On the Global Convergence of Trust Region Algorithms Using Inexact Gradient Information

R.G. Carter

Technical Report 87-6, June 1987.

Mathematical Sciences Department, Rice University, Houston, Texas, 77251-1892. Research sponsored by DOE DE-FG05-86ER25017, SDIO/IST/ARO DAAG-03-86-K-0113, and AFOSOR 85-0243.

On the Global Convergence of Trust Region Algorithms Using Inexact Gradient Information

R.G. Carter

Abstract. Trust region algorithms are an important class of methods that can be used to solve unconstrained optimization problems. Moré [10] has proven a global convergence result for a class of trust region methods where the gradient values are approximated rather than computed exactly, provided the approximations are consistent. We show that the assumption of consistency can be replaced by a simple condition on the relative error in the gradient approximation. This new condition has both practical and theoretical advantages. First, it provides a practical test for judging the adequacy of a given gradient approximation, and does not require new approximations to be computed for unsuccessful iterations. Second, it leads to stronger convergence results than obtained in [10].

Key words. Unconstrained optimization, trust region methods, inexact gradients, global convergence.

Acknowledgements. This research was sponsored by DOE DE-FG05-86ER25017, SDIO/IST/ARO DAAG-03-86-K-0113, and AFOSOR 85-0243. I would also like to thank John Dennis and Richard Tapia for their suggestions and support, and Vivian Choi for helping prepare the manuscript.

CONTENTS

1. Introduction

1.1. Trust region algorithms using inexact gradients. This paper considers trust region methods for the solution of the *unconstrained optimization problem*

$$\underset{x}{\text{minimize}}\, f(x), \tag{1.1}$$

with $f:\mathbb{R}^n \to \mathbb{R}^1$. These methods generate 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

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\, \psi_k(x_k + s):\ \|D_k s\| \le \Delta_k \tag{1.2}$$

where $D_k \in \mathbb{R}^{n\times n}$ is a scaling matrix, Δ_k is a positive variable known as the *trust radius*, and ψ_k is a quadratic model of f about the point x_k:

$$\psi_k(x_k + s) = f(x_k) + g_k{}^T s + \tfrac{1}{2} s^T B_k s\ . \tag{1.3}$$

The vector $g_k \in \mathbb{R}^n$ is thus the gradient of ψ_k at x_k and the symmetric matrix $B_k \in \mathbb{R}^{n\times n}$ is the Hessian of ψ_k. Ideally, g_k should be identical to $\nabla f(x_k)$ (the gradient of f at x_k) while B_k should be identical to $\nabla^2 f(x_k)$ (the Hessian of f at x_k), but it may not be practical to compute these quantities exactly.

Strong global convergence results have been shown for trust region algorithms which take $g_k \equiv \nabla f(x_k)$ (see, for example [1], [3], [7], [13], and [14]). If the sequence of Hessian approximations $\{B_k\}$ is uniformly bounded, mild conditions of f and $\{D_k\}$ are sufficient to establish that

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0 \tag{1.4}$$

for most implementations.

Moré [10] considers the global convergence of a class of trust region algorithms in which the condition $g_k \equiv \nabla f(x_k)$ is relaxed. Instead of requiring exact gradient values, Moré allows g_k to be an approximation to $\nabla f(x_k)$ provided the sequence of approximations satisfies the consistency property

$$x_k \to x^* \Rightarrow \lim_{k\to\infty} \|g_k - \nabla f(x_k)\| = 0\ . \tag{1.5}$$

Using (1.5) as a primary assumption, he is able to establish[1]

$$\liminf_{k\to\infty} \|g_k\| = 0 \tag{1.6}$$

but provides no suggestions about how (1.5) should be enforced.

In this paper, we show that the consistency assumption (1.5) can be replaced by a condition on the relative error in the gradient approximation. In simplest form,[2] this condition is

$$\frac{\|g_k - \nabla f(x_k)\|}{\|g_k\|} \le \varsigma \qquad \forall\ k \tag{1.7}$$

for some constant[3] $\varsigma < 1$. Since error estimates are often available when approximate gradients are calculated, (1.7) provides a practical test for judging the adequacy of a given gradient approximation. We argue that this is a more natural approach than trying to enforce (1.5) directly. Furthermore, (1.7) leads to the global convergence result

$$\lim_{k\to\infty} \|g_k\| = 0\,, \tag{1.8}$$

which is much stronger than (1.6) under certain conditions.[4] Moreover, (1.7) and (1.8) imply

$$\lim_{k\to\infty} \|g_k - \nabla f(x_k)\| = 0\,, \tag{1.9}$$

which is an even stronger consistency property than (1.5). Consistency of the gradient approximations is therefore a *consequence* of our theory rather than an assumption.

1.2. Structure of trust region algorithms. Before presenting justification for our claim that (1.7) is a more practical condition to directly enforce than (1.5), the structure of trust region algorithms must be described in more detail. Authors typically describe trust region algorithms by

---

[1] More specifically, he establishes $\liminf_{k\to\infty} \|g_k\|_{(D_k{}^T D_k)^{-1}} = 0$, where the elliptical norm $\|x\|_A$ is defined to be $\|x\|_A \equiv (x^T A x)^{1/2}$ for symmetric positive definite $A \in \mathbb{R}^{n\times n}$. For implementations which require $\|D_k{}^T D_k\| \le \sigma_1{}^2$ and $\|(D_k{}^T D_k)^{-1}\| \le \sigma_2{}^2$ for constants σ_1, σ_2, More's result is thus equivalent to $\liminf_{k\to\infty} \|g_k\| = 0$.

[2] This form is valid for $D_k = I$; slightly different forms of this condition will be used for the more general case $D_k \neq I$.

[3] The value of ς will depend upon some of the other parameters in the trust region implementation, but will typically be about 0.9.

[4] Notice that (1.7) and (1.8) imply $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$, the same strong global convergence result obtained when $g_k \equiv \nabla f(x_k)$. If $\{x_k\}$ converges, then (1.5) and (1.6) also imply this strong result, but if $\{x_k\}$ is unbounded or has more than one limit point, then (1.5) and (1.6) do not even imply $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$.

using either a *single loop* indexing system (see, for example, [10] and [13]) or a *nested loop* indexing system (see, for example [1], [14], and [15].) For methods that take $g_k \equiv \nabla f(x_k)$, the differences between these two indexing systems are purely semantic, but for $g_k \neq \nabla f(x_k)$ it is instructive to consider them separately. The single loop structure used in [10] is as follows.

Algorithm (1). Single loop structure for the trust region method.

Let $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < 1 < \gamma_2$ be prespecified.[5] Select an initial guess $x_0 \in \mathbf{R}^n$ and trust radius $\Delta_0 > 0$. Compute $f(x_0)$, and compute or initialize g_0, B_0, and D_0.

For $k = 0, 1, \ldots$, until "convergence" do:

(a) Determine an approximate solution s_k to problem (1.2).

(b) Compute $\rho_k = (f(x_k) - f(x_k + s_k))/(\psi_k(x_k) - \psi_k(x_k + s_k))$.

(c) If $\rho_k < \eta_1$ then set $s_k = 0$ and $\Delta_{k+1} \in (0, \gamma_1 \Delta_k]$.

(d) If $\eta_1 \leq \rho_k < \eta_2$ then set $\Delta_{k+1} \in [\gamma_1 \Delta_k, \Delta_k]$.

(e) If $\eta_2 \leq \rho_k$ then set $\Delta_{k+1} \in [\Delta_k, \gamma_2, \Delta_k]$.

(f) Set $x_{k+1} = x_k + s_k$ and update g_k, B_k, and D_k.

End loop.

In this structure, trial steps s_k are rejected and the trust radius is reduced if $\rho_k < \eta_1$. Such an iteration is called *unsuccessful* since $x_{k+1} = x_k$; iterations for which $\rho_k \geq \eta_1$ are called *successful*. Clearly, step (c) is designed to prevent an infinite series of unsuccessful iterations, while steps (d) and (e) are designed to pick a trust radius for the next iteration that is small enough to have a good chance of producing a successful step yet large enough to permit rapid convergence.

The structure of Algorithm (1) neither requires nor prohibits updates of g_k, B_k, and D_k at unsuccessful iterations. In implementations which take $g_k \equiv \nabla f(x_k)$, such updates are rarely found

[5] Typical values for these parameters ar $\eta_1 = 0.001$, $\eta_2 = 0.1$, $\gamma_1 = 0.25$, $\gamma_2 = 4.0$.

in the literature.[6] The following algorithm uses a nested loop structure in which the outer loop indexes only successful iterations and updates to g_k, B_k, and D_k are *not* allowed during the inner loop.

Algorithm (2). Nested loop structure for the trust region method.

Let $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < 1 < \gamma_2$ be prespecified. Select an initial guess $x_0 \in \mathbb{R}^n$ and trust radius $\Delta_0 > 0$. Compute $f(x_0)$, and compute or initialize g_0, B_0, and D_0.

For $k = 0, 1, \cdots$ until "convergence" do:

(a) Repeat until $\rho_k \geq \eta_1$:

(a.1) Determine an approximate solution to s_k to problem (1.2)

(a.2) Compute $\rho_k = (f(x_k) - f(x_k + s_k))/(\psi_k(x_k) - (\psi_k(x_k + s_k))$

(a.3) If $\rho_k < \eta_1$, then set $\Delta_k \in (0, \gamma_1 \Delta_k]$.

End loop.

(b) If $\rho_k < \eta_2$ then set $\Delta_{k+1} \in (0, \Delta_k]$, else set $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$.

(c) Set $x_{k+1} = x_k + s_k$ and update g_k, B_k, and D_k.

End loop.

The form of Algorithm (2) raises the possibility that at some iteration k, the inner loop (a) may fail to generate an acceptable new iterate. Consider, for example,[7] an initial gradient approximation $g_0 = -\nabla f(x_0)$ with $B_0 = D_0 = I$. Since every descent direction for f is an ascent direction for ψ_0, ρ_0 will be negative[8] no matter how much Δ_0 is reduced in the inner loop (a). Such failures

[6]This is not to say they are unimportant. The well known algorithm NL2SOL [6] for the solution of the nonlinear least squares problem owes much of its success to its capability of switching between alternate Hessian approximations. Global convergence theory for such switching is given in [1] and [4] in sufficient generality to provide a framework for an expert systems approach to optimization. However, [1], [4], and [6] all take $g_k \equiv \nabla f(x_k)$, which makes the question of updating g_k at unsuccessful iterations moot.

[7]This example is presented in greater detail in Section 3 of this paper.

[8]Although this example depends on the angle between g_k and $\nabla f(x_k)$ being greater than ninety degrees, another example will be presented in Section 3 that demonstrates the possibility of failure in the inner loop even if the angle between

of the inner loop to converge can occur at any iteration unless:

(i) additional conditions are imposed on g_k, or

(ii) g_k is successively improved in the inner loop so that $\lim_{k\to\infty} ||g_k - \nabla f(x_k)|| = 0$.

The latter approach is implicit in the formal statement of More's algorithm, but we prefer imposing the additional condition

$$\frac{||g_k - \nabla f(x_k)||_{(D_k{}^T D_k)^{-1}}}{||g_k||_{(D_k{}^T D_k)^{-1}}} \leq \varsigma. \tag{1.10}$$

We show in this paper that if $\varsigma \in [0, 1-\eta_1)$, then (1.10) is sufficient to ensure the success of the inner loop. If error estimates are available, (1.10) can be checked at the *start* of every iteration k and g_k can be recomputed if necessary. Trying to use approach (ii) so that the analysis of [10] holds is less practical because it involves recomputing g_k with successively greater accuracy as Δ_k decreases in the inner loop without regard for whether error in g_k is the problem or whether Δ_k is really too large. Since unsuccessful steps are quite common even with $g_k \equiv \nabla f(x_k)$ and since recomputing g_k with successively greater accuracy is generally very expensive computationally, this approach is much less satisfying than using (1.10).

Even if all the iterates are acceptable, directly enforcing the consistency condition (1.5) presents a practical difficulty in that no specification is made about how fast to force $\{||g_k - \nabla f(x_k)||\}$ to converge to zero. One might enforce the condition

$$||g_k - \nabla f(x_k)|| \leq c\ ||s_k|| \tag{1.11}$$

for some constant $c \in (0, \infty)$, but (1.5) provides no suggestions for selecting a reasonable value for c. On the other hand, a reasonable value for ς in (1.10) is much easier to select since we show that strong global convergence results can be obtained for any $\varsigma \in [0, 1-\eta_2)$.

1.3. Synopsis. In Section 2 of this paper, we briefly discuss the techniques generally used to compute trial steps for a given model, scaling matrix, and trust radius. In Section 3, we present two detailed examples of how the inner loop of Algorithm (2) can indeed fail to produce a solution.

---

g_k and $\nabla f(x_k)$ is zero.

We then show that condition (1.10) with $\varsigma < 1 - \eta_1$ is sufficient to ensure the success of the inner loop. In Section 4, we show that (1.10) with $\varsigma < 1 - \eta_2$ is sufficient to establish $\liminf_{k \to \infty} ||g_k|| = \liminf_{k \to \infty} ||\nabla f(x_k)|| = 0$ provided $\{B_k\}$, $\{D_k{}^T D_k\}$ and $\{(D_k{}^T D_k)^{-1}\}$ are uniformly bounded. We then demonstrate two ways that the stronger convergence result $\lim_{k \to \infty} ||\nabla f(x_k)|| = 0$ can be obtained using (1.10) with $\varsigma < 1 - \eta_2$ given that $\{B_k\}$ is uniformly bounded and $\{D_k\}$ satisfies some mild assumption. The final section of this paper summarizes our results and suggests some possibilities for future study.

1.4. Nomenclature and standard assumptions. In addition to the notation already introduced, the following definitions and conventions are used throughout this paper. Unless otherwise specified, $||\cdot||$ denotes the Euclidean norm (or the matrix norm induced by the Euclidean norm), while $||x||_A$ is the elliptical norm $(x^T A x)^{\frac{1}{2}}$ for A a symmetric positive definite matrix in $\mathbf{R}^{n \times n}$. A function $h: \mathbf{R}^n \to \mathbf{R}^m$ is said to be *Lipschitz* with constant L in an open convex region Ω if $||h(x) - h(y)|| \leq L ||x - y|| \ \forall \ x, y \in \Omega$. The *level set* of a function f at a point $x_k \in \mathbf{R}^n$ is the set of all $x \in \mathbf{R}^n$ *such that* $f(x) \leq f(x_k)$.

Let Ω be an open convex set containing the level set of f *at* x_0. The function $f: \mathbf{R}^n \to \mathbf{R}$ is said to satisfy the *standard assumptions* if

f *is continuously differentiable on* Ω, (1.12a)

f *is bounded below, and* (1.12b)

∇f *is Lipschitz with constant* L *in* Ω. (1.12c)

It is frequently convenient to represent the trust region subproblem in local coordinates. We define the predicted function reduction $pred_k(s)$ as

$$\begin{aligned} pred_k(s) &\equiv \psi_k(x_k) - \psi_k(x_k + s) \\ &= -g_k{}^T s - \tfrac{1}{2} s^T B_k s , \end{aligned} \tag{1.13}$$

and the actual function reduction $ared_k(s)$ is defined

$$ared_k(s) \equiv f(x_k) - f(x_k + s) . \tag{1.14}$$

Although the notation used in Algorithm (2) is usually convenient, a rigorous treatment of the success or failure of the inner loop requires indexing the trial steps and trial trust radii within the loop. The following algorithm is completely equivalent to Algorithm 2, but introduces some additional nomenclature. Specifically, $\{s^i\}$ represents the complete sequence of trial steps generated and $\{\Delta^i\}$ represents the corresponding trial trust radii, so that $\{s_k\} \subset \{s^i\}$ and $\{\Delta_k\} \subset \{\Delta^i\}$.

Algorithm (3). Trust region method with full notation.

Let $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < 1 < \gamma_2$ be prespecified. Select an initial guess $x_0 \in \mathbb{R}^n$ and trust radius $\Delta^0 > 0$. Compute $f(x_0)$, and compute or initialize g_0, B_0, and D_0. Set $i = -1$.

For $k = 0, 1, \ldots$, until "convergence" do:

(a) Repeat until $\rho^i \geq \eta_1$:

(a.1) Increment i and determine an approximate solution s^i to

$$\underset{s \in \mathbb{R}^n}{minimize}\ \psi_k(x_k + s):\ ||D_k s|| \leq \Delta^i \ . \tag{1.15}$$

(a.2) Compute

$$\rho^i = ared_k(s^i)/pred_k(s^i) \tag{1.16}$$

(a.3) If $\rho^i < \eta_1$ then set $\Delta^{i+1} \in (0, \gamma_1 \Delta^i]$,

Else set $s_k = s^i$, $\Delta_k = \Delta^i$, and $\rho_k = \rho^i$.

End loop.

(b) If $\rho^i < \eta_2$ then set $\Delta^{i+1} \in (0, \Delta^i]$,

Else set $\Delta^{i+1} \in [\Delta^i, \gamma_2 \Delta^i]$.

(c) Set $x_{k+1} = x_k + s_k$ and update g_k, B_k, and D_k.

End loop.

2. Computation of trial steps.

2.1. Introduction. In order to establish our results, we must use several properties satisfied by standard techniques for computing trial steps. This section summarizes these properties, but is not intended to be a comprehensive discussion of methods of step computation. An excellent survey (with an extensive bibliography) of the many step computation strategies is presented in [10]. Readers familiar with these techniques may wish to proceed directly to Section 3.

2.2. Scaling matrices and preconditioning. For any nonsingular $D_k \in \mathbf{R}^{n \times n}$ consider the change of variables

$$\tilde{x} = D_k x \tag{2.1}$$

so that $\tilde{s} = D_k s$ and $\tilde{x}_k = D_k x_k$. Then the definitions

$$\tilde{\psi}_k(\tilde{x}_k + \tilde{s}) \equiv \psi_k(x_k + s)\,, \tag{2.2}$$

$$\tilde{pred}_k(\tilde{s}) \equiv pred_k(s)\,, \tag{2.3}$$

and

$$\tilde{ared}_k(\tilde{s}) \equiv ared_k\,(s) \tag{2.4}$$

lead to

$$\tilde{pred}_k(\tilde{s}) = -\,\tilde{g}_k{}^T \tilde{s} - \tfrac{1}{2}\tilde{s}^T \tilde{B}_k \tilde{s}\,, \tag{2.5}$$

$$\tilde{ared}_k(\tilde{s}) = f(x_k) - f(x_k + D_k{}^{-1}\tilde{s})\,, \tag{2.6}$$

$$\tilde{g}_k \equiv D_k{}^{-T} g_k\,, \tag{2.7}$$

and

$$\tilde{B}_k = D_k{}^{-T} B_k\, D_k{}^{-1}\,. \tag{2.8}$$

In this notation, (1.15) becomes the simpler problem of finding $\tilde{s}^i$ that approximately solves

$$\underset{\tilde{s} \in \mathbf{R}^n}{minimize}\ \tilde{\psi}_k(\tilde{x}_k + \tilde{s}):\ \|\tilde{s}\| \le \Delta^i\,. \tag{2.9}$$

The step s^i can then be recovered by inverting transformation (2.1) to give

$$s^i = D_k{}^{-1}\tilde{s}^i\,. \tag{2.10}$$

Typically, methods for calculating s^i use (2.9) and (2.10) rather than (1.15), although the change of variables (2.1) need not be explicitly performed. Consider the relationship between the *method of steepest descent*, which takes

$$s^i = -\alpha g_k \tag{2.11}$$

for some positive α, and the *preconditioned steepest descent* method, which uses a positive definite *preconditioning matrix* $C_k \in \mathbf{R}^{n \times n}$ and sets

$$s^i = -\alpha C_k^{-1} g_k \tag{2.12}$$

for some positive α. Although this preconditioning does not explicitly use scaling (2.1), applying the method of steepest descent to the scaled problem (2.9) yields

$$\tilde{s}^i = -\alpha \tilde{g}_k \tag{2.13}$$

or

$$s^i = -\alpha (D_k^T D_k)^{-1} g_k \tag{2.14}$$

so that (2.12) implicitly uses a change of variables for which $D_k^T D_k = C_k$.

The matrices D_k are often assumed to be diagonal in trust region literature. Because of the relationship between scaling and preconditioning, we prefer not to make this assumption, as nondiagonal preconditioners are widely used in conjugate direction methods for large scale problems.

2.3. Asymptotic behavior of step directions. The first property that we will need concerns the *direction* that trial steps s^i tend toward as the trial trust radii tend toward zero. This property is, obviously, directly dependent on the method used to compute the trial steps. Let Θ^i be defined to be the angle between $\tilde{s}^i$ and $-\tilde{g}_k$ so that

$$\cos \Theta^i = -(\tilde{s}^i)^T \tilde{g}_k / (\,||\tilde{s}^i||\;||\tilde{g}_k||\,). \tag{2.15}$$

We will show that for the two major classes of solution techniques, if $\Delta^i \to 0$ in the inner loop of Algorithm (3) and $g_k \neq 0$, then $\cos \Theta^i \to 1$. Furthermore, let Θ_k be the angle between $\tilde{s}_k$ and $-\tilde{g}_k$. If an infinite sequence of successful iterates are generated and $\limsup_{k \to \infty} ||\tilde{B}_k|| < \infty$, then $\Delta_k \to 0$ and $\liminf_{k \to \infty} ||\tilde{g}_k|| > 0$ imply that $\cos \Theta_k \to 1$.

One of the major classes of solution methods is based on the following powerful result.[9]

THEOREM 2.1. Let g be a vector in $\mathbf{R}^n$, let $B \in \mathbf{R}^{n \times n}$ be symmetric, and let $D \in \mathbf{R}^{n \times n}$ be nonsingular. A vector $s \in \mathbf{R}^n$ is a global solution to

$$\textit{minimize } g^T s + \tfrac{1}{2} s^T B s : \; ||Ds|| \leq \Delta \tag{2.16}$$

if and only if s and Δ obey the following relations for some $\mu \geq 0$.

$$(B + \mu D^T D)s = -g \,, \tag{2.17}$$

$$||Ds|| \leq \Delta \,, \tag{2.18}$$

$$\mu(\Delta - ||Ds||) = 0 \,, \tag{2.19}$$

and

$$B + \mu D^T D \textit{ is positive semidefinite.} \tag{2.20}$$

Furthermore, if $B + \mu D^T D$ is positive definite, then (2.16) has a unique global solution.

Theorem 2.1 is unusually strong in that it completely characterizes all of the global solutions of problem (2.16), and simultaneously suggests an approach to approximately solving the trust region subproblem for a prespecified Δ^i. Consider any $\mu \geq 0$ which is sufficiently large to make $B + \mu D^T D$ positive semidefinite, and let $s(\mu)$ be a solution to (2.17). Furthermore, define $\Delta(\mu) = ||Ds(\mu)||$ so that (2.18) and (2.19) are satisfied. We see that $s(\mu)$ exactly solves (2.16) for[10]

$\Delta = \Delta(\mu)$ and hence one possible approach to solving the trust region subproblem is to use some sort of procedure to find a μ^i for which $\Delta(\mu^i) \approx \Delta^i$. If, for example, a μ^i is found for which $\Delta(\mu^i) = (1 + \epsilon)\Delta^i$ for some small ϵ, then $s(\mu^i)$ is an *exact* solution to the problem

$$\textit{minimize } g^T s + \tfrac{1}{2} s^T B s : \; ||Ds|| \leq (1 + \epsilon)\Delta^i \,. \tag{2.21}$$

Methods of this type are sometimes called *optimal locally constrained* [8], or *OLC* methods. Such methods approximately solve the trust region subproblem (1.15) by *exactly* solving the *nearby*

[9]This well known result is founded on work done by Goldfeld, Quandt, and Trotter [9], and was first stated in modern form by Gay [8] and Sorensen [15]. The reader is referred to [10] for a more complete history and discussion.

[10]In fact, if $\mu = 0$ and B is symmetric positive definite, $s(\mu)$ exactly solves (2.16) for every $\Delta \geq \Delta(\mu)$. That is, $\mu = 0$ corresponds to the constraint not being binding.

problem (expressed here in scaled form):

$$\text{minimize } \tilde{g}_k \tilde{s} + \tfrac{1}{2} \tilde{s}^T \tilde{B}_k \tilde{s} : \; ||\tilde{s}|| \leq \overline{\Delta}^i \tag{2.22}$$

with

$$c_1 \Delta^i \leq \overline{\Delta}^i \leq c_2 \Delta^i \tag{2.23}$$

for some constants[11] $c_1 \in (0, 1]$ and $c_2 \in [1, 2)$. Trial steps therefore satisfy

$$||\tilde{s}^i|| \leq c_2 \Delta^i \; . \tag{2.24}$$

A large number of efficient techniques can be found in the literature for finding a satisfactory μ^i. Experimental results have been published for several implementations [10] in which the average number of matrix factorizations (of $B + \mu D^T D$) required to find an acceptable μ^i is roughly 1.5.

We have now characterized OLC methods sufficiently to examine the directional behavior of s^i as $\Delta^i \to 0$.

THEOREM 2.2. For $k = 1, 2, \ldots, k_{\max} \leq \infty$, let $\{\tilde{g}_k\}$ be a set of vectors in $\mathbf{R}^n$ and let $\{\tilde{B}_k\}$ be a set of symmetric matrices in $\mathbf{R}^{n \times n}$. Let $\{\Delta^i\}$ be a sequence of positive numbers with $\{\Delta_k\} \subset \{\Delta^i\}$, let $\tilde{s}^i \in \mathbf{R}^n$ be calculated by an OLC method, and define Θ^i to be the angle between the vectors $\tilde{s}^i$ and $-\tilde{g}_k$. We then have the following.

(i) For fixed k, either $\tilde{g}_k = 0$ or

$$\lim_{i \to \infty} \Delta^i = 0 \;\Rightarrow\; \lim_{i \to \infty} \cos(\Theta^i) = 1 \; . \tag{2.25}$$

(ii) Suppose $\{\Delta_k\}$ is an infinite sequence and let $\{\tilde{s}_k\}$ be the subsequence of $\{\tilde{s}^i\}$ associated with $\{\Delta_k\}$. Define Θ_k to be the angle between $\tilde{s}_k$ and $-\tilde{g}_k$. If $\limsup_{k \to \infty} ||\tilde{B}_k|| < \infty$, then either $\liminf_{k \to \infty} ||\tilde{g}_k|| = 0$ or

$$\lim_{i \to \infty} \Delta^i = 0 \;\Rightarrow\; \lim_{k,i \to \infty} \cos(\Theta^i) = 1 \;\Rightarrow\; \lim_{k \to \infty} \cos \Theta_k = 1 \; . \tag{2.26}$$

The proof of this theorem is given in the appendix of this paper since it is rather unenlightening. It should be pointed out that (i) above is well known but is seldom stated in the literature,

[11] A typical choice is $c_1 = 0.9$ and $c_2 = 1.1$.

since standard convergence theory with $g_k \equiv \nabla f(x_k)$ can be established without invoking (2.25).[12]

The major alternative to OLC methods for computing approximate solutions to (2.9) is a class of techniques that we will refer to as *generalized dogleg methods*. The oldest and simplest such method is Powell's *dogleg* algorithm [11]. This method[13] defines a piecewise linear path $\tilde{s}(\alpha)$ starting at $\tilde{s} = 0$, extending to the *Cauchy step*

$$\tilde{s}_k^{cs} = -\frac{\tilde{g}_k^T \tilde{g}_k}{\tilde{g}_k \tilde{B}_k \tilde{g}_k} \tilde{g}_k , \tag{2.27}$$

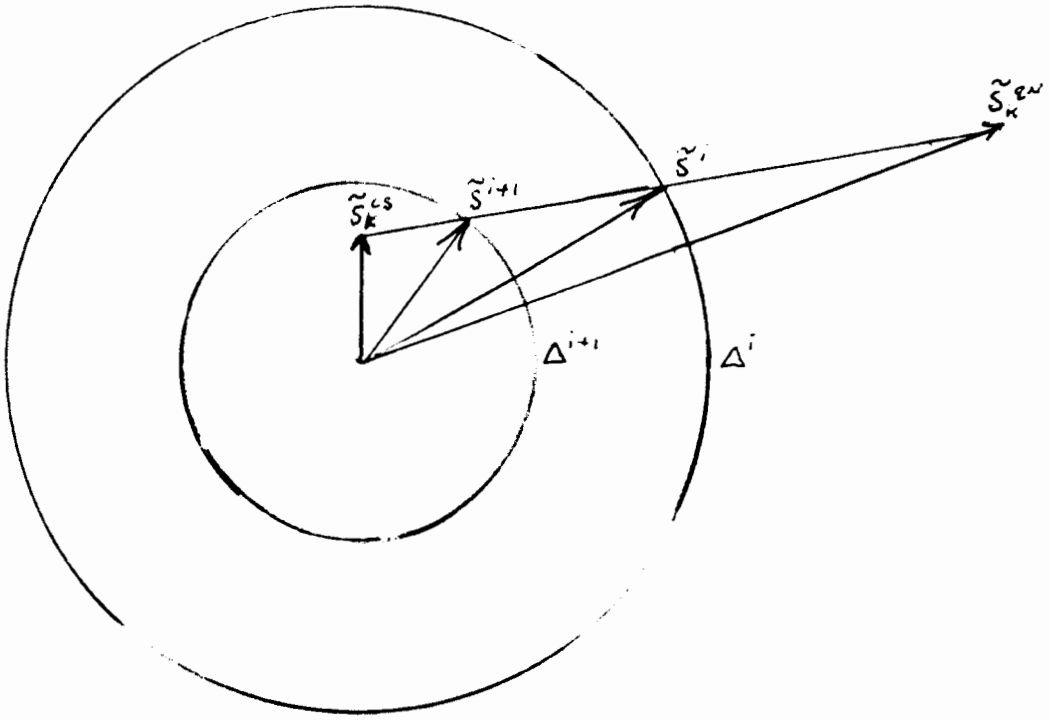

Figure 1. Trial steps computed by the dogleg method for two different values of Δ.

[12]Although it is possible to prove many of our results without using Theorem 2.2, such analysis requires $\varsigma << 1 - \eta_2$ in some cases.

[13]Powell originally only considered $D_k \equiv I$. The more general form given here is sometimes called the *preconditioned dogleg*.

and then proceeding to the *quasi-Newton step*

$$\tilde{s}_k^{qn} = -\tilde{B}_k^{-1}\tilde{g}_k , \tag{2.28}$$

(assuming for the moment that $\tilde{B}_k$ is positive definite.)

If $\tilde{s}_k^{qn}$ is inside the trust region, then $\tilde{s}^i$ is taken to be $\tilde{s}_k^{qn}$. Otherwise the dogleg method sets $\tilde{s}^i$ as the intersection of $\tilde{s}(\alpha)$ with the surface of the trust region. In either event, $\tilde{s}^i$ minimizes $\tilde{\psi}_k(\tilde{x}_k + \tilde{s}(\alpha))$: $||\tilde{s}(\alpha)|| \leq \Delta^i$. This method has the advantage of requiring only one matrix factorization per major iteration: Once $\tilde{s}_k^{qn}$ has been calculated, computation of s^i for any given Δ^i is trivial. Figure 1 illustrates trial steps computed by the dogleg method for different values of Δ. It is clear that, for sufficiently small Δ^i, the trial step $\tilde{s}^i$ is a positive multiple of $-\tilde{g}_k$, the direction of steepest descent for the model.

Other methods exist in the literature which compute approximate solutions to the trust region subproblems by minimizing ψ over a piecewise linear path. The *double dogleg* of Dennis and Mei [5] uses a path with one extra "leg" in order to give a larger bias toward the quasi-Newton direction $-\tilde{B}_k^{-1}\tilde{g}_k$. Steihaug [17] uses a dogleg path defined by the steps generated by a conjugate gradient method (with preconditioner $D_k^T D_k$) applied to the problem $B_k s = -g_k$. Other dogleg methods exist (see, for example, [14]) that take advantage of negative curvature in $\tilde{\psi}$ (i.e., B_k need not be positive definite.) All of these methods use $\tilde{s}_k^{cp}$ as the initial segment of the dogleg and define $\tilde{s}(\alpha)$ such that $||\tilde{s}(\alpha)||$ is increasing so that the intersection of $\tilde{s}(\alpha)$ and the surface of a trust region will be unique. We can therefore state the following.

PROPOSITION 2.3. The conclusions of Theorem 2.2 remain valid if each $\tilde{s}^i$ is computed by a generalized dogleg method rather than an OLC method.

Proof. This proposition follows immediately from (2.27), the uniqueness of $\tilde{s}(\alpha) \cap \{\tilde{s}: ||\tilde{s}|| = \Delta^i\}$, and the hypotheses of Theorem 2.2. □

2.4. The uniform predicted decrease condition. A technical condition concerning trial steps computed by OLC or generalized dogleg methods that is of great use in proving global convergence is the *uniform predicted decrease*[14] (UPD) condition:

$$pred_k(s^i) \geq \tfrac{1}{2} c_3 \, ||\tilde{g}_k|| \min\left\{\Delta^i, \frac{||\tilde{g}_k||}{\sigma_1}\right\} \tag{2.29}$$

for some constants $c_3 \in (0,1]$ and $\sigma_1 \in (0,\infty)$. A complete discussion of this condition is not necessary for the purposes of this paper, and we merely give the well known[15] result that OLC and generalized dogleg methods satisfy (2.29) provided[16]

$$||\tilde{B}_k|| \leq \sigma_1 \quad \forall \; k \; . \tag{2.30}$$

3. Successful termination of the inner loop.

3.1. Introduction. Using the properties of trial steps described in the last section, it is easy to generate examples for which the inner loop of Algorithm (3) will fail to find an acceptable step in a finite number of iterations.

Example 3.1 : g_k not a descent direction. Define $f(x) = \tfrac{1}{2} x^T x$ and select any nonzero x_0. We have $ared_k(s) = \tfrac{1}{2} x_0{}^T x_0 - \tfrac{1}{2}(x_0+s)^T(x_0+s) = -\nabla f\,(x_0)^T s - \tfrac{1}{2} s^T s$. Now suppose that $g_0 = -\nabla f\,(x_0)$, $B_0 = I$, and $D_0 = I$. We have that $pred_k(s) = \nabla f\,(x_0)^T s - \tfrac{1}{2} s^T s$ and

$$\rho^i = \frac{ared_k(s^i)}{pred_k(s^i)} = \frac{-\nabla f\,(x_0)^T s^i - \tfrac{1}{2}(s^i)^T(s^i)}{\nabla f\,(x_0)^T s^i - \tfrac{1}{2}(s^i)^T(s^i)} \; . \tag{3.1}$$

For simplicity, suppose that each s^i is being computed by a dogleg procedure so that $s^i = -\Delta^i g_0 / ||g_0||$ for any $\Delta^i \leq ||g_0||$. Substituting into (3.1) gives

---

[14] The term "uniform" is used because of the uniform bound $||\tilde{B}_k|| \leq \sigma_1 \; \forall \; k$ as opposed to, say, bounds of the form $||\tilde{B}_k|| \leq \sigma_1(1+k) \; \forall \; k$.

[15] See for example, [2], [8], [10], [11], and [14].

[16] In [2] it is argued that the weaker assumption of a uniform upper bound on $\{\tilde{g}_k \tilde{B}_k \tilde{g}_k / \tilde{g}_k{}^T \tilde{g}_k\}$ is to be preferred, since: (a) this also implies the UPD condition, (b) natural methods exist for enforcing this weaker condition, and (c) numerical testing of these safeguarding techniques has shown that they can dramatically improve the reliability of a standard method without decreasing the overall efficiency of the overall algorithm. The best one of these methods is probably of limited utility for models with $g_k \neq \nabla f(x_k)$ because it makes use of first order differences in "$g(x)$" to safeguard the model Hessian, but an alternate safeguarding technique using second order differences in f is also shown in [2] to improve reliability.

$$\rho^i = \frac{-\|g_0\| - \tfrac{1}{2}\Delta^i}{\|g_0\| - \tfrac{1}{2}\Delta^i}. \tag{3.2}$$

Hence, if $\Delta^0 \leq \|g_0\|$, $\rho^i < 0$ for any $\Delta^i \leq \Delta^0$, and the inner loop of Algorithm (3) will never terminate. Similar examples can be shown for any gradient approximation and scaling matrix which do not satisfy $(D_k^{-T} g_k)^T (D_k^{-T} \nabla f(x_k)) > 0$.

Example 3.2 : $\|\mathbf{g_k}\| >> \|\nabla \mathbf{f}(\mathbf{x_k})\|$. Even if $-(D_k{}^T D_k)^{-1} g_k$ is always a descent direction, the inner loop of Algorithm (3) is not assured of success. Consider the last example with g_0 taken to be $\frac{2}{\eta_1}\nabla f(x_0)$. Again taking $s^i = -\Delta^i g_0 / \|g_0\|$, we have

$$\rho^i = \frac{-\nabla f(x_0)^T s^i - \tfrac{1}{2}(s^i)^T (s^i)}{-g_0{}^T s^i - \tfrac{1}{2}(s^i)^T (s^i)} \tag{3.3}$$

Then for any $\Delta^i \leq \Delta^0 \leq \min\{\|g_0\|, \|\nabla f(x_0)\|\}$ we have

$$\rho^i = \tfrac{1}{2}\eta_1 \frac{\|\nabla f(x_0)\| - \tfrac{1}{2}\Delta^i}{\|\nabla f(x_0)\| - 1/4\,\eta_1 \Delta^i} < \eta_1, \tag{3.4}$$

so that the inner loop of Algorithm (3) will never find a successful iterate.

These examples, although rather extreme, clearly demonstrate that additional conditions must be imposed on the gradient approximation to assure the finite termination of the inner loop at every major iteration. It should be pointed out that this in no way contradicts More's result that consistency of the gradient approximations implies $\liminf_{k\to\infty} \|g_k\|_{(D_k{}^T D_k)^{-1}} = 0$ for Algorithm (1). Since his notation includes both inner and outer loops, hypothesis (1.5) becomes

$$(x_k \to x^*) \;\; or \;\; (s^i \to 0 \;\; for\ fixed \;\; k) \Rightarrow \lim_{i\to\infty} \|g^i - \nabla f(x_k)\| = 0 \tag{3.5}$$

in our notation, where $\{g^i\}$ is the set of approximations to $\nabla f(x_k)$ used in the inner loop. However, we prefer algorithms which keep a fixed approximation during the inner loop, and More's hypothesis cannot be applied directly to Algorithms (2) or (3).

3.2. Ensuring successful termination of the inner loop. In this section we show that if the relative error in the gradient approximation is less than $1 - \eta_1$, at a given iteration, then the inner loop of Algorithm (3) is assured of finding a successful new iterate. The following lemma will prove useful.

LEMMA 3.1. Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be continuously differentiable on an open convex set Ω containing a point x_k, and let ∇f be Lipschitz continuous on Ω with constant $L \in (0, \infty)$. Let the functions $ared_k(s)$ and $pred_k(s)$ be defined as in (1.13) and (1.14). Let $\lambda_k^{\min} \in (-\infty, \infty)$ be the smallest eigenvalue of B_k and let $\lambda_k^{\max} \in [\lambda_k^{\min}, \infty)$ be the largest. If the error in g_k is defined to be

$$e_k = g_k - \nabla f(x_k), \tag{3.6}$$

then for all $s \in \mathbf{R}^n$ such that $x_k + s \in \Omega$, we have

$$-\tfrac{1}{2}\,||s||^2(L + \lambda_k^{\max}) - e_k^T s \leq pred_k(s) - ared_k(s) \leq \tfrac{1}{2}\,||s||^2(L - \lambda_k^{\min}) - e_k^T s\ . \tag{3.7}$$

Proof. We first use an integral representation of $ared_k(s)$ to establish

$$\begin{aligned} pred_k(s) - ared_k(s) &= -\,g_k{}^T s - \tfrac{1}{2}\,s^T B_k s + \int_0^1 \nabla f(x_k + \lambda s)^T s\ d\lambda \\ &= -\,e_k^T s - \tfrac{1}{2}\,s^T B_k s + \int_0^1 (\nabla f(x_k + \lambda s) - \nabla f(x_k))^T s\ d\lambda\ . \end{aligned} \tag{3.8}$$

Now, $\lambda_k^{\min}\,||s||^2 \leq s^T B_k s \leq \lambda_k^{\max}\,||s||^2$ and

$$\begin{aligned} \left| \int_0^1 (\nabla f(x_k + \lambda s) - \nabla f(x_k))^T\ s\ d\lambda \right| &\leq \int_0^1 ||\nabla f(x_k + \lambda s) - \nabla f(x_k)||\ ||s||\, d\lambda \\ &\leq \int_0^1 L\,||\lambda s||\ ||s||\, d\lambda = \tfrac{1}{2} L\ ||s||^2\ . \end{aligned} \tag{3.9}$$

Substituting these bounds into (3.8) immediately establishes (3.7).

We can now establish the main result of this section. It ensures that a successful step can always be found provided the relative error in g_k is less than $1 - \eta_1$.

THEOREM 3.2. Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be continuously differentiable on an open convex set Ω containing a point x_k, and let ∇f be Lipschitz continuous on Ω with constant $L \in (0, \infty)$. Let $ared_k(s)$, $pred_k(s)$, e_k, and $\lambda_k^{\min}$ be defined as in Lemma 3.1, and let D_k be any nonsingular matrix. Consider a sequence of trial iterates $\{s^i\}$ and associated trust radii $\{\Delta^i\}$ satisfying $\Delta^i \rightarrow 0$, $pred_k(s^i) > 0\ \forall\ i$, $||D_k s^i|| \leq c_2 \Delta^i\ \forall\ i$, and $\lim_{\Delta^i \rightarrow 0} \cos\Theta^i = 1$, with Θ^i defined to be the angle between $D_k s^i$ and $-D_k^{-T} g_k$. If $g_k \neq 0$ and

$$\frac{||e_k||_{(D_k{}^T D_k)^{-1}}}{||g_k||_{(D_k{}^T D_k)^{-1}}} \leq \varsigma \tag{3.10}$$

for some $\varsigma \in [0, 1-\eta_1)$, then for sufficiently small Δ^i we have

$$\rho^i = \frac{ared_k(s^i)}{pred_k(s^i)} > \eta_1 \,. \tag{3.11}$$

Proof. Assume without loss of generality that each Δ^i is sufficiently small to imply $x_k + s^i \in \Omega$. Since $pred_k(p^i) > 0 \;\forall\; i$, Lemma 3.1 allows us to write

$$\begin{aligned} 1-\rho^i &= \frac{pred(s^i) - ared(s^i)}{pred(s^i)} \\ &\leq \frac{\tfrac{1}{2}(L-\lambda_k^{\min})||s^i||^2 - e_k^T s^i}{-g_k{}^T s^i - \tfrac{1}{2}(s^i)^T B_k(s^i)} \\ &\leq \frac{-(D_k{}^{-T} e_k)^T (D_k s^i) + \tfrac{1}{2}(L-\lambda_k^{\min})||s^i||^2}{-(D_k{}^{-T} g_k)^T (D_k s^i) - \tfrac{1}{2}(s^i)^T B_k(s^i)} \,. \end{aligned} \tag{3.12}$$

Using the Cauchy Schwarz inequality, some algebraic manipulations, and the definition

$$\cos(\Theta^i) = \frac{-(D_k{}^{-T} g_k)^T (D_k s^i)}{||D_k{}^{-T} g_k||\;||D_k s^i||} \,, \tag{3.13}$$

we can rewrite (3.12) as

$$\begin{aligned} 1-\rho^i &\leq \frac{||D_k{}^{-T} e_k||\;||D_k s^i|| + \tfrac{1}{2}(L-\lambda_k^{\min})||s^i||^2}{-(D_k{}^{-T} g_k)^T (D_k s^i) - \tfrac{1}{2}(s^i)^T B_k(s^i)} \\ &\leq \frac{1}{||D_k{}^{-T} g_k||} \frac{||D_k{}^{-T} e_k|| + \tfrac{1}{2}(L-\lambda_k^{\min})||s^i||^2/(||D_k s^i||)}{\cos(\Theta^i) - \tfrac{1}{2}(s^i)^T B_k(s^i)/(||D_k{}^{-T} g_k||\;||D_k s^i||)} \,. \end{aligned} \tag{3.14}$$

Now,

$$\lim_{\Delta^i \to 0} \frac{||s^i||^2}{||D_k s^i||} = \lim_{\Delta^i \to 0} \frac{||s^i||^2}{[(s^i)^T D_k{}^T D_k (s^i)]^{\frac{1}{2}}} = 0 \tag{3.15}$$

and

$$\lim_{\Delta^i \to 0} \frac{(s^i)^T B_k(s^i)}{||D_k s^i||} = \lim_{\Delta^i \to 0} \frac{(s^i)^T B_k(s^i)}{[(s^i)^T D_k{}^T D_k (s^i)]^{\frac{1}{2}}} = 0 \tag{3.16}$$

so that by combining (3.14), (3.15), (3.16) and the hypothesis $\lim_{\Delta^i \to 0} \cos\Theta^i = 1$ we obtain

$$\lim_{\Delta^i \to 0} (1-\rho^i) \leq \frac{||D_k^{-T} e_k||}{||D_k^{-T} g_k||} = \frac{||e_k||_{(D_k{}^T D_k)^{-1}}}{||g_k||_{(D_k{}^T D_k)^{-1}}} \leq \varsigma . \tag{3.17}$$

Since $\varsigma < 1-\eta_1$, we therefore have $1-\rho^i < 1-\eta_1$ for sufficiently small Δ^i. This establishes our result since $1-\rho^i < 1-\eta_1$ if and only if $\rho^i > \eta_1$. □

An immediate consequence of Theorem 3.2 is that Algorithm (3) will either generate an infinite sequence of iterates or terminate with $\nabla f(x_k) = g_k = 0$ provided (3.10) holds at every iteration. This can be formally stated as follows.

COROLLARY 3.3. Let $f: \mathbf{R}^n \to \mathbf{R}$ satisfy the standard assumptions and let $\{D_k\}$ be a sequence of nonsingular diagonal matrices. Then Algorithm (3), using any of the step computation techniques of Section 2, will either produce an infinite sequence of iterates satisfying $f(x_k) < f(x_{k-1})$ or will terminate at some iterate x_k with $\nabla f(x_k) = 0$ provided the relative error in the gradient approximation satisfies

$$\frac{||g_k - \nabla f(x_k)||_{(D_k{}^T D_k)^{-1}}}{||g_k||_{(D_k{}^T D_k)^{-1}}} \leq \varsigma < 1-\eta_1 \tag{3.18}$$

at every iteration.

Proof. Since any acceptable iterate satisfies $pred_k(p_k) > 0$ and $\rho_k > 0$, $f(x_k) < f(x_{k-1})$ for all (existing) iterates, and hence $x_k \in \Omega$ for all (existing) x_k. Now suppose Algorithm (3) succeeds in generating $x_0, x_1, \ldots, x_k$. If $g_k = 0$, (3.18) implies that $\nabla f(x_k) = 0$. Otherwise, the algorithm generates a trial step s^i by the methods of Section 2. If this step satisfies $\rho^i \geq \eta_1$, then x_{k+1} exists. If not, the inner loop of Algorithm (3) will try $\Delta^{i+1} \in (0, \gamma_1 \Delta^i]$, $\Delta^{i+2} \in (0, \gamma_1 \Delta^{i+1}]$, etc. as per step (a3) of the inner loop. Since $\gamma_1 < 1$, the conditions of Theorem 3.2 are satisfied, and hence x_{k+1} exists. Our result follows by induction. □

Some remarks should be made concerning the possibility of $g_k = 0$ or $\nabla f(x_k) = 0$ for some iteration k. If $g_k = 0$, then (3.1) requires that $g_k = \nabla f(x_k)$. This is quite reasonable, in that if the approximate gradient indicates that x_k is a stationary point of f, then the sensible procedure is to recompute g_k with sufficient accuracy to verify or contradict that $\nabla f(x_k) = 0$. We also include no

theory to affirm or deny the implementability of the algorithm beyond any iteration with $g_k = \nabla f(x_k) = 0$. Methods exist (see, for example, [14]) which are guaranteed to be repelled from such a stationary point if and only if it is a saddle point, but these methods assume that the model Hessian B_k is the exact Hessian $\nabla^2 f(x_k)$. Since the use of a model with an approximate gradient and an exact Hessian appears somewhat unlikely, we prefer (for this paper) to say nothing about the existence of x_{k+1} if $g_k = \nabla f(x_k) = 0$.

3.3. Relative error bound as an auxiliary to consistency. In one sense, Theorem 3.2 might be considered the main result of this paper in that using (3.10) as an *auxiliary* condition to (1.5) eliminates the major practical difficulty in directly enforcing (1.5). That is, (3.10) assures us that no further increases in the accuracy of the gradient approximation will be required in the inner loop. Enforcing (1.5) for the successful iterates is a lesser problem (even though conditions like (1.11) are still somewhat unsatisfying). Moreover, (3.10) is a sufficient condition for $\liminf_{k\to\infty} ||g_k||_{(D_k{}^T D_k)^{-1}} = 0$ to imply $\liminf_{k\to\infty} ||\nabla f(x_k)||_{(D_k{}^T D_k)^{-1}} = 0$. Consistency alone is not sufficient to establish this unless $\{x_k\}$ converges.

In Section 4, we show that consistency can be *entirely replaced as a primary assumption* by conditions on the relative error. However, for completeness we conclude this section by showing how using (3.10) as an auxiliary assumption to consistency allows the results of Moré to be strengthened.

We first state the following lemma.

LEMMA 3.4. Let $\{D_k\}$ be a sequence of nonsingular matrices in $\mathbf{R}^{n\times n}$ satisfying $||D_k{}^T D_k|| \le (\sigma_2)^2$ and $||(D_k{}^T D_k)^{-1}|| \le (\sigma_3)^2$ for some constants $\sigma_2, \sigma_3 < \infty$. Let $\{g_k\}$ and $\{\nabla f(x_k)\}$ be sequences in $\mathbf{R}^{n\times n}$ that satisfy either

$$\frac{||e_k||}{||g_k||} \le \varsigma < 1 \quad \forall\ k \tag{3.19}$$

or

$$\frac{||e_k||_{(D_k{}^T D_k)^{-1}}}{||g_k||_{(D_k{}^T D_k)^{-1}}} \leq \varsigma < 1 \quad \forall \ k \ . \tag{3.20}$$

Then the following are equivalent.

$$\liminf_{k\to\infty} ||g_k|| = 0 \ . \tag{3.21}$$

$$\liminf_{k\to\infty} ||g_k||_{(D_k{}^T D_k)^{-1}} = 0 \ . \tag{3.22}$$

$$\liminf_{k\to\infty} ||\nabla f(x_k)|| = 0 \ . \tag{3.23}$$

$$\liminf_{k\to\infty} ||\nabla f(x_k)||_{(D_k{}^T D_k)^{-1}} = 0 \ . \tag{3.24}$$

Furthermore, the following are also equivalent.

$$\lim_{k\to\infty} ||g_k|| = 0 \ . \tag{3.25}$$

$$\lim_{k\to\infty} ||g_k||_{(D_k{}^T D_k)^{-1}} = 0 \ . \tag{3.26}$$

$$\lim_{k\to\infty} ||\nabla f(x_k)|| = 0 \ . \tag{3.27}$$

$$\lim_{k\to\infty} ||\nabla f(x_k)||_{(D_k{}^T D_k)^{-1}} = 0 \ . \tag{3.28}$$

Proof. We first notice that the conditions on $\{D_k\}$ imply

$$\frac{1}{\sigma_2} ||y|| \leq ||y||_{(D_k{}^T D_k)^{-1}} \leq \sigma_3 ||y|| \tag{3.29}$$

and

$$\frac{1}{\sigma_3} ||y||_{(D_k{}^T D_k)^{-1}} \leq ||y|| \leq \sigma_2 ||y||_{(D_k{}^T D_k)^{-1}} \tag{3.30}$$

for all $y \in \mathbb{R}^n$. This immediately implies (3.21) $\Leftrightarrow$ (3.22), (3.23) $\Leftrightarrow$ (3.24), (3.25) $\Leftrightarrow$ (3.26), and (3.27) $\Leftrightarrow$ (3.28). Now, if $||e_k||/||g_k|| \leq \varsigma < 1 \ \forall \ k$, we have that (3.21) $\Leftrightarrow$ (3.23) and (3.25) $\Leftrightarrow$ (3.27). If $||e_k||_{(D_k{}^T D_k)^{-1}}/||g_k||_{(D_k{}^T D_k)^{-1}} \leq \varsigma < 1$, then (3.22) $\Leftrightarrow$ (3.24) and (3.26) $\Leftrightarrow$ (3.28). Linking all of these equivalences immediately establishes the lemma. □

A hybrid of Theorem 3.2 and the global convergence results of Moré can now be stated.

THEOREM 3.5. Let $f: \mathbb{R}^n \to \mathbb{R}$ satisfy the standard assumptions, let $\{B_k\}$, $\{D_k{}^T D_k\}$ and $\{(D_k{}^T D_k)^{-1}\}$ be uniformly bounded, and let $\{x_k\}$ be the set of iterates produced by Algorithm (1)

using any of the step computation techniques of Section 2. Let the gradient approximations satisfy the relative error bound (3.18), and assume that if the set of successful iterates is an infinite sequence, then the consistency condition (1.5) holds. Further assume that $x_{k+1} = x_k \Rightarrow g_{k+1} = g_k$. We then have that either

$$\liminf_{k \to \infty} ||g_k|| = \liminf_{k \to \infty} ||\nabla f(x_k)|| = 0 , \tag{3.31}$$

or

$$g_k = \nabla f(x_k) = 0 \tag{3.32}$$

for some iterate x_k.

Proof. We first note from Corollary 3.3 that either $\nabla f(x_k) = g_k = 0$ for some iteration x_k, or Algorithm (1) generates an infinite sequence of successful iterates. If $\{x_k\}$ is an infinite sequence, then by hypothesis $\{g_k\}$ is consistent and More's [10] result that $\liminf_{k \to \infty} ||g_k||_{(D_k{}^T D_k)^{-1}} = 0$ applies (our assumptions on f, $\{B_k\}$, $\{D_k\}$, and the step computation procedure are more than sufficient to imply the hypotheses used in [10]). Hence Lemma 3.4 implies that either (3.31) or (3.32) holds.

□

4. Global convergence.

4.1. Introduction. Although Theorem 3.5 shows that applying (3.10) as an auxiliary condition bypasses the largest practical difficulty with directly enforcing consistency, this theory is still less than satisfying because nothing is specified about how fast $||e_k||$ should be forced to zero as $\{x_k\}$ converges. If a condition such as $||e_k|| \leq c\ ||s_k||$ is used, c can be chosen to be any value in $[0, \infty)$. In Section 4.2, we establish the same global convergence results as in Theorem 3.5 without using consistency as a primary hypothesis. We instead use the condition

$$\frac{||e_k||_{(D_k{}^T D_k)^{-1}}}{||g_k||_{(D_k{}^T D_k)^{-1}}} \leq \varsigma < 1 - \eta_2 . \tag{4.1}$$

Since typical values for η_2 usually fall in [0.1, 0.25] and typical values for η_2 usually fall in [0.001, 0.1], condition (4.1) is only slightly more restrictive than (3.10).

In order to establish the strong result $\lim_{k\to\infty} ||\nabla f(x_k)|| = 0$, Moré [10] returns to the assumption that $g_k = \nabla f(x_k)$. This stronger assumption is not necessary. We show in Section 4.3 that bounding the relative error in the gradient approximation to be less than or equal to any constant $\varsigma \in [0,1)$ is sufficient to establish $\lim_{k\to\infty} ||\nabla f(x_k)|| = 0$. This strong global convergence result is often called first order stationary point convergence.

4.2. Replacing the consistency assumption with a relative error bound. We now show that the results of Theorem 3.5 remain true if the consistency assumption is replaced by (4.1).

THEOREM 4.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ satisfy the standard assumptions, let $\{B_k\}$, $\{D_k{}^T D_k\}$ and $\{(D_k{}^T D_k)^{-1}\}$ be uniformly bounded, and let $\{x_k\}$ be the set of iterates produced by Algorithm (3) using any of the step computation techniques of Section 2. Let the gradient approximation satisfy the relative error bound (4.1). We then have that either

$$\liminf_{k\to\infty} ||g_k|| = \liminf_{k\to\infty} ||\nabla f(x_k)|| = 0 \tag{4.2}$$

or

$$g_k = \nabla f(x_k) = 0 \tag{4.3}$$

for some iterate x_k.

Proof. The central ideas in this proof are largely due to Powell [12], but we also draw heavily on the ideas used to prove Theorem 3.2.

(a) We first note that since $\varsigma < 1 - \eta_2 < 1 - \eta_1$, by Corollary 3.3 we have that either (4.3) is true or the algorithm generates an infinite sequence of successful iterates satisfying $f(x_k) < f(x_{k-1})$. Hence $x_k \in \Omega$ for all k.

(b) Suppose $\{x_k\}$ is an infinite sequence but

$$\liminf_{k\to\infty} ||g_k||_{(D_k{}^T D_k)^{-1}} > \epsilon > 0 . \tag{4.4}$$

From (1.16), (2.29), and the bounds on $\{B_k\}$ and $\{D_k\}$ we have

$$ared_k(s_k) \geq \tfrac{1}{2}\eta_1 c_3 \|g_k\|_{(D_k^T D_k)^{-1}} \min\left\{\Delta_k, \frac{1}{\sigma_1}\|g_k\|_{(D_k^T D_k)^{-1}}\right\} \tag{4.5}$$

for some $\sigma_1 \in (0,\infty)$, $\eta_1 \in (0,1)$ and $c_3 \in (0,1]$. Since f is bounded below, (4.5) implies that $\Delta_k \to 0$ and hence $\Delta^i \to 0$. Assume without loss of generality that k is sufficiently large to imply $\|g_k\|_{(D_k^T D_k)^{-1}} \geq \epsilon$ and $x_k + s^i \in \Omega$. From (3.14) we have

$$1 - \rho^i \leq \frac{\|D_k^{-T} e_k\| / \|D_k^{-T} g_k\| + \tfrac{1}{2}(L - \lambda_k^{\min})\|s^i\|^2 / (\|D_k^{-T} g_k\| \, \|D_k s^i\|)}{\cos\Theta^i - \tfrac{1}{2}(s^i)^T B_k(s^i)/(\|D_k^{-T} g_k\| \, \|D_k s^i\|)} . \tag{4.6}$$

From Theorem 2.2 and Proposition 2.3 we have that $\Delta^i \to 0 \Rightarrow \cos\Theta^i \to 1$ and since $\{B_k\}$ and $\{(D_k{}^T D_k)^{-1}\}$ are bounded, we can write

$$\lim_{k,i\to\infty} \frac{\|s^i\|^2}{\|D_k^{-T} g_k\| \, \|D_k s^i\|} \leq \frac{1}{\epsilon} \lim_{k,i\to\infty} \frac{\|s^i\|^2}{[(s^i)^T D_k{}^T D_k (s^i)]^{1/2}} = 0 . \tag{4.7}$$

and

$$\lim_{k,i\to\infty} \frac{(s^i)^T B_k(s^i)}{\|D_k^{-T} g_k\| \, \|D_k s^i\|} \leq \frac{1}{\epsilon} \lim_{k,i\to\infty} \frac{(s^i)^T B_k(s^i)}{[(s^i)^T D_k{}^T D_k (s^i)]^{1/2}} = 0 . \tag{4.8}$$

Furthermore, $\lambda_k^{\min}$ is bounded away from $-\infty$, so combining (4.6), (4.7), and (4.8) gives

$$\lim_{i\to\infty}(1-\rho^i) \leq \lim_{k\to\infty} \frac{\|D_k^{-T} e_k\|}{\|D_k^{-T} g_k\|} \leq \varsigma < 1 - \eta_2 , \tag{4.9}$$

therefore there exists $\bar{i}$ such that $i > \bar{i} \Rightarrow 1 - \rho^i < 1 - \eta_2$, and hence $\rho^i > \eta_2$. But since no trust radius reduction is allowed if $\rho^i > \eta_2$, we have that $\liminf_{i\to\infty} \Delta^i > 0$, which implies $\liminf_{k\to\infty} \Delta_k > 0$, which is a contradiction. Thus if $\{x_k\}$ is an infinite sequence,

$$\liminf_{k\to\infty} \|g_k\|_{(D_k{}^T D_k)^{-1}} = 0 . \tag{4.10}$$

The result (4.2) follows from (4.10) and Lemma 3.4. □

4.3. First order stationary point convergence.

4.3.1. Relative error measured in the Euclidean norm. The following theorem establishes first order stationary point convergence provided the sequence $\{x_k\}$ satisfies the weaker property $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$, the uniform predicted decrease condition holds, and

$||e_k||/||g_k|| \leq \varsigma < 1$. This is a very powerful result, as it allows us to obtain the strong result $\lim_{k\to\infty} ||\nabla f(x_k)|| = 0$ directly from the previously established weak convergence property (4.2) without using any information concerning the trust radius updating procedure.

THEOREM 4.2. Let $f: \mathbb{R}^n \to \mathbb{R}$ satisfy the standard assumptions. Let $\{x_k\}$ be an infinite sequence of vectors which satisfy the (UPD) condition

$$pred_k(s_k) \geq \tfrac{1}{2} c_3 ||g_k||_{(D_k{}^T D_k)^{-1}} \min\left\{\Delta_k, \frac{1}{\sigma_1} ||g_k||_{(D_k{}^T D_k)^{-1}}\right\} \tag{4.11}$$

and

$$ared_k(s_k) \geq \eta_1 pred_k(s_k), \tag{4.12}$$

where Δ_k is a positive number satisfying

$$||s_k||_{D_k{}^T D_k} \leq c_2 \Delta_k \tag{4.13}$$

with $c_3 \in (0, 1]$, $\sigma_1 \in (0, \infty)$, $\eta_1 \in (0, 1)$ and $c_2 \in [1, 2)$.

Let the sequence of scaling matrices $\{D_k\}$ satisfy

$$||D_k{}^T D_k|| \leq (\sigma_2)^2 \tag{4.14}$$

and

$$||(D_k{}^T D_k)^{-1}|| \leq (\sigma_3)^2 \tag{4.15}$$

for $\sigma_2, \sigma_3 \in (0, \infty)$. If

$$\frac{||g_k - \nabla f(x_k)||}{||g_k||} \leq \varsigma \tag{4.16}$$

for all k with $\varsigma \in [0, 1)$, then

$$\liminf_{k\to\infty} ||g_k|| = 0 \Rightarrow \lim_{k\to\infty} ||g_k|| = 0 \Rightarrow \lim_{k\to\infty} ||\nabla f(x_k)|| = 0 . \tag{4.17}$$

Proof. Define $\epsilon = \tfrac{1}{2}(1-\varsigma)/(1+\varsigma)$ and consider any iterate x_m with nonzero g_m. Since $\liminf_{k\to\infty} ||g_k|| = 0$, there exists $\bar{m} \geq m$ for which $||g_{\bar{m}+1}|| \leq \epsilon ||g_m||$ and $||g_k|| > \epsilon ||g_m||$ for all $k \in [m, \bar{m}]$. Now, from (4.11) and (4.12)

$$
\begin{aligned}
f(x_m) - f(x_{\bar{m}+1}) &= \sum_{k=m}^{\bar{m}} ared_k(s_k) \geq \sum_{k=m}^{\bar{m}} \eta_1 pred_k(s_k) \\
&\geq \sum_{k=m}^{\bar{m}} \tfrac{1}{2}\eta_1 c_3 \, ||g_k||_{(D_k^T D_k)^{-1}} \min\left\{\Delta_k, \frac{1}{\sigma_1} ||g_k||_{(D_k^T D_k)^{-1}}\right\}.
\end{aligned} \tag{4.18}
$$

Using the facts that $||g_k||_{(D_k^T D_k)^{-1}} \geq \frac{1}{\sigma_2} ||g_k||$, $\Delta_k \geq \frac{1}{c_2} ||s_k||_{(D_k^T D_k)} \geq \frac{1}{\sigma_3 c_2} ||s_k||$, and $||g_k|| > \epsilon ||g_m||$ for all $k \in [m, \bar{m}]$, (4.18) can be transformed into

$$
f(x_m) - f(x_{\bar{m}+1}) \geq \tfrac{1}{2}\eta_1 c_3 \frac{\epsilon}{\sigma_2} ||g_m|| \sum_{k=m}^{\bar{m}} \min\left\{\frac{||s_k||}{\sigma_3 c_2}, \frac{\epsilon ||g_m||}{\sigma_1 \sigma_2}\right\}. \tag{4.19}
$$

This can be divided into two cases.

(i) If $||g_m|| \geq \frac{\sigma_1 \sigma_2}{\epsilon \sigma_3 c_2} ||s_k||$ for at least one $k \in [m, \bar{m}]$, we have

$$
f(x_m) - f(x_{\bar{m}+1}) \geq \tfrac{1}{2}\eta_1 c_3 (\frac{\epsilon}{\sigma_2})^2 \frac{||g_m||^2}{\sigma_1}. \tag{4.20}
$$

(ii) Otherwise,

$$
f(x_m) - f(x_{\bar{m}+1}) \geq \tfrac{1}{2}\eta_1 c_3 \frac{\epsilon}{\sigma_2} \frac{||g_m||}{\sigma_3 c_2} \sum_{k=m}^{\bar{m}} ||s_k||. \tag{4.21}
$$

Now, in order to merge case (i) and case (ii), we need to establish a lower bound on $\sum_{k=m}^{\bar{m}} ||s_k||$. From the triangle inequality we can write $||g_m|| \leq ||g_{\bar{m}+1} - g_m|| + ||g_{\bar{m}+1}||$ and hence $||g_m|| \leq ||g_{\bar{m}+1} - g_m|| + \epsilon ||g_m||$. By rearranging terms, again applying the triangle inequality, invoking the Lipschitz continuity of ∇f, and substituting in $e_k = g_k - \nabla f(x_k)$, we can obtain

$$
\begin{aligned}
(1-\epsilon) ||g_m|| &\leq ||g_{\bar{m}+1} - g_m|| \\
&\leq ||\nabla f(x_{\bar{m}+1}) - \nabla f(x_m)|| + ||e_{\bar{m}+1} - e_m|| \\
&\leq L ||x_{\bar{m}+1} - x_m|| + ||e_{\bar{m}+1} - e_m|| \\
&\leq L \sum_{k=m}^{\bar{m}} ||s_k|| + ||e_{\bar{m}+1}|| + ||e_m||.
\end{aligned} \tag{4.22}
$$

Substituting (4.22) into (4.21) yields

$$
\begin{aligned}
f(x_m) - f(x_{\bar{m}+1}) &\geq \tfrac{1}{2} \frac{\eta_1 c_3}{\sigma_2 \sigma_3 c_2} \frac{1}{L} \epsilon \, \|g_m\| [(1-\epsilon)\|g_m\| - \|e_{\bar{m}+1}\| - \|e_m\|] \\
&\geq \tfrac{1}{2} \frac{\eta_1 c_3 \epsilon}{\sigma_2 \sigma_3 c_2 L} \|g_m\|^2 \left[(1-\epsilon) - \frac{\|e_m\|}{\|g_m\|} - \frac{\|e_{\bar{m}+1}\| \, \|g_{\bar{m}+1}\|}{\|g_{\bar{m}+1}\| \, \|g_m\|} \right] \\
&\geq \tfrac{1}{2} \frac{\eta_1 c_3 \epsilon}{\sigma_2 \sigma_3 c_2 L} \|g_m\|^2 [1 - \epsilon - \varsigma - \varsigma\epsilon] \\
&\geq \tfrac{1}{2} \frac{\eta_1 c_3 \epsilon}{\sigma_2 \sigma_3 c_2 L} \|g_m\|^2 [\tfrac{1}{2}(1-\varsigma)] \; .
\end{aligned}
\tag{4.23}
$$

Hence for either case (i) or case (ii) we have that

$$
f(x_m) - f(x_{\bar{m}+1}) \geq \bar{\epsilon} \, \|g_m\|^2 \tag{4.24}
$$

where $\bar{\epsilon}$ is the positive constant $\bar{\epsilon} = \tfrac{1}{2} \eta_1 c_3 \frac{\epsilon}{\sigma_2} \min \left\{ \frac{\epsilon}{\sigma_2 \sigma_1}, \frac{1-\varsigma}{2L \sigma_3 c_2} \right\}$.

Now, by hypothesis, f is nonincreasing and bounded below, so $\{f(x_k)\}$ must converge to some limit, say f^*. Thus, for any m, either $g_m = 0$ or

$$
\begin{aligned}
\|g_m\| &\leq (f(x_m) - f(x_{\bar{m}+1})) / \bar{\epsilon} \\
&\leq (f(x_m) - f^*) / \bar{\epsilon} \, .
\end{aligned}
\tag{4.25}
$$

Therefore $g_k \longrightarrow 0$ and by Lemma 3.4, $\nabla f(x_k) \longrightarrow 0$. □

Condition (4.16) is a fairly natural condition, but it is slightly different from the condition used previously because it measures the relative error in the Euclidean norm while (1.10) measures it in the elliptical norm induced by $(D_k^T D_k)^{-1}$. In the next section we introduce a variation of Theorem 4.2 which uses (1.10).

4.3.2. Relative error measured in the norm induced by the scaling matrices. The following theorem establishes first order stationary point convergence under conditions similar to those of Theorem 4.2. There are, however, two differences. First, we assume $\|e_k\|_{(D_k^T D_k)^{-1}} / \|g_k\|_{(D_k^T D_k)^{-1}} \leq \varsigma < 1$ to be consistent with the theory in Section 4.2. Second, we impose an extra condition on the sequence of scaling matrices.

THEOREM 4.3. Let the hypotheses of Theorem 4.2 be satisfied, with the exception that (4.16) is replaced by

$$\frac{\| g_k - \nabla f(x_k) \|_{(D_k^T D_k)^{-1}}}{\| g_k \|_{(D_k^T D_k)^{-1}}} < \varsigma \tag{4.26}$$

for all k with $\varsigma \in [0, 1)$. Let us further assume that there exists a constant $\bar{L} \in (0, \infty)$ such that

$$\| D_{k+1}^{-T} \nabla f(x_{k+1}) - D_k^{-T} \nabla f(x_k) \| \leq \bar{L} \| s_k \| \tag{4.27}$$

for all k. We then have that

$$\liminf_{k \to 0} \| g_k \| = 0 \Rightarrow \lim_{k \to \infty} \| g_k \| = 0 \Rightarrow \lim_{k \to \infty} \| \nabla f(x_k) \| = 0 . \tag{4.28}$$

The proof of Theorem 4.3 is quite similar to that of Theorem 4.2, so we shall defer it until the Appendix.

Condition (4.27) is quite interesting. If a fixed scaling matrix D is used rather than an adaptive scaling technique, (4.27) is automatically implied by the Lipschitz condition on ∇f. Furthermore, simply assuming that $\{D_k{}^T D_k\}$ and $\{(D_k^T D_k)^{-1}\}$ are bounded is definitely *not* sufficient to imply (4.27). For example, if $\nabla f(x_{k+1}) = \nabla f(x_k)$, $\| D_{k+1}{}^{-T} \nabla f(x_{k+1}) - D_k^{-T} \nabla f(x_k) \| = \| (D_{k+1}{}^{-T} - D_k^{-T}) \nabla f(x_k) \|$.

Adaptive scaling is poorly understood at present. Most implementations that make use of it generate $\{D_k\}$ by heuristic methods rather than procedures with a firm theoretical basis. Given this lack of understanding, theoretical conditions such as (4.27) are important because they suggest guidelines to be used in designing methods for generating scaling matrices $\{D_k\}$.

An extension of our theory which might seem desirable would be a result analogous to Theorem 4.2 but with the relative error expressed in the Euclidean norm. Such a theorem would increase the symmetry between the results of Sections 4.2 and 4.3. Unfortunately, this conjecture is not true. Say, for example, $\nabla f(x_k) = (-\frac{1}{2}, 1)^T$, $g_k = (\frac{1}{2}, 1)^T$, $\eta_2 = 0.1$, and $D_k = \begin{pmatrix} 1/4 & 0 \\ 0 & 1 \end{pmatrix}$. Now, $e_k = (1, 0)$ and $\| e_k \| / \| g_k \| = \sqrt{4/5} < .9$, so that our condition $\| e_k \| / \| g_k \| \leq \varsigma < 1 - \eta_2$ is satisfied. However, the preconditioned steepest descent direction, $-(D_k^T D_k)^{-1} g_k$, is $-(8, 1)^T$. This is not a descent direction for f since $(-\nabla f(x_k))^T (-8, -1)^T < 0$. Therefore, since s^i tends in

direction toward $-(D_k^T D_k)\, g_k^{-1}$ as $\Delta^i \to 0$, a sufficiently small Δ^0 will imply $ared_k(s^i)/pred_k(s^i) < 0 \;\forall\; \Delta^i \leq \Delta^0$.

5. Conclusion.

5.1. Summary of results. The global convergence result $\liminf_{k\to\infty} ||g_k||_{(D_k^T D_k)^{-1}} = 0$ has previously been shown for trust region algorithms that use inexact gradient values provided these approximations are consistent. We demonstrate, however, that for implementations that do not update g_k on unsuccessful iterations, the algorithm may fail at a point x_k with $g_k \neq 0$. This failure cannot occur if

$$\frac{||g_k - \nabla f(x_k)||_{(D_k^T D_k)^{-1}}}{||g_k||_{(D_k^T D_k)^{-1}}} < \varsigma \tag{5.1}$$

and $\varsigma \in [0, 1-\eta_1)$. Furthermore, if (5.1) holds with $\varsigma \in [0, 1-\eta_2)$, the result $\liminf_{k\to\infty} ||g_k|| = \liminf_{k\to\infty} ||\nabla f(x_k)|| = 0$ can be established without using consistency as a primary assumption[17] : consistency is instead a *consequence* of our theory. Finally, (5.1) also allows us to obtain the strong global convergence result $\lim_{k\to\infty} ||\nabla f(x_k)|| = 0$ provided (4.27) holds.

Since many of the procedures used for generating gradient approximations simultaneously provide an error estimate, our results provide a practical criteria for deciding whether a given approximation is adequate.

5.2. Final remarks. Several possibilities suggest themselves for future study. One is to establish our results using alternative assumptions. Rather than taking $||e_k||/||g_k|| \leq \varsigma$, we might try assumptions like

$$\frac{g_k{}^T \nabla f(x_k)}{||g_k||\;||\nabla f(x_k)||} \geq \varsigma \tag{5.2}$$

or

[17]This result uses mild assumptions on f and assumes that $\{B_i\}$, $\{D_i{}^T D_i\}$ and $\{(D_i^T D_i)^{-1}\}$ are uniformly bounded.

$$\eta_1 < \frac{||\nabla f(x_k)||}{||g_k||} < \frac{1}{\eta_1} . \tag{5.3}$$

The two examples in Section 3.1 show that neither (5.2) nor (5.3) taken alone is sufficient to imply implementability, but some combination of similar assumptions might work. The existence (and utility) of such alternative assumptions is an open question.

Another topic for future research is to examine the local convergence rates of these methods. Steihaug [16] establishes q-superlinear convergence for a class of trust region algorithms assuming (among other things) that $\lim_{k\to\infty} \frac{||B_k s_k + \nabla f(x_k)||}{||\nabla f(x_k)||} = 0$, or equivalently $\lim_{k\to\infty} \frac{||(B_k s_k + g_k) - e_k||}{||g_k - e_k||} = 0$. The structural similarity between Steihaug's analysis and that of this paper suggests that q-superlinear convergence can be obtained if $\lim_{k\to\infty} \frac{||e_k||}{||g_k||} = 0$. This is probably an unrealistic assumption since gradient approximations are generally used only when exact (or almost exact) values are extremely expensive computationally, so an important question is the existence of less restrictive assumptions which imply fast local convergence.

6. Appendix.

6.1. Proof of Theorem 2.2. First notice that case (i) can be treated as a special instance of case (ii) by defining $\tilde{g}_{\bar{k}} = \tilde{g}_k$ and $\tilde{B}_{\bar{k}} = \tilde{B}_k \ \forall\ \bar{k} \geq k$. To prove case (ii), we recall that by Theorem 2.1, there exists a sequence of nonnegative numbers $\{\mu^i\}$ such that

$$(\tilde{B}_k + \mu^i I)\, \tilde{s}^i = -\tilde{g}_k \tag{6.1}$$

and

$$||\tilde{s}^i|| \leq c_2 \Delta^i \tag{6.2}$$

with $\tilde{B}_k + \mu^i I$ positive semidefinite. Applying the Cauchy Schwarz inequality to (6.1) gives

$$||\tilde{s}^i|| \geq ||\tilde{g}_k|| / ||\tilde{B}_k + \mu^i I|| . \tag{6.3}$$

Suppose there exists $\epsilon > 0$ such that $||\tilde{g}_k|| \geq \epsilon$ for all k sufficiently large. Equation (6.3) and the hypothesis that $\{\tilde{B}_k\}$ is bounded establishes that

$$||\tilde{s}^i|| \to 0 \Rightarrow \mu^i \to \infty . \tag{6.4}$$

Now,

$$\cos\Theta^i = \frac{-(\tilde{s}^i)^T \tilde{g}_k}{||\tilde{s}^i|| \, ||\tilde{g}_k||} \tag{6.5}$$

so that by substituting $(\tilde{B}_k + \mu^i I)\,\tilde{s}^i$ for $-\tilde{g}_k$ in (6.5) and expanding the resulting terms we get

$$\begin{aligned}\cos\Theta^i &= \frac{(\tilde{s}^i)^T(\tilde{B}_k + \mu^i I)\,\tilde{s}^i}{||\tilde{s}^i|| \, ||(\tilde{B}_k + \mu^i I)\,\tilde{s}^i||} \\ &= \frac{1 + \frac{1}{\mu^i}(\tilde{s}^i)^T \tilde{B}_k(\tilde{s}^i)/\,||\tilde{s}^i||^2}{\left[1 + \frac{2}{\mu^i}(\tilde{s}^i)^T \tilde{B}_k(\tilde{s}^i)/\,||\tilde{s}||^2 + (\frac{1}{\mu^i})^2(\tilde{s}^i)^T \tilde{B}_k{}^T \tilde{B}_k(\tilde{s}^i)/\,||\tilde{s}^i||^2\right]^{\frac{1}{2}}} .\end{aligned} \tag{6.6}$$

Hence by (6.2), (6.4), (6.6) and the hypotheses $\limsup_{k\to\infty} ||\tilde{B}_k|| < \infty$ and $\lim_{i\to\infty} \Delta^i = -0$, we have $\lim_{k,i\to\infty} \cos(\Theta^i) = 1$. □

6.2. Proof of Theorem 4.3. The proof of this theorem is quite similar to that of Theorem 4.2. Define

$$\epsilon = \tfrac{1}{2}(1-\varsigma)/(1+\varsigma) \tag{6.7}$$

and consider any iterate x_m with nonzero g_m.

Since $\liminf_{k\to\infty} ||g_k|| = 0$, by Lemma 3.4 we have that $\liminf_{k\to\infty} ||D_k{}^{-T} g_k|| = 0$, and thus there exists $\bar{m} \geq m$ for which $||D_{k+1}{}^{-T} g_{\bar{m}+1}|| \leq \epsilon\, ||D_m^{-T} g_m||$ and $||D_k{}^{-T} g_k|| > \epsilon\, ||D_m^{-T} g_m||$ for all $k \in [m, \bar{m}]$. Using equation (4.18) and the facts that $\Delta_k \geq \frac{1}{c_2}\, ||s_k||_{D_k{}^T D_k} \geq \frac{1}{\sigma_3 c_2}\, ||s_k||$ and $||D_k{}^{-T} g_k|| > \epsilon\, ||D_m^{-T} g_m|| \;\forall\; k \in [m, \bar{m}]$, we can write

$$f(x_m) - f(x_{\bar{m}+1}) \geq \frac{1}{2}\eta_1 c_3 \epsilon\, ||D_m^{-T} g_m|| \sum_{k=m}^{\bar{m}} \min\left\{\frac{||s_k||}{\sigma_3 c_2}, \epsilon\, \frac{||D_m^{-T} g_m||}{\sigma_1}\right\}. \tag{6.8}$$

We then use the triangle inequality to show

$$\begin{aligned}||D_m^{-T} g_m|| &\leq ||D_m^{-T} g_m - D_{\bar{m}+1}^{-T} g_{\bar{m}+1}|| + ||D_{\bar{m}+1}^{-T} g_{\bar{m}+1}|| \\ &\leq ||D_m^{-T} g_m - D_{\bar{m}+1}^{-T} g_{\bar{m}+1}|| + \epsilon\, ||D_m^{-T} g_m||\end{aligned} \tag{6.9}$$

so that

$$(1-\epsilon)\,||D_m^{-T}g_m\,|| \leq\; ||D_m^{-T}g_m - D_{\bar{m}+1}^{-1}g_{\bar{m}+1}\,|| \;. \tag{6.10}$$

Rearranging terms, defining $e_k = g_k - \nabla f(x_k)$, and again applying the triangle inequality allows us to write

$$\begin{aligned}
(1-\epsilon)\,||D_m^{-T}g_m\,|| &\leq\; ||D_m^{-T}\nabla f(x_m) + D_m^{-T}e_m - D_{\bar{m}+1}^{-T}\nabla f(x_{\bar{m}+1}) - D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,|| \\
&\leq\; ||D_m^{-T}\nabla f(x_m) - D_{\bar{m}+1}^{-T}\nabla f(x_{\bar{m}+1})\,|| + ||D_m^{-T}e_m\,|| + ||D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,|| \\
&=\; ||\sum_{k=m}^{\bar{m}} (D_k{}^{-T}\nabla f(x_k) - D_{k+1}{}^{-1}\nabla f(x_{k+1}))\,|| + ||D_m^{-T}e_m\,|| + ||D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,|| \\
&\leq\; \sum_{k=m}^{\bar{m}} ||D_k{}^{-T}\nabla f(x_k) - D_{k+1}{}^{-1}\nabla f(x_{k+1})\,|| + ||D_m^{-T}e_m\,|| + ||D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,|| \\
&\leq\; \bar{L}\sum_{k=m}^{\bar{m}} ||s_k\,|| + ||D_m^{-T}e_m\,|| + ||D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,|| \;.
\end{aligned} \tag{6.11}$$

Using (6.7), (6.11), and the inequality $||D_{\bar{m}+1}^{-T}g_{\bar{m}+1}\,|| \leq \epsilon\;||D_m^{-T}g_m\,||$ gives

$$\begin{aligned}
\sum_{k=m}^{\bar{m}} ||s_k\,|| &\geq \bar{L}^{-1}\left[(1-\epsilon)\,||D_m^{-T}g_m\,|| - ||D_m^{-T}e_m\,|| - ||D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,||\right] \\
&\geq \bar{L}^{-T}\;||D_m^{-T}g_m\,||\left[(1-\epsilon) - \frac{||D_m^{-T}e_m\,||}{||D_m^{-T}g_m\,||} - \frac{||D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,||}{||D_m^{-T}g_m\,||}\right] \\
&\geq \bar{L}^{-1}\;||D_m^{-T}g_m\,||\left[1-\epsilon-\varsigma - \frac{||D_{\bar{m}+1}^{-T}e_{\bar{m}+1}\,||}{||D_{\bar{m}+1}^{-T}g_{\bar{m}+1}\,||}\,\frac{||D_{\bar{m}+1}^{-T}g_{\bar{m}+1}\,||}{||D_m^{-T}g_m\,||}\right] \\
&\geq \bar{L}^{-1}\;||D_m^{-T}g_m\,||\left[1-\varsigma-\epsilon(1+\varsigma)\right] \\
&\geq\; ||D_m^{-T}g_m\,||\,\frac{1}{2}(1-\varsigma)/\bar{L} \;.
\end{aligned} \tag{6.12}$$

Substituting this into (6.8) yields

$$f(x_m) - f(x_{\bar{m}+1}) \geq \bar{\epsilon}\;||D_m^{-T}g_m\,||^2 \tag{6.13}$$

where

$$\bar{\epsilon} = \frac{1}{2}\eta_1 c_3\,\epsilon\,\min\left\{\frac{\epsilon}{\sigma_1}, \frac{1-\varsigma}{2\sigma_3 c_2 \bar{L}}\right\} > 0 \;. \tag{6.14}$$

By hypothesis, f is nonincreasing and bounded below so that $f(x_k) \rightarrow f^*$ for some f^*. Thus for

any m, either $g_m = 0$, or

$$\begin{aligned} ||D_m^{-T} g_m||^2 &\leq (f(x_m) - f(x_{\bar{m}+1}))/\bar{\epsilon} \\ &\leq (f(x_m) - f^*)/\bar{\epsilon}. \end{aligned} \tag{6.15}$$

Therefore, $\lim_{k\to\infty} ||D_k^{-T} g_k|| = 0$, and by Lemma 3.4, $\lim_{k\to\infty} ||g_k|| = 0$ and $\lim_{k\to\infty} ||\nabla f(x_k)|| = 0$. □

REFERENCES

[1] CARTER, R.G. [1986]. Multi-model algorithms for optimization, TR86-3, Rice University, May 1986.

[2] CARTER, R.G. [1987a]. Safeguarding Hessian approximations in trust region algorithms, TR87-12, Rice University, June 1987.

[3] CARTER, R.G. [1987b]. Global convergence theory for linesearch and trust region algorithms, TR87-16, Rice University, July 1987.

[4] CARTER, R.G., and J.E.DENNIS, Jr. [1987c] A globally convergent framework for applying an expert systems approach to optimization, in preparation.

[5] DENNIS, J.E. Jr. and H.H. MEI [1979]. Two new unconstrained optimization algorithms which use function and gradient values, *J. Optim. Theory Appl.*, 28, pp. 453-482.

[6] DENNIS, J.E. Jr., D.M. GAY and R.E. WELSCH [1981]. An adaptive nonlinear least-squares algorithm, *TOMS*, 7, pp. 348-368.

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

[8] GAY, D.M. [1981]. Computing optimal locally constrained steps, *SIAM J. Sci. Statist. Comput.*, 2, pp. 186-197.

[9] GOLDFELDT, S.M., R.E. QUANDT, and H.F. TROTTER [1966]. Maximization by quadratic hill-climbing, *Econometrica*, 34, pp. 541-551.

[10] MORE. J. [1982]. Recent developments in algorithms and software for trust region methods, Argonne National Labs Report ANL/MCS-TM-2.

[11] POWELL, M.J.D. [1970]. A new algorithm for unconstrained optimization, in *Nonlinear Programming*, J.B. Rosen, O.L. Mangasarian and K. Ritter, eds., Academic Press, New York, pp. 31-65.

[12] POWELL, M.J.D. [1975]. Convergence properties of a class of minimization algorithms, *Nonlinear Programming*, 2, O.L. Mangasarian, R.R. Meyer, S.M. Robinson, eds., Academic Press, New York, pp. 1-27.

[13] POWELL, M.J.D. [1984]. On the global convergence of trust region algorithms for unconstrained minimization, *Math. Programming*, 29, pp. 297-303.

[14] SCHULTZ, G.A., R.B. SCHNABEL and R.H. BYRD [1985]. A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties, *SIAM J. Numer. Anal.*, 22, pp. 47-67.

[15] SORENSEN, D.C. [1982]. Newton's method with a model trust region modification, *SIAM J. Numer. Anal.*, 19, pp. 409-426.

[16] STEIHAUG, T. [1980]. Quasi-Newton Methods for Large Scale Nonlinear Problems, Ph.D dissertation, SOM Technical Report #49, Yale University.

[17] STEIHAUG, T. [1981]. The conjugate gradient method and trust regions in large scale optimization, Department of Mathematical Sciences, TR81-1, Rice University.

[18] THOMAS, S.W. [1975]. Sequential estimation techniques for quasi-Newton algorithms, Technical Report TR75-227, Department of Computer Science, Cornell University.