JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 34, No. 1, MAY 1981

Generalized Descent for Global Optimization[1]

A. O. GRIEWANK[2]

Communicated by D. G. Luenberger

Abstract. This paper introduces a new method for the global unconstrained minimization of a differentiable objective function. The method is based on search trajectories, which are defined by a differential equation and exhibit certain similarities to the trajectories of steepest descent. The trajectories depend explicitly on the value of the objective function and aim at attaining a given target level, while rejecting all larger local minima. Convergence to the global minimum can be proven for a certain class of functions and appropriate setting of two parameters.

Key Words: Global optimization, generalized descent, search trajectories, target level.

1. Introduction

We consider the problem of unconstrained minimization of a differentiable objective function $f \in C^1(\mathbb{R}^n)$ with several local minima $\hat{f}_i$ and corresponding local minimizers in $f^{-1}(\hat{f}_i)$. The local minimization problem (i.e., the location of a local minimum in the neighborhood of any initial point) can be solved by algorithms of high reliability. The evaluation of f is required only on a path of sample points, and rapid convergence can be achieved if the gradient ∇f and possibly the Hessian $\nabla^2 f$ are available.

Unless f satisfies subsidiary assumptions, the search for the global minimum

$$f^* = \min\{\hat{f}_i\}$$

is mathematically ill-posed in that a lower bound for f^* cannot be given after any finite number of evaluations. Provided the objective function satisfies a

[1] The author wishes to thank Professor R. P. Brent for making helpful suggestions and acknowledges the financial support of an Australian National University Postgraduate Scholarship.

[2] Visitor, Department of Applied Mathematics and Theoretical Physics, Cambridge, United Kingdom.

0022-3239/81/0500-0011$03.00/0

Lipschitz condition with known constant L and the search area is bounded, f^* and some $x^* \in f^{-1}(f^*)$ can be approximated by *space covering techniques* [e.g., Schubert (Ref. 1), Aird and Rice (Ref. 2), and Dickson, Gomulka, and Szegö (Ref. 3)]. For more than two variables, these methods tend to exceed computational limitation as they have to sample the function on a set of points that is sufficiently dense (depending on L and the tolerance) to *cover* the search area. All other global minimization methods are *heuristic*, in that they return the lowest local minimum found as an estimate of f^* or they depend on more specific conditions on f. The classes of problem functions for which known methods can be guaranteed to converge to the global minimum are as yet very restricted. Rather than giving rise to a particular strategy for the global search, conditions like unimodality or convexity merely assert that there is only one local minimum, which may then be located in a conventional downhill fashion.

Methods for the solution of simultaneous equations are considered to be *globally convergent* if they are guaranteed to reach some solution from all initial points in some *large* set S, if not the full space $\mathbb{R}^n$. Under fairly general assumptions on the properties of the simultaneous equations in some compact domain Ω, this is true for the global Newton method, with S consisting of almost all points on the boundary $\partial\Omega$. This method was originally suggested by Branin (Ref. 4), later on theoretically justified on the basis of Sard's theorem (Ref. 5) by Smale (Ref. 6), and has recently been cast into a numerically appropriate form by Keller (Ref. 7).

Before applying the global Newton method to the gradient system of an objective function f, we have to ensure that certain assumptions are satisfied. This is most easily done by adding a quadratic penalty term to f outside the interesting research area and by using a sufficiently large ball as Ω. Then, we know that, from almost all initial points on the sphere $\partial\Omega$, the method will reach some stationary point of the objective function, which need not be a minimum; and if it is a minimum, it may be an artefact of the modification described above. In any case, the continuation curves are completely independent of the actual value of the objective function and would be exactly the same whether we wanted to minimize or maximize.

Since conventional downhill minimization techniques are globally convergent to *some* local minimum, the global Newton method only represents an improvement toward the location of the global minimum to the extent that it can be expected to find *all* stationary points. Unfortunately, this cannot be guaranteed under the assumptions used by any of the authors.

To see this, we note that, by a construction of John (Ref. 8), the objective function can be modified in an arbitrarily small neighborhood of any nonstationary point x to introduce two new stationary points of which one may be a new global minimum. By suitable smoothing, we can maintain any finite order of differentiability and ensure that the set of points where the Hessian of the modified function is singular has zero measure, so that results like Sard's theorem still apply. Of the general assumptions that are frequently used, only such strong conditions as analyticity or unimodality are violated by the modification. After any calculation in which the function or its derivatives have not been sampled in the neighborhood of x, the question of whether all stationary points and in particular the global minimum has been found is undecidable. Consequently, all methods that are not space-covering methods must be heuristic, unless the assumptions on f are strong enough to prohibit modifications of the kind described above.

This situation is theoretically very unsatisfactory, but there is a great practical need for numerical procedures which locate comparatively low local minima, if not the global minimum itself. For more than two variables, *global optimization* should not be viewed as the theory of locating the global solution, since this is either impossible or reduces essentially to local minimization. Instead, *global optimization* is considered here as the design and numerical comparison of heuristic procedures which locate low local minima most successfully. General results, such as global convergence, cannot be expected except for space-covering techniques.

2. Motivation and Derivation of the Method

To develop the basic idea of the proposed method, we consider a function f, which is the sum of a well-behaved function u having a unique minimum $u^* = u(y^*)$ and a bounded oscillatory perturbation v. We can imagine that some sort of smoothing process would make f approximately the same as u. For instance, f could be a truncation of an infinite mathematical expression for u or an interpolation of noisy data representing u. Provided the amplitude of v is small compared to the range of values of u, f attains its global minimum $f(x^*) = f^*$ and all other local minima with values similar to f^* in the neighborhood of y^*. Under these circumstances, the various versions of the global Newton method or multistart techniques are likely to approach many stationary points or local minima outside the interesting area around y^*.

Desirable Properties of a Search Trajectory. Using a *target level* c, which is ideally slightly larger than the global minimum f^*, we formulate the following desirable properties of a search trajectory.

(i) The trajectory cannot converge to minima with values greater than c.

(ii) As long as $f \gg c$, the trajectory is little affected by the perturbation v and tries to follow a descent direction with respect to the negative gradient of u.

(iii) As f tends to c, the trajectory minimizes more thoroughly and finally reduces to a local minimization technique when $f \leq c$.

(iv) The trajectory does not depend explicitly on the Hessian $\nabla^2 f$.

(v) The trajectory is invariant with respect to translations of the variables and the multiplication of $f-c$ by a positive scalar.

Defining Equation and Its Justification. These requirements are satisfied by the solutions of the second-order differential equation

$$x''(t) = -e(I - x'(t)x'^T(t))\nabla f(x(t))/[f(x(t)) - c], \qquad e > 0, \tag{1}$$

from any initial point (x_0, x_0') with

$$f_0 \equiv f(x_0) > c, \qquad \|x_0'\| = 1.$$

This system is autonomous; thus, the parameter t, which gives the distance along the trajectory, can be omitted. Equation (1) was originally obtained by a variational method (Ref. 9) as a necessary condition for a once-differentiable solution of the problem

$$\min_{x(t) \in X} \int_{t_0}^{t_1} [f(x(\tau)) - c]^{-e} \, d\tau,$$

where X is the set of continuous and rectifiable paths connecting two fixed endpoints

$$x_0 = x(t_0), \qquad x_1 = x(t_1),$$

and satisfying

$$\|x'\| \equiv 1.$$

As the relationship between this problem and the original task of minimizing $f - c$ is unclear, we do not develop the variational analysis, but justify (1) directly from the requirements listed above.

Since the trajectory should pick up the general descent direction, it must correct its current tangent $dx/d\tau$ with a certain sensitivity $\tilde{s}(f(x)) > 0$ toward the local direction of steepest descent $-\nabla f(x)$. Thus, we have

$$d^2x/d\tau^2 = -\nabla f(x)\tilde{s}(f(x)).$$

This second-order differential equation is autonomous and has the independent variable τ. By an appropriate parameter transformation from τ

to t, we can ensure that the tangent $x' \equiv dx/dt$ has always unit length. Then, we obtain the system

$$x'' = -(I - x'x'^T)\nabla f(x) s(f(x)), \qquad \|x'\| = 1,$$

with the new sensitivity

$$s(f) \equiv \tilde{s}(f) \Big/ \left\{ [\|dx/d\tau\|^2]_{\tau=0} - 2\int_{f_0}^{f} \tilde{s}(\tilde{f})\, d\tilde{f} \right\},$$

which depends only on the fixed initial conditions and the current value of f.

Because $\|x'\| = 1$, the curvature ρ of the trajectory is simply the norm of x'' (Ref. 10), so that

$$\rho = \|x''\| = s(f) \cdot [\|\nabla f\|^2 - (\nabla f^T x')^2]^{1/2} \le s(f)\|\nabla f\|.$$

If $s(f)$ were bounded as f approaches c, the trajectory would be restricted in its response to the local shape of the objective function by a bounded curvature. Hence, $s(f)$ must have a pole for $f = c$. Finally, $1/s(f)$ must be linear in $f - c$ to ensure the required invariance with respect to scalar multiplications of $f - c$. Thus, we have $s(f) = e/(f - c)$ with a *sensitivity* parameter $e > 0$, which leads to Eq. (1).

Comparison to Similar Approaches. The use of second-order differential equations for global searches has been suggested by a number of authors, for instance Inomata and Cumada (Ref. 11), Zidkov and Siedrin (Ref. 12), and Incerti, Parisi, and Zirilli (Ref. 13). The search trajectories are based by analogy with classical mechanics on the differential equation

$$\mu(t)x''(t) + \nu(t)x'(t) = -\nabla f(x(t)),$$

which represents Newton's law for a particle of mass $\mu(t)$ in a potential f subject to a dissipative force $-\nu(t)x'(t)$. Given suitable assumptions on f and the parameters $\mu(t) > 0 < \nu(t)$, any trajectory converges to a local minimum of f. Since

$$(d/dt)[\tfrac{1}{2}\|x'(t)\|^2] = -(\nabla f^T(x(t))x'(t) + \nu(t)\|x'(t)\|^2)/\mu(t),$$

there is a tendency to slow down in uphill situations, $\nabla f^T(x)x' > 0$, but to gather momentum while going downhill, thus actually decreasing the influence of the local gradient. This behavior is quite the opposite of our third requirement and would appear disadvantageous as the trajectory may race through a low minimum, but settle in a small dent on top of a hill. With the rather artificial choices

$$\mu(t) = [f(x(t)) - c]/e, \qquad \nu(t) = -\nabla f^T(x(t))x'(t),$$

we obtain (1).

3. Mathematical Properties of the Defining Differential Equation

Throughout this section we impose two conditions on f:

(i) f is twice continuously differentiable, $f \in C^2(\mathbb{R}^n)$; (2)

(ii) $f-c$ is nondegenerate, $f=c \Rightarrow \nabla f \neq 0$. (3)

Because of (2), there is, for any pair (x_0, x_0') with

$$f(x_0) \neq c, \qquad \|x_0'\| = 1,$$

a neighborhood where the corresponding initial-value problem of system (1) has a unique solution $x(t)$. As the differential equation satisfies our requirements only when the function value is greater than the target c, we consider only the positive, forward solutions with

$$\phi(t) \equiv f(x(t)) > c, \qquad t \geq 0,$$

below. The gradient and Hessian of f will be denoted as functions of t by

$$g(t) \equiv \nabla f(x(t)), \qquad H(t) \equiv \nabla^2 f(x(t)).$$

To analyze the properties of the trajectory $x(t)$, we consider the distance

$$r(t) = \|x(t) - \hat{x}\|,$$

where $\hat{x}$ may be any point of particular interest. The following lemma shows how the current search direction $x' \equiv x'(t)$ and the derivative $r' = dr/dt$ depend on the values $\phi(\tau)$ and $g(\tau)$ for $\tau \leq t$, i.e., the accumulated information about the objective function.

Lemma 3.1. Let f satisfy the assumption (2), and let $x(t)$ be a solution of system (1). Then, we have, for any point on the trajectory,

$$x'/(\phi-c)^e = x_0'/(\phi_0-c)^e - e\int_0^t [g/(\phi-c)^{e+1}]\, d\tau, \tag{4}$$

$$\|x'\| = \|x_0'\| = 1, \tag{5}$$

$$r'r/(\phi-c)^e = r_0'r_0/(\phi_0-c)^e + \int_0^t \{[\phi-c-eg^T(x-\hat{x})]/(\phi-c)^{e+1}\}\, d\tau,$$

$$r'^2r^2/(\phi-c)^{2e} = r_0'^2r_0^2/(\phi_0-c)^{2e} \tag{6}$$

$$+2\int_0^t \{[\phi-c-eg^T(x-\hat{x})]rr'/(\phi-c)^{2e+1}\}\, d\tau, \tag{7}$$

$$r^2(1-r'^2)/(\phi-c)^{2e} = r_0^2(1-r_0'^2)/(\phi_0-c)^{2e}$$

$$-2e\int_0^t \{[r\phi' - r'g^T(x-\hat{x})]r/(\phi-c)^{2e+1}\}\, d\tau. \tag{8}$$

Comments on Lemma 3.1. Equation (4) shows that the current direction x' is an average of previous negative gradients weighted by $(\phi - c)^{-e-1}$. Equation (5) confirms that the parameter transformation in the previous section normalizes $\|x'\|$ as desired. The formulas (6)–(8) enable us to establish convergence to or divergence from a particular local minimum $\hat{x}$.

Proof. (4)–(7) follow from a more general result. For any twice continuously differentiable function $h \in C^2(\mathbb{R}^n, \mathbb{R}^m)$ with some $m \in \mathbb{N}$, we find, along the trajectory $x(t)$, with $\psi(t) \equiv h(x(t))$,

$$\psi'/(\phi - c)^e = \psi_0'/(\phi_0 - c)^e + \int_0^t w \, d\tau, \tag{9}$$

$$\psi'^T\psi'/(\phi - c)^{2e} = \psi_0'^T\psi_0'/(\phi_0 - c)^{2e} + 2\int_0^t [w^T\psi'/(\phi - c)^e] \, d\tau, \tag{10}$$

where

$$w \equiv x'^T\nabla^2 h(x)x'/(\phi - c)^e - e\nabla h^T(x)g/(\phi - c)^{e+1}.$$

The validity of these equations can be established by differentiation. Integrating by parts, we can obtain (8) from (7).

Because of assumptions (2) and (3), the *target set* $f^{-1}(c)$ is a smooth $(n-1)$-dimensional manifold. Any trajectory may stay in the interior of one of the open connected subsets of $\{x \in \mathbb{R}^n \,|\, f(x) > c\}$, for all $t \geq 0$, or approach the boundary $f^{-1}(c)$. There is a maximal $t_c \in (0, \infty]$, such that $x(t)$ is defined on $[0, t_c)$. The next theorem examines the conditions under which t_c is finite, and how $x(t)$ behaves in this case.

Theorem 3.1. *Final Approach.* Let $x(t)$ be a solution of system (1) on $[0, t_c)$, and let f satisfy the assumptions (2) and (3). Then, the following conditions are equivalent.

$$\text{(i)} \begin{cases} \liminf_{t \to t_c} \phi(t) = c, \\ \liminf_{t \to t_c} \|g(t)\| > 0, \\ \limsup_{t \to t_c} \|H(t)\| < \infty; \end{cases}$$

$$\text{(ii)} \quad t_c < \infty;$$

$$\text{(iii)} \begin{cases} \lim_{t \to t_c} x(t) = x_c \in f^{-1}(c), \\ \lim_{t \to t_c} x'(t) = -g_c/\|g_c\|, \end{cases}$$

where $g_c \equiv \nabla f(x_c)$ in (iii).

Comments on Theorem 3.1. The theorem says that, if ϕ comes arbitrarily close to c, and if $\|H\|$ is bounded and $\|g\|$ is bounded away from zero, then the trajectory has only finite length t_c and converges to a limit point x_c, such that $\phi(t)$ and $x(t)$ are once continuously differentiable on $[0, t_c]$. This final approach to the target set is perpendicular, in that $x'(t)$ converges to $-g_c/\|g_c\|$, which is the normal at x_c to $f^{-1}(c)$.

Proof. (i)$\rightarrow$(ii). For a sufficiently large $t_1 < t_c$, we have

$$0 < \sigma \equiv \inf_{t_1 < t < t_c} \|g(t)\|, \qquad \beta \equiv \sup_{t_1 < t < t_c} \|H(t)\| < \infty.$$

We examine the derivatives of ϕ and bound them by

$$\phi' = g^T x', \qquad -\|g\| \leq \phi' \leq \|g\|, \qquad \text{as } \|x'\| = 1,$$
$$\phi'' = x'^T H x' + g^T x'' = x'^T H x' - e(\|g\|^2 - \phi'^2)/(\phi - c),$$
$$\phi'' \leq \beta - e(\sigma^2 - \phi'^2)/(\phi - c), \qquad \text{for } t \in [t_1, t_c].$$

As $\phi'' < 0$, if

$$(\phi - c) < \tfrac{1}{2} e\sigma^2/\beta, \qquad \phi' = 0,$$

the function $\phi(t) - c$ can have no minima below $\frac{1}{2}e\sigma^2/\beta$ and must decrease monotonically beyond a certain $t_2 \in [t_1, t_c)$. Furthermore, $\phi'' < -\beta/2$, if

$$t > t_2, \qquad |\phi'| < \sigma/2,$$

so that

$$\phi' < -\sigma/2, \qquad \text{for } t \in [t_2 + \sigma/\beta, t_c),$$

which ensures that

$$t_c < \infty, \qquad \limsup_{t \to t_c} \phi' < 0.$$

(ii)$\rightarrow$(i). Because $\|x'\| = 1$, the path $x(t)$ is Lipschitz continuous on $[0, t_c)$ and can be extended continuously to $[0, t_c]$, with

$$x_c = \lim_{t \to t_c} x(t).$$

Since ϕ, g, H are continuous in t, the three cluster points in (i) are in fact limits, which satisfy the given conditions by definition of t_c and because of (3).

(i), (ii)$\rightarrow$(iii). From the first section of this proof, we know that

$$-\|g_c\| \leq \limsup_{t \to t_c} \phi'(t) < 0.$$

As $\phi'' \leq \beta$ and t_c is finite, ϕ' must have a limit

$$\phi'_c \equiv \lim_{t \to t_c} \phi'(t) < 0.$$

According to the mean-value theorem, there are $\delta(t)$, $\epsilon(t) \in (t, t_c)$, such that

$$\begin{aligned} 0 &= \lim_{t \to t_c} [\phi'(t) - \phi'_c] = \lim_{t \to t_c} [\phi''(\delta(t))(t - t_c)] \\ &\geq \lim_{t \to t_c} \{\beta(t - t_c) - e[\|g(\delta(t))\|^2 - \phi'^2(\delta(t))](t - t_c)/[\phi(\delta(t)) - c]\} \\ &\geq -e(\|g_c\|^2 - \phi_c'^2) \lim_{t \to t_c} [\phi'(\epsilon(t))]^{-1} = -e(\|g_c\|^2 - \phi_c'^2)/\phi'_c \leq 0. \end{aligned}$$

As $\phi'_c < 0$, we must have

$$\phi'_c = -\|g_c\|,$$

so that

$$\lim_{t \to t_c} \|x'(t) + g(t)/\|g(t)\|\|^2 = \lim_{t \to t_c} 2[1 + \phi'(t)/\|g(t)\|] = 0,$$

which proves the second equation of (iii).

(iii) → (i). As $x(t)$ is continuous and has a limit, so do $\phi(t)$, $g(t)$, $H(t)$. Obviously, we have

$$\phi_c = c, \qquad \|g_c\| > 0, \qquad \|H_c\| < \infty.$$

4. Method as Generalized Descent

Theorem 3.1 (iii) ensures that the solutions of (1) transform gradually to the trajectory of steepest descent as f goes to c. This behavior satisfies our third requirement, and we can extend any finite solution of (1) in a continuously differentiable way by the trajectory

$$x' = -\nabla f(x)/\|\nabla f(x)\|, \qquad x(t_c) = x_c.$$

The resulting complete search trajectories are uniquely determined by the initial point (x_0, x'_0) and solve the implicit system

$$\max[0, f(x) - c]x'' + e(I - x'x'^T)\nabla f(x) = 0. \tag{11}$$

Unless the target level is never attained, the solutions of (11) converge to a minimum or possible saddle point with a value not greater than c.

The combination with the steepest-descent trajectory is only of theoretical interest as there are much more efficient methods for local minimization available. Practical experience with functions of the form

$$f = u + v,$$

as specified in the introduction of Section 2, confirmed that the solutions of (1) behave like the steepest-descent trajectories generated by the smooth function u (see Figs. 1 and 2). Hence, we can use the new trajectories to explore the basic shape of the objective function.

Ideally, this search leads into the regions where the general level of the objective function is lower than all previously obtained local minima. However, this attempt can fail completely, for instance if c is so low that the target set $f^{-1}(c)$ is empty, or if the trajectory is trapped in the region of attraction of an *unacceptable local minimum* (i.e., $\hat{f} > c$) with steep surrounding slopes. The first problem can be overcome by a sufficient increase of c, while the second problem requires a suitable decrease of e. The choice of these two free parameters is of great importance to the method, and we shall discuss their effect on the trajectories in Section 5.

Symmetry and Unbounded Growth. The major difference between local minimization trajectories and the solutions of (1) is that the latter climb occasionally uphill. Ideally, this enables the method to leave the neighborhood of unacceptable local minima and to cross small ridges. However, since the trajectory is supposed to minimize the function, one would like to give a bound on $\phi(t)$ in terms of the initial configuration (x_0, x_0', c, e). Unfortunately, such a bound does not exit, because of the symmetry of system (1).

We call (1) symmetric, because it is unaltered by a reversed parametrization. Let $x(t)$ be a solution of (1) emanating from x_0, x_0'. Then, the trajectory

$$y(s) \equiv x(1-s)$$

solves the same differential equation, since

$$y'' = x'', \qquad y'y'^T = x'x'^T,$$

and both paths are identical if we choose

$$y_0 = x(1), \qquad y_0' = -x'(1).$$

If $x(t)$ is a successful minimizing trajectory, in that

$$f(x_1) \ll f(x_0),$$

we find that $y(s)$ does exactly the opposite and increases the function value significantly.

Scaling Dependence and Zig-Zagging. Because of the similarity between steepest descent and the new trajectories, they have some properties in common. In particular, we have the undesirable property that the defining system (1) is not independent of linear transformations on the variables of f. This corresponds to the unfortunate zig-zagging of the steepest-descent method in valleys, which can be observed similarly for the new trajectories.

For local minimization, this problem has been overcome by variable-metric techniques, which normalize the contours of f locally by a suitable linear transformation A^{-1} and define the search direction as $-B\nabla f$, with $B \equiv AA^T$. The corresponding transformation of (1) gives

$$x'' = -e(B - x'x'^T)\nabla f(x)/[f(x) - c], \qquad x_0'^T B^{-1} x_0' = 1, \qquad f_0 > c. \tag{12}$$

At the final approach to the target level, the solutions of (12) satisfy

$$x_c' = -B\nabla f/(\nabla f^T B \nabla f)^{1/2}.$$

Given suitable updates of B, we can thus achieve a continuous transition from the transformed trajectory to a corresponding variable-metric method. However, such concepts are not discussed further in this paper.

5. Convergence and Divergence Properties

As we have indicated in the introduction, any method which evaluates f and ∇f only along trajectories must be heuristic, as it is not space covering. For arbitrary objective functions, we cannot prove global results, and in general there is little that we can say about the behavior of the trajectories. In fact, it can be shown that, for any nonself-intersecting, thrice differentiable path $x(t)$ of length $t_1 < \infty$ and any positive function

$$0 < \phi(t) - c \in C^1[0, t_1],$$

there is an objective function $f \in C^1(\mathbb{R}^n)$, such that Eq. (1) is satisfied with

$$f(x(t)) \equiv \phi(t).$$

Since we do not want to impose any restrictive conditions on the objective function, we confine our examination of the trajectories to the proximity of local minima.

Ideally, any trajectory which meets the region of attraction of a local minimum $\hat{f} = f(\hat{x})$ should converge to $\hat{x}$ if $\hat{f} \le c$ and diverge from $\hat{x}$ (i.e., leave the region again) if $\hat{f} > c$. Divergence from *unacceptable minima*, $\hat{f} > c$, is as important as convergence to *acceptable minima*, $\hat{f} \le c$; and we attempt to ensure both by a suitable choice of the sensitivity e. There can be no uniformly satisfactory choice of e, since the modified objective function $(f - c)^{1/d}$ with $d > 0$ gives, for $e = 1$, exactly the same trajectory as $f - c$ itself for $e = 1/d$.

To obtain some insight into the behavior of the trajectories in the neighborhood of local minima, we consider for the remainder of this section *separable objective functions* of the form

$$f(x) = f(\hat{x}) + \gamma(r) h((x - \hat{x})/r) r^d, \tag{13}$$

where

$$r = \|x - \hat{x}\|, \qquad d > 0,$$

and the functions $\gamma \in C^2(\mathbb{R})$ and $h \in C^2(\mathbb{R}^n)$ such that, for all arguments,

$$0 < \mu_0 \leq \gamma \leq \mu_1 \equiv \eta\mu_0, \qquad 0 < \lambda_0 \leq h \leq \lambda_1 \equiv \kappa\lambda_0.$$

For $\gamma \equiv 1$, we have the class of homogeneous functions of degree d satisfying the equivalent conditions

$$d(f - \hat{f}) \equiv \nabla f^T (x - \hat{x}), \qquad f - \hat{f} \equiv f((x - \hat{x})/r) r^d.$$

For $d \neq 1$, homogeneity is a property that holds only with respect to one particular center point $\hat{x}$. All quadratic functions are homogeneous with $d = 2$. In this case, λ_0 and λ_1 represent the smallest and the largest eigenvalues, and the ratio $\kappa = \lambda_1/\lambda_0$ is the condition number of the Hessian $\nabla^2 f$.

As γ and h are always positive, f attains its global minimum at the unique minimizer $\hat{x}$, but may have arbitrarily many further local minima. Obviously, f is twice continuously differentiable in the open set $\mathbb{R}^n - \{\hat{x}\}$. At $\hat{x}$ itself, f is twice differentiable only if $d > 2$, and is not differentiable at all if $d < 1$. The latter possibility may, for instance, occur when f is the norm of a residual in a parameter estimation problem and is therefore of some interest. We see from the integrated form (4) of the differential equation (1) that the trajectories are still well defined if f is nondifferentiable at a discrete set of points. Thus, we can consider separable functions with arbitrary exponent $d > 0$.

The next theorem contains two parts applying to homogeneous and general separable functions, respectively. The latter represents the only case where we can prove convergence to the global solution in the presence of several isolated local minima. To achieve this result, we have to assume that the sensitivity e is exactly reciprocal to the order of growth d of f and that c equals the global minimum $\hat{f}$. In the homogeneous case, we obtain convergence and divergence results under more general conditions on e and c. These results allow us to draw conclusions regarding the suitable choice of the two parameters in the nonseparable case.

Provided they exist, the limits x_c, f_c, r_c of $x(t)$, $\phi(t)$, $r(t) = \|x(t) - \hat{x}\|$, as $t \to t_c$, are used to describe the following situations:

$x_c = \hat{x}$, convergence to global minimum with $t_c < \infty$;

$f_c = c$, convergence to target level with $t_c < \infty$;

$r_c = \infty$, divergence towards infinity with $t_c = \infty$.

Theorem 5.1. *Convergence and Divergence in Separable Case.* Let f be separable, as specified by (13); then, these results hold.

(i) For homogeneous f with $\gamma \equiv 1$, the implications in Table 1 hold; note that the division into triangles is defined in the top left-hand corner. In the table, S^+ and S^- denote the following sets:

$$S^+ \equiv \{x \in \mathbb{R}^n \mid \lambda_0(1 - ed)r^d > c - \hat{f}\},$$
$$S^- \equiv \{x \in \mathbb{R}^n \mid \lambda_1(1 - ed)r^d < c - \hat{f}\}.$$

(ii) Let $\gamma \in C^1(0, \infty)$ have k_0^- minima in $(0, r_0)$ and k_0^+ minima in (r_0, ∞). Then, we find that, for $ed = 1$ and $c = \hat{f}$,

$$r_0' < -\sqrt{\{(\eta - 1)[(\kappa - 1)k_0^- + \kappa]\}} \Rightarrow x_c = \hat{x},$$
$$r_0' > \sqrt{\{(\eta - 1)[(\kappa - 1)k_0^+ + \kappa]\}} \Rightarrow r_c = \infty.$$

Proof. (i) by definition of f, Eq. (6) reduces to

$$r'r/(\phi - c)^e = r_0'r_0/(\phi_0 - c)^e + \int_0^t [(1 - ed)(\phi - \hat{f}) + \hat{f} - c]/(\phi - c)^{e+1}\, d\tau, \tag{14}$$

Table 1. Implications for case (i).

$r_0' < 0$ $r_0' > 0$	$c > \hat{f}$	$c = \hat{f}$	$c < \hat{f}$
$ed > 1$	$f_c = c$ $r_c = \infty$ or $f_c = c$	$x_c = \hat{x}$ $r_c = \infty$ or $x_c = \hat{x}$	—
$ed = 1$	$f_c = c$ $r_c = \infty$ or $f_c = c$	$x_c = \hat{x}$ $r_c = \infty$	$r_c = \infty$
$ed < 1$	$x_0 \in S^- \Rightarrow f_c = c$ $x_0 \in S^+ \Rightarrow r_c = \infty$	$x_c = \hat{x}$ or $r_c = \infty$ $r_c = \infty$	$r_c = \infty$

which yields after differentiation

$$r'' = [(1-ed)(\phi-\hat{f})+\hat{f}-c]/r(\phi-c)+[e\phi'/(\phi-c)-r'/r]r'. \tag{15}$$

The distance $r(t)$ is twice continuously differentiable in $(0, t_c)$, unless it has a minimum $r(t_1)=0$ for some $t_1 \in (0, t_c)$. This can only occur in the last column of the table, where $c<\hat{f}$, as otherwise $\hat{x}$ lies at or below the target level. At all other extrema, $r(t)$ must be differentiable with $r'=0$ and $r>0$, so that by (15)

$$\operatorname{sign}(r'') = \operatorname{sign}[(1-ed)(\phi-\hat{f})+\hat{f}-c]. \tag{16}$$

Except for the three boxes in the cross diagonal of Table 1, we have

$$\operatorname{sign}(ed-1) \neq \operatorname{sign}(\hat{f}-c). \tag{17}$$

This inequality implies with (16) that $r(t)$ has at most one extremum; this extremum is a minimum if

$$ed<1 \qquad \text{or} \qquad c<\hat{f},$$

and is a maximum if

$$ed>1 \qquad \text{or} \qquad \hat{f}<c.$$

Consequently, there must be a limit

$$r_c = \lim_{t\to t_c} r(t) \geq 0.$$

Clearly, we have convergence if

$$t_c<\infty \qquad \text{or} \qquad r_c=0$$

and divergence if

$$t_c=\infty=r_c.$$

We can exclude that third possibility,

$$t_c=\infty, \qquad 0<r_c<\infty,$$

as it would require that

$$\lim_{t\to\infty} r'(t)=0, \qquad \lim_{t\to\infty} \inf|r''(t)|=0,$$

and consequently, by (15),

$$\lim_{t\to\infty} \inf|(1-ed)[\phi(t)-\hat{f}]+\hat{f}-c|=0.$$

This is impossible, because of (17) and $r_c>0$. Now, the assertions for all cases where (17) is satisfied follow directly from the additional assumptions on $\operatorname{sign}(r_0')$.

The results for $ed < 1$ and $\hat{f} < c$ can be shown in a similar fashion, as $x(t)$ cannot leave S^+ or S^- if $r'_0 \geq 0$ or $r'_0 \leq 0$, respectively. In the last case,

$$ed - 1 = 0 = \hat{f} - c,$$

it follows directly from (14) that

$$\kappa^{-e} \leq r'/r'_0 \leq \kappa^e,$$

which implies the assertions in the center box.

(ii) The lengthy but elementary proof is omitted.

Comments on Theorem 5.1. Considering Table 1, we note that all trajectories for which

$$(ed - 1)(\hat{f} - c) \leq 0$$

either converge to their target set or diverge toward infinity. Provided this inequality is satisfied, any trajectory that has an attainable target level ($c \geq \hat{f}$) and moves at some stage toward the minimum ($r'_0 < 0$) must converge, and any trajectory with an unattainable target level ($c < \hat{f}$) must diverge.

If, on the other hand,

$$(ed - 1)(\hat{f} - c) > 0,$$

which is only the case in the bottom left-hand and top right-hand box of Table 1, the behavior of the trajectories cannot be predicted and they may wander around endlessly. Theoretically, this undesirable situation can occur if the minimum $\hat{x}$ is acceptable ($\hat{f} < c$), but the sensitivity e is too small to ensure convergence ($e < 1/d$), or if the minimum is unacceptable ($\hat{f} > c$), but the sensitivity is too large to ensure divergence ($e > 1/d$). In practice, the former possibility, which corresponds to the bottom left-hand box in Table 1, is less serious than the latter, which corresponds to the top right-hand box.

Depending on the *condition number*

$$\kappa = \lambda_1/\lambda_0,$$

there is a gap $\mathbb{R}^n - S^- - S^+$ between S^- and S^+. Since any trajectory that enters S^- or S^+ must converge or diverge, respectively, a trajectory can only stay undetermined if it never leaves that gap. This is unlikely to occur and has not been observed in practical calculations. In contrast, trajectories with $c < \hat{f}$ and $e > 1/d$ cannot converge to the unacceptable minimum $\hat{x}$, but are often unable to escape its attraction. This kind of orbiting around an unacceptable minimum can be quite stable and is not easily detected.

If $ed = 1$, any trajectory that moves at some time toward the minimum $\hat{x}$ will converge if $\hat{x}$ is acceptable, and will diverge toward infinity if $\hat{x}$ is unacceptable. Since smooth functions can be closely approximated by homogeneous functions in the neighborhoods of their local minima, the trajectories exhibit locally the same selective behavior in the general case. However, they cannot be expected actually to diverge toward infinity, but merely to leave that region about some unacceptable minimum in which the homogeneous approximation is valid. If f is locally of the form

$$f = r^d h(x/r) v,$$

with some bounded perturbation term v, we find that, for $ed = 1$, instead of (14),

$$r'r/(\phi - c)^e = r_0' r_0/(\phi_0 - c)^e - \int_0^t [c/(\phi - c)^{e+1}]\, d\tau$$

$$- e\int_0^t \{x^T \nabla v(x)/(\phi - c)^e [v(x) - cr^{-d}/h(x/r)]\}\, d\tau.$$

Unless there is a special correlation between v and a particular trajectory $x(t)$, one would expect that the last integral is close to zero and does not grow with t. The remaining terms are identical to those for a purely homogeneous function and their equality would lead to the same conclusions as in the theorem.

Case (ii) deals with the general separable case, but assumes

$$c = \hat{f}, \qquad de = 1.$$

This result is only applicable if both κ and η are close to 1, as otherwise the condition on r_0' may not be satisfiable for any r_0' with $|r_0'| \leq 1$. If the values of f deviate from the quadratic term r^2 at most by 20% and if h and γ contribute evenly to this deviation, we find that

$$\kappa \approx 1.2 \approx \eta.$$

Then,

$$|r_0'| > 0.8$$

ensures that r' cannot change its sign before the trajectory has passed at least through 10 minima of the radial term $\gamma(r)$, which implies convergence or divergence if

$$|k_0^-| < 10 > |k_0^+|.$$

As to the choice of e, we draw the following conclusions. Suppose that an arbitrary objective function f has an *average order of growth* d in the neighborhoods U_j of its local minima $\hat{x}_j$, in that, for $\hat{x}_j + z \in U_j$,

$$f(\hat{x}_j + z) - f(\hat{x}_j) \approx [f(\hat{x}_j + z/\|z\|) - f(\hat{x}_j)]\|z\|^d;$$

then, we can expect successful minimization by means of the search trajectories, if we have chosen e such that

$$e \approx 1/d.$$

If $ed < 1$, a trajectory may fail to converge toward an acceptable minimum; and, if $ed > 1$, it can be trapped in the neighborhood of an unacceptable minimum. If $\nabla^3 f$ is of moderate magnitude, we can assume that f is locally quadratic and use $e = \frac{1}{2}$, at least as an initial value. If f is the Euclidean norm of a residual which vanishes at the solution, we should use $e = 1$ instead.

6. Incorporation in a Minimization Procedure

Based on the theoretical results for separable objective functions and computational experience with other test functions, a minimization procedure has been developed and implemented in a FORTRAN program. The procedure generates a sequence of search trajectories from various initial points (x_0, x_0'), where the initial search direction x_0' is usually taken as the direction of steepest descent $-\nabla f_0/\|\nabla f_0\|$. Even more important than the choice of the initial points is the initial setting and subsequent readjustment of the target level c and the sensitivity e. This can be done in many different ways, and as yet there is no fully reliable implementation that could handle all problems automatically. Therefore, it is essential that the user understands the role of the two parameters c and e and can adjust them properly. The following strategy has been adopted for the current implementation.

We distinguish between the free local minima of $f(x)$ in $\mathbb{R}^n$ and the constrained minima of $\phi(t)$ on the trajectories. All evaluated free and constrained minima are ordered according to their function values and stored in two separate arrays. Whenever a restart becomes necessary, we choose the new initial point from the set of constrained minima.

Successive Choice of the Target Level. Let f_l be the value of the lowest free minimum located so far, and let f_p be the lowest evaluated constrained minimum of the current trajectory, with $f_p \equiv f_0$ at its initial point. If f is known to attain values at or below the user-supplied target level c, we initialize $f_l = c$. As we will see later, this guarantees that c is never raised above its initial value. In particular, we leave c unchanged if it is known to equal the global minimum exactly. For instance, c should be zero throughout the computation if one solves nonlinear equations by minimizing a norm of the residual. Otherwise, we set $f_l \equiv f_0$, and similarly $c \equiv f_0$, if the target level is not specified at all. After a new local minimum $\hat{f} < c$ has been found, we lower the target level and f_l according to

$$c \leftarrow 3\hat{f} - 2f_l < \hat{f}, \qquad f_l \leftarrow \hat{f},$$

in an attempt to avoid a linear decrease of the minimal values $\hat{f}$. As this new target level may be lower than the actual global minimum, we increase c gradually toward the attainable value $\min(f_l, f_p)$, applying, at each step of length $\|s\|$,

$$c \leftarrow c + [\min\{f_l, f_p\} - c][1 - \exp(-\delta\|s\|)],$$

with some constant $0 < \delta \ll 1$. Since c is always less than f_l, the trajectory can never converge to a local minimum with a value obtained previously.

Adjustment of the Sensitivity. If a trajectory fails to attain the target level, it may orbit around a free minimum or leave the search area, which is assumed to be compact. The latter possibility is easily detected, and its occurrence suggests that the sensitivity e has been chosen too small. Thus, we multiply e by some factor $\epsilon \sim 1.1$ and restart in the interior. It is much more difficult to detect a trajectory that is orbiting about a particular minimum or wanders around in a wider area where $f > f_l$. Practical experience indicates that trajectories are only likely to be trapped by a minimim with a value less than f_p. If this is so, the free minimum x_s, adjacent to the lowest constrained minimum x_p of the current trajectory, is most likely to prevent the trajectory from leaving its region of attraction. Therefore, we store x_p and call a local minimization routine to locate x_s when the value of the next constrained minimum at x_p^+ turns out to be larger than $f_p = f(x_p)$. If x_s is found to be acceptable, in that $f(x_s) < c$, the sensitivity e is apparently too small, and we multiply e by ϵ, in addition to lowering the target as described above. Whether x_s was acceptable or not, we continue the integration from $(x_p^+, x_p^{+\prime})$. Until a new lowest constrained minimum is found, we minimize locally at each subsequent constrained minimum whose distance to x_s is less than $\|x_s - x_p^+\|$. If these local minimizations lead several times to the same unacceptable minimum x_s, we assume that the trajectory is trapped and restart after dividing e by ϵ.

Stopping Criterion. Finally, we limit the number of function and gradient evaluations per trajectory by a specified bound. If this bound has been reached, we restart without changing e or c. When the minimal value f_l has not been improved after a certain number of restarts, the procedure stops and returns the associated point x_l, with $f_l = f(x_l)$, as an estimate for the global minimum.

7. Numerical Treatment of the Defining Differential Equation

In the form (1), the differential equation is not suitable for numerical integration, since both the second derivative x'' and the Jacobian $\partial(x', x'')/\partial(x, x')$ can become arbitrarily large when f approaches the target level c. Since the integration should be economical as long as $f \gg c$, but accurate for small f-c, we apply a parameter transformation from t to τ, such that $\dot{x} \equiv dx/d\tau$ satisfies

$$\|\dot{x}\| = dt/d\tau = f(x) - c.$$

Instead of (4), we have, with $\psi(\tau) \equiv f(x(\tau))$,

$$\dot{x}/(\psi - c)^{1+e} = \dot{x}_0/(\psi_0 - c)^{1+e} - e\int_0^\tau \nabla f(x(\tilde{\tau}))/[\psi(\tilde{\tau}) - c]^e \, d\tilde{\tau}, \tag{18}$$

which yields after differentiation

$$\ddot{x} = -[eI - (1+e)\dot{x}\dot{x}^T/\|\dot{x}\|^2]\nabla f(x) \cdot [f(x) - c]. \tag{19}$$

It can be shown that the norm of the new Jacobian $\partial(\dot{x}, \ddot{x})/\partial(x, \dot{x})$ is bounded by

$$1 + (1+e)^2\|\nabla f\|^2 + (f - c)\|\nabla^2 f\|$$

and that $\dot{x}$ is as often differentiable in τ as f is in x. Therefore, integration schemes of arbitrary order can be used, provided f is sufficiently smooth. The integration is particularly accurate in the neighborhood of the target set $f^{-1}(c)$, since all higher derivatives of $x(\tau)$ can be bounded by multiples of $f - c$.

Stability, Choice of Stepsize, and Termination. The accuracy and stability of methods for numerical integration depend on the stepsize h multiplied by the eigenvalues of the Jacobian. In our case, the $2n$ eigenvalues $\lambda_j^\pm$ of $\partial(\dot{x}, \ddot{x})/\partial(x, \dot{x})$ are approximately given by

$$\lambda_j^\pm \simeq \tilde{\lambda}_j^\pm \equiv \tfrac{1}{2}\phi'(1+e) \pm \sqrt{[\tfrac{1}{4}\phi'^2(1+e)^2 - e\rho_j]},$$

where the ρ_j's are the real eigenvalues of the symmetric matrix

$$(f - c)\nabla^2 f + \nabla f \nabla f^T$$

and

$$\phi' = \dot{\psi}/(\psi - c).$$

It can be shown that, for all j,

$$|\lambda_j^\pm - \tfrac{1}{2}(1+e)\phi'| \leq \tfrac{1}{2}(1+e)\|\nabla f\| + [\|\nabla^2 f\|(f-c)(2e+1) + \|\nabla f\|^2(1+e)(2+e)]^{1/2}.$$

Since the eigenvalues are clustered about the real point $\frac{1}{2}(1+e)\phi'$ in the complex plane, we can expect that the trajectories are more stable while descending ($\phi'<0$) than while ascending ($\phi'>0$).

Since, for $e \leq 1$,

$$|\tilde{\lambda}_i^{\pm}| \leq |\phi'|(1+e)+\sqrt{(e|\rho_i|)} \leq 3\|\nabla f\|+\sqrt{[(f-c)\|\nabla^2 f\|]},$$

we choose the stepsize to be

$$h \equiv \sigma/\{3\|\nabla f\|+\sqrt{[(f-c)\mu]}\}$$

with μ an estimated bound on $\|\nabla^2 f\|$ and σ a positive constant. Up to second-order terms in h, this definition implies that the angle between two subsequent search directions $\dot{x}_k$ and $\dot{x}_{k+1}$ is smaller than $\sigma/3$ radians and that the difference between $\psi(\tau)$ and its linear interpolant in the interval $[\tau_k, \tau_{k+1}]$ is less than $\sigma^2(\psi-c)/32$.

As the distance between two subsequent points is approximately equal to

$$(f-c)h=\sigma\sqrt{(f-c)}/[\sqrt{\mu}+3\|\nabla f\|/\sqrt{(f-c)}],$$

it could diverge toward infinity if the search area were not bounded. This condition has already been imposed in Section 6. Since $|\phi'|$ is bounded if t_c is finite, we find that, for any trajectory that does not leave the search area,

$$\tau_c=\int_0^{t_c} dt/[\phi(t)-c]=\infty.$$

Therefore, the target set $f^{-1}(c)$ cannot be reached in a finite number of steps with bounded length h. However, we can terminate the global search and commence the local minimization as soon as

$$\tfrac{1}{2}\|\nabla f\|^2 \geq (f-c)\mu.$$

If μ is in fact a global bound on $\|\nabla^2 f\|$, this inequality guarantees that the target level is attainable in the neighborhood of the current point.

Special-Purpose Integration Method. Exploiting the special structure of the differential equation (19), we can design a one-step method of convergence order three. Differentiation of (19) yields the third derivative

$$\dddot{x}=-(2+e)e\dot{\psi}\nabla f(x)+(1+e)[(2e+1)\phi'^2-e\|\nabla f(x)\|^2]\dot{x}$$
$$-[f(x)-c][eI-(1+e)\dot{x}\dot{x}^T/\|\dot{x}\|^2]\nabla^2 f(x)\dot{x}.$$

Substituting the Hessian $\nabla^2 f$ by an approximation B that satisfies

$$B\dot{x}_0=\nabla^2 f_0\dot{x}_0+O(h),$$

we can calculate from the Taylor expansion of $x(\tau)$ at $x_0 \equiv x(0)$ an approximate solution $\tilde{x}_h$, such that

$$\|\tilde{x}_h - x(h)\| = O(h^4).$$

After the evaluation of

$$\tilde{f}_h = f(\tilde{x}_h), \qquad \nabla\tilde{f}_h = \nabla f(\tilde{x}_h),$$

B can be updated to B_h, such that the quasi-Newton equation

$$B_h s \equiv B_h(\tilde{x}_h - x_0) = y \equiv \nabla\tilde{f}_h - \nabla f_0$$

is satisfied. In choosing a particular update formula, we look for a good approximation to the Hessian, which need not be positive definite. Also, because the inner product $y^T s$ may be zero or very small, the Powell–Broyden symmetric update (Ref. 15)

$$B_h = B + [(y - Bs)s^T + s(y - Bs)^T]/s^T s - s^T(y - Bs)ss^T/(s^T s)^2$$

is preferable to formulas of the Broyden family, which all have $y^T s$ in the denominator.

In order to obtain an approximation $\tilde{\dot{x}}_h$ to $\dot{x}(h)$ from (18), we evaluate the integral on the right by the fourth-order quadrature formula

$$\int_0^h F(\tau)\, d\tau = (h/3)(2F_0 + F_h) + (h^2/6)\dot{F}_0 - (h^4/72)[(d^4F/d\tau^4)]_{\delta\in[0,h]}.$$

Given a second-order approximation

$$w = \nabla^2 f_0 \dot{x}_0 + O(h^2), \tag{20}$$

we find, for

$$F(\tau) \equiv \nabla f(x(\tau))/[\psi(\tau) - c]^e,$$

that

$$\int_0^h \{\nabla f(\tau)/[\psi(\tau) - c]^e\} = [h\nabla f_0/3(f_0 - c)^e][2 - eh\dot{f}_0/2(f_0 - c)]$$
$$+ h\nabla f_h/3(\tilde{f}_h - c)^e + h^2 w/6(f_0 - c)^e + O(h^4). \tag{21}$$

Provided special measures are taken at the first step, the condition (20) is satisfied by the weighted average

$$w = 1/(1 + p)B_h\dot{x}_0 + p/(1 + p)B\dot{x}_0,$$

where p is the ratio of the current and the previous stepsize. Using (21), we obtain from (18) the computed tangent $\tilde{\dot{x}}_h$, such that the truncation error between the true solution $(x(h), \dot{x}(h))$ and the approximation $(\tilde{x}_h, \dot{\tilde{x}}_h)$ is of fourth order. Consequently, the global error is of order three. It can be shown that this one-step method is convergent and stable in the sense of the Definitions 4.2 and 4.3 of Gear (Ref. 16).

In test calculations with several test functions, the method was found to achieve a given accuracy with fewer function and gradient evaluations than the general purpose routine DVOGER (Ref. 17). Like Eq. (1), the system (19) is symmetric, as discussed in Section 4, so that the correctness and accuracy of integration codes can be tested conveniently by the comparison between forward and backward integrations along the same path.

8. Experimental Results

This section has three parts. Firstly, we examine the behavior of individual trajectories with fixed c and e as displayed in the two figures for the case of a perturbed quadratic objective function. Secondly, we consider the success of trajectories with various settings of c and e in minimizing the camel-back function from a grid of 18 starting points. Finally, we report the performance of the program described in Section 6 for an objective function in 10 variables with several thousand local minima.

In Figs. 1 and 2, the objective function

$$f \equiv u + v \equiv (x_1^2 + x_2^2)/200 + 1 - \cos x_1 \cos(x_2/\sqrt{2})$$

is depicted by contours at the levels $\{f = 0.1 + 0.4j\}$, for $j \geq 0$. This function attains the global minimum $f^* = 0$ at $x = 0$ and has some 500 further local minima. The lowest of these have values of about 0.15 units and are separated from the global solution x^* by ridges with a minimal height of little less than one unit.

Figure 1 shows three trajectories starting at the point $(40, -35)$ with the same target $c = 0$, but with different sensitivities

$$e \in \{0.4, 0.5, 0.67\}.$$

Figure 2 shows three trajectories starting at $(35, -30)$ with the same sensitivity $e = 0.5$, but different targets

$$c \in \{-0.4, 0, 0.4\}.$$

As long as $f \gg c$, the quadratic term u determines the behavior of the trajectories directing them gradually toward its minimizer $\overset{*}{x}$. At this stage, the trigonometric perturbation v has little effect, and several local minima are passed without any change in the search direction. As f comes closer to the target, the trajectories come under increasing influence of the local

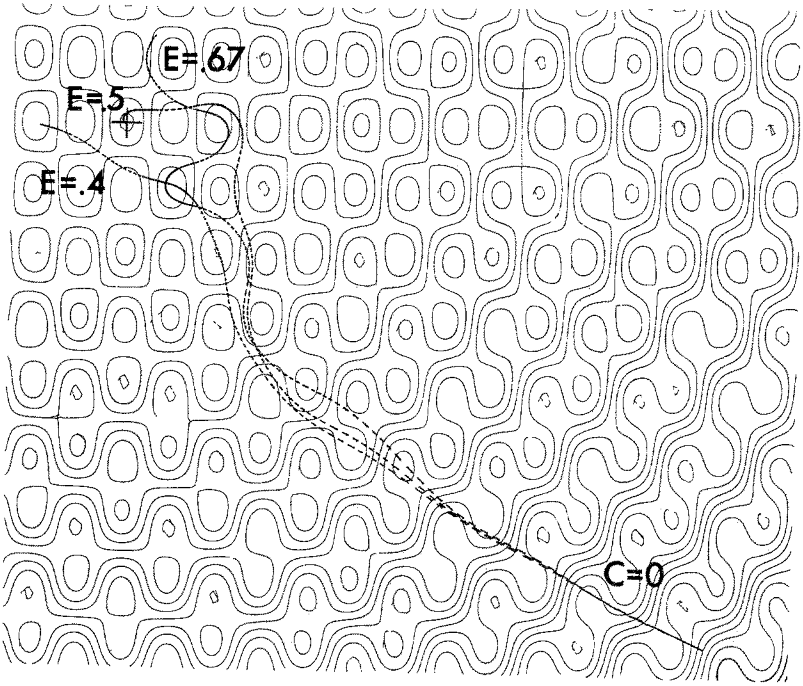

Fig. 1. Search trajectories with target $c = 0$ and sensitivity $e \in \{0.4, 0.5, 0.67\}$ on the objective function $f = (x_1^2 + x_2^2)/200 + 1 - \cos x_1 \cos(x_2/\sqrt{2})$. Initial point $(40, -35)$. Global minimum at origin marked by $+$.

gradient and explore the objective function more thoroughly. Simultaneously, the stepsize, which equals the length of the dashes, becomes smaller to ensure an accurate integration.

The behavior of the individual trajectories confirms in principle the results of Theorem 5.1(i) applied to the quadratic term u, with d being equal to 2. The combination

$$e = \tfrac{1}{2} = 1/d$$

and

$$c = 0 = f^*$$

seems optimal, even though the corresponding trajectory converges to the global solution x^* only from the initial point $(40, -35)$, but not from $(35, -30)$. In the latter case, as shown in Fig. 2, the trajectory is distracted

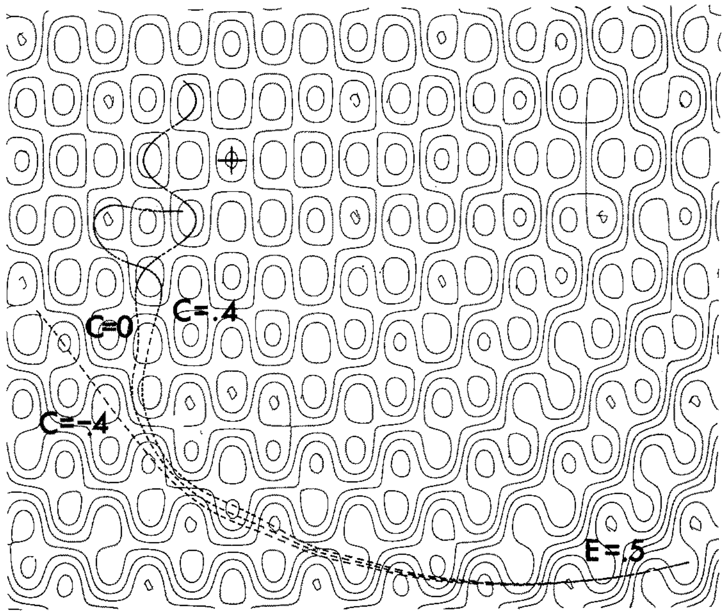

Fig. 2. Search trajectories with sensitivity $e=0.5$ and target $c\in\{-0.4, 0, 0.4\}$ on the objective function $f=(x_1^2+x_2^2)/200+1-\cos x_1 \cos(x_2/\sqrt{2})$. Initial point $(35, -30)$. Global minimum at origin marked by +.

from x^* by a sequence of suboptimal minima and eventually diverges toward infinity. Trajectories with sensitivities larger than 0.5, like the one with $e=0.67$ in Fig. 1, usually lack the penetration to reach x^* and wander around endlessly, as they cannot escape the attraction of the quadratic term u. On the other hand, trajectories with sensitivities less than 0.5, like the one with $e=0.4$ in Fig. 1, are likely to pass the global solution x^* at some distance before diverging toward infinity. The same is true of trajectories with appropriate sensitivity $e=\frac{1}{2}$, but having an unattainable target, as we can see from the case $c=-0.4$ in Fig. 2. Trajectories whose target is attainable are likely to achieve their goal, like the one with $c=0.4$ in Fig. 2, which attains its target close to the suboptimal minimum

$$\hat{x}_1 \approx -\pi(1, \sqrt{2})^T,$$

with value

$$f(\hat{x}_1) \approx 0.15$$

after passing through the neighborhood of two unacceptable minima

$$\hat{x}_2 \approx -2\pi(1, \sqrt{2})^T, \qquad \hat{x}_3 = -\pi(3, \sqrt{2})^T,$$

with values

$$f(\hat{x}_2) \approx 0.6, \qquad f(\hat{x}_3) \approx 0.55,$$

respectively.

On the whole, we draw the following conclusions. Compared to the high function values at the initial points and their surrounding minima, all trajectories locate low local minima, with values only slightly above the optimal solution f^*. Convergence to the global minimum requires the appropriate setting of target level and sensitivity and is even then a matter of chance.

Survey Calculations on the Camel-Back Function. The method was applied to the inverted six-hump, camel-back function

$$f(x_1, x_2) \equiv 4x_1^2 - 2.1x_1^4 + \tfrac{1}{3}x_1^6 + x_1 \cdot x_2 - 4x_2^2 + 4x_2^4,$$

introduced by Branin (Ref. 4). This function is symmetric about the origin and has three conjugate pairs of local minima with the values

$$\hat{f}_1 = -1.0316, \qquad \hat{f}_2 = -0.2154, \qquad \hat{f}_3 = 2.1042.$$

Because of the leading terms $\frac{1}{3}x_1^6$ and $4x_2^4$, the function grows very rapidly with the distance from the origin, so that the area outside the ball

$$\mathcal{B} \equiv \{x \in \mathbb{R}^2 \,|\, \|x\| \le 8\}$$

is of little interest. Numerical experience shows that any trajectory with

$$\|x_0\| > 8, \qquad x_0' = -\nabla f_0 / \|\nabla f_0\|$$

will eventually enter the ball $\mathcal{B}$ provided its sensitivity is not smaller than $\frac{1}{4}$.

Trajectories with the sensitivities and target settings

$$e \in \{\infty, 2, 1, 0.7, 0.5, 0.33, 0.25\}, \qquad c \in \{1, -1, -3\},$$

were started from each point on the grid

$$\mathcal{G} \equiv \{(x_1, x_2)^T \in \mathbb{R}^2 \,|\, x_1 \in \{1, 3, 5\}, x_2 \in \{-5, -3, -1, 1, 3, 5\}\}.$$

The case $e = \infty$ corresponds to the trajectory of steepest descent, which was used to determine the success of local downhill minimizations from the 18 points in $\mathcal{G}$. It was found that each of the values $\hat{f}_1, \hat{f}_2, \hat{f}_3$ was attained by six (i.e., one-third) of the steepest descent trajectories; this indicates that the initial points in $\mathfrak{G}$ are evenly distributed in the regions of attraction of the various local minimizers.

For the target setting

$$c = 1 < \hat{f}_3,$$

all trajectories with

$$\tfrac{1}{4} \leq e \leq 2$$

attained the target level, i.e., reached the neighborhood of one of the four lowest minima from every point in $\mathfrak{G}$. Similarly, almost all trajectories with

$$c = -1 < \hat{f}_2 < \hat{f}_3, \qquad \tfrac{1}{2} \leq e \leq 2$$

reached the neighborhood of a global minimizer with only three exceptions, which had $e = \frac{1}{3}$. Two of these three trajectories did actually attain values less than $\hat{f}_2$, but did not respond strongly enough to the attraction by the global minimum. The same applies to most trajectories with $c = -1$ and $e = \frac{1}{4}$, of which only three converged to the target set $f^{-1}(-1)$.

The unattainable target level $c = -3$ was chosen to test the ability of the trajectories to leave the area around the six minima, which are now all unacceptable. While all trajectories with

$$e \in \{0.25, 0.33\}$$

left the ball $\mathscr{B}$, only five of those with

$$e \in \{0.5, 0.7\}$$

were able to do so in less than 800 steps. All other trajectories, especially those with

$$e \in \{1.0, 2.0\}$$

were trapped as their high sensitivity prevented them from climbing the steep walls created by the fourth-order and sixth-order terms in f.

On the whole, we can conclude that, for the wide range of sensitivities

$$e \in [\tfrac{1}{3}, 2],$$

convergence is almost certain, if the target level is attainable, $c \geq \hat{f}_1$. On the other hand, if $c < \hat{f}_1$, divergence seems possible only for $e \leq \frac{1}{3}$, so that a sensitivity of about one-third appears to be the optimal compromise between the quadratic character of f in the immediate neighborhood of its minima and the rapid growth which prevails further out because of the higher-order terms.

The procedure outlined in Section 6 was applied to the function $f \in C^2(\mathbb{R}^{10})$,

$$f(x) = \sum_{k=1}^{10} x_k^2/4000 - \prod_{k=1}^{10} \cos(x_k/\sqrt{k}) + 1,$$

from the initial point

$$(100, 50, -5, 40, 30, -20, 60, -70, 80, -90)^T.$$

The initial target was chosen well above the initial function vale, so that it did not convey any information about the global minimum $f(0) = 0$. The procedure located eight local minima ranging from 9.3 to 0.015, the last value being attained at one of the four suboptimal minimizers

$$\hat{x} \approx (\pm\pi, \pm\pi\sqrt{2}, 0, 0, 0, 0, 0, 0, 0, 0)^T.$$

The program stopped after 6600 evaluations, of which 3000 were used after the lowest minimum had been found, but failed to improve the result. The procedure restarted six times, and the total length of the trajectories was 4400 units. The search area was defined as the ball with radius 600 about the origin, in which the given objective function attains several thousand local minima. Since each local minimization with a quasi-Newton scheme took on average 20 steps, the 6600 evaluations would not have been sufficient for 1024 local minimizations from a grid with two different values in each coordinate leaving gaps of at least 400 units between the initial points. This example illustrates that even coarse grids have an excessive number of points in higher dimensions, whereas trajectory methods maintain essentially the characteristics of a linear search.

9. Summary and Conclusions

In an attempt to locate comparatively low minima of an objective function f with a multitude of local minima, we considered the search trajectories defined by the differential equation (1), depending on the two parameters c and e. The trajectories aim at attaining the target level c, while rejecting unacceptable minima with values greater than c. It was shown both theoretically (Section 5) and in practical calculations (Section 8) that this aim can be achieved if the sensitivity e is approximately reciprocal to the average growth order d of f in the neighborhoods of its local minima.

It was found that trajectories may bypass acceptable minima or orbit around unacceptable minima if ed is too small or too large, respectively. The avoidance and detection of these difficulties was of major concern in the design of the minimization procedure (Section 6), which uses the solutions of (1) for global searches with successively lower target levels, followed by conventional local minimizations. After a suitable parameter transformation, the defining system (1) can be integrated by any general purpose routine or by the problem-oriented third-order, one-step method developed in Section 7.

Test calculations in the case of the six-hump, camel-back function and a perturbed quadratic form in 10 variables show that the method is quite successful in locating comparatively low local minima, if not the global solution. The expense in terms of function and gradient evaluations is high, but seems comparable to that of the few available alternatives. Improvements in the method can be expected by the incorporation of a variable-metric approach, a more economical integration scheme, and a better understanding of the parameters c and e.

References

1. SCHUBERT, B. O., *Sequential Optimization of Multimodal Discrete Function with Bounded Rate of Change*, Management Science, Vol. 18, pp. 687–693, 1972.
2. AIRD, T. J., and RICE, J. R., *Systematic Search in High Dimensional Sets*, SIAM Journal of Numerical Analysis, Vol. 14, pp. 296–312, 1977.
3. DICKSON, L. C. W., GOMULKA, J., and SZEGÖ, G. P., *Toward a Global Optimization Technique*, Toward Global Optimization, Edited by L. Dickson and G. Szegö, North-Holland Publishing Company, Amsterdam, Holland, 1975.
4. BRANIN, F. H., *Widely Convergent Method for Finding Multiple Solutions of Nonlinear Equations*, IBM Journal of Research and Development, Vol. 16, pp. 504–522, 1972.
5. HIRSCH, M., *Differential Topology*, Springer-Verlag, New York, New York, 1976.
6. SMALE, S., *A Convergent Process of Price Adjustment and Global Newton Methods*, Journal of Mathematical Economics, Vol. 3, pp. 107–220, 1976.
7. KELLER, H. B., *Global Homotopics and Newton Methods*, Recent Advances in Numerical Analysis, Edited by C. DeBoor and G. H. Golub, Academic Press, New York, New York, 1978.
8. JOHN, F., *Ueber die Vollstaendigkeit der Relationen von Morse fuer die Anzahlen Kritischer Punkte*, Mathematische Annalen, Vol. 109, pp. 381–394, 1934.
9. GRIEWANK, A., *A Generalized Descent Method for Global Optimization*, Australian National University, MS Thesis, 1977.
10. WILLMORE, T. J., *An Introduction to Differential Geometry*, Clarendon Press, Oxford, England, 1959.
11. INOMATA, S., and CUMADA, M., *On the Golf Method*, Denshi Gijutso Sogo Kenkyujo, Tokyo, Japan, Bulletin of the Electrotechnical Laboratory, Vol. 25, pp. 495–512, 1964.
12. ZIDKOV, N. P., and SIEDRIN, B. M., *A Certain Method of Search for the Minimum of a Function of Several Variables*, Computing Methods and Programming, Izdat Moscow University, Vol. 10, pp. 203–210, 1968.
13. INCERTI, S., PARISI, V., and ZIRILLI, F., *A New Method for Solving Nonlinear Simultaneous Equations*, SIAM Journal of Numerical Analysis, Vol. 16, pp. 779–789, 1979.
14. GRIEWANK, A., *Analysis and Modification of Newton's Method at Singularities*, Australian National University, PhD Thesis, 1980.

15. POWELL, M. J. D., *A New Algorithm for Unconstrained Optimization*, Nonlinear Programming, Edited by J. B. Rosen, O. L. Mangasarian, and K. Ritter, Academic Press, New York, New York, 1970.
16. GEAR, C. W., *Numerical Initial-Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.
17. ANONYMOUS, International Mathematical and Statistical Libraries, Library 2, Edition 6, Vol. 1, Houston, Texas, 1977.