ and $C = A_i(S)$:

$$\Phi(R+S) = \sum_{1 \le i \le M} \varphi(A_i(R+S)) = \sum_{1 \le i \le M} \varphi(A_i(R) + A_i(S)) = \Phi(R) + \sum_{1 \le i \le M} (A_i(S), \nabla\varphi(A_i(R)))$$

$$+ \frac{1}{2}\sum_{1 \le i \le M} (\mathbf{A}_i(S))^{\mathrm{T}}[\nabla^2\varphi(A_i(R) + \theta_i A_i(S))]\mathbf{A}_i(S) \le \Phi(R) + (S, \nabla\Phi(R))$$

$$+\frac{1}{2}\sum_{1\le i\le M}\omega(\varphi(A_i(R)))\|A_i(S)\|_E^2 \le \Phi(R) + (S, \nabla\Phi(R))$$

$$+\frac{\omega(\nu^{-1}\Phi(R))}{2}\left(\sum_{1\le i\le M}\|A_i(S)\|_E^2\right) \le \Phi(R) + (S, \nabla\Phi(R)) + \frac{\omega(\nu^{-1}\Phi(R))}{2}c_0^2\|S\|_E^2.$$

The two latest inequalities follow from (50) and Lemma 5, respectively.

2. ITERATIVE MINIMIZATION OF THE FUNCTIONAL

An admissible set of grids described by an n-by-N matrix R formed of the coordinates of grid nodes is defined as follows. For a sufficiently large $\Gamma > 0$, define the *feasible region* as $\mathcal{R}(\Gamma) = \{\Phi(R) \le \Gamma\}$.

2.1. The General Case

Let $R^{(k)} \in \mathcal{R}(\Gamma)$. Following (8), we examine the iterative process

$$R^{(k+1)} = R^{(k)} - \alpha_k\nabla\Phi(R^{(k)}), \quad k = 0, 1, \dots, \tag{53}$$

$$\alpha_k = \frac{1}{c_0\omega(\nu^{-1}\Phi(R^{(k)}))}\min\left(\frac{1}{c_0}, \frac{1}{\|\nabla\Phi(R^{(k)})\|_E}\right). \tag{54}$$

To this end, we use Theorem 3 with $R = R^{(k)}$ and $S = -\alpha_k\nabla\Phi(R^{(k)})$. It is easy to verify that condition (51) is satisfied in view of (54); therefore, (52) takes the form

$$\Phi(R^{(k+1)}) \le \Phi(R^{(k)}) - \left(\alpha_k - \frac{\alpha_k^2}{2}c_0^2\omega(\nu^{-1}\Phi(R^{(k)}))\right)\|\nabla\Phi(R^{(k)})\|_E^2.$$

Applying again (54) and taking into account the fact that $\alpha - c\alpha^2/2$ is an increasing function on the interval $0 \le \alpha \le 1/c$, we obtain the bound

$$\Phi(R^{(k+1)}) \le \Phi(R^{(k)}) - \frac{1}{2\omega(\nu^{-1}\Phi(R^{(k)}))}\min\left(\frac{\|\nabla\Phi(R^{(k)})\|_E}{c_0}, \frac{\|\nabla\Phi(R^{(k)})\|_E^2}{c_0^2}\right). \tag{55}$$

We sum the resulting inequalities for all the iteration steps up to the kth step inclusive and take into account the facts that the functional is bounded below (recall that $\Phi(R) \ge 0$), $\omega(\cdot)$ is a monotone function, and the values of Φ decrease at each step, as shown in (55). Then, it is easy to derive the bound

$$\Phi(R^{(0)}) \ge \frac{k+1}{2\omega(\nu^{-1}\Phi(R^{(0)}))}\min_{1\le i\le k}\min\left(\frac{\|\nabla\Phi(R^{(l)})\|_E}{c_0}, \frac{\|\nabla\Phi(R^{(l)})\|_E^2}{c_0^2}\right), \tag{56}$$

which directly implies the following assertion.

Theorem 4. *All the approximations in iterative process* (53), (54) *belong to the feasible region* $\mathcal{R}(\Phi(R^{(0)}))$, *and* $\|\nabla\Phi(R^{(k)})\|_E \longrightarrow 0$ *as* $k \longrightarrow 0$; *moreover, if*

$$k + 1 \ge 2\Phi(R^{(0)})\omega(\nu^{-1}\Phi(R^{(0)})), \tag{57}$$

then it holds that

$$\min_{0\le l\le k}\|\nabla\Phi(R^{(l)})\|_E^2 \le \frac{2c_0^2\Phi(R^{(0)})\omega(\nu^{-1}\Phi(R^{(0)}))}{k+1}. \tag{58}$$

Proof. If (57) is true, then (56) implies that

$$\min_{1\le l\le k}\min\left(\frac{\|\nabla\Phi(R^{(l)})\|_E}{c_0}, \frac{\|\nabla\Phi(R^{(l)})\|_E^2}{c_0^2}\right) \le 1.$$

Then,

$$\min_{1\le l\le k}\min\left(\frac{\|\nabla\Phi(R^{(l)})\|_E}{c_0}, \frac{\|\nabla\Phi(R^{(l)})\|_E^2}{c_0^2}\right) = \min_{1\le l\le k}\left(\frac{\|\nabla\Phi(R^{(l)})\|_E^2}{c_0^2}\right).$$

Substituting this equality in (56), we obtain (58). The fact that $\|\nabla\Phi(R^{(l)})\|_E$ converges to zero follows from the convergence of the series

$$\sum_{l=0}^{\infty}\min\left(\frac{\|\nabla\Phi(R^{(l)})\|_E}{c_0}, \frac{\|\nabla\Phi(R^{(l)})\|_E^2}{c_0^2}\right) \le 2\Phi(R^{(0)})\omega(\nu^{-1}\Phi(R^{(0)})),$$

where the latest bound can easily be derived from (55).

The result just obtained means that the norm of the gradient monotonically approaches zero for a certain subsequence of iteration steps.

2.2. The Case of a Polyconvex Function φ

Let $\varphi(A)$ be a polyconvex function (see [1]); i.e., $\varphi(A) = \beta_2(A, \det A)$ if $d = 2$ and $\varphi(A) = \beta_3(A, \operatorname{cof} A, \det A)$ if $d = 3$, where β_2 and β_3 are convex functions of 5 and 19 variables, respectively. It is well known that, in this case, $\varphi(A)$ also has the property of rank-one convexity, i.e.,

$$\varphi(\vartheta A + (1-\vartheta)B) \le \vartheta\varphi(A) + (1-\vartheta)\varphi(B),$$

if $\operatorname{rank}(B - A) = 1$. Note that functions (9) and (10) and the other stored energy functions are polyconvex.

In this case, Property 3 takes the following form:

Property 3'. There exist nondecreasing positive functions $\omega_1(t)$, $1/\omega_-(t)$, and $\omega_+(t)$ such that if $\det A > 0$ and C is an arbitrary perturbation matrix for which

$$\|C\|_E \le \frac{1}{\omega_+(\varphi(A))} \quad \text{and } \operatorname{rank} C = 1, \tag{59}$$

then

$$\|\nabla\varphi(A + C)\| \le \omega_1(\varphi(A)) \tag{60}$$

and

$$\omega_-(\varphi(A))p^{\mathrm{T}}p \le p^{\mathrm{T}}\nabla^2\varphi(A + C)p \le \omega_+(\varphi(A))p^{\mathrm{T}}p. \tag{61}$$

For instance, we can show that Property 3' holds for functional (10) with

$$\omega_-(t) = 1/\mu^2, \quad \omega_+(t) = \omega(t),$$

where

$$\omega(t) = \mu^{-2}\max[2\mu\exp(-1/2), 1 + 2\exp(1 + 2t)].$$

Also, in this case, we can easily obtain an appropriate modification of Theorem 3 in which, under the assumption $\operatorname{rank}(S) = 1$, not only an upper bound is given for the expression $\Phi(R + S) - \Phi(R) - (S, \nabla\Phi(R))$ but also a positive lower bound is derived for this expression.

Thus, in the presence of rank-one convexity, we can extend the iterative process indicated above by replacing the exact gradient of the functional $\nabla\Phi(R^{(k)})$ by its appropriate approximation

$$G_k = p_k q_k^{\mathrm{T}}, \quad p_k \in \mathcal{R}^d, \quad q_k \in \mathcal{R}^N,$$

which is arranged as a d-by-N matrix of rank 1. Moreover, it is not difficult to choose G_k so that

$$(G_k, \nabla\Phi(R^{(k)}) - G_k) = 0, \quad d^{-1}\|\nabla\Phi(R^{(k)})\|^2 \le \|G_k\|^2 \le \|\nabla\Phi(R^{(k)})\|^2. \tag{62}$$

In the simplest case, G_k is obtained by replacing all the rows in $\nabla\Phi(R^{(k)})$ by the zero ones except for the row with the greatest Euclidean length. The resulting minimization method reduces to the gradient descent along the direction that corresponds to a variation in only, say, the first coordinate of each grid point (if the first row in $\nabla\Phi(R^{(k)})$ turned out to be not shorter than the second). We can also use the singular value decomposition of the gradient matrix; in this case, the descent is effected by the displacement of each grid point along the left singular vector of the matrix $\nabla\Phi(R^{(k)})$ by a distance that is proportional to the corresponding component in its right singular vector. The vectors p_k and q_k are then taken as scalar multiples of the singular pair associated with the greatest singular value. Thus, we have

$$G_k = B_k \nabla\Phi(R^{(k)}),$$

where B_k is an orthoprojector.

Up to the expression for c_0, the result of Theorem 4 remains valid in this case as well. Indeed, using Theorem 3 with $R = R^{(k)}$ and $S = -\alpha_k G_k$, we find that condition (1) is equivalent to the relation

$$\alpha_k \|G_k\| \le 1/c_0 \omega(\nu^{-1}\Phi(R^{(k)})), \tag{63}$$

while bound (52) implies the inequality

$$\Phi(R^{(k+1)}) \le \Phi(R^{(k)}) - \left(\frac{\alpha_k}{d} - \frac{\alpha_k^2}{2} c_0^2 \omega(\nu^{-1}\Phi(R^{(k)}))\right) \|\nabla\Phi(R^{(k)})\|_E^2$$

in view of (62). Thus, if we define the iterative process by the formulas

$$R^{(k+1)} = R^{(k)} - \alpha_k G_k, \quad k = 0, 1, \dots, \tag{64}$$

$$\alpha_k = \frac{1}{c_0 \omega(\nu^{-1}\Phi(R^{(k)}))} \min\left(\frac{1}{c_0 d}, \frac{1}{\|\nabla\Phi(R^{(k)})\|_E}\right), \tag{65}$$

then, in view of (65) and (62), the relation (63) is fulfilled and the bound for the decrease in the functional value takes the form

$$\Phi(R^{(k+1)}) \le \Phi(R^{(k)}) - \frac{1}{2\omega(\nu^{-1}\Phi(R^{(k)}))} \min\left(\frac{\|\nabla\Phi(R^{(k)})\|_E}{c_0 d}, \frac{\|\nabla\Phi(R^{(k)})\|_E^2}{c_0^2 d^2}\right).$$

Thus, the assertion of Theorem 4 (with c_0 replaced by $c_0 d$) remains true if a rank-one approximation of the gradient is used. At the same time, it is sufficient here that Property 3 holds only for rank-one increments C, which may significantly facilitate the verification of condition (20) or extend the class of admissible functionals or improve the resulting bound (due to the use of a lesser function ω).

2.3. *The Use of Preconditioning*

It is well known that if iterative approximations are sufficiently close to the solution and the Hessian matrix $\nabla^2\Phi$ is not too ill-conditioned, then the rate of convergence can be considerably improved by applying Newton's method

$$R^{(k+1)} = R^{(k)} - \alpha_k (\nabla^2\Phi(R^{(k)}))^{-1} \nabla\Phi(R^{(k)}), \quad k = 0, 1, \dots.$$

Here, $0 < \alpha_k \longrightarrow 1$ is the step size parameter chosen so that the norm $\|\Phi(R^{(k+1)})\|_E$ is decreased.

However, in practice, a sufficiently good approximation to the desired grid is not available and the dN-by-dN Hessian matrix turns out to be not only ill-conditioned but also indefinite. Hence, the costly Newton iteration may simply not converge.

The situation is much more favorable if the property of rank-one convexity is used. As before, we restrict ourselves to analyzing variations in the coordinates of grid nodes along a single coordinate axis. Then, it can be verified that the corresponding N-by-N Hessian matrix (i.e., the appropriate diagonal block in $\nabla^2\Phi$) is positive definite, which ensures the convergence of the corresponding modification of Newton's method. The resulting method is very close to the algorithmic construction employed in [7, 3], where, instead of the exact Hessian matrix in Newton's method, the authors used its block diagonal part consisting of d blocks of size N-by-N corresponding to the same coordinatewise arrangement and partition.

A theoretical analysis of the resulting method for minimization of a functional can be performed analogously to the analysis given above. The use of approximate inverses for the blocks in the Hessian matrix can be justified in the same way as in [8].

REFERENCES

1. J. M. Ball, "Convexity Conditions and Existence Theorems in Nonlinear Elasticity," Arch. Rational Mech. Anal. **63**, 337–403 (1977).
2. P. G. Ciarlet, *Mathematical Elasticity* (Elsevier, New York, 1988/1997; Mir, Moscow, 1992), Vols. 1, 2.
3. V. A. Garanzha, "Barrier Method for Quasi-Isometric Grid Generation," Zh. Vychisl. Mat. Mat. Fiz. **40**, 1685–1705 (2000) [Comput. Math. Math. Phys. **40**, 1617–1637 (2000)].
4. P. J. Blatz, "On the Thermostatic Behavior of Elastomers," in *Polymer Networks, Structure, and Mechanical Properties* (Plenum, New York, 1971), pp. 23–45.
5. P. G. Ciarlet and G. Geymonat, "Sur les lois de comportement en elasticite non-lineaire compressible" C.R. Acad. Sci. Paris, Ser. II **295**, 423–426 (1982).
6. R. W. Ogden, "Large Deformation Isotropic Elasticity: On the Correlation of Theory and Experiment for Compressible Rubber-Like Solids," Proc. R. Soc. London A **328**, 567–583.
7. V. A. Garanzha and I. E. Kaporin, "Regularization of the Barrier Variational Method of Grid Generation," Zh. Vychisl. Mat. Mat. Fiz. **39**, 1489–1503 (1999) [Comput. Math. Math. Phys. **39**, 1426–1440 (1999)].
8. I. E. Kaporin, "Using Inner Conjugate Gradient Iterations in Solving Large-Scale Sparse Nonlinear Optimization Problems," Zh. Vychisl. Mat. Mat. Fiz. **43**, 802–807 (2003) [Comput. Math. Math. Phys. **43**, 766–771 (2003)].