Nonlinear Coordinate Transformations for Unconstrained Optimization II. Theoretical Background*

TAMÁS RAPCSÁK
Computer and Automation Institute, Hungarian Academy of Sciences, Budapest, Hungary

and

TIBOR CSENDES
Kalmár Laboratory, József Attila University, Szeged, Hungary

(Received: 14 December 1990; accepted: 23 March 1992)

Abstract. In this two-part article, nonlinear coordinate transformations are discussed in order to simplify global unconstrained optimization problems and to test their unimodality on the basis of the analytical structure of the objective functions. If the transformed problems can be quadratic in some or all the variables, then the optimum can be calculated directly, without an iterative procedure, or the number of variables to be optimized can be reduced. Otherwise, the analysis of the structure can serve as a first phase for solving global unconstrained optimization problems.

The first part treats real-life problems where the presented technique is applied and the transformation steps are constructed. The second part of the article deals with the differential geometrical background and the conditions of the existence of such transformations.

Key words. Global optimization, smooth unconstrained optimization, geodesic convexity, unimodal function, nonlinear coordinate transformation.

Mathematics Subject Classifications (1991). 65K05, 90C30, 54H00.

1. Introduction

The global unconstrained optimization problem is considered in the following from:

$$\min f(\mathbf{x}), \quad \mathbf{x} \in R^n, \tag{1.1}$$

where R^n is the n-dimensional Euclidean space and $f: R^n \to R$ is a twice continuously differentiable function. The exact solution of global optimization problems with even a few variables should be extremely difficult and time-consuming, e.g., Horst and Tuy (1990), Ratschek and Rokne (to appear). In applications, one frequently finds only local optimum or unproved global one.

*The paper was presented at the II. IIASA Workshop on Global Optimization, Sopron (Hungary), December 9–14, 1990.

Journal of Global Optimization **3**: 359–375, 1993.

These numerical experiences verify the importance of Tuy's conclusion: "To construct suitable algorithms, one has to understand the mathematical structure of problems."

This paper is devoted to the analysis of the structure of smooth global unconstrained optimization problems. The results can serve as the basis of the first phase for solving global optimization problems, which is needed to clarify the possibilities of simplification, the reduction of variables and the nature of the problem, providing useful information for choosing and/or developing convenient algorithms. Without this analysis it is impossible to imagine expert systems including global optimization methods. Analytical tools used earlier in global optimization were reported by Hansen *et al.* (1989, 1991).

Nonlinear coordinate transformations are widely used in nonlinear optimization, but a systematic description taking into account the improvement of algorithms is not available. There are two subclasses. The first one consists of the nonlinear coordinate transformations completely preserving the structure of the problem from the optimization point of view, while the second subclass involves the remaining ones. This paper describes the structure of the smooth global unconstrained optimization problem such that the resulting properties do not depend on nonlinear coordinate transformations belonging to the first subclass.

In the second part, the smooth global unconstrained optimization will be considered as an optimization problem defined on a Riemannian manifold. This approach enables us to extend both the theoretical results investigated in detail by Rapcsák (1989) and the algorithmic ones.

On the basis of this approach, the optimization problem can be handled as a tensor field optimization problem introduced by Rapcsák (1990). In differential geometry, in theoretical physics and in several applications of mathematics, the concept of tensor proved to be instrumental. In optimization theory, a new class of methods, called tensor methods, was introduced for solving systems of nonlinear equations (Schnabel and Frank, 1984) and for unconstrained optimization using second derivatives (Schnabel and Chow, 1991). Tensor methods are general-purpose methods intended especially for problems where the Jacobian matrix at the solution is singular or ill-conditioned. The description of a linear programming problem in the tensor notation is proposed in order to study the integrability of vector and multivector fields associated with interior point methods by Iri (1991). The most important feature of tensors is that their values do not change when they cause nonlinear coordinate transformations, and thus this notion seems to be useful for the characterization of structural properties. This motivated the idea of using this notion within the framework of nonlinear optimization. The third part contains the extension of the tensor approach to global unconstrained optimization problems.

In the next part, the local-global property (every local optimum is global) of a smooth nonlinear unconstrained optimization problem is investigated. This property is related to the concept of generalized convexity (unimodality) which plays an important role in the mathematical optimization theory.

The usual set convexity in linear topological spaces is based upon the possibility of connecting any two points of the space, which has led to the convex and generalized convex functions as well as to convex optimization. Since convexity is often not enjoyed by real problems, various approaches to the generalizations of the usual line segment have recently been proposed by Ortega and Rheinboldt (1970), Ben-Tal (1977), Prenowitz and Jantosciak (1979), Avriel and Zang (1980), Horst (1982, 1984), Martin (1982), Singh (1983), Castagnoli and Mazzoleni (1987) and Horst and Thach (1988). In this conception, a generalization was proposed which differs from the others in the use of a Riemannian manifold as a definition domain by Rapcsák (1986, 1987, 1991). In this case, the linear space is replaced by a Riemannian manifold, and the line segment by a geodesic arc. The advantage of this approach, motivated first of all by Luenberger's works (1972, 1973), is the recognition of the geometrical structure of optimization problems which can lead to new theoretical and algorithmic results.

When geodesic convexity has been proved, it is concluded that a stationary point is a global optimum point too, and consequently every algorithm which gives a stationary point gives a global minimum point too. The geodesic convex optimization problems contain the convex ones as a special case.

First, the characterization of geodesic convexity was elaborated in the case of a submanifold of R^n Rapcsák (1987, 1991) by using the tools of the immersion. On the basis of this result, in order to check the geodesic convexity property of a function on the feasible region it is necessary and sufficient to state the positive semidefiniteness of a suitable matrix in this domain. For nonlinear optimization problems containing only equalities, such a matrix is constructed by means of the gradients and the Hessian matrices of the objective and the constraint functions. The corresponding computational complexity is of the same order as in the convexity and less than in the pseudo-convexity.

By developing these results, a unified, coordinate-free framework based on the notion of tensor and tensor calculus was obtained in order to formulate the statements independently from the immersion, yet ensuring the possibility of symbolic computation Rapcsák (1990).

The main aim of the following part is to clarify the geometrical background and to characterize a subclass of the global unconstrained optimization problems endowed with the geodesic convexity property by applying the previous results. It seems that this generalization enlarges the class of functions with a local-global property.

After characterizing the structure of global unconstrained optimization problems and the properties (optimality and geodesic convexity) which do not depend on nonlinear coordinate transformations, it is possible to give conditions for using nonlinear coordinate transformations in order to obtain a quadratic function (Morse theorem) and for improving the original structure without changing the optimality properties.

Finally, the symbolic computation aspects of the results are discussed.

2. Smooth Unconstrained Optimization Problem

Consider the following optimization problem:

$$\min f(\mathbf{x}), \quad \mathbf{x} \in R^n, \tag{2.1}$$

where R^n is the n-dimensional Euclidean space and $f: R^n \rightarrow R$ is a twice continuously differentiable function.

An $\mathbf{x}_0 \in R^n$ having the property $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for all $\mathbf{x} \in R^n$ is called a global minimum for f. If $f(\mathbf{x}_0) \leq f(\mathbf{x})$ holds for a neighbourhood of $\mathbf{x}_0$, then $\mathbf{x}_0$ is called a local minimum for f.

The space R^n is the product space of ordered n-tuples of real numbers which forms an n-dimensional differentiable manifold. Consider two copies of a Euclidean space: R^n with Cartesian coordinates $\mathbf{x} = (x_1, \ldots, x_n)$ and R'^n with Cartesian coordinates $\mathbf{u} = (u_1, \ldots, u_n)$.

DEFINITION 2.1. A continuous coordinate system in a domain of Euclidean space R^n is said to be a system of functions $\mathbf{u}(\mathbf{x})$ which maps this domain continuously and bijectively onto a certain domain of R'^n.

The system of functions $\mathbf{u}(\mathbf{x})$ is a homeomorphism between the two domains which defines the coordinates of the domain of R^n relative to this homeomorphism. Let the system of functions $\mathbf{x}(\mathbf{u})$ denote the inverse mapping of $\mathbf{u}(\mathbf{x})$ which is the parametrization of the original domain in a differential geometric sense relative to the inverse mapping. Among all continuous coordinate mappings, those are of special interest that define a smooth mapping between the domains.

DEFINITION 2.2. A curvilinear coordinate system in a domain of Euclidean space R^n is a system of smooth functions $\mathbf{u}(\mathbf{x})$ which maps bijectively the domain of R^n onto a domain in R'^n such that the determinant of Jacobian matrix $\det(J\mathbf{u}(\mathbf{x})) = \det(\partial\mathbf{u}/\partial\mathbf{x})$ is not zero at all points of the domain of R^n.

If the set of smooth functions $\mathbf{u}(\mathbf{x})$ has the property that the determinant of the Jacobian matrix is not zero in a domain of R^n, then for each point of this domain there exists an open neighbourhood such that $\mathbf{u}(\mathbf{x})$ defines a local coordinate system in this neighbourhood. In this case, the inverse mapping also defines local curvilinear coordinates.

By introducing curvilinear coordinate systems in problem (2.1), we may consider it as nonlinear coordinate transformations. From our point of view, the local nonlinear coordinate transformations of parametrizations in a differential geometric sense will be interesting because in this case problem (2.1) can be expressed in a neighbourhood U of R'^n as follows:

$$\min f(\mathbf{x}(\mathbf{u})), \quad \mathbf{u} \in U \subset R'^n. \tag{2.2}$$

This is the standard way to handle differentiable manifolds.

In problem (2.1), the manifold R^n is endowed with the Euclidean metric which is a special Riemannian one.

DEFINITION 2.3. A Riemannian metric is said to be given in a Euclidean domain if, in any curvilinear coordinate system, there is defined a set of smooth functions g_{ij}, $i, j = 1, \dots, n$ (a matrix function G) such that the following conditions hold:

1. $g_{ij} = g_{ji}$, $i, j = 1, \dots, n$, (the matrix function G is symmetric),
2. G is nonsingular and positive definite,
3. under curvilinear coordinate transformations, the value of the quadratic forms defined by G does not change.

If the indefiniteness is required instead of the positive definiteness, then the Riemannian metric is indefinite.

A differentiable manifold endowed with a Riemannian metric is called a Riemannian manifold.

It is possible to deduce from (2.1) an equivalent problem in the form

$$\min f(\mathbf{x}), \quad \mathbf{x} \in M, \tag{2.3}$$

where $f \in C^2$ and M is a Riemannian C^2 n-manifold.

3. Tensors in Optimization

In this part, the object is to have a unified, coordinate-free framework based on the notion of tensor and tensor calculus for both theoretical and algorithmical aspects of the global unconstrained optimization problems. This approach was introduced in Rapcsák (1990).

First the notion of the tensor is recalled and thereafter the tensor (field) optimization problem will be defined.

Let M be a C^2 n-manifold and let m be a point in M. The tangent space TM_m at m is an n-dimensional vector space. Let TM_m^* be the dual space of TM_m. TM_m^* is endowed with its natural vector space structure. The theory of linear algebra can now be applied to define tensors (e.g. Hicks, 1965).

DEFINITION 3.1. A p-covariant tensor at m (for $p > 0$) is a real-valued p-linear function on $TM_m \times TM_m \times \cdots \times TM_m$ (p-copies).

A q-contravariant tensor at m (for $q > 0$) is a real-valued q-linear function on $TM_m^* \times TM_m^* \times \cdots \times TM_m^*$ (q-copies).

A p-covariant and a q-contravariant tensor at m is a $(p+q)$-linear real-valued function on $(TM_m)^p \times (TM_m^*)^q$.

A tensor is symmetric if its value remains the same for all possible permutations of its arguments.

A 0-covariant tensor at m is a real number.

A tensor field on M is a mapping that assigns a tensor at m to each m in M.

DEFINITION 3.2. A nonlinear optimization problem is said to be a tensor optimization one at $m \in M$ if the objective function and all the constraints are tensors, i.e., the problem is given in the following form:

$$\begin{aligned}
&\min T(\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{v}_1^*, \ldots, \mathbf{v}_q^*) \\
&T_j(\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{v}_1^*, \ldots, \mathbf{v}_q^*) = 0, \quad j = 1, \ldots, k_1, \\
&T_i(\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{v}_1^*, \ldots, \mathbf{v}_q^*) \geq 0, \quad i = 1, \ldots, m_1, \\
&(\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{v}_1^*, \ldots, \mathbf{v}_q^*) \in (TM_m)^p \times (TM_m^*)^q,
\end{aligned} \tag{3.1}$$

where T, T_j, T_i, $j = 1, \ldots, k_1$, $i = 1, \ldots, m_1$ are p-covariant and q-contravariant tensors.

If all the tensors are replaced by tensor fields in the definition, then the tensor field optimization problem is obtained.

DEFINITION 3.3. A second-order covariant tensor is positive semidefinite (definite) at a point $m \in M$ if the corresponding matrix is positive semidefinite (definite) on $TM_m \times TM_m$ in any coordinate representation.

A second-order covariant tensor field is positive semidefinite (definite) on $A \subset M$ if it is positive semidefinite (definite) at every point of A.

If there are no constraints, then the problem is an unconstrained tensor (field) optimization problem.

It follows from the definitions that neither the values of the tensors nor the optimum value of problem (3.1) change on a nonlinear coordinate transformation, and thus this problem class can become an adequate tool to study the structure of problem (2.3).

To build a tensor optimization problem, the operations of tensor algebra (addition, subtraction, multiplications and contraction) and of tensor analysis (covariant differentiation) can be applied, as can other operations which preserve the tensor character.

REMARK 3.1. As the objective function $f(\mathbf{x})$ of problem (2.3) is a 0-covariant tensor field on M, problem (2.3) is an unconstrained tensor field optimization problem on A.

In this paper, covariant differentiation will be the most important tool. In order to explain covariant differentiation, we follow Milnor (1969) and Gabay (1982). Define a vector field V on the manifold M as a smooth map $V: M \rightarrow R^n$ such that $V(m) \in TM_m$ for all $m \in M$.

Let $m \in M$. Given a vector $\mathbf{v} \in TM_m$ and a vector field W on M, we define a new vector $D_{\mathbf{v}}W \in TM_m$, called the covariant derivative of W along $\mathbf{v}$.

The application $D_{\mathbf{v}}W: TM_m \to TM_m$ must be linear in $\mathbf{v}$ and satisfy the chain rule given by

$$D_{\mathbf{v}}(fW) = f(m)D_{\mathbf{v}}W + \nabla f(m)W\,, \tag{3.2}$$

where $f(m)$ is any real-valued smooth function on M. It specifies an affine connection on M at m.

Let V and W now be vector fields on M. We define the field $D_{\mathbf{v}}W$, the covariant derivative of W with respect to V on M, by its values

$$D_{V(m)}W = D_{\mathbf{v}}W\,, \tag{3.3}$$

where $\mathbf{v} = V(m) \in TM_m$.

The affine connection is thus specified globally on M (e.g. Milnor, 1969).

The covariant differentiation can be extended to arbitrary tensor fields. In a system of local coordinates, the coefficient functions of the covariant differentiation (affine connection) $\Gamma^{l_3}_{l_1 l_2}$, $l_1, l_2, l_3 = 1, \ldots, n$ define the covariant derivative for all the tensor fields. Let $V\Gamma = \Sigma^n_{l_3=1} V_{l_3}\Gamma^{l_3}_{l_1 l_2}$, $l_1, l_2 = 1, \ldots, n$, where V_{l_3}, $l_3 = 1, \ldots, n$ are the component functions of a covariant vector field, and let $V\Gamma = \Sigma^n_{l_1=1} V^{l_1}\Gamma^{l_3}_{l_1 l_2}$, $l_2, l_3 = 1, \ldots, n$, where V^{l_1}, $l_1 = 1, \ldots, n$ are the component functions of a contravariant vector field. The following result is well-known in differential geometry (e.g. Mischenko and Fomenko, 1988):

PROPOSITION 3.1. *On a covariant vector field V, the covariant derivative is equal to*

$$DV = JV - V\Gamma\,, \tag{3.4}$$

while on a contravariant vector field

$$DV = JV + V\Gamma\,, \tag{3.5}$$

where JV denotes the Jacobian matrix of the corresponding vector field and $V\Gamma$ *is the multiplication of the vector field and the three-dimensional* Γ *matrix at each point of an arbitrary coordinate neighbourhood.*

For an arbitrary tensor field, the covariant derivative forms a tensor field.

If the tensor field is scalar (i.e. a smooth function on M), then the covariant derivative is equal to the gradient.

DEFINITION 3.4. If

$$\Gamma^{l_3}_{l_1 l_2} = \Gamma^{l_3}_{l_2 l_1} \quad \text{for all} \quad l_1, l_2, l_3 = 1, \ldots, n\,, \tag{3.6}$$

in every system of local coordinates, then the connection is symmetric.

The fundamental theorem of Riemannian manifolds is as follows (e.g. Mischenko and Fomenko, 1988):

THEOREM 3.1. *Let G be a symmetric matrix function defining the metric on a Riemannian manifold M in any system of local coordinates. Then there exists a unique symmetric connection such that*

$$\Gamma_{l_1 l_2}^{l_3} = \sum_{\alpha=1}^{n} 1/2\left(\frac{\partial g_{l_1\alpha}}{\partial x_{l_2}} + \frac{\partial g_{l_2\alpha}}{\partial x_{l_1}} - \frac{\partial g_{l_1 l_2}}{\partial x_\alpha}\right)(G^{-1})_{l_3\alpha} \tag{3.7}$$

for all $l_1, l_2, l_3 = 1, \ldots, n$.

Here, the coefficient functions $\Gamma_{l_1 l_2}^{l_3}$, $l_1, l_2, l_3 = 1, \ldots, n$ uniquely determined by the Riemannian metric are called the second Christoffel symbols.

If the Riemannian metric is Euclidean, then $\Gamma_{l_1 l_2}^{l_3} = 0$, $l_1, l_2, l_3 = 1, \ldots, n$.

4. Geodesic Convexity for Unconstrained Optimization Problems

In this part, the object is the characterization of the geodesic convexity property for global unconstrained optimization problems by using tensor calculus. The approach given here can be found in details in the paper of Rapcsák (1991), and therefore the proofs of the first statements are omitted. Geodesic convexity was investigated under a different definition of the geodesic convex set (totally geodesic convex set) in Udriste (1976, 1977, 1979, 1984).

Let M be a connected Riemannian C^2 n-manifold. As is usual in differential geometry, a curve of M is called a geodesic if its tangent is parallel along the curve (e.g. Hicks, 1965).

DEFINITION 4.1. A set $A \subset M$ is said to be geodesic convex if any two points of A are joined by a geodesic belonging to A.

EXAMPLE 4.1. A connected, complete Riemannian manifold is geodesic convex (e.g. Hicks, 1965).

EXAMPLE 4.2. For every point m in M, there is a neighbourhood U of m which is geodesic convex, and for any two points in U, there is a unique geodesic which joins the two points and lies in U (e.g. Hicks, 1965).

DEFINITION 4.2. *Let $A \subset M$* be a geodesic convex set. Then it is said that a function $f: A \rightarrow R$ is geodesic convex if its restrictions to all geodesic arcs belonging to A are convex in the arc length parameter.

By the definition, the following inequalities hold for every geodesic $\gamma(s) \in A$, $s \in [0, b]$ joining the two arbitrary points $m_1, m_2 \in A$:

$$f(\gamma(tb)) \leq (1-t)f(\gamma(0)) + tf(\gamma(b)), \quad 0 \leq t \leq 1, \tag{4.1}$$

where $\gamma(0) = m_1$, $\gamma(b) = m_2$ and s is the arc length parameter.

If $M \subset R^n$ is a connected Euclidean manifold, then the geodesic convex set $A \subset M$ is a convex set and the geodesic convex function $f: A \to R$ is a convex function on **A**, where

$$\gamma(tb) = m_1 + t(m_2 - m_1), \tag{4.2}$$

$b = |m_2 - m_1|$ and $||$ means the Euclidean norm of a vector.

From (4.1), we obviously conclude the following

LEMMA 4.1. *Let $A \subset M$ be a geodesic convex set and let $f: A \to R$ be a geodesic convex function. Then the level sets*

$$\{m \mid f(m) \leq f(m_0), \quad m, m_0 \in A\} \tag{4.3}$$

are geodesic convex.

THEOREM 4.1. *Let $A \subset M$ be a geodesic convex set and let $f: A \to R$ be a geodesic convex function. Then a local minimum point is also a global minimum point.*

THEOREM 4.2. *Let $A \subset M$ be an open geodesic convex set. Then a function $f: A \to R$ is geodesic convex if and only if it is geodesic convex in a geodesic convex neighbourhood of every point of A.*

DEFINITION 4.3. The point $m \in A \subset M$ is a stationary (critical) point of the continuously differentiable function $f: A \to R$ if the gradient $\nabla f(m)$ is equal to zero.

COROLLARY 4.1. *Let $A \subset M$ be an open geodesic convex set and let $f: A \to R$ be a continuously differentiable geodesic convex function. Then every stationary point of f is a global minimum point too. Moreover, the set of global minimum points is geodesic convex.*

The geodesic convexity is a natural generalization of the classic convexity notion of functions, because every Riemannian metric provides a geodesic convexity property. The geodesic convexity of unconstrained optimization problems will now be characterized by tensors.

Let $M = R^n$ endowed with a Riemannian metric, and let Df and D^2f denote the first- and second-order covariant derivative of f on M with respect to the Riemannian metric, respectively.

THEOREM 4.3. *Let $A \subset M$ be an open geodesic convex set and let $f: A \to R$ be a twice continuously differentiable function. Then f is (strictly) geodesic convex on A if and only if $D^2 f$ is a (strictly) positive semidefinite tensor at every point.*

Proof. If $A \subset M$ is a geodesic convex set, then a function $f: A \to R$ is geodesic convex if its restrictions to all geodesic arcs belonging to A are convex in the arc length parameter. By Theorem 4.2, a function $f: A \to R$ is geodesic convex if and only if it is geodesic convex in a geodesic convex neighbourhood of every point of A. Thus, it is sufficient to prove the statement only in an arbitrary geodesic convex neighbourhood.

Consider an arbitrary coordinate representation of the manifold M in any geodesic convex neighbourhood of A. Then a geodesic can be given in the form $\mathbf{x}(\mathbf{u}(s))$, $s \in (s_1, s_2)$, where this function is a twice continuously differentiable function and s means the arc length parameter. Now the geodesic convexity of the function $f(\mathbf{x}(\mathbf{u}(s)))$, $s \in (s_1, s_2)$ is equivalent to the non-negativeness of the second derivative at every point.

Let us introduce the following notations and operations:

$$H\mathbf{x}(\mathbf{u}) = \begin{pmatrix} Hx_1(\mathbf{u}) \\ \vdots \\ Hx_n(\mathbf{u}) \end{pmatrix}, \quad J\mathbf{x}(\mathbf{u}) = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{u}}\right),$$

where $Hx_i(\mathbf{u})$, $i = 1, \ldots, n$ are $n \times n$ Hessian matrices and $J\mathbf{x}(\mathbf{u})$ is an $n \times n$ matrix,

$$\mathbf{y}^T = (y_1, \ldots, y_n) \in R^n, \quad \mathbf{w}^T = (w_1, \ldots, w_k) \in R^n,$$

$$\mathbf{y}^T H\mathbf{x}(\mathbf{u}) = \sum_{i=1}^{n} y_i Hx_i(\mathbf{u}), \quad \mathbf{w}^T H\mathbf{x}(\mathbf{u})\mathbf{w} = \begin{pmatrix} \mathbf{w}^T Hx_1(\mathbf{u})\mathbf{w} \\ \vdots \\ \mathbf{w}^T Hx_n(\mathbf{u})\mathbf{w} \end{pmatrix},$$

$$\lambda \mathbf{w}^{\mathrm{T}} H\mathbf{x}(\mathbf{u})\mathbf{w} = \mathbf{w}^{\mathrm{T}} \lambda H\mathbf{x}(\mathbf{u})\mathbf{w}, \quad \lambda \in \mathbf{R}.$$

By differentiating twice the function $f(\mathbf{x}(\mathbf{u}(s)))$, $s \in (s_1, s_2)$, we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}s} f(\mathbf{x}(\mathbf{u}(s))) = \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u}(s))) J\mathbf{x}(\mathbf{u}(s))\mathbf{u}'(s),$$

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} f(\mathbf{x}(\mathbf{u}(s))) = \mathbf{u}'(s)^T J\mathbf{x}(\mathbf{u}(s))^T H_{\mathbf{x}} f(\mathbf{x}(\mathbf{u}(s))) J\mathbf{x}(\mathbf{u}(s))\mathbf{u}'(s)$$
$$+ \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u}(s)))\mathbf{u}'(s)^T H_{\mathbf{u}} \mathbf{x}(\mathbf{u}(s))\mathbf{u}'(s) + \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u}(s))) J\mathbf{x}(\mathbf{u}(s))\mathbf{u}''(s). \quad (4.4)$$

As the curve $\mathbf{x}(\mathbf{u}(s))$, $s \in (s_1, s_2)$ is a geodesic, we can substitute the following system of differential equations for $\mathbf{u}''(s)$:

$$u''(s) = -\mathbf{u}'(s)^T \Gamma \mathbf{u}'(s), \quad (4.5)$$

where the $n \times n \times n$ matrix Γ contains the second Christoffel symbols and $\mathbf{u}'(s)$, $s \in (s_1, s_2)$ are the tangent vectors. Considering only geodesics at each point and in every direction, we obtain that the geodesic Hessian matrix

$$H^g_{\mathbf{u}} f(\mathbf{x}(\mathbf{u})) = J\mathbf{x}(\mathbf{u})^T H_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u}) + \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) H\mathbf{x}(\mathbf{u}) - \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u})\Gamma , \quad (4.6)$$

where the matrix multiplication $J\mathbf{x}(\mathbf{u})\Gamma$ is defined by the rule related to the multiplication of a row vector and a space matrix, applied consecutively for every row vector of $J\mathbf{x}(\mathbf{u})$. (Note that the result does not change if first the multiplication $\nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u})$ is performed.)

The right-hand side of this expression is exactly the second-order covariant derivative of $f(\mathbf{x}(\mathbf{u}))$, i.e.

$$D^2 f(\mathbf{x}(\mathbf{u})) = J\mathbf{x}(\mathbf{u})^T H_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u}) + \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) H\mathbf{x}(\mathbf{u}) - \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u})\Gamma , \quad (4.7)$$

which proves the statement. □

COROLLARY 4.2. *The geodesic convexity property is invariant under nonlinear coordinate transformations.*

REMARK 4.1. The gradient $\nabla_{\mathbf{u}} f(\mathbf{x}(\mathbf{u}))$ is equivalent to the expression

$$Df(\mathbf{x}(\mathbf{u})) = \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u}) . \quad (4.8)$$

Rosenbrock's banana function $f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ defined on R^2 is a continuously differentiable nonconvex function having a unique minimum at $(1, 1)$. However, it does not belong to any family of generalized convex functions such as pseudo- or quasiconvex functions, because its level sets are nonconvex "banana-shaped" sets. In the first part of our article, a nonlinear coordinate transformation $y_1 = x_1$, $y_2 = x_1^2 - x_2$ was proposed in order to obtain a convex function. By Theorem 4.3 and Corollary 4.2, the geodesic convexity property does not depend on nonlinear coordinate transformations, and thus Rosenbrock's function is included in this class. This approach is more general than that of Avriel (1976).

For the tensor field optimization problem (2.3), the characterization of the local optimality is a direct consequence of the preceding theorem, and thus the connection between the optimality and convexity properties becomes clearer.

THEOREM 4.4. *If $m_0 \in M$ is a local minimum point of* (2.3), *then*

$$Df(m_0) = 0 \quad (4.9)$$

and

$$D^2 f(m_0) \quad \text{is positive semidefinite} . \quad (4.10)$$

If, at $m_0 \in M$, (4.9) *holds and*

$$D^2 f(m_0) \quad \text{is positive definite} , \quad (4.11)$$

then m_0 is a strict local minimum of (2.3).

The optimality conditions and geodesic convexity characterized by tensors do not depend on nonlinear coordinate transformations, so it is possible to use this analytical tool in order to attain a better realization of the problem from an optimization point of view which preserves the structure.

We have seen above that the covariant derivative depends on the coefficient functions of the covariant differentiation (affine connection) and for a Riemannian manifold on the Riemannian metric (Theorem 3.1). It follows that the geodesic convexity property characterized by Theorem 4.3 has a one-to-one correspondence with the Riemannian metric (every Riemannian metric generates a geodesic convexity property), and thus a wide class of generalized convex functions having the local-global property can be introduced in optimization theory. For example, the Euclidean metric of R^n determines the straight lines as the geodesics and the well-known convexity notion of smooth functions defined in a linear vector space.

5. Improvement of the Structure

In the first part of our article, nonlinear parameter transformations are discussed to simplify the nonlinear objective function and to test whether it is unimodal. Here, the conditions of the existence of such transformations are investigated on the basis of the differential geometrical approach. Because of the manifold structure, the coordinate transformations are defined only in neighbourhoods and are generally nonlinear. For this case, the Morse theorem providing sufficient conditions is fundamental.

Let $Hf_{|TM}$ denote the Hessian matrix of the function f restricted to the tangent space TM of M.

DEFINITION 5.1. The matrix $Hf_{|TM}$ is called the Hessian of f at a critical point. If the Hessian matrix is nondegenerate at a critical point, i.e., if $Hf_{|TM}$ is a nonsingular matrix, then we call the critical point nondegenerate.

The Hessian of f is defined only at a critical point because it coincides with the second covariant derivative at this point. Outside this point the covariant derivative depends on the Riemannian metric. This fact is demonstrated in (4.6). This formula shows that the type of the critical point should not be changed by introducing a new Riemannian metric, as the covariant derivative is equal to zero.

DEFINITION 5.2. If m is a nondegenerate critical point of f on M, then the index of the Hessian matrix of f at m is called the index of the nondegenerate critical point m.

MORSE THEOREM. *Let m be a nondegenerate critical point of f on M. Then there is a local coordinate system in a neighbourhood of m satisfying* $\mathbf{u}(m) = 0$ *and*

$$f = f(m) + (u_1)^2 + \cdots + (u_r)^2 - (u_{r+1})^2 - \cdots - (u_n)^2 . \tag{5.1}$$

Here, $n - r$ is equal to the index of m.

The Morse theorem is the generalization of Theorem 1 of the first article for the case of a Riemannian manifold and a nondegenerate critical point.

The question now is: How to improve the structure of the problem without changing the optimization character? The next theorem investigates the possibility of the introduction of a Riemannian metric on a differentiable manifold such that the function $f: M \to R$ becomes geodesic convex on M. The proof needs the following lemma (e.g. Matsushima, 1972):

LEMMA 5.1. *Let $a_{ij}(\mathbf{u})$, $i, j = 1, \ldots, n$ be C^2 functions defined on a neighbourhood U of the origin of R^n satisfying $a_{ij}(\mathbf{u}) = a_{ji}(\mathbf{u})$, $i, j = 1, \ldots, n$, $\det A(\mathbf{u}) \neq 0$ for $\mathbf{u} \in U$, where we set the matrix function $A(\mathbf{u}) = (a_{ij}(\mathbf{u}))$. Then, there exist $n \times n$ nonsingular matrices $T(\mathbf{u}) = (t_{ij}(\mathbf{u}))$ such that the elements of $T(\mathbf{u})$ are C^2 functions defined on some neighbourhood V ($V \subset U$) of 0 and*

$$T(\mathbf{u})^T A(\mathbf{u})\, T(\mathbf{u}) = \begin{pmatrix} \varepsilon_1 & 0 & \cdots & 0 \\ 0 & \varepsilon_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varepsilon_n \end{pmatrix}, \tag{5.2}$$

$\varepsilon_i = 1$ or $\varepsilon_i = -1$, $i = 1, \ldots, n$ holds at each point $\mathbf{u} \in V$.

REMARK 5.1. In the original lemma, the elements of the matrices $A(\mathbf{u})$ and $T(\mathbf{u})$, $\mathbf{u} \in V$ are C^∞ functions (analytic), but this condition can be replaced in the proof by the twice differentiability.

THEOREM 5.1. *Let $A \subset M$ be the interior of a compact, geodesic convex set and let $f: A \to R$ be a twice continuously differentiable function. Assume that f has only one critical point on A which is nondegenerate and whose index is zero. Then there exists a Riemannian metric G on A such that f becomes a strictly geodesic convex function on A with respect to the metric G.*

Proof. By Theorem 4.2, it is sufficient to prove the geodesic convexity property for a covering coordinate neighbourhood system of A. As A is the interior of a compact set, there exists a finite number of neighbourhoods which cover it. Because of the assumptions, f is a geodesic convex function in a neighbourhood of the critical point, so we take a finite number of coordinate neighbourhoods not containing the critical point into account.

By Theorem 4.3, the function f is geodesic convex on A if and only if $D^2 f$ is a positive definite tensor field on A. The second-order covariant derivative $D^2 f$ depends on the Riemannian metric at every point which is different from the critical point, and thus in the following steps we will be able to ensure the positive definiteness of D^2f.

1. Introduce a nonlinear coordinate transformation in every coordinate neighbourhood such that the matrix

$$J\mathbf{x}(\mathbf{u})^T H_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u}) + \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) H_{\mathbf{u}} \mathbf{x}(\mathbf{u}) \tag{5.3}$$

becomes a constant diagonal matrix.

This is always possible by Lemma 5.1 and by the fact that the matrix is not a tensor field, and therefore we can introduce a previous nonlinear coordinate transformation (if it is necessary) such that its determinant will not be singular. Of course, the resulting diagonal matrices can be different in the different coordinate neighbourhoods.

Let $\mathbf{x}(\mathbf{u})$ denote the new coordinate representation, too.

2. Introduce an orthogonal transformation in R'^n such that all components of the vector $\nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J(\mathbf{x}(\mathbf{u}))$ become greater than zero.

As $\nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) \neq 0$ and $J(\mathbf{x}(\mathbf{u}))$ is a nonsingular matrix, such an orthogonal transformation always exists.

Denote $\mathbf{x}(\mathbf{u})$ the new coordinate representation, too.

3. Introduce the following Riemannian metric in every coordinate neighbourhood:

$$\begin{pmatrix} e^{-2cu_1} & 0 & \cdots & 0 \\ 0 & e^{-2cu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-2cu_n} \end{pmatrix}, \quad c > 0 \text{ (constant)}\,. \tag{5.4}$$

By (3.7), we obtain that the 3-dimensional matrix Γ contains only the components $\Gamma^i_{ii} = -c$, $i = 1, \ldots, n$ different from zero.

Thus the $n \times n$ matrix

$$-\nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{u})) J\mathbf{x}(\mathbf{u}) \Gamma \tag{5.5}$$

is a positive definite diagonal matrix in every coordinate neighbourhood with the elements multiplied by c.

4. Choose the value $c > 0$ in every coordinate neighbourhood such that the matrix $D^2 f(\mathbf{x}(\mathbf{u}))$ becomes a positive definite matrix. Let c^* denote the maximum of the c values in the finite number neighbourhoods and introduce the Riemannian metric

$$G = \begin{pmatrix} e^{-2c^*u_1} & 0 & \cdots & 0 \\ 0 & e^{-2c^*u_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-2c^*u_n} \end{pmatrix}, \quad c^* > 0 \text{ (constant)} \tag{5.6}$$

on A.

Then it turns out that the function f is geodesic convex on A with respect to the metric G. □

REMARK 5.2. The introduction of the metric G on A is equivalent to a nonlinear coordinate transformation.

Theorem 5.1 is not a local, but a global result, which ensures a general method for the improvement of the structure of smooth problems and algorithms. When the second derivatives or their approximations are used for solving unconstrained problems, then the change in the Riemannian metric can imply the positive definiteness of the geodesic Hessian matrices and the matrix updating formulas of variable metric methods. In this way, the Newton-like and quasi-Newton-like methods can be involved in this framework for a singular Hessian matrix too. Redkovskii (1989) and Perekatov and Redkovskii (1989) described nonlinear coordinate transformations to achieve a positive Hessian matrix for the Newton method. The change of the Riemannian metric can replace this type of nonlinear coordinate transformations.

6. Concluding Remarks

In this paper, nonlinear coordinate transformations are discussed in order to clarify some structural properties of global unconstrained optimization problems. The analysis of the structure can serve as a tool for the simplification of the problems and the reduction of the variables, and for choosing and developing convenient algorithms. The tensor approach is a global, coordinate-free description which can be connected with symbolic computation.

Symbolic computation is a new and promising field in mathematics and computer science, covering all aspects of algorithmic solutions to problems dealing with symbolic (i.e. non-numerical) objects. Important subareas of symbolic computation form the basis for many high-tech application areas such as CAD/CAM, robotics, geometric modelling, expert systems, etc. Symbolic operations for optimization (Stoutemyer, 1978) can be made by the softwares MATHEMATICA or REDUCE. From a global optimization point of view, the first and second covariant differentiation with respect to the Riemannian metric and the nonlinear coordinate transformations were proposed in this paper. In algorithms, nonlinear coordinate transformations (symbolic computation), e.g. Redkovskii (1989) can be replaced by changing the metric (numerical computation) on the basis of Theorem 5.1.

Acknowledgements

This work was supported by Grants OTKA 1074/1987, 2568/1991 and 2879/1991.

References

Avriel, M. (1976), *Nonlinear Programming, Analysis and Methods*, Prentice-Hall, New Jersey.
Avriel, M. and Zang, I. (1980), Generalized Arcwise Connected Functions and Characterizations of Local-Global Properties, *Journal of Optimization Theory and Applications* **32**, 407–425.

Ben-Tal, A. (1977), On Generalized Means and Generalized Convex Functions, *Journal of Optimization Theory and Applications* **21**, 1–13.

Castagnoli, E. and Mazzoleni, P. (1987), Generalized Connectedness for Families of Arcs, *Optimization* **18**, 3–16.

Gabay, D. (1982), Minimizing a Differentiable Function over a Differentiable Manifold, *Journal of Optimization Theory and Applications* **37**, 117–219.

Hansen, P., Jaumard, B., and Lu, S.-H. (1989), An Analytical Approach to Global Optimization, Rutcor Research Report 4-89, Rutgers University, New Brunswick.

Hansen, P., Jaumard, B., and Lu, S.-H. (1991) An Analytical Approach to Global Optimization, Mathematical Programming **52**, 227–254.

Hicks, N. J. (1965) *Notes on Differential Geometry*, Van Nostrand Publishing Company, Princeton, New Jersey.

Horst, R. (1982), A Note on Functions Whose Local Minima Are Global, *Journal of Optimization Theory and Applications* **36**, 457–463.

Horst, R. (1984) Global Optimization in Arcwise Connected Metric Spaces, *Journal of Optimization Theory and Applications* **104**, 481–483.

Horst, R. and Thach, P. T. (1988) A Topological Property of Limes-Arcwise Strictly Quasiconvex Functions, *Journal of Mathematical Analysis and Applications* **134**, 426–430.

Horst, R. and Tuy, H. (1990), *Global Optimization*, Springer-Verlag, Berlin, Heidelberg, New York.

Iri, M. (1991), Integrability of Vector and Multivector Fields Associated with Interior Point Methods for Linear Programming, *Mathematical Programming* **52**, 511–525.

Luenberger, D. G. (1972), The Gradient Projection Methods Along Geodesics, *Management Science* **18**, 620–631.

Luenberger, D. G. (1973), *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Reading.

Martin, D. H. (1982), Connected Level Sets, Minimizing Sets and Uniqueness in Optimization, *Journal of Optimization Theory and Applications* **36**, 71–93.

Matsushima, Y. (1972), *Differentiable Manifolds*, Marcel Dekker, Inc., New York.

Milnor, J. W. (1969), *Morse Theory*, Princeton University Press, Princeton, New Jersey, 1969.

Mishchenko, A. and Fomenko, A. (1988), *A Course of Differential Geometry and Topology*, Mir Publishers Moscow, Moscow.

Ortega, J. M. and Rheinboldt, W. C. (1970), *Iterative Solution of Nonlinear Equations*, Academic Press, New York.

Perekatov, A. E. and Redkovskii, N. N. (1989), Method for Minimization of Unimodal Non-Convex Functions, *Dokladi AN USSR Ser. A Physical-Mathematics and Technical Sciences* **10**, 36–38 (in Russian).

Prenowitz, W. and Jantosciak, J. (1979), *Join Geometries*, Springer, New York.

Rapcsák, T. (1986), Convex Programming on Riemannian Manifold, System Modelling and Optimization, *Proceedings of 12-th IFIP Conference*, Edited by A. Prékopa, J. Szelezsán and B. Strazicky, Springer Verlag, Berlin, Heidelberg, 733–741.

Rapcsák, T. (1987), Arcwise-Convex Functions on Surfaces, *Publicationes Mathematicae* **34**, 35–41.

Rapcsák, T. (1989), Minimum Problems on Differentiable Manifolds, *Optimization* **20**, 3–13.

Rapcsák, T. (1990), Tensor Optimization, MTA SZTAKI Report, 34/1990.

Rapcsák, T. (1991), Geodesic Convexity in Nonlinear Optimization, *Journal of Optimization Theory and Applications* **69**, 169–183.

Ratschek, H. and Rokne, J. (1993), Solving a Global Optimization Problem, *J. of Global Optimization* (to appear).

Redkovskii, N. N. (1989), Nonlinear Transformations of Coordinates in Unconstrained Optimization Problems, *Issledovanii Operatsii UAS* **34**, 60–65 (in Russian).

Schnabel, R. B. and Frank, P. (1984), Tensor Methods for Nonlinear Equations, *SIAM Journal on Numerical Analysis* **21**, 815–843.

Schnabel, R. B. and Chowe, T.-T. (1991) Tensor Methods for Unconstrained Optimization Using Second Derivatives, *SIAM Journal on Optimization* **1**, 293–315.

Singh, C. (1983), Elementary Properties of Arcwise Connected Sets and Functions, *Journal of Optimization Theory and Applications* **41**, 377–387.

Stoutemyer, D. R. (1978), Automatic Categorization of Optimization Problems: An Application of Computer Symbolic Mathematics, *Operations Research* **26**, 773–788.

Udriste, C. (1976) Convex Functions on Riemannian Manifolds, *St. Cerc. Mat.* **6**, 735–745 (in Roumanian).

Udriste, C. (1977), Continuity of Convex Functions on Riemannian Manifolds, *Bulletine Mathematique de Roumanie* **21**, 215–218.

Udriste, C. (1979), Directional Derivatives of Convex Functions on Riemannian Manifolds, *Revue Roumaine de Mathématiques Pures et Appliqées* **24**, 1385–1388.

Udriste, C. (1984), Kuhn–Tucker Theorem on Riemannian Manifolds, in *Colloquia Mathematica Societatis János Bolyai, 46. Topics in Differential Geometry*, Debrecen, Hungary, 1247–1259.