NUMERICAL EXPERIENCE WITH A CLASS OF ALGORITHMS FOR NONLINEAR OPTIMIZATION USING INEXACT FUNCTION AND GRADIENT INFORMATION

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
Numerical Experience With a Class of Algorithms For Nonlinear Optimization Using Inexact Function and Gradient Information

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

Abstract

For optimization problems associated with engineering design, parameter estimation, image reconstruction, and other optimization/simulation applications, low accuracy function and gradient values are frequently much less expensive to obtain than high accuracy values. We investigate the computational performance of trust region methods for nonlinear optimization when high accuracy evaluations are unavailable or prohibitively expensive, and confirm earlier theoretical predictions than the algorithm is convergent even with relative gradient errors of 0.5 or more. The proper choice of the amount of accuracy to use in function and gradient evaluations can result in orders-of-magnitude savings in computational cost.

*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

Consider the nonlinear optimization problem

\[
\text{minimize } f(x) \\
\text{subject to } x \in \mathbb{R}^n
\]  

where the function \( f \) has gradient \( \nabla f \), with \( \nabla f \) assumed to be Lipschitz continuous. We are concerned with numerically solving this problem when function and gradient values are not known exactly.

Problems of this nature frequently occur in engineering design, parameter estimation, and many other situations. Consider, for example, the design of a heat sink for transferring excess heat away from an electronic component. Given the geometry of the sink (expressed, perhaps, as the spacing, thickness, and length of each cooling fin) and the heat flux from the component, one can mathematically model the temperature distribution in the sink and the surrounding medium by a system of partial differential equations. If we wish to find the design geometry which minimizes the time-averaged temperature of the component, we must numerically solve this system of PDE's at each iteration of the optimization algorithm to determine a value for the objective function \( f \). Furthermore, a value for the gradient of \( f \) must be computed at each iteration either through successively perturbing each component of \( x \) and recomputing \( f \) to obtain a finite difference approximation, or through directly solving a larger system of PDEs. Clearly, exact function and gradient values are not attainable, and the computational expense of any approximation at a given iteration increases very rapidly as the required accuracy is increased. Let \( h \) be the discretization mesh size selected and \( m_d \) be the number of spatial dimensions in the PDE, so that the number of elements in the discretization is given by \( k = o((1/h)^{m_d}) \). If the order of the solution method is given by error = \( o(h^{m_o}) \), and the computational expense for the linear algebra associated with the problem is CPU = \( o(k^{m_l}) \), we have
CPU/iteration = \(o(error^{-l})\) \quad (2)

for \(l = \frac{m_d m_l}{m_0}\). Although this estimate is admittedly crude, it seems to hold for many applications and indicates that computational expense per iteration can rise extremely rapidly as more accurate solutions are required. In our example, if a full three-dimensional model of the sink is being used with a direct linear algebra solver and an \(o(h^2)\) solution method, we would have \(m_d = 3, m_l = 3, m_0 = 2\); hence \(l\) would be 3 and a thousand-fold increase in computational time would be needed to increase the accuracy of any given approximation by one digit. Problems involving systems of ODEs tend to be more benign with much smaller values of \(l(\frac{1}{4} \text{ to} \frac{1}{10})\), but computational expense still increases geometrically with accuracy.

Since low accuracy function and gradient evaluations can be orders of magnitude less expensive than high accuracy evaluations, it behooves us to consider optimization algorithms that do not require the maximum possible accuracy at each iteration. Trust region methods are a natural candidate for investigation because of their reputation for robustness and efficiency. A number of authors have established global convergence results for trust region methods using inexact gradients ([12], [3], [17]), and inexact function evaluations have also been treated [4]. In this paper, we investigate the numerical behavior of the algorithms presented in [3] and [4] in order to answer the following questions.

1. How much error can we allow in our evaluations before the algorithm fails? Does this level agree with theoretical predictions?

2. The performance of the algorithm will certainly decrease when less accurate evaluations are used. How fast does performance degrade and how problem-dependent is the rate of degradation?

3. How does this lessened performance balance with the greatly decreased computational cost associated with less accurate evaluations?

The remainder of this paper is organized as follows.

In Section 2, we present the trust region algorithm and review the conditions on permissible levels of error established in [3]. These conditions depend on some of the parameters of the trust region method but are remarkably relaxed: typically relative errors in the gradient of 0.5 or more are permissible. In Section 3, we examine the performance of the algorithm on the set of standard test problems from Moré, Garbow, and Hillstrom [13] when synthetically generated errors are added to the gradients at each iteration. Our results confirm the theoretical predictions for the algorithm, and we note that the number of iterations required by the algorithm tends to increase exponentially with the relative error induced in the gradient. When balanced against (2), however, we find that allowing low accuracy evaluations is still attractive. In Section 4, we examine the performance of the algorithm on a parameter identification problem found in the literature in order to confirm our results without resorting to synthetically induced errors or invoking (2). We also verify a technique for estimating and controlling error when the gradient is approximated by finite differences. Section 5 summarizes our results.

2 The trust region algorithm and permissible error in function and gradient evaluations

The trust region method for solving (1) generates a sequence of iterates $\{x_k\}$ by approximately solving a sequence of constrained quadratic model problems. Each local quadratic model is of the form

$$\Psi_k(x_k + s) = f_k + g_k^T s + \frac{1}{2} s^T B_k s$$

(3)
where \( f_k \) is an approximation to \( f(x_k) \), \( g_k \) is an approximation to the gradient \( \nabla f(x_k) \), and the symmetric matrix \( B_k \in \mathbb{R}^{n \times n} \) is an approximation to the Hessian matrix \( \nabla^2 f(x_k) \). At each iteration, we take \( x_{k+1} = x_k + s_k \), where \( s_k \) is an approximate solution to the trust region subproblem

\[
\minimize \quad \Psi_k(x_k + s) \quad s \in \mathbb{R}^n
\]

The positive scalar \( \Delta_k \) is known as the trust radius, and the nonsingular matrix \( D_k \in \mathbb{R}_N \) is the scaling or preconditioning matrix (often taken to be a fixed diagonal matrix). At each iteration, \( \Delta_k \) is adjusted so that the ball \( \|D_k s_k\| \leq \Delta_k \) represents the region over which we expect (3) to adequately model the function \( f \).

A number of techniques are available for computing an approximate solution to (4). An excellent survey of the main classes of these methods can be found in [12]. In our computations, we chose to use an optimal locally constrained, or OLC technique [6]. Similarly, a number of techniques can be used to generate the Hessian approximation \( \{B_t\} \), but we selected the popular BFGS secant update

\[
B_{k+1} = B_k + \frac{(g_{k+1} - g_k)(g_{k+1} - g_k)^T}{(g_{k+1} - g_k)^T s_k} - \frac{B_k s_k s_k^T B_k^T}{s_k^T B_k s_k}
\]

when \( (g_{k+1} - g_k)^T s_k \geq 10^{-6}(g_{k+1} - g_k)^T (g_{k+1} - g_k) \), and \( B_{k+1} = B_k \) otherwise. Since \( g_k \) is only an approximation to \( \nabla f(x_k) \), [4] and [3] suggest that upper bounds of the form

\[
\|B_k\| \leq c_1
\]

or

\[
g_k^T B_k g_k \leq c_1 g_k^T g_k
\]

be directly enforced for some appropriately large \( c_1 \). This could be easily done by using the replacement operation
\[ B_k := \min\{1, \frac{c_1}{\|B_k\|}\} B_k \quad \text{(8)} \]

at each iteration. Although bounds such as (6) and (7) are needed to establish convergence results, in practice we found (8) unnecessary.

A simple version demonstrating the salient features of the trust region algorithm is as follows.

**Algorithm (1): The trust region method.**

Let the constants \(0 < \eta_1 < \eta_2 < \eta_3 < 1\) be prespecified. Select an initial guess \(x_0 \in \mathbb{R}^N\) and an initial trust radius \(\Delta_0\). Compute \(f_0\) and \(g_0\), and compute or initialize \(B_0\). For \(k = 0, 1, \ldots\) until "convergence" do:

(a) Determine an approximate solution \(s_k\) to problem (4).

(b) Calculate the predicted function reduction.

\[ \text{pred}_k(s_k) = -g_k^T s_k - \frac{1}{2} s_k^T B_k s_k \quad \text{(9)} \]

and the computed function reduction

\[ \text{cred}_k(s_k) = f_k - f_{k+1}. \quad \text{(10)} \]

If necessary, recompute \(f_{k+1}\) and/or \(f_k\) to greater accuracy.

(c) Compute the ratio

\[ \rho_k = \frac{\text{cred}_k(s_k)}{\text{pred}_k(s_k)}. \quad \text{(11)} \]

(d) If \(\rho_k < \eta_1\), then the step is unacceptable. Set \(\Delta_k := \frac{1}{10} \Delta_k\) and return to (a).
(e) If \( \eta_1 \leq \rho_k < \eta_2 \), then set \( \Delta_{k+1} = \frac{1}{2} \Delta_k \).

Else if \( \eta_3 < \rho_k \leq (2 - \eta_3) \), then set \( \Delta_{k+1} = 2 \Delta_k \).

Else set \( \Delta_{k+1} = \Delta_k \).

(f) Set \( x_{k+1} = x_k + s_k \), compute \( g_{k+1} \), and compute or update \( B_{k+1} \).

End loop

End algorithm

Typical values for the step acceptance/trust radius update parameters are \( \eta_1 = 0.001 \), \( \eta_2 = 0.1 \), and \( \eta_3 = 0.75 \). Notice that no step is accepted unless \( \rho_k \geq \eta_1 \), and that the trust radius is never reduced unless \( \rho_k \leq \eta_2 \). Further notice that \( g_k \) is only computed once per major iteration.

Two conditions are required of the approximate function values. Define

\[
\text{ared}_k(s_k) = f(x_k) - f(x_k + s_k).
\]

We then require

\[
|\text{ared}_k(s_k) - \text{cred}_k(s_k)| \geq \zeta_{f,1} \text{pred}_k(s_k)
\]

and

\[
|\text{ared}_k(s_k) - \text{cred}_k(s_k)| \leq \zeta_{f,2} |\text{cred}_k(s_k)|
\]

for some constants \( \zeta_{f,1} \) and \( \zeta_{f,2} \). A stronger variation of these conditions that is typically more practical is

\[
|f_{k+1} - f(x_{k+1})| + |f_k - f(x_k)| \leq \zeta_{f,1} \text{pred}_k(s_k)
\]

and

\[
|f_{k+1} - f(x_{k+1})| + |f_k - f(x_k)| \leq \zeta_{f,2} |\text{cred}_k(s_k)|
\]
Assuming error estimates are available for $|f_k - f(x_k)|$, [3] suggests that (15) and (16) be enforced by using the following procedure in place of step (b) of Algorithm (1).

**Procedure 2**

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

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

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

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

(iv) Compute $\text{cred}_k(s_k)$. If (16) is satisfied, then exit procedure, else reduce $\text{emax}$ and return to (ii).

**End procedure**

The permissible level of error in the gradient evaluations can be characterized in two different ways. The preferable condition is

$$\frac{\|D_k^{-1}e_k\|}{\|D_k^{-1}g_k\|} \leq \zeta_g$$

for some constant $\zeta_g$, with

$$e_k = g_k - \nabla f(x_k).$$

Under appropriate assumptions on $f$ and $\{D_k\}$, equation (17) leads to the strong global convergence result $\lim_{k \to \infty} \|g_k\| = \lim_{k \to \infty} \|\nabla f(x_k)\| = 0$. The weaker convergence result $\liminf_{k \to \infty} \|g_k\| = 0$ can be obtained using the condition

$$\frac{(D_k^{-1}e_k)^T (D_k^{-1}g_k)}{(D_k^{-1}g_k)^T (D_k^{-1}g_k)} \leq \zeta_g$$

(19)
If each scaling matrix $D_k$ is taken to be the identity, (17) and (19) become simply

\[
\frac{\|e_k\|}{\|g_k\|} \leq \zeta_g
\]  

(20)

and

\[
\frac{e_k^T g_k}{g_k^T g_k} \leq \zeta_g.
\]  

(21)

Finally, we must specify the allowed values for the error bounds $\zeta_{f,1}, \zeta_{f,2}$ and $\zeta_{f,3}$. These values are given by the inequalities

\[
\zeta_g + \zeta_{f,1} < 1 - \eta_2
\]  

(22)

and

\[
0 \leq \zeta_{f,2} < 1,
\]  

with $\zeta_g \geq 0$ and $\zeta_{f,1} \geq 0$. These limits are remarkably generous. If $\eta_2$ (the parameter controlling trust radius reduction) is 0.1, we could select $\zeta_{f,1} = 0.05, \zeta_{f,2} = 0.99$, and $\zeta_g = 0.8$ --- less than one significant bit of accuracy in the components of the gradient approximation.

3 Algorithm Performance on Moré-Garbow-Hillstrom Test Problems With Synthetically Induced Errors

The Moré-Garbow-Hillstrom test set [13] contains eighteen "typical" optimization problems. These problems are all algebraically defined, and thus function and gradient values are both inexpensive to compute and available to high accuracy. In order to test the effects of gradient error, we synthetically induced a random error into each gradient computed. More specifically, we computed a vector $w_k$ with each component selected from a uniform distribution on $[-1,1]$ and then set $g_k = \nabla f(x_k) + e_k$ with

\[
e_k = 100 * w_k * \|\nabla f(x_k)\|/2^m,
\]  

(24)
where \( m \) is the smallest positive integer for which (20) is satisfied. For our first set of tests no errors were induced in the computation of function values.

Our optimization code, we used an implementation of the Dennis-Schnabel routines for an OLC step computation technique. The model Hessians \( B_k \) were computed using standard BFGS procedure (5). We emphasize that no special techniques were used to control gradient error when computing \( B_k \), nor were safeguards such as (8) enforced. The value for \( \eta_2 \) in the optimization code was 0.1, so that the theory in [4] and [3] suggests a limit of 0.9 for \( \zeta_g \).

For each problem in the test set, we considered twenty different values of \( \zeta_g \) ranging from 0.1 to 0.95. For each value of \( \zeta_g \), we ran the algorithm with a number of different realizations for the random number generator used to compute \( e_k \) values, and tabulated the minimum, median, and maximum number of iterations required to converge to a local minimum. In all, over 5000 test cases were run on a network of SUN 3/50 workstations in AN 77 double precision.

Figures (1) through (5) show plots of our results for selected problems. Let \( K(\zeta_g) \) denote the number of iterations required for convergence as a function of relative gradient error. The vertical axis in each plot represents the natural log of \( K(\zeta_g) \), while the three traces in each plot represent the observed minimum, median, and maximum values.

Figures (1) through (3) are typical of most of our test cases, in which performance degrades exponentially:

\[
K(\zeta_g) \approx K(0) \exp(b\zeta_g), \tag{25}
\]

with observed decay coefficients \( b \) ranging from roughly 2 to 6 for the various problems. Figures (4) and (5) represent anomalous cases where \( \ln(K(\zeta_g)) \) has significant variation from linearity at small and large values of \( \zeta_g \).

Although the exponential performance degradation given by (25) is a telling argument

\(^1\) Depending on the computational expense associated with a problem and the observed variability of results, between 5 and 100 test cases were run for each value of \( \zeta_g \).
Figure 1: Number of iterations versus $\zeta_g$ for the Watson Test Function.

Figure 2: Number of iterations versus $\zeta_g$ for the Brown and Dennis Test Function.
Figure 3: Number of iterations versus $\zeta_s$ for the Extended Powell Singular Test Function.

Figure 4: Number of iterations versus $\zeta_s$ for the Gaussian Test Function.
against using low accuracy gradients if high accuracy evaluations are obtainable without greatly increased computational expense, low accuracy evaluations are still attractive in cases where the computational expense increases rapidly with increasing accuracy. Suppose, for instance, that

\[
\text{CPU/iteration} \approx c_2 \zeta^{-l}
\]

for some constants \(c_2\) and \(l\). The total computational expense for solving a given problem will then be proportional to \(\zeta^{-l}K(\zeta)\). Figures (6) through (10) show the predicted total computational cost of the median curves for \(K(\zeta)\) in figures (1) through (5) for the values \(l = \frac{1}{2}, 1,\) and \(2\) (with each curve normalized so that the minimum value is 1.0).

Notice that each curve of total computational cost increases very rapidly as \(\zeta \to 0\) or \(\zeta \to 1\), and has a relatively large, flat minima. This behavior holds for both the "typical" cases (figures (6) through (8)) and the anomalous cases (figures (9) and (10)). Interestingly,
Figure 6: Computational cost profiles for the Watson Test Function.

Figure 7: Computational cost profiles for the Brown and Dennis Test Function.
Figure 8: Computational cost profiles for the Extended Powell Singular Test Function.

Figure 9: Computational cost profiles for the Gaussian Test Function.
combining the idealizations (25) and (26) into

$$\text{CPU} \approx K(0)c_\delta \zeta_\delta^{-1}\exp(b\zeta_\delta)$$  \hspace{1cm} (27)

yields the theoretical "best" value

$$\zeta_\delta^* = \frac{l}{b}. \hspace{1cm} (28)$$

Using the observed "typical" value $b = 4$ yields the rule of thumb choice

$$\zeta_\delta^* = \frac{l}{4}, \hspace{1cm} (29)$$

which worked quite well for all of our test problems.

A number of variations on these numerical tests were also tried. Rather than (26) we also considered the idealization
CPU/iteration $\approx c_2 \zeta_g^{-i} + c_3,$ \hfill (30)

where $c_3$ represents a fixed "overhead" cost per iteration. For moderate values of $c_3,$ the character of the overall computational cost behavior remained unchanged. Errors in function values were also considered with $\zeta_g$ fixed at 0.1. The algorithm proved to be quite insensitive to these errors for $\zeta_{f,1} < 0.5.$ Indeed, the algorithm works in practice even if $\zeta_{f,1} > 1$ provided the average value of $|\text{cred}_k(s_k) - \text{cred}_k(s_k)| / \text{pred}_k(s_k)$ is sufficiently less than $1 - \eta_2 - \zeta_g.$ As pointed out in [4], this is a very reasonable result since function values are only used to update the trust radii, and a mistake at any given iteration will not cause the algorithm to fail.

4 Algorithm performance on a parameter identification problem

The test problems of the last section are widely recognized as representing "typical" optimization problems, and because they were algebraically defined, we were able to run an enormous number of test cases to examine the ranges of possible behavior in the presence of errors. It should be remembered, however, that our interpretation of these results rests both on the character of the synthetic noise added to each gradient evaluation and on idealization (26). In order to verify our results, we also tested our algorithm on the following parameter identification problem from [10] and [11].

Consider the accidental release of the radioactive gas tritium into an enclosure surrounding a nuclear reactor. The tritium will react with water vapor in the containment to produce other tritium-based species via

$$T_2 + H_2O \rightleftharpoons HT + HTO,$$ \hfill (31)

and some of the HTO may be adsorbed into the surface of the containment. This adsorbed tritium species represents a significant clean-up problem.
Given the reaction rate constant in (31) and the adsorption and release rates of HTO on the surfaces of the containment, the physical problem can be modeled by a system of four coupled initial value ODEs:

\[ Y'(t; x) = h(t, Y(t; x)), Y(0; x) = Y_0, \]  

with the components \( Y : \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{R}^4 \) being species concentrations. Unfortunately, the rate constants are not directly measurable. Maroni et al performed an experiment in which a known amount of tritium was introduced into a small enclosure and the total tritium concentration \( Y_1 + Y_2 + Y_3 \) was measured at \( m \) discrete time points. The rate constants \( x_1, x_2, \) and \( x_3 \) can then be estimated by minimizing \( f : \mathbb{R}^4 \rightarrow \mathbb{R} \) with

\[
f(x) = \frac{1}{2} \sum_{i=1}^{m} \left( \sum_{j=1}^{3} Y_j(t_i, x) - O_i + x_4^4 \right)^2 / (O_i)^2, \]  

where \( O_i \) is the observed experimental concentration

\[ O_i = Y_1(t_i) + Y_2(t_i) + Y_3(t_i), \]  

and \( x_4^4 \) is an additional variable representing an unknown experimental bias in the instrument for measuring \( O_i \).

Equation (33) is a classical inverse problem. Note that each function evaluation for a given iterate \( x_k \) involves the numerical solution of four coupled ODEs. Gradient values can be computed via finite differences, or by the numerical solution of a system of sixteen coupled ODEs. Although the latter technique is usually preferable in practice, we used the more difficult approach of estimating \( g_k \) by finite differences so that we could investigate the techniques suggested in [3] for estimating and controlling gradient error.

Our numerical experiments with (33) were designed as follows. In order to approximate \( f \) at a given point \( x_k \), (32) was solved using ODEPACK [7], which uses an adaptive solution technique. An important feature of ODEPACK is that it allows a desired level of accuracy (either absolute or relative) to be prespecified for each component of \( Y \). In order to achieve
an accuracy of, say, $\varepsilon_k$ in $f(x_k)$, we specified a desired accuracy of $\alpha_{k-1} \varepsilon_k$ in each component of $Y$, where $\alpha_{k-1}$ was the amount of accuracy lost due to cancellation in evaluating (33) at the last iteration. This simple procedure worked remarkably well: the actual error in $f_k$ was typically in the range $[1/10 \varepsilon_k, 2 \varepsilon_k]$ in our preliminary tests of this technique.

Each gradient was initially approximated by a central difference formula using $2n$ extra function evaluations, where each function evaluation was computed with a specified desired relative accuracy of $\varepsilon_k$. Denote this approximation $\tilde{g}_k$. We then computed a more accurate estimate of the directional derivative of $f$ in the direction $g_k$ (or $(D_k^T D_k)^{-1} \tilde{g}_k$ if the scaling matrix is not the identity) as suggested in [3], using the formula

$$d_k = \frac{1}{2 \delta_k} (f(x_k + \delta_k \tilde{g}_k) - f(x_k - \delta_k \tilde{g}_k)).$$

Each function evaluation in (35) was computed with a specified desired relative accuracy of $\varepsilon_k/10$. The perturbation length $\delta_k$ was taken to be

$$\delta_k = \left( \frac{\varepsilon_k}{10} \right)^{1/2} |f_k|/g_k^T g_k,$$

a value expected to perturb two-thirds of the accurate digits of $f$. Using (18) and (35), we can then estimate the error term in (21) via

$$\frac{e_k^T g_k}{g_k^T g_k} = 1 - \frac{\nabla f(x_k)^T \tilde{g}_k}{g_k^T g_k} \approx 1 - \frac{d_k}{||\tilde{g}_k||^2}.$$

If this error was significantly larger than the desired gradient error level $\zeta_g$, then $\varepsilon_k$ was decreased before the next iteration; if it was significantly smaller $\varepsilon_k$ was immediately decreased and $\tilde{g}_k$ was recomputed (in practice this seldom occurred except at the first gradient computation). Figure (11) shows the agreement between actual and requested gradient error. Even with the simple procedures used to adjust $\varepsilon_k$, the error estimate given by (35) and (37) allows us to control, with reasonable certainty, the level of accuracy in $\tilde{g}_k$.

The approximation $\tilde{g}_k$ can be further improved at no additional cost by setting
Figure 11: Actual versus Commanded error in the gradient for the Tritium Parameter Estimation Problem.

\[ g_k = \frac{d_k}{\|g_k\|^2} g_k \]  \hspace{1cm} (38)

so that

\[ \frac{\epsilon_k^T g_k}{g_k^T g_k} = 1 - \frac{\nabla f(x_k)^T g_k}{d_k} \]  \hspace{1cm} (39)

Figure (12) shows the CPU time required to achieve a given level of accuracy using (38) in addition to the previously discussed procedure for adjusting \( \tau_k \). We see that computational expense increases geometrically as accuracy increases.\(^2\)

Given these methods of evaluating \( f_k \) and \( g_k \) to some specified accuracy, we recorded the

\(^2\)The idealised cost profile (26) yields a very close fit to this plot if \( l \) is taken to be \( \frac{1}{10} \). However, tests with different values of \( z_k \) showed that better empirical form this problem is

\[ \text{CPU/gradient evaluation} \approx c_2 \|g_k\|^{-1} \zeta^{-1/10}. \]

Nonetheless, the rule-of-thumb choice (29) with \( l = \frac{1}{10} \) proved to be close to the optimal selection in our numerical tests.
Figure 12: Solution time required for one gradient evaluation in Tritium parameter identification problem.

computational time required to solve (33) for a number of different values of $\zeta_\phi$, and for the following 3 cases:

1. Each $f_k$ value was computed to high relative accuracy ($10^{-8}$), and each $g_k$ value was computed as previously described including the correction (38).

2. Each $f_k$ was computed to high relative accuracy ($10^{-8}$), and each $g_k$ was computed as described previously but without doing correction (38).

3. Each $f_k$ was computed using Procedure 2 with $\zeta_{f,1} = 0.1$ and $\zeta_{f,2} = 0.99$, and each $g_k$ value was computed as previously described including correction (38).

A somewhat different implementation of the trust region method was used rather than the Dennis-Schnabel code used in the last section. First, we included nonnegativity constraints on the first three components of $x$ to be consistent with the physics of the problem. This
Figure 13 shows the results of our tests.

Case 1 was tested for 15 different values of $\zeta_\varphi$. Note that the total computational time required is less than 8000 seconds for $\zeta_\varphi = 0.15$, but rises to almost 16000 and 24000 seconds for $\zeta_\varphi = 1.5 \times 10^{-5}$ and $\zeta_\varphi = 0.25$, respectively. Fewer data were collected for cases 2 and 3, but note that the correction (38) appears to make little difference to the algorithm when $e_k^T g_k / g_k^T g_k$ is small. On the other hand, using Procedure 2 rather than computing each $f_k$ to a fixed accuracy of $10^{-8}$ resulted in a moderately faster algorithm.
In addition to the above 3 cases, a number of other numerical tests were made. Rather than keeping $\zeta_g$ fixed throughout the algorithm, we tried setting

$$\zeta_g^{k+1} = \begin{cases} 1/10 & \text{if } k = 0 \\ \max\{\zeta_g^k/2, 10^{-5}\} & \text{otherwise,} \end{cases}$$ \hspace{1cm} (40)$$

or conversely

$$\zeta_g^{k+1} = \begin{cases} 10^{-5} & \text{if } k = 0 \\ \min\{2\zeta_g^k, 1/10\} & \text{otherwise.} \end{cases}$$ \hspace{1cm} (41)$$

The idea behind (40) is to try to obtain the fast local convergence properties of the BFGS method when accurate gradients are available, while the idea behind (41) is to avoid highly accurate gradient approximations near the solution where they are likely to be most expensive. Interestingly, both of these approaches performed similarly, requiring 8887 and 10647 seconds, respectively.

Although the correction (38) appears to be of little use when $e_k^T g_k / g_k^T g_k$ is already small, it does appear to be useful in preventing the algorithm from failing due to occasionally encountering highly inaccurate gradient evaluations. In tests where a large synthetic error was added to the gradient approximation every $p$ iterations, the algorithm was much more robust when (38) was used (although the algorithm did still have problems involving convergence to a point with $g_k = 0$ and $\nabla f(x_k) \neq 0$, as predicted in [4].) In a similar vein, the gradient accuracy test given by (35) and (37) should be useful in verifying the accuracy of analytically derived gradients.

5 Summary

We have examined the numerical behavior of trust region algorithms for nonlinear optimization when function and gradient values are not computed exactly. This class of algorithms has proven remarkably robust, and can be successfully implemented even with very large errors in the function and gradient evaluations.
In a large number of tests using standard test problems with synthetically induced gradient errors, we observed that the algorithm performance, as measured by the number of iterations required for convergence, tends to degrade exponentially as the relative gradient error increases. This is a telling argument for using accurate evaluations provided they can be obtained at reasonable expense. For many optimization/simulation problems, however, the computational expense of these evaluations rises sharply with increasing accuracy, and low accuracy evaluations are again attractive. A good choice for the amount of relative gradient error allowed in the algorithm can result in orders-of-magnitude savings in computational cost. If (26) holds, then the choice $\zeta_g = 1/4$ was nearly optimal for all of our test problems.

Using a parameter estimation problem based on the numerical solution of a system of ODEs, we tested a technique for estimating and controlling the amount of error in a gradient approximation. This technique was very successful when used in conjunction with the "user specified accuracy" feature in the numerical differential equation solver ODEPACK. Actual computational costs for various values of relative gradient error were examined to confirm the behavior observed in the test problems with synthetically induced errors.

References


Figure 1: Number of iterations versus $\zeta_g$ for the Watson Test Function.

Figure 2: Number of iterations versus $\zeta_g$ for the Brown and Dennis Test Function.
Figure 3: Number of iterations versus $\zeta_g$ for the Extended Powell Singular Test Function.

Figure 4: Number of iterations versus $\zeta_g$ for the Gaussian Test Function.
against using low accuracy gradients if high accuracy evaluations are obtainable without greatly increased computational expense, low accuracy evaluations are still attractive in cases where the computational expense increases rapidly with increasing accuracy. Suppose, for instance, that

\[ \text{CPU/iteration} \approx c_2 \zeta_g^{-l} \]  

(26)

for some constants \( c_2 \) and \( l \). The total computational expense for solving a given problem will then be proportional to \( \zeta_g^{-l}K(\zeta_g) \). Figures (6) through (10) show the predicted total computational cost of the median curves for \( K(\zeta_g) \) in figures (1) through (5) for the values \( l = \frac{1}{2}, 1, \) and \( 2 \) (with each curve normalized so that the minimum value is 1.0).

Notice that each curve of total computational cost increases very rapidly as \( \zeta_g \to 0 \) or \( \zeta_g \to 1 \), and has a relatively large, flat minima. This behavior holds for both the “typical” cases (figures (6) through (8)) and the anomalous cases (figures (9) and (10)). Interestingly,
Figure 6: Computational cost profiles for the Watson Test Function.

Figure 7: Computational cost profiles for the Brown and Dennis Test Function.
Figure 8: Computational cost profiles for the Extended Powell Singular Test Function.

Figure 9: Computational cost profiles for the Gaussian Test Function.
Figure 10: Computational cost profiles for the Trigonometric Test Function.

Combining the idealizations (25) and (26) into

\[ CPU \approx K(0)c_2\zeta_g^{-l}\exp(b\zeta_g) \]  

yields the theoretical "best" value

\[ \zeta_g^* = \frac{l}{b} \]  

Using the observed "typical" value \( b = 4 \) yields the rule of thumb choice

\[ \zeta_g^* = \frac{l}{4} \]  

which worked quite well for all of our test problems.

A number of variations on these numerical tests were also tried. Rather than (26) we also considered the idealization
Figure 11: Actual versus Commanded error in the gradient for the Tritium Parameter Estimation Problem.

\[ g_k = \frac{d_k}{\|g_k\|^2} \quad (38) \]

so that

\[ \frac{e_k^T g_k}{g_k^T g_k} = 1 - \frac{\nabla f(x_k)^T g_k}{d_k}. \quad (39) \]

Figure (12) shows the CPU time required to achieve a given level of accuracy using (38) in addition to the previously discussed procedure for adjusting \( \varepsilon_k \). We see that computational expense increases geometrically as accuracy increases. \(^2\)

Given these methods of evaluating \( f_k \) and \( g_k \) to some specified accuracy, we recorded the

\(^2\)The idealized cost profile (26) yields a very close fit to this plot if \( l \) is taken to be \( \frac{1}{10} \). However, tests with different values of \( x_k \) showed that better empirical form this problem is

\[ \text{CPU/gradient evaluation} \approx c_2 \|g_k\|^{-1} \zeta_{g}^{-1/10}. \]

Nonetheless, the rule-of-thumb choice (29) with \( l = \frac{1}{10} \) proved to be close to the optimal selection in our numerical tests.
Figure 12: Solution time required for one gradient evaluation in Tritium parameter identification problem.

computational time required to solve (33) for a number of different values of \( \zeta_g \), and for the following 3 cases.

1. Each \( f_k \) value was computed to high relative accuracy \( (10^{-8}) \), and each \( g_k \) value was computed as previously described including the correction (38).

2. Each \( f_k \) was computed to high relative accuracy \( (10^{-8}) \), and each \( g_k \) was computed as described previously but without doing correction (38).

3. Each \( f_k \) was computed using Procedure 2 with \( \zeta_{f,1} = 0.1 \) and \( \zeta_{f,2} = 0.99 \), and each \( g_k \) value was computed as previously described including correction (38).

A somewhat different implementation of the trust region method was used rather than the Dennis-Schnabel code used in the last section. First, we included nonnegativity constraints on the first three components of \( x \) to be consistent with the physics of the problem. This
Figure 13: Solution time for numerical optimization of the Tritium parameter identification problem.

was done by replacing the ellipsoidal trust region in (4) with the rectangular trust region \( \|D_k \delta \|_\infty \leq \Delta_k \), and by using a quadratic programming code to exactly solve the trust region subproblem subject to the nonnegativity constraints on \( z \). Second, we used the Hessian safeguarding techniques proposed in [2] in addition to the standard BFGS update (5). Third, we used a simplified stopping criteria for comparison purposes. Each test case was terminated when \( f_k - f^* \leq \frac{1}{100}(f_0 - f^*) \), where \( f^* \) was the optimal value of \( f \).

Figure 13 shows the results of our tests.

Case 1 was tested for 15 different values of \( \zeta_g \). Note that the total computational time required is less than 8000 seconds for \( \zeta_g = 0.15 \), but rises to almost 16000 and 24000 seconds for \( \zeta_g = 1.5 \times 10^{-5} \) and \( \zeta_g = 0.25 \), respectively. Fewer data were collected for cases 2 and 3, but note that the correction (38) appears to make little difference to the algorithm when \( e_k^T g_k / g_k^T g_k \) is small. On the other hand, using Procedure 2 rather than computing each \( f_k \) to a fixed accuracy of \( 10^{-8} \) resulted in a moderately faster algorithm.

34
For optimization problems associated with engineering design, parameter estimation, image reconstruction, and other optimization/simulation applications, low accuracy function and gradient values are frequently much less expensive to obtain than high accuracy values. We investigate the computational performance of trust region methods for nonlinear optimization when high accuracy evaluations are unavailable or prohibitively expensive, and confirm earlier theoretical predictions that the algorithm is convergent even with relative gradient errors of 0.5 or more. The proper choice of the amount of accuracy to use in function and gradient evaluations can result in orders-of-magnitude savings in computational cost.

**Key Words (Suggested by Author(s))**
- nonconvex unconstrained optimization
- inexact functions, inexact gradients
- trust region methods, global convergence

**Distribution Statement**
64 - Numerical Analysis
Unclassified - Unlimited