NASA Contractor Report 181926

ICASE Report No. 89–45

ICASE

NUMERICAL OPTIMIZATION IN HILBERT SPACE USING INEXACT FUNCTION AND GRADIENT EVALUATIONS

Richard G. Carter

Contract No. NAS1–18605
June 1989

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, Virginia 23665–5225

Operated by the Universities Space Research Association

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23665-5225

Numerical Optimization in Hilbert Space Using Inexact Function and Gradient Evaluations

Richard G. Carter*
Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, VA 23665

Abstract

Trust region algorithms provide a robust iterative technique for solving nonconvex unconstrained optimization problems, but in many instances it is prohibitively expensive to compute high accuracy function and gradient values for the method. Of particular interest are inverse and parameter estimation problems, since function and gradient evaluations involve numerically solving large systems of differential equations.

We present global convergence theory for trust region algorithms in which neither function nor gradient values are known exactly. The theory is formulated in a Hilbert space setting so that it can be applied to variational problems as well as the finite dimensional problems normally seen in trust region literature. The conditions concerning allowable error are remarkably relaxed: relative errors in the gradient values of 0.5 or more are allowed by the theory. One form of the gradient error condition is automatically satisfied if the error is orthogonal to the gradient approximation. A technique for estimating gradient error and improving the approximation is also presented.

*This research was supported by the National Aeronautics and Space Administration under NASA Contract No. NAS1-18605 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23665.

1 Introduction

An increasingly important area of computational mathematics involves problems requiring both numerical simulation and numerical optimization techniques. For example, in the design of a large flexible structure such as the space station, engineers may derive an ODE or PDE model of the structure based on a number of design parameters, define an objective (cost) function for the possible designs based on some criteria of interest (weight, flexibility, controllability, cost), and use a numerical optimization routine to find the "best" set of design parameters. Often the differential equations involved in the model are not amenable to analytic solutions, and therefore the calculation of objective function and gradient values in the optimization routine will involve the numerical solution of a system of differential equations. Problems of this type are common not only in structural design, but also in control, parameter estimation, and image reconstruction, to name but a few areas. A number of such problems are surveyed by Minkoff [11]. Among the points stressed in his study are the very wide range of applications in which these problems are encountered, the computationally intense nature of the simulations, and the wide availability of numerical packages such as ODEPACK [7] which allow the user to specify the amount of computational accuracy desired in the simulation. Clearly, "exact" function and gradient evaluations are not feasible in such situations, and low accuracy evaluations may even be desirable in cases where computational expense increases very rapidly with increased accuracy in the simulation. Equally clear is the fact that sufficiently large errors will cause the optimization algorithm to fail.

Trust region algorithms for nonlinear optimization have been an increasingly popular choice in recent years because of their elegance, efficiency, and robust convergence properties. In this paper, we establish global convergence results for a class of these algorithms when neither function nor gradient values are computed exactly. The conditions concerning allowable error are both natural and exceedingly mild. Although trust region methods are most commonly applied to finite dimensional problems, in this paper we emulate Toint [19] and present our analysis in a general Hilbert space setting so that the trust region algorithm

can, in principle, be applied directly to a variational (distributed parameter) problem rather than to a finite dimensional discretization of the problem developed at an early stage of the design process. A comparison of the relative merits of these two approaches is an interesting research issue, but is beyond the scope of this paper.

Synopsis. In Section 2, we define our problem and present the trust region algorithm. Our conditions for admissible error in function and gradient values are introduced and briefly discussed. In Section 3, we present several properties associated with the computation of trial steps from sequences of quadratic models. Using these properties and our conditions concerning function and gradient error, we first establish that at least a subsequence of the gradient approximations converges to zero, and then the stronger result that both the sequence of gradients and the sequence of gradient approximations converge to zero. In Section 4, we discuss implementation of the gradient error conditions, and suggest a technique for directly estimating the gradient error if other estimates are not available. This technique is particularly appropriate if gradients are computed using a finite difference procedure, and can also be used to *improve* the accuracy of a given approximation. In Section 5, we discuss a few of the possible generalizations of our theory. In Section 6, we summarize our results.

2 Preliminaries

Let H denote a real Hilbert space, and consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x), \\ &x \in H \end{aligned} \tag{1}$$

for some functional $f : H \to \Re$. For a given vector $x_0 \in H$, let Ω be an open convex subset of H containing the level set of f at x_0. We assume

A.1 f is Fréchet differentiable on Ω,

A.2 f is bounded below, and

A.3 the Fréchet derivative of f, denoted f', is Lipschitz continuous on Ω.

Our trust region algorithm for solving (1) generates a sequence of iterates $\{x_k\}$ by producing and approximately solving a sequence of constrained quadratic model problems. That is, $x_{k+1} = x_k + s_k$ for a step s_k that approximately solves

$$\text{minimize}\, \psi_k(x_k + s) : \|s\| \leq \Delta_k \tag{2}$$

where Δ_k is a positive variable known as the trust radius and ψ_k is a quadratic model of the objective functional f about the point x_k. Let $\langle \cdot, \cdot \rangle$ denote the inner product on H, and let $\|\cdot\|$ denote the associated norm or the induced operator norm. Our quadratic model ψ_k will then be of the form

$$\psi_k(x_k + s) = f_k + \langle g_k, s \rangle + \frac{1}{2}\langle B_k s, s \rangle, \tag{3}$$

where f_k is our approximation to $f(x_k)$, g_k is our approximation to $\nabla f(x_k)$, the gradient of f at x_k, and B_k is a self-adjoint operator from H into H approximating $\nabla^2 f(x_k)$.

If $f_k \neq f(x_k)$, we must specify conditions on how much error is allowable. These conditions will apply to the difference between successive function values rather than to errors in the values themselves. Define the *actual* function reduction

$$\text{ared}_k(s_k) = f(x_k) - f(x_k + s_k), \tag{4}$$

the *computed* function reduction

$$\text{cred}_k(s_k) = f_k - f_{k+1}, \tag{5}$$

and the *predicted* function reduction

$$\text{pred}_k(s_k) = \psi_k(x_k) - \psi_k(x_k + s_k). \tag{6}$$

We then require two conditions to be satisfied at every iteration for some appropriately chosen constants $\xi_{f,1}$ and $\xi_{f,2}$:

$$|\text{ared}_k(s_k) - \text{cred}_k(s_k)| \leq \xi_{f,1}\, \text{pred}_k(s_k), \tag{7}$$

and

$$|\text{ared}_k(s_k) - \text{cred}_k(s_k)| \leq \xi_{f,2} |\text{cred}_k(s_k)|. \tag{8}$$

Since direct estimates of $|\text{ared}_k(s_k) - \text{cred}_k(s_k)|$ are probably not available in most applications, in practice (7) and (8) should be replaced by

$$|f_{k+1} - f(x_{k+1})| + |f_k - f(x_k)| \leq \xi_{f,1}\text{pred}_k(s_k) \tag{9}$$

and

$$|f_{k+1} - f(x_{k+1})| + |f_k - f(x_k)| \leq \xi_{f,2}|\text{cred}_k(s_k)|. \tag{10}$$

If $g_k \neq \nabla f(x_k)$, we must similarly specify conditions on how much error is allowable in the gradient. Define

$$e_k = g_k - \nabla f(x_k). \tag{11}$$

We will show that the condition

$$\frac{\langle e_k, g_k \rangle}{\langle g_k, g_k \rangle} \leq \xi_g, \tag{12}$$

will lead to the global convergence result $\liminf_{k\to\infty} \|g_k\| = 0$ for appropriately chosen constant ζ_g, while the stronger result $\lim_{k\to\infty} \|g_k\| = \lim_{k\to\infty} \|\nabla f(x_k)\| = 0$ can be obtained by using the stronger condition

$$\frac{\|e_k\|}{\|g_k\|} \leq \xi_g. \tag{13}$$

Our algorithm is structured as follows.

Algorithm(1): Trust region method using inexact function and gradient evaluations.

Let the constants $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < 1 < \gamma_2$ be prespecified, and select the error control constants $\xi_{f,1}, \xi_{f,2}$, and ξ_g such that

$$\xi_g + \xi_{f,1} < 1 - \eta_2, \tag{14}$$

and

$$\xi_{f,2} < 1. \tag{15}$$

Select an initial guess $x_0 \epsilon H$ and an initial trust radius Δ_0. Compute initial function and gradient values f_0 and g_0, and compute or initialize B_0.

For $k = 0, 1, ...$ **until** "convergence" **do:**

(a) Determine an approximate solution s_k to problem (2).

(b) Calculate $\text{pred}_k(s_k)$ and $\text{cred}_k(s_k)$. If necessary, recompute f_{k+1} and/or f_k to greater accuracy until (7) and (8) are satisfied.

(c) Compute the ratio

$$\rho_k = \frac{\text{cred}_k(s_k)}{\text{pred}_k(s_k)} \tag{16}$$

(d) If $\rho_k < \eta_1$, then set $\Delta_{k+1} \in (0, \gamma_1 \Delta_k]$,

else if $\rho_k < \eta_2$, then set $\Delta_{k+1} \in (0, \Delta_k]$,

else set $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$.

(e) If $\rho_k < \eta_1$, then the current step is unacceptable. Set $x_{k+1} = x_k$.

Otherwise, the iteration is **successful**. Set $x_{k+1} = x_k + s_k$.

(f) If $x_{k+1} \neq x_k$, then compute g_{k+1} and compute or update B_{k+1}.

Otherwise, retain the current values by setting $f_{k+1} = f_k$, $g_{k+1} = g_k$, $B_{k+1} = B_k$.

End Loop.

At this point, a number of comments should be made concerning the algorithm.

1. The maximum, error levels given by (14) and (15) are extremely mild. A typical value for η_2 in an algorithm might be 0.1, in which case we could select $\xi_g = 0.5, \xi_{f,1} = 0.3$, and $\xi_{f,2} = 0.99$ so that we are allowed a relative error of one half in the gradient approximation and a similarly large error in the difference between successive function values.

2. Conditions (7), (8), (12) and (13) are different than the conditions used in [12], [19] and [3]. All of these papers consider only the case $f_k = f(x_k)$. Instead of (12) and (13), [12] uses the consistency condition

$$\{x_k \to x^*\} \Longrightarrow \lim_{k\to\infty} \|e_k\| = 0 \tag{17}$$

for the case $H = \Re^n$, while [19] uses the condition

$$\|e_k\| \leq \min\{\kappa_1, \kappa_2 \Delta_k\} \tag{18}$$

in the very general setting of Hilbert space with simple bounds. Condition (13) is used in [3], but the weaker condition (12) is not considered.

3. In practice, the trust region defined in (2) is often replaced by the scaled trust region $\|D_k s\| \leq \Delta_k$ for some invertible linear operator D_k. For simplicity we take D_k to be the identity in this paper, but note that our results can still be established provided (12) and (13) are replaced by

$$\frac{\langle D_k^{-1} e_k \,,\, D_k^{-1} g_k \rangle}{\langle D_k^{-1} g_k \,,\, D_k^{-1} g_k \rangle} \leq \zeta_g \tag{19}$$

and

$$\frac{\|D_k^{-1}\, e_k\|}{\|D_k^{-1}\, g_k\|} \leq \xi_g, \tag{20}$$

respectively, and some restrictions are placed on the sequence of scalings $\{D_k\}$. A complete treatment of (20) for the case $f_k = f(x_k)$ and $H = \Re^n$ can be found in [3].

4. In algorithm (1), no requirement is made that g_k be recomputed to greater accuracy following unsuccessful iterations. This is an important property, since even for the case where function and gradient values are known exactly, unsuccessful iterations are quite common and merely indicate that the trust radius should be adjusted.

5. On the other hand, conditions (7) and (8) may require function values to be recomputed to greater accuracy at any iteration. Although these conditions are therefore less

elegant than (12) and (13), they are still fairly natural and can be relatively easily implemented if error in f is controllable.

The optimal strategy for enforcing (7) and (8) will of course depend on such factors as the reliability of error estimates and the amount of work involved in recomputing f_k to a greater accuracy. For example, the following simple procedure could be used to implement step(b) of algorithm(1).

Procedure 2.

Let $\alpha \in (0,1)$ be prespecified. Given x_k, s_k, ψ_k, and an estimate for $|f_k - f(x_k)|$, do the following.

(i) Calculate $\text{pred}_k(s_k)$, and set emax $= \xi_{f,1}\,\text{pred}_k(s_k)$.

(ii) If necessary, recompute f_k to greater accuracy so that the inequality $|f_k - f(x_k)| \leq (1-\alpha)\,\text{emax}$ holds .

(iii) Compute f_{k+1} so that $|f_{k+1} - f(x_{k+1})| \leq \alpha\,\text{emax}$.

(iv) Compute $\text{cred}_k(s_k)$. If condition (10) is satisfied, then exit procedure, else reduce emax and return to (ii).

End procedure.

Clearly, if error estimates are unreliable $\xi_{f,1}$ should be chosen fairly small. On the other hand, $\xi_{f,2}$ should usually be selected close to one to avoid unnecessary recomputations of f_k and f_{k+1}. If such recomputations are very expensive, one might consider taking α close to one.

3 Convergence Results

In order to establish our results, we will need several properties regarding our trial steps s_k. These properties are

$$\text{pred}_k(s_k) \geq \frac{1}{2} c_1 \|g_k\| \min\{\Delta_k, \|g_k\|/c_2\}, \tag{21}$$

$$\|s_k\| \leq c_3 \Delta_k, \tag{22}$$

and

$$\{\liminf_{k\to\infty} \|g_k\| > 0 \text{ and } \lim_{k\to\infty} \Delta_k = 0\} \Longrightarrow \lim_{k\to\infty} -\frac{\langle s_k, g_k\rangle}{\|s_k\|\|g_k\|} = 1, \tag{23}$$

for some constants $c_1 \in (0,1), c_2 \in (0,\infty)$, and $c_3 \in [1,2]$. Obviously, (21), (22) and (23) are directly dependent on both the methods used to compute trial steps and the properties of the sequence of quadratic models chosen. For the special case $H = \Re^n$, a few comments are in order.

Condition (21) is well knows (see, for example, [18]) and is usually established by assuming an upper bound of the form

$$\|B_k\| \leq c_2 \tag{24}$$

However (21) can also be established [2] given an upper bound of the form

$$\langle B_k g_k\, ,\, g_k\rangle \leq c_2 \langle g_k, g_k\rangle \tag{25}$$

Condition (23) can be interpreted geometrically as stating that steps s_k tend in direction toward $-g_k$ as $\|s_k\|/\|g_k\|$ goes to zero. This property is established in [3] by assuming an upper bound on $\{\|B_k\|\}$, but can also be established using the milder condition (25) for one of the popular classes of techniques of computing trial steps (generalized dogleg methods).

In the context of inexact function and gradient values, assumptions such as (24) or (25) are quite reasonable: if first order information is not known accurately it is only natural to directly enforce an upper limit on our approximation to second order information. In this paper, however, we make no assumptions about how (21), (22), and (23) are obtained. The only assumption we directly use concerning the sequence $\{B_k\}$ is that

$$-c_4\langle s, s\rangle \leq \langle B_k s, s\rangle \tag{26}$$

at every iteration for all $s \in H$ and some constant c_4. First note the following simple results.

Lemma 3.1. Let f satisfy assumptions A.1 and A.3 and let c_5 be the Lipshitz constant associated with ∇f. Then we have

$$\text{pred}_k(s) - \text{ared}_k(s) \leq \frac{1}{2}(c_4 + c_5)\|s\|^2 - \langle e_k, s\rangle. \tag{27}$$

Proof. Using an integral representation of $\text{pred}_k(s) - \text{ared}_k(s)$, we have

$$\begin{aligned} \text{pred}_k(s) - \text{ared}_k(s) &= -\langle g_k, s\rangle - \frac{1}{2}\langle B_k s, s\rangle + \int_0^1 \langle \nabla f(x_k + \lambda s), s\rangle d\lambda \\ &= -\langle e_k, s\rangle - \frac{1}{2}\langle B_k s, s\rangle + \int_0^1 \langle \nabla f(x_k + \lambda s) - \nabla f(x_k), s\rangle d\lambda. \quad (28) \end{aligned}$$

Using (26), the Cauchy-Schwarz inequality, and the Lipshitz continuity of ∇f, we have

$$\begin{aligned} \text{pred}_k(s) - \text{ared}_k(s) &\leq -\langle e_k, s\rangle + \frac{1}{2}c_4\|s\|^2 + \int_0^1 \|\nabla f(x_k + \lambda s) - \nabla f(x_k)\|\|s\| d\lambda \\ &\leq -\langle e_k, s\rangle + \frac{1}{2}c_4\|s\|^2 + \int_0^1 c_5\|\lambda s\|\,\|s\|\, d\lambda, \quad (29) \end{aligned}$$

which immediately establishes (27). □

Lemma 3.2. Let f satisfy assumptions A.1, A.2, and A.3, and let Ω_0 be the interior of the level set of f at x_0. We then have that ∇f is bounded on Ω_0.

Proof. Let c_5 be the Lipschitz constant associated with ∇f, and let c_6 be such that $f(x) \geq c_6 \forall x \in \Omega_0$. Now, suppose ∇f is unbounded on Ω_0 so that $\exists\, \overline{x} \in \Omega_0$ with $\|\nabla f(\overline{x})\|^2 > 8c_4(f(x_0) - c_6)$. Define $\overline{s} = \frac{-\alpha}{2c_5}\nabla f(\overline{x})$. For all α sufficiently small that $\overline{x} + \overline{s} \in \Omega$, we have

$$\begin{aligned} f(\overline{x}) - f(\overline{x} + \overline{s}) &= -\int_0^1 \langle \nabla f(\overline{x}), \overline{s}\rangle d\lambda - \int_0^1 \langle \nabla f(\overline{x} - \lambda\overline{s}) - \nabla f(\overline{x}), \overline{s}\rangle d\lambda \\ &\geq \frac{\alpha}{4c_5}\|\nabla f(\overline{x})\|^2 - \frac{1}{2}c_5\|\overline{s}\|^2 \quad (30) \\ &\geq \frac{1}{4}\frac{\alpha}{c_5}\|\nabla f(\overline{x})\|^2(1 - \frac{\alpha}{2}) \\ &> 2\alpha(1 - \frac{\alpha}{2})(f(x_0) - c_6). \end{aligned}$$

Now, the final term in (30) is positive for all $\alpha \in (0, 2)$ and hence $\overline{x} + \overline{s} \in \Omega$ for $\alpha = 1$. But this leads to the contradiction $f(\overline{x}) - f(\overline{x} + \overline{s}) > f(x_0) - c_6$, so ∇f cannot be unbounded on Ω_0. □

We now establish that $\{g_k\}$ is not bounded away from zero.

Theorem 3.3. Let f satisfy assumptions A.1, A.2, and A.3, let the steps generated in Algorithm (1) satisfy (21), (22), and (23), let $\{B_k\}$ satisfy (26), and let the function evaluations

satisfy (7) and (8). Then our algorithm generates a sequence of iterates satisfying

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{31}$$

provided condition (12) holds.

Proof. Let K_s denote the set of successful iterations. First notice that $\text{pred}_k(s_k) > 0$ and from (16), $\text{cred}_k(s_k) \geq \eta_1 \text{pred}_k(s_k)\ \forall k \in K_s$. From (8) we have

$$(1 - \zeta_{f,2})\text{cred}_k(s_k) \leq \text{ared}_k(s_k) \leq (1 + \zeta_{f,2})\text{cred}_k(s_k) \tag{32}$$

Combining (32) with (16) and (21) yields.

$$\begin{aligned} \text{ared}_k(s_k) &\geq (1 - \zeta_{f,2})\eta_1 \,\text{pred}_k(s_k) \\ &\geq \frac{1}{2} c_1 \eta_1 (1 - \xi_{f,2}) \,\|g_k\| \min\{\Delta_k, \|g_k\|/c_2\} \end{aligned} \tag{33}$$

for all $k \in K_s$.

Next, define θ_k such that

$$\cos\,\theta_k = \frac{\langle -g_k, s_k\rangle}{\|g_k\|\,\|s_k\|} \tag{34}$$

and $w_k \in H$ such that $w_k = 0$ if $\sin\,\theta_k = 0$ and

$$w_k = \frac{1}{\sin\,\theta_k}\left(\frac{s_k}{\|s_k\|} + \cos\theta_k \,\frac{g_k}{\|g_k\|}\right) \tag{35}$$

otherwise. Notice that $\langle g_k, w_k\rangle = 0$, $\langle e_k, w_k\rangle = \langle -\nabla f(x_k), w_k\rangle$, and $\|w_k\| = 1$ for $\sin\,\theta_k \neq 0$, and that

$$s_k = \|s_k\|(-\cos\theta_k \,\frac{g_k}{\|g_k\|} + \sin\theta_k\, w_k). \tag{36}$$

Now, from (7) and (16) and the fact that $\text{pred}_k(s_k) > 0$, we have

$$\begin{aligned} 1 - \rho_k &= \frac{\text{pred}_k(s_k) - \text{cred}_k)s_k)}{\text{pred}_k(s_k)} \\ &= \frac{\text{pred}_k(s_k) - \text{ared}_k(s_k) + \text{ared}_k(s_k) - \text{cred}_k(s_k)}{\text{pred}_k(s_k)} \\ &\leq \frac{\text{pred}_k(s_k) - \text{ared}_k(s_k)}{\text{pred}_k(s_k)} + \zeta_{f,1}. \end{aligned} \tag{37}$$

Using (3), (6), (26), and (27) gives

$$\begin{aligned} 1-\rho_k &\leq \zeta_{f,1} + \frac{\frac{1}{2}(c_4+c_5)\|s_k\|^2 - \langle e_k, s_k\rangle}{-\langle g_k, s_k\rangle - \frac{1}{2}\langle B_k s_k, s_k\rangle} \\ &\leq \zeta_{f,1} + \frac{-\langle e_k, s_k\rangle + \frac{1}{2}(c_4+c_5)\|s_k\|^2}{-\langle g_k, s_k\rangle + \frac{1}{2}c_4\|s_k\|^2} \end{aligned} \tag{38}$$

Substituting (36) into (38) yields

$$\begin{aligned} 1-\rho_k &\leq \zeta_{f,1} + \frac{\frac{\|s_k\|}{\|g_k\|}\cos\theta_k\langle e_k, g_k\rangle - \|s_k\|\sin\theta_k\langle e_k, w_k\rangle + \frac{1}{2}(c_4+c_5)\|s_k\|^2}{\frac{\|s_k\|}{\|g_k\|}\cos\theta_k\langle g_k, g_k\rangle - \|s_k\|\sin\theta_k\langle g_k, w_k\rangle + \frac{1}{2}c_4\|s_k\|^2} \\ &= \zeta_{f,1} + \frac{\cos\theta_k\frac{\langle e_k, g_k\rangle}{\langle g_k, g_k\rangle} + \frac{\sin\theta_k}{\|g_k\|}\langle\nabla f(x_k), w_k\rangle + \frac{1}{2}(c_4+c_5)\frac{\|s_k\|}{\|g_k\|}}{\cos\theta_k + \frac{1}{2}c_4\|s_k\|/\|g_k\|} \end{aligned} \tag{39}$$

Now, suppose $\liminf_{k\to\infty} \|g_k\| > 0$. Since f is bounded below, (33) implies that $\lim_{k\in K_s} \Delta_k = 0$ and hence $\lim_{k\to\infty}\Delta_k = 0$. But if $\lim_{k\to\infty} \Delta_k = 0$, from (23) we have $\lim_{\Delta_k\to 0}\cos\theta_k = 1$ and $\lim_{\Delta_k\to 0}\sin\theta_k = 0$. By Lemma 3.2 and the Cauchy Schwarz inequality, $\langle\nabla f(x_k), w_k\rangle$ is bounded, and hence from (14) and (22) we have

$$\begin{aligned} \lim_{\Delta_k\to 0}(1-\rho_k) &\leq \zeta_{f,1} + \lim_{\Delta_k\to 0}\frac{\cos\theta_k\frac{\langle e_k, g_k\rangle}{\langle g_k, g_k\rangle} + \frac{\sin\theta_k}{\|g_k\|}\langle\nabla f(x_k), w_k\rangle + \frac{1}{2}(c_4+c_5)\frac{c_3\Delta_k}{\|g_k\|}}{\cos\theta_k + \frac{1}{2}c_4(c_3\Delta_k)^2/\|g_k\|^2} \\ &\leq \zeta_{f,1} + \frac{\langle e_k, g_k\rangle}{\langle g_k, g_k\rangle} \\ &\leq \zeta_{f,1} + \zeta_g < 1-\eta_2, \end{aligned} \tag{40}$$

and hence $\rho_k > \eta_2$ for all sufficiently large k. But this is a contradiction since $\rho_k > \eta_2 \Rightarrow \Delta_{k+1} \geq \Delta_k$. Hence, $\liminf_{k\to\infty}\|g_k\|$ cannot be greater than zero. □

The final result of this section uses the stronger error condition (13) to establish the stronger result $\lim_{k\to\infty}\|\nabla f(x_k)\| = 0$.

Theorem 3.4. Let f satisfy assumptions A.1, A.2, and A.3, let the steps generated in Algorithm (1) satisfy (21) and (22), and the function evaluations satisfy (8). Then (13) and (31) imply

$$\lim_{k\to\infty}\|g_k\| = \lim_{k\to\infty}\|\nabla f(x_k)\| = 0. \tag{41}$$

Proof: First note that from (13) with any $\zeta_g < 1$ we can immediately obtain the equivalences ($\liminf_{k\to\infty}\|g_k\| = 0$) $\Longleftrightarrow$ ($\liminf_{k\to\infty}\|\nabla f(x_k)\| = 0$) and ($\lim_{k\to\infty}\|g_k\| = 0$) $\Longleftrightarrow$

($\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$). Define $\varepsilon = \frac{1}{2}(1-\zeta_g)/(1+\zeta_g)$. Since $\liminf_{k\to\infty} \|g_k\| = 0$, for any m with $\|g_m\| \neq 0$ there exists $\overline{m} \geq m$ for which $\|g_{\overline{m}+1}\| \leq \varepsilon\|g_m\|$ and $\|g_k\| > \varepsilon\|g_m\|$ $\forall k \in [m, \overline{m}]$. Using (22) and (33) we have

$$\begin{aligned} f(x_m) - f(x_{\overline{m}+1}) &\geq \sum_{k=m}^{\overline{m}} \text{ared}_k(s_k) \\ &\geq \sum_{k=m}^{\overline{m}} \frac{1}{2} c_1 \eta_1 (1-\zeta_{f,2}) \|g_k\| \min\{\Delta_k, \|g_k\|/c_2\} \qquad (42) \\ &\geq \frac{1}{2} c_1 \eta_1 (1-\zeta_{f,2}) \varepsilon \|g_m\| \sum_{k=m}^{\overline{m}} \min\left\{\frac{\|s_k\|}{c_3}, \frac{\varepsilon}{c_2}\|g_m\|\right\} \end{aligned}$$

From the triangle inequality we have

$$\begin{aligned} \|g_m\| &\leq \|g_m - g_{\overline{m}+1}\| + \|g_{\overline{m}+1}\| \\ &\leq \|g_m - g_{\overline{m}+1}\| + \varepsilon\|g_m\|. \qquad (43) \end{aligned}$$

Rearranging terms, substituting $g_k = e_k + \nabla f(x_k)$, and again applying the triangle inequality, we have

$$\begin{aligned} (1-\varepsilon)\|g_m\| &\leq \|g_m - g_{\overline{m}+1}\| \\ &\leq \|\nabla f(x_m) - \nabla f(x_{\overline{m}+1})\| + \|e_m\| + \|e_{\overline{m}+1}\| \\ &\leq \sum_{k=m}^{\overline{m}} \|\nabla f(x_{k+1}) - \nabla f(x_k)\| + \|e_m\| + \|e_{\overline{m}+1}\| \qquad (44) \\ &\leq c_5 \sum_{k=m}^{\overline{m}} \|s_k\| + \|e_m\| + \|e_{\overline{m}+1}\| \end{aligned}$$

Using (13) in this equation with $\varepsilon = \frac{1}{2}(1-\zeta_g)/(1+\zeta_g)$ and $\|g_{\overline{m}+1}\| \leq \varepsilon\|g_m\|$ yields

$$\begin{aligned} \sum_{k=m}^{\overline{m}} \|s_k\| &\geq \frac{1}{c_5}[(1-\varepsilon)\|g_m\| + \|e_m\| + \|e_{\overline{m}+1}\|] \\ &\geq \frac{\|g_m\|}{c_5}\left[(1-\varepsilon) - \frac{\|e_m\|}{\|g_m\|} - \frac{\|e_{\overline{m}+1}\|}{\|g_{\overline{m}+1}\|}\frac{\|g_{\overline{m}+1}\|}{\|g_m\|}\right] \\ &\geq \frac{\|g_m\|}{c_5}[1-\varepsilon-\zeta_g-\zeta_g\varepsilon] \qquad (45) \\ &\geq \frac{\|g_m\|}{c_5}[1-\zeta_g-\varepsilon(1+\zeta_g)] \\ &\geq \frac{\|g_m\|}{c_5}\frac{1}{2}(1-\zeta_g). \end{aligned}$$

Substituting into (42) gives

$$f(x_m) - f(x_{\overline{m}+1}) \geq \varepsilon_1 \|g_m\|^2, \tag{46}$$

where

$$\varepsilon_1 = \frac{1}{2} c_1 \eta_1 (1 - \zeta_{f,2}) \varepsilon \min \left\{ \frac{1}{2} \frac{(1-\zeta_g)}{c_3 c_5}, \frac{\varepsilon}{c_2} \right\}, \tag{47}$$

and hence

$$\|g_m\|^2 \leq \frac{1}{\varepsilon_1} (f(x_m) - f(x_{\overline{m}+1})). \tag{48}$$

Now, $\{f(x_k)\}$ is nonincreasing and bounded below so that $f(x_k) \to f^*$ for some f^*. Hence for any m, either $g_m = 0$ or $\|g_m\| \leq \frac{1}{\varepsilon_1}(f(x_m) - f^*)$, and $\lim_{k\to\infty} \|g_k\| = \lim_{k\to\infty} \|\nabla f(x_k)\| = 0$. □

4 Implementing the gradient error conditions

In terms of global convergence results, (13) is clearly superior to (12) and should be enforced whenever possible. The availability of error estimates will of course depend on the application, but we point to the increasing availability and use of high quality software such as ODEPACK [7] which allow the user to prespecify desired levels of accuracy in each component of the differential equation being solved.

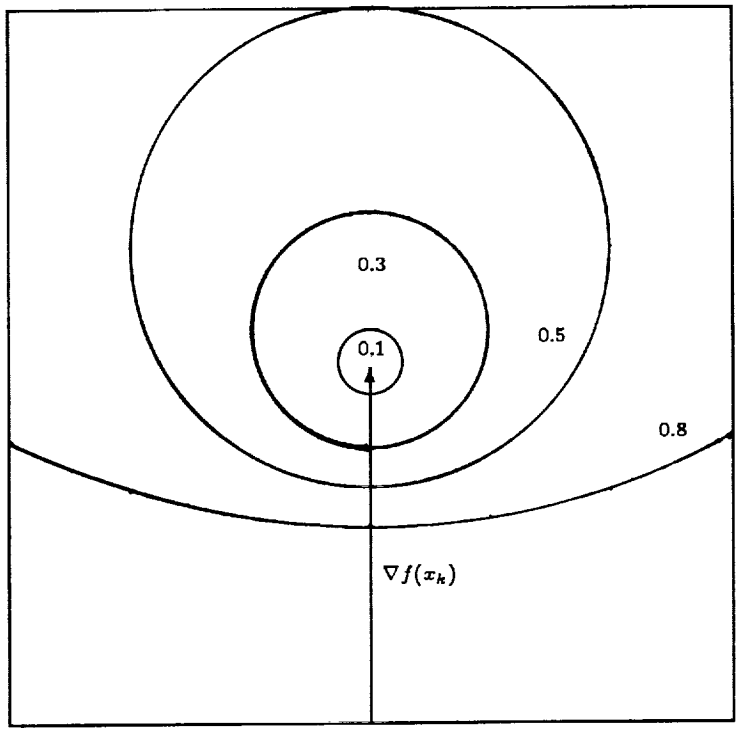

Figure 1 : S_r for $\zeta_g = 0.1$, 0.3, 0.5, and 0.8.

Condition (12), although leading to the weaker convergence result $\liminf_{k\to\infty} \|g_k\| = 0$, has a number of interesting properties.

First notice that if g_k is a Galerkin approximation to $\nabla f(x_k)$, condition (12) is automatically satisfied since $\langle e_k, g_k \rangle = 0$. Second, (12) is a much milder condition than (13) unless ζ_g is close to one. For example, define

$$S_r(\zeta_g, \nabla f(x_k)) = \{g_k : \|e_k\| \leq \zeta_g \|g_k\|\} \tag{49}$$

and

$$S_p(\zeta_g, \nabla f(x_k)) = \{g_k : \langle e_k, g_k \rangle \leq \zeta_g \langle g_k, g_k \rangle\}. \tag{50}$$

These sets can be interpreted geometrically for $\nabla f \in \Re^2$. Figure 1 shows S_r for a variety of values of ζ_g while figure 2 shows S_p for the same values of ζ_g. Figures 3 and 4 directly compare S_r and S_p for $\zeta_g = 0.1$ and 0.5. For small values of ζ_g, condition (12) is significantly milder than (13). For larger values of ζ_g, the difference is less pronounced.

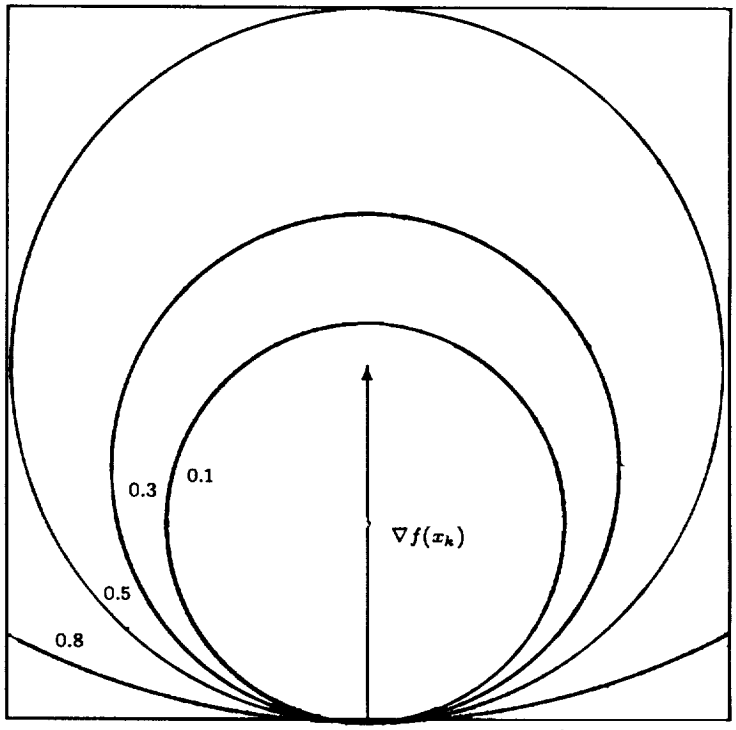

Figure 2 : S_p for $\zeta_g = 0.1$, 0.3, 0.5, and 0.8.

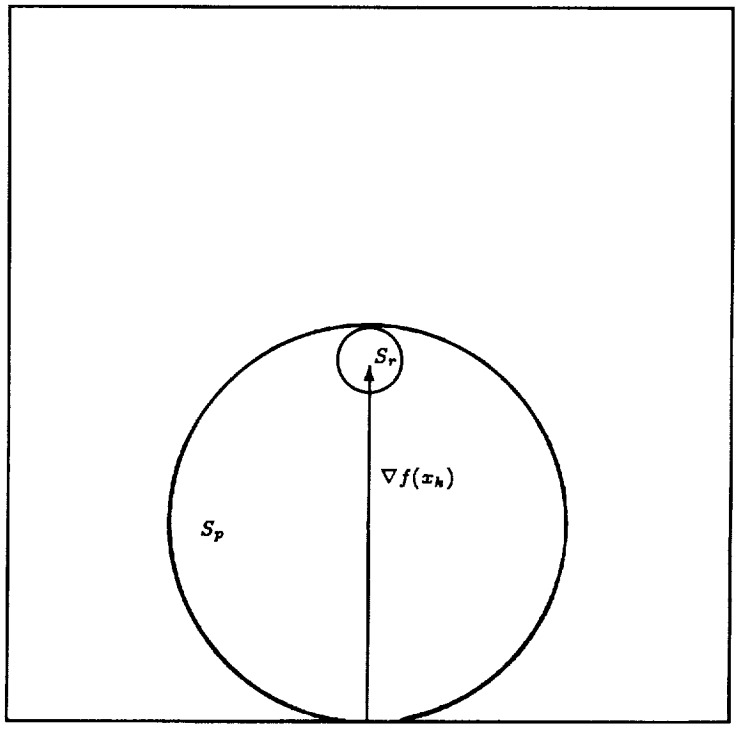

Figure 3 : S_p and S_r for $\zeta_g = 0.1$.

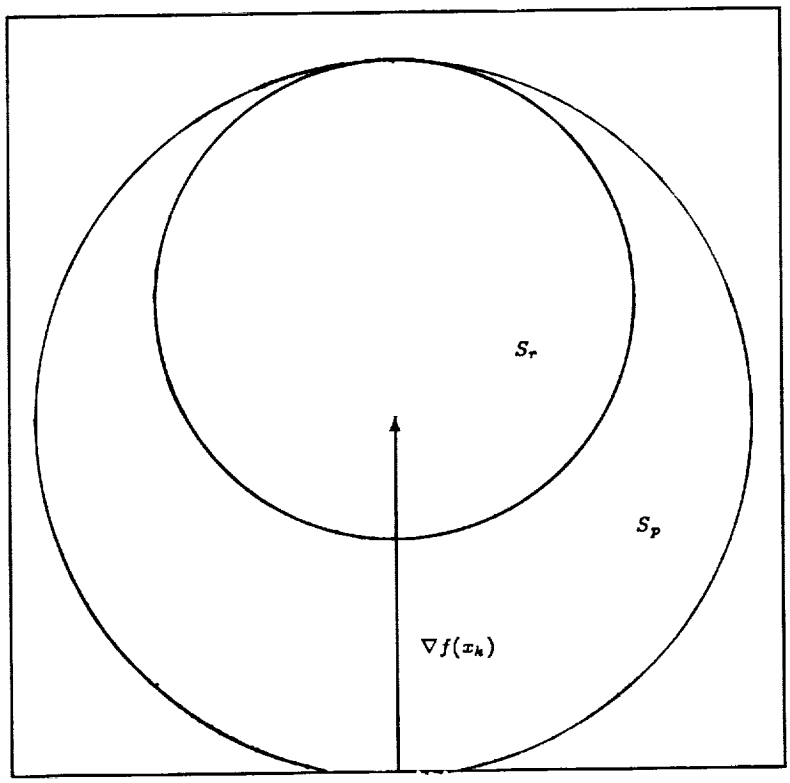

Figure 4 : S_p and S_r for $\zeta_g = 0.5$.

Third, consider that (12) can be written as

$$1 - \frac{\langle \nabla f(x_k), g_k \rangle}{\langle g_k, g_k \rangle} \leq \zeta_g. \tag{51}$$

But $\langle \nabla f(x_k), g_k \rangle$ is simply the Gateaux differential of f at x_k with increment g_k:

$$\langle \nabla f(x_k), g_k \rangle = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} [f(x_k + \varepsilon g_k) - f(x_k)]. \tag{52}$$

Hence, if error estimates are not available from other sources, $\langle \nabla f(x_k), g_k \rangle$ may be approximated by a finite difference formula and substituted into (51). Moreover, if the estimate for $\langle \nabla f(x_k), g_k \rangle$ is sufficiently accurate, we can improve our gradient approximation g_k by replacing it with the scaled gradient

$$\overline{g}_k = \frac{\langle \nabla f(x_k), g_k \rangle}{\langle g_k, g_k \rangle} g_k \tag{53}$$

so that

$$\frac{\langle \overline{e}_k, \overline{g}_k \rangle}{\langle \overline{g}_k, \overline{g}_k \rangle} = 1 - \frac{\langle \nabla f(x_k), \overline{g}_k \rangle}{\langle \overline{g}_k, \overline{g}_k \rangle} = 0. \tag{54}$$

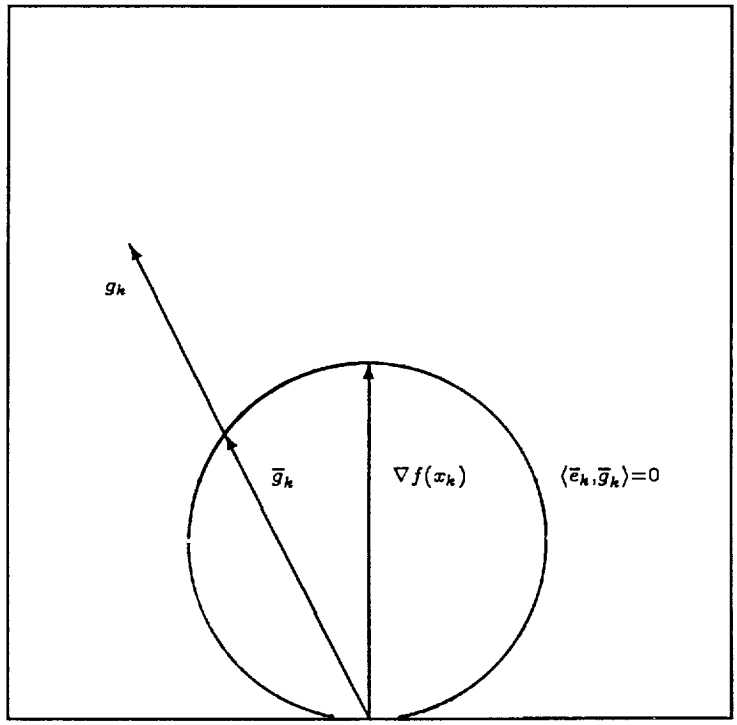

Figure 5 : Gradient correction through the projection operation (53)

This approach seems particularly attractive in the event that $H = \Re^n$, the f_k values are known very accurately, and g_k is being approximated by finite differences. After computing a first approximation to g_k by, say, forward differences using n extra function evaluations, a central difference approximation to $\langle \nabla f(x_k), g_k \rangle$ using two additional function evaluations can be computed to validate the accuracy of g_k. If (12) is violated, then the algorithm can henceforth use central difference approximations to g_k. Notice that the normally difficult problem of selecting an appropriate perturbation size ε in the finite difference procedure is simple in this case, since $\langle g_k, g_k \rangle$ can be used as a rough scale estimate for $\langle \nabla f(x_k), g_k \rangle$. Following Dennis and Schnabel [5] we write

$$\langle \nabla f(x_k), g_k \rangle \approx \frac{1}{2\varepsilon}(f(x_k + \varepsilon g_k) - f(x_k - \varepsilon g_k)) \tag{55}$$

and chose an ε that we expect will perturb two-thirds of the accurate digits of f. If σ is the relative error in function evaluations, we want $f(x_k + \varepsilon g_k) - f_k \approx \sigma^{1/3}|f_k|$, and an appropriate

ε is then

$$\varepsilon = \sigma^{1/3} |f_k| / \langle g_k, g_k \rangle. \tag{56}$$

In addition to the possibility of estimating $\langle \nabla f(x_k), g_k \rangle$ via (55), some applications may admit a direct computation of this quantity. Whenever the action of f' on a single vector can be computed with less expense and/or more accuracy than g_k, such an approach should be considered.

We wish to reemphasize that the convergence results using (12) are significantly weaker than the results using (13), and that (13) should be used whenever possible. If (12) is used, then whenever g_k becomes sufficiently close to zero to trigger the convergence tests used by the algorithm, it should be recomputed to the maximum attainable accuracy to affirm that $\nabla f(x_k)$ has also converged to zero.

5 Extensions To Theory

The conditions for allowable error in our algorithm have been formulated to be as simple and lucid as possible, and we again point out that the upper limits (14) and (15) are exceptionally mild. Due to the very broad range of potential applications, however, we should consider whether any of the details of our theory can be further relaxed.

One extension that can be made to our theory is to enforce (7) only in expectation rather than at every iteration. Similarly, condition (8) could be enforced only in the limit provided the set Ω is large enough to include all the iterates (some of which may be uphill from $f(x_0)$.) Such a result is quite reasonable in that function values are only used to update the trust radius Δ_k, and mistakes in this procedure can be tolerated as long as they balance out in the long run. Cognizant of the practical fact that computed error estimates in a simulation may occasionally be very poor, such a stochastic theory might seem attractive, but we prefer not to include it in this paper because it adds little insight to the analysis.

In contrast to the situation for function evaluations, a single sufficiently bad gradient evaluation can cause the algorithm to fail. For example, if $g_k = -\nabla f(x_k)$ at some iteration k,

then our algorithm can decrease Δ_k indefinitely without ever finding an acceptable step. One might speculate that a condition such as $\langle g_k, \nabla f(x_k) \rangle > 0$ might be sufficient to guarantee an acceptable step will eventually be found (assuming for the moment that function values are computed exactly), but this turns out to be slightly too general a condition. As pointed out in [3], the approximation $g_k = \frac{4}{\eta_1} \nabla f(x_k)$ may also cause the algorithm to fail. What then is the most general condition guaranteeing an acceptable step will always be found? Taking $g_k, \nabla f(x_k)$, and B_k to be fixed, taking $\zeta_{f,1} = \zeta_{f,2} = 0$, and following the same general approach as the proof Theorem 3.3, it is straightforward to establish that $\lim_{\Delta_k \to \infty}(1 - \rho_k) = \lim_{\Delta_k \to 0} \frac{\langle e_k, g_k \rangle}{\langle g_k, g_k \rangle}$. Hence our algorithm is assured of finding an acceptable step for sufficiently small Δ_k if and only if g_k is in the interior of $S_p(1 - \eta_1, \nabla f(x_k))$. Note that $S_p(1 - \eta_1, \nabla f(x_k))$ is a sphere with center $\frac{1}{2\eta_1} \nabla f(x_k)$ and diameter $\frac{1}{\eta_1} \| \nabla f(x_k) \|$, and recall that a typical value for η_1 is 0.001. Our "worst case" limit for a single bad gradient evaluation is therefore only slightly more restrictive (from a practical point of view) than the requirement $\langle g_k, \nabla f(x_k) \rangle > 0$. Of course, one should still strive for a more accurate approximation satisfying $g_k \in S_p(1 - \eta_2, \nabla f(x_k))$ or $g_k \in S_r(1 - \eta_2, \nabla f(x_k))$, but it is reassuring to know that the algorithm has such a large margin for recovery from occasional mistakes in gradient error control.

A final slight extension to our theory comes from the observation that Theorem 3.4 is proven using only the bound $\zeta_g < 1$ rather than $\zeta_g < 1 - \eta_2$. Hence, we could obtain our same strong convergence results $\lim_{k \to \infty} \| g_k \| = \lim_{k \to \infty} \| \nabla f(x_k) \| = 0$ using two different bounds in (12) and (13). That is, we could require $g_k \in S_p(\zeta_{g,1}, \nabla f(x_k)) \bigcap S_r(\zeta_{g,2}, \nabla f(x_k))$ with $\zeta_{g,1} < 1 - \eta_2$ and $\eta_{g,2} < 1$. This is a slightly larger set of admissible approximations than $S_r(\zeta_{g,1}, \nabla f(x_k))$.

6 Summary

Using the error conditions (7), (8), and (12), we have established the result $\liminf_{k \to \infty} \| g_k \| = 0$ for our trust region algorithm using very mild assumptions. In particular, the requirement that $\zeta_g + \zeta_{f,1} \leq 1 - \eta_2$ is exceptionally generous, as $\zeta_g = 0.5$ would typically correspond

to only one significant bit in each component of g_k for $H = \Re^n$. Condition (12) may also be automatically satisfied if g_k is computed by a projection technique such as a Galerkin method.

If error estimates for g_k are not available through other means, (12) can be evaluated through a one-dimensional finite difference test such as (53). In some applications, a separate numerical formulation might allow the action of f' on g_k to be computed directly. Besides allowing us to evaluate error condition (12), such approaches may allow us to improve each approximate gradient using (51) provided our estimate of $\langle \nabla f(x_k), g_k \rangle$ is accurate enough.

The stronger convergence result $\lim_{k\to\infty} \|g_k\| = \lim_{k\to\infty} \|\nabla f(x_k)\| = 0$ can be obtained by using condition (13) rather than (12). We recommend that this condition be used whenever possible. If a scaled version of the trust region algorithm is being used, gradient errors *must* be measured in the norm induced by the rescaling.

References

[1] T.M. Apostol. *Mathematical Analysis.* Addison-Wesley, Reading, Massachusetts, 1957.

[2] R.G. Carter. Safeguarding Hessian approximations in trust region algorithms. Technical Report TR87-06, Rice University, Dept. of Mathematical Sciences, Revised October 1988.

[3] R.G. Carter. On the global convergence of trust region algorithms using inexact gradient information. Technical Report TR87-12, Rice University, Dept. of Mathematical Sciences, Revised April 1989.

[4] R.G. Carter. Numerical experience with a class of algorithms for nonlinear optimization using inexact function and gradient information. Technical Report 89-46 , Institute for Computer Applications in Science and Engineering, September 1989.

[5] J.E. Dennis Jr. and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[6] D.M. Gay. Computing optimal locally constrained steps. *SIAM J. Sci. Statist. Comput.*, 2:186–197, 1981.

[7] A.C. Hindmarsh. ODEPACK, a systematized collection of ODE solvers. In *Scientific Computing*, R. S. Stepleman et al, editor, pages 55–64. North-Holland, Amsterdam, 1983.

[8] J.D. Lambert. *Computational methods in ordinary differential equations.* John Wiley and Sons, New York, 1973.

[9] D.G. Luenberger. *Optimization by vector space methods.* John Wiley and Sons, New York, 1969.

[10] J.N. Lyness. Remarks about performance profiles. Technical Memorandum 369, Applied Mathematics Division, Argonne National Laboratory, 1981.

[11] M. Minkoff. Approaches to optimization/simulation problems. *Appl. Numer. Math.*, 3:453–466, 1987.

[12] J.J. Moré. Recent developments in algorithms and software for trust region methods. In *Mathematical Programming: State of the Art*, A. Bachem, M. Grőtschel, and B.Korte, editors, pages 258–287. Springer Verlag, Berlin, 1983.

[13] J.J. Moré and D.C. Sorensen. Newton's method. Technical Report ANL-82-8, Argonne National Labs, 1982.

[14] J.J. Moré and D.C. Sorensen. Computing a trust region step. *SIAM J. Sci. Statist. Comput.*, 4:553–572, 1983.

[15] M.J.D. Powell. A new algorithm for unconstrained optimization. In *Nonlinear Programming*,J.B. Rosen, O.L. Mangasarian, and K. Ritter, editors, pages 31–65. Academic Press, London, 1970.

[16] M.J.D. Powell. Some global convergence properties of a variable metric algorithm without exact line searches. In *Nonlinear Programming*, R. Cottle and C. Lemke, editors, pages 53–72. AMS, Providence, Rhode Island, 1976.

[17] M.J.D. Powell. On the global convergence of trust region algorithms for unconstrained minimization. *Math. Prog.*, 29:297–303, 1984.

[18] G.A. Schultz, R.B. Schnabel, and R.H. Byrd. A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties. *SIAM J. Numer. Anal.*, 22:47–67, 1985.

[19] Ph. L. Toint. Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space. Technical Report 87/6, Department of Mathematics, Facultés Universitaires ND de la Paix, Namur, Belgium, 1987.

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-181926 ICASE Report No. 89-45		

4. Title and Subtitle	5. Report Date
NUMERICAL OPTIMIZATION IN HILBERT SPACE USING INEXACT FUNCTION AND GRADIENT EVALUATIONS	June 1989
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Richard G. Carter	89-45
	10. Work Unit No.
	505-90-21-01

9. Performing Organization Name and Address	11. Contract or Grant No.
Institute for Computer Applications in Science and Engineering Mail Stop 132C, NASA Langley Research Center Hampton, VA 23665-5225	NAS1-18605
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Langley Research Center Hampton, VA 23665-5225	Contractor Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Langley Technical Monitor:
Richard W. Barnwell

SIAM J. on Control and Optimization

Final Report

16. Abstract

Trust region algroithms provide a robust iterative technique for solving non-convex unstrained optimization problems, but in many instances it is prohibitively expensive to compute high accuracy function and gradient values for the method. Of particular interest are inverse and parameter estimation problems, since function and gradient evaulations involve numerically solving large systems of differential equations.

We present global convergence theory for trust region algorithms in which neither function nor gradient values are known exactly. The theory is formulated in a Hilbert space setting so that it can be applied to variational problems as well as the finite dimensional problems normally seen in trust region literature. The conditions concerning allowable error are remarkably relaxed: relative errors in the gradient error condition is automatically satisfied if the error is orthogonal to the gradient approximation. A technique for estimating gradient error and improving the approximation is also presented.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
nonconvex unconstrained optimization, inexact functions, inexact gradients, trust region methods, global convergence	64 - Numerical Analysis Unclassified - Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	24	A03

NASA FORM 1626 OCT 86