Nonlinear Coordinate Transformations for Unconstrained Optimization I. Basic Transformations*

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

and

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

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

Abstract. In this two-part article, nonlinear coordinate transformations are discussed to simplify unconstrained global optimization problems and to test their unimodality on the basis of the analytical structure of the objective functions. If the transformed problems are 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 unconstrained global 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, nonlinear parameter transformation, unconstrained optimization, unimodal function.

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

1. Introduction

In [6], a parameter estimation problem was studied, with an objective function of the sum-of-squares form

$$F(R_{aw}, I_{aw}, B, \tau) = \left[\frac{1}{m}\sum_{i=1}^{m} |Z_L(\omega_i) - Z_L'(\omega_i)|^2\right]^{1/2}, \tag{1}$$

where $Z_L(\omega_i) \in C$ is the measured complex impedance value; $Z_L'(\omega_i)$ is the modelled impedance at frequencies ω_i for $i = 1, 2, \ldots, m$ and R_{aw}, I_{aw}, B, τ are the model parameters. The model function was

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

Journal of Global Optimization **3**: 213–221, 1993.

$$Z_L'(\omega) = R_{aw} + \frac{B\pi}{4.6\omega} - j\left(I_{aw}\omega + \frac{B\log(\gamma\tau\omega)}{\omega}\right), \tag{2}$$

where $\gamma = 10^{1/4}$ and j is the imaginary unit. For this model identification problem, a different yet equivalent model function was found:

$$Z_L'(\omega) = R_{aw} + \frac{B\pi}{4.6\omega} - j\left(I_{aw}\omega + \frac{A + 0.25B + B\log(\omega)}{\omega}\right), \tag{3}$$

where the model parameters are R_{aw}, I_{aw}, A and B. This model function is linear in the parameters, the objective function value goes to infinity as the absolute value of any parameter grows, and thus the optimization problem has a unique minimum. The minimum can be calculated directly, without an iterative procedure (the zero of the gradient can be determined by a finite procedure, that is much easier than the solution of the global optimization problem in general). The reason for the existence of (3) is that the parameter B can be obtained via the fitting of the real part of the model function, and this value can be used to identify the product $B\log(\gamma\tau\omega)$. The variable transformation that gives (3) from (2) is $A = B\log(\tau)$. This simple finding and many other practical problems [1, 3] where analytical relations are *a priori* known between the variables motivated the present study to investigate nonlinear variable transformations for testing the unimodality of unconstrained nonlinear minimization problems.

In contrast with traditional numerical methods [2, 4], the technique to be presented utilizes the analytical expression of the objective function. Similar methods are rather rare in the literature. Stoutemyer discussed an analytical optimization method [9], and a program that can yield a problem class (such as linear, quadratic, separable, etc.) for a problem given in symbolic form [10]. The algorithm studied by Hansen *et al.* [5] is intended to solve constrained global optimization problems, and it also uses the expression of the constraints in a branch-and-bound framework. This algorithm gave the exact solution of many global optimization problems. The method presented in this paper can be applied as a further test procedure in this branch-and-bound algorithm. Redkovskii and Perekatov described nonlinear coordinate transformations [7, 8] to achieve positive definite Hessian matrices for the Newton local minimization method.

In the first part of our article, nonlinear variable transformations are discussed to simplify the nonlinear objective function and to test whether it is unimodal. The conditions of the existence of such transformations are also studied. The second part treats the differential geometrical background of these transformations.

2. Unimodality and Parameter Transformations

Consider the unconstrained nonlinear optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{4}$$

where $f(x): \mathbb{R}^n \to \mathbb{R}$ is a nonlinear function with continuous second derivatives. We study homeomorph coordinate transformations, i.e., invertible continuous one-to-one mappings $\mathbb{R}^n \to Y \subseteq \mathbb{R}^n$, for which the inverse mapping $Y \to \mathbb{R}^n$ is also continuous. We wish to find a transformation $y = h(x)$ such that $f(x) = f(h^{-1}(y)) = g(y)$, where $g(y)$ is a quadratic function, and thus its optimum can easily be calculated.

For the sake of simplicity, we consider here only the unconstrained nonlinear optimization problem, although practical problems are usually constrained. In parameter estimation problems such as those in [6], a unique global minimum may be of interest, even if it is outside the bounds (which typically give only a region of reasonable starting points).

A one-dimensional function $f(x)$ is defined to be unimodal [4] if there exists a point x^* such that for any points x_1 and x_2 for which $x_1 < x_2$: if $x_2 < x^*$ then $f(x_1) > f(x_2)$, and if $x^* < x_1$ then $f(x_1) < f(x_2)$. If $f(x)$ is continuous, then $f(x^*)$ is the unique local minimum, and thus the global minimum. What is more, the statements that a one-dimensional continuous function is unimodal, and that it has a unique local minimum and no local maximum, are equivalent (for the unconstrained case considered here). The unimodality is usually not defined for n-dimensional functions. The main reason for this is that the simple minimization algorithms based on the unimodal property of a function [4] cannot be generalized for the multidimensional case.

We now define an n-dimensional continuous function to be unimodal on an open set $X \subseteq \mathbb{R}^n$ if there exists a set of infinite continuous curves such that the curve system is a homeomorphic mapping of the polar coordinate system of the n-dimensional space, and the function $f(x)$ grows strictly monotonically along the curve. This definition means in other words that a function is unimodal if it has a single region of attraction on the given set X. For a smooth function, unimodality clearly ensures that it has only one local minimizer point, and there exists no maximizer point in X. This definition fits the optimization usage well, since multimodal and multiextremal functions are regarded as synonyms. An n-dimensional unimodal function does not necessarily satisfy the conditions of one-dimensional unimodality along every line in the n-dimensional space. Consider, for example, the well-known Rosenbrock or "banana" function [4]: it is clearly unimodal, yet the one-dimensional unimodality does not hold along the line $x_2 = 1$, for instance.

Theorem 1 discusses the relationship between homeomorph variable transformations and the unimodality of nonlinear functions.

THEOREM 1. *The continuous function $f(x)$ is unimodal in the n-dimensional real space if and only if there exists a homeomorph variable transformation $y = h(x)$ such that $f(x) = f(h^{-1}(y)) = y^T y + c$, where c is a real constant, and the origin is in the range S of $h(x)$.*

Proof. If $f(x)$ is unimodal with the only local minimizer point of x^*, then the constant c in Theorem 1 is equal to $f(x^*)$ and the point x have to be transformed to a y for which $|y| = \sqrt{f(x) - f(x^*)}$. If $x \neq x^*$, then the points satisfying this equation form a sphere around the origin. Choose now that candidate point y, which is the intersection of the sphere and the half-line that is the homeomorphic transform of the curve passing through x. The global minimizer point x^* must be transformed to the origin. The range S of the presented transformation may be a subset of $\mathbb{R}^n$ (e.g. if the range of $f(x)$ is finite). For each point y in S, there is exactly one half-line that starts at the origin and goes through y. The respective curve determines the inverse of y. The construction of the transformation ensures the necessary properties.

Consider now the case when there exists a variable transformation satisfying the requirements of Theorem 1. The global minimizer point of $y^T y + c$ on the range of $h(x)$ is then the origin, $x^* = h^{-1}(0)$ must be a global minimizer point of $f(x)$. The half-lines starting from the origin form the set of curves for which the conditions of unimodality are fulfilled. The inverse transformation h^{-1} is continuous according to the assumption, hence it gives the set of curves that prove the unimodality of $f(x)$. □

The transformation introduced in the proof cannot be used directly to test the unimodality of a nonlinear function and to determine the global minimum, since, e.g., the global minimizer point must be known to construct the transformation. Nevertheless, because of the uniqueness of this transformation (up to an appropriate rotation), we may make use of the form of this transformation.

3. Basic Transformation Steps

The main difference between model functions (2) and (3) is that $B \log \tau$ is replaced by a new parameter A. This was possible since the objective function is separable, it can be written as a sum of squared functions, and thus the optimal value of B is determined by fitting the term $B\pi/4.6\omega$. Having an optimal value for B, the fitting of the product $B \log \tau$ merely involves fitting here a single parameter A and then calculating τ from $A = B \log \tau$.

Assume that a variable x_i occurs in the expression $f(x)$ everywhere in the form of $h(x)$. Change every occurrence of $h(x)$ in the expression of $f(x)$ to y_i, and rename the remaining variables x_j as y_i providing the transformed function $g(y)$. Some variables may disappear from the expression via the substitution, and in this case $g(y)$ is constant as a function of these variables. The transformed function may be independent of a variable y_j, even if $\partial f(x)/\partial x_j$ is not zero. Consider for instance $f(x) = (x_1 + x_2)^2$, $h(x) = x_1 + x_2$, $g(y) = y_1^2$. Such a case was reported for a biomedical parameter estimation problem in [1]. The transformation between the two $\mathbb{R}^n$ spaces is not a differmorphism, yet it makes sense to omit the variables that does not affect the objective function value. The recognition of such a redundant parameter is very important for the user, it is not trivial (see again [1] and Example 4), and in general it is not detectable by usual numerical techniques. Computer algebra systems (such as Derive, Mathematica, or Reduce) can facilitate the variable transformations discussed in this paper.

The following statement is true for the introduced transformation:

THEOREM 2. *If $h(x)$ is smooth and strictly monotonic in x_i, then the corresponding transformation simplifies the function in the sense that each occurrence of $h(x)$ in the expression of $f(x)$ is padded by a variable in the transformed function $g(y)$, while every local minimizer (or maximizer) point of $f(x)$ is transformed to a local minimizer (maximizer) point of the function $g(y)$.*

Proof. Assume that x^* is a local minimizer point with the function value $f^* = f(x^*)$. Denote its transform by y^*. The neighborhood $N(x^*)$ is transformed to the neighborhood $N'(y^*)$ (possibly in a lower-dimensional space), since $h(x)$ is a smooth function. The strict monotonicity ensures that only the points of $N(x^*)$ are transformed to the points of $N'(y^*)$. The transformation does not change the function values, and hence $g(y) \geq g(y^*)$ holds for each $y \in N'(y^*)$. This indicates that the transform of a local minimizer point is a local minimizer point of the transformed problem. The same procedure can be carried out for local maximizer points. □

It would be sufficient to prove Theorem 2 that $h(x)$ has different values for different x_i variables. However, this property together with the smoothness of $h(x)$ implies its monotonicity. The following theorem discusses the case when the range of $h(x)$ is the whole real space.

THEOREM 3. *If $h(x)$ is smooth, strictly monotonic as a function of x_i, and its range is equal to $\mathbb{R}$, then for every local minimizer (or maximizer) point y^* of the transformed function $g(y)$ there exists an x^* such that y^* is the transform of x^*, and x^* is a local minimizer (maximizer) point of $f(x)$.*

Proof. Set $x_j^* = y_j^*$ for all y_j variables of $g(y)$ except from y_i. If some of the original variables x_k (for $k \neq i$) of $f(x)$ were deleted by the transformation then set these variables to zero. This setting of the deleted variables defines a hyperplane S in $\mathbb{R}^n$. Having fixed every x_j^* for $j = 1, 2, \ldots, i-1, i+1, \ldots, n$, the value of x_i^* can be determined for every y_i^* value, since $h(x)$ is strictly monotonic and its range is equal to $\mathbb{R}$. The function $h(x)$ is invertible for the coordinate x_i, and thus its inverse is continuous.

Consider a neighborhood $N(y^*)$ of the local minimizer point y^*. A neighborhood $N'(x^*)$ is obtained with the presented transformation in which x^* is a local minimizer point. If no variable was deleted in the transformation $y_i = h(x)$ then this conveys the proof of the Theorem. Otherwise, $N'(x^*)$ is a neighbourhood in the hyperplane S, and the hypersurfaces $h(x) = c$ where c is a real constant cover the original n-dimensional variable space. The function $f(x)$ is obviously constant on such a hypersurface. The point x^* is thus a local minimizer point in the region specified by $N'(x^*)$ and the hypersurfaces that intersect $N'(x^*)$. This region is not of zero measure, and hence x^* is a local minimizer point of $f(x)$ in $\mathbb{R}^n$. This train of thoughts can be repeated for local maximizer points. □

If the transformation suggested by Theorem 2 is invertible and smooth, then this transformation can be regarded as a transformation step, a sequence of which may form the transformation discussed in Theorem 1 (as in the case of the problem shown in the Introduction).

The following assertion (that can readily be proved with the rules of differentiation) facilitates the automatic recognition of possible variable transformations.

ASSERTION 1. *If a variable x_i appears everywhere in the expression of a smooth function $f(x)$ in a term $h(x)$, then the partial derivative $\partial f(x)/\partial x_i$ can be written in the form $(\partial h(x)/\partial x_i)p(x)$, where $p(x)$ is continuously differentiable.*

Obviously, a number of other factorizations exist besides the one mentioned in this Assertion, yet a canonical form can be derived for a wide class of functions (e.g. for the algebraic functions). Even if $\partial h(x)/\partial x_i$ is known, it may be difficult to determine a good transformation function $h(x)$. If $\partial f(x)/\partial x_i$ is not factorizable, then $h(x)$ is linear in x_i for the only possible transformation (satisfying the conditions of Theorem 2).

The condition of the existence of a variable transformation $h(x)$ that decreases the number of variables of an unconstrained nonlinear optimization problem is given in the following statement:

ASSERTION 2. *If the variables x_i and x_j appear everywhere in the expression of a smooth function $f(x)$ in a term $h(x)$, then the partial derivatives $\partial f(x)/\partial x_i$ and $\partial f(x)/\partial x_j$ can be factorized in the forms $(\partial h(x)/\partial x_i)p(x)$ and $(\partial h(x)/\partial x_j)q(x)$, respectively, and $p(x) = q(x)$.*

Again, the proof can be obtained via simple calculations. Nonlinear optimization problems are usually multiextremal and they can be simplified only in rare cases. However, the presented symbolic manipulation techniques may transform the original problem to a new one that is substantially easier to solve. The cases referred to indicated that a surprisingly large proportion of practical problems could be simplified by the presented procedure.

4. Examples

1. Consider first the Rosenbrock function [4]: $f(x) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2$. The usual bounds of the variables are $-2 \leq x_1 \leq 2$ and $-2 \leq x_2 \leq 2$. Factorization of the gradient of $f(x)$ showed that only linear transformations (in the transforma-

tion variable) are possible. There is no matching pattern for all occurrences of x_1; it appears in the expression of $f(x)$ as x_1^2, $1-x_1$, $x_1^2-x_2$, $(1-x_1)^2$ and $(x_1^2-x_2)^2$. The variable x_2 appears in the forms $x_1^2-x_2$, $(x_1^2-x_2)^2$ and $100(x_1^2-x_2)^2$, and these forms characterize all occurrences of x_2. The last two terms are not monotonic as functions of the variable x_2, and hence they cannot fulfil the conditions of Theorems 2 and 3.

With the transformation $y_1 = x_1$ and $y_2 = x_1^2 - x_2$, the function $g(y) = (1-y_1)^2 + 100y_2^2$ is obtained. This transformation is obviously smooth, its range is $\mathbb{R}^2$, and it is also invertible. The transformed function $g(y)$ has a diagonal Hessian, and thus it is separable. The function $g(y)$ is now quadratic and the only local minimum is easy to calculate: $y_1 = 1$, $y_2 = 0$. The inverse of the applied transformation is $x_1 = y_1$ and $x_2 = y_1^2 - y_2$. Hence, the unique local minimizer point of the original unconstrained problem is $x_1^* = 1$, $x_2^* = 1$. This point is inside the bounds, and thus x^* is a global minimizer of the Rosenbrock problem, and the uniqueness of this minimum is guaranteed by the applied procedure.

2. Consider the function $f(x) = \cos(e^{x_1} + x_2) + \cos(x_2)$. The symbolic derivation and factorization now indicated that a transformation in the form $e^{x_1} + h(x_2)$ might be possible. With the smooth but not invertible variable transformation $y_1 = e^{x_1} + x_2$, $y_2 = x_2$, we obtained $g(y) = \cos(y_1) + \cos(y_2)$. This has local maximizers (e.g. $y_1 = -2\pi$, $y_2 = 0$) that are not transforms of local maximizers of $f(x)$ (cf. Theorem 2). However, $g(y)$ is separable and simpler to optimize, and the transformation clearly indicates the structure of the optimizers of the original problem.

3. Problem (1), presented in the Introduction, proved to be simplifiable. For the sake of simplicity, delete the square root and the term $1/m$ in (1), and use the squares of the real and imaginary differences instead of the absolute values of complex differences. The only partial derivative which proved to be factorizable is $\partial F/\partial \tau$, and a factor of B/τ appeared. By integrating this, we obtain the transformation function $B\log(\tau)$, which was evaluated in the Introduction with the help of the understanding of the problem structure.

4. Consider a parameter estimation problem [1] as the last example. The underlying Otis model of respiratory mechanics was earlier widely studied. In [1], by utilizing the properties of the particular electric model, it was proved that the model function has a redundant parameter. The objective function again has the least-squares form as in (1), but we now concentrate on the model function

$$Z_m(s) = \frac{Ds^3 + Cs^2 + Bs + 1}{Gs^2 + Fs}, \tag{5}$$

where $s = j\omega$ and

$$B = R_C(C_1 + C_2) + R_1C_1 + R_2C_2\,, \tag{6}$$

$$C = I_C(C_1 + C_2) + [R_C(R_1 + R_2) + R_1R_2]C_1C_2\,, \tag{7}$$

$$D = I_C(R_1 + R_2)C_1C_2\,, \tag{8}$$

$$F = C_1 + C_2\,, \tag{9}$$

$$G = (R_1 + R_2)C_1C_2\,. \tag{10}$$

Although some simplifications are possible (e.g. $R_1' = R_1C_1$) by applying the technique presented in this paper, the number of model parameters cannot be decreased in this way. However, equations (6)–(10) also represent a smooth nonlinear variable transformation. Each new parameter is a linear function of a given old one (having the other old parameters fixed), and thus the conditions of Theorem 3 are fulfilled. According to the arguments of the proof of Theorem 3, a local minimizer point of the transformed problem can be transformed back to a hypersurface of the original parameter space, every point of which is a local minimizer with the same function value. This example indicates further methods for the recognition of simplifying parameter transformations.

Acknowledgements

The authors thank Dr B. Daróczy for valuable suggestions and two anonymous referees for their helpful criticism. This work was supported by Grants OTKA 1074/1987, 2568/1991 and 2879/1991.

References

1. Avanzolini, G. and Barbini, P. (1982), Comment on 'Estimating Respiratory Mechanical Parameters in Parallel Compartment Models', *IEEE Transactions on Biomedical Engineering* **29**, 772–774.
2. Csendes, T. (1988), Nonlinear Parameter Estimation by Global Optimization – Efficiency and Reliability, *Acta Cybernetica* **8**, 361–370.
3. Csendes, T., Daróczy, B., and Hantos, Z. (1992), Optimal Measuring Signal Generation, *J. of Global Optimization* **2**, 225–226.
4. Gill, P. E., Murray, W., and Wright, M. H. (1981), *Practical Optimization*, Academic, London.
5. Hansen, P., Jaumard, B., and Lu, S. H. (1991), An Analytical Approach to Global Optimization, *Mathematical Programming* **52**, 227–254.
6. Hantos, Z., Daróczy, B., Csendes, T., Suki, B., and Nagy, S. (1990), Modeling of Low-frequency Pulmonary Impedance in the Dog, *J. of Applied Physiology* **68**, 849–860.
7. Perekatov, A. E. and N. N. Redkovskii (1989), Method for Minimization of Unimodal Non-Convex Functions (in Russian), *Dokladi AN USSR Ser. A Physical-Mathematics and Technical Sciences* **10**, 36–38.

8. Redkovskii, N. N. (1989), Nonlinear Transformations of Coordinates in Unconstrained Optimization Problems (in Russian), *Issledovanii Operatsii UAS* **34**, 60–65.
9. Stoutemyer, D. R. (1975), Analytical Optimization Using Computer Algebraic Manipulation, *ACM Transactions on Mathematical Software* **1**, 147–164.
10. Stoutemyer, D. R. (1978) Automatic Categorization of Optimization Problems: An Application of Computer Symbolic Mathematics, *Operations Research* **26**, 773–788.