# IBM Research Report

## A gradient-dependent trust region algorithm for the minimization of unconstrained nonlinear partially separable functions

**Johara Shahabuddin**
IBM Research Division
IBM India Research Lab
Block I, I.I.T. Campus, Hauz Khas
New Delhi - 110016. India.

**Abstract.** Trust region algorithms are strongly convergent, typically restricting the step to lie within a spherical trust region. Structured trust region algorithms attempt greater efficiency by allowing differing trust region radii in different partially separable subspaces. However, the unpredictable shape of this trust region takes away some convergence strength for naive implementations. Restrictions on the step have been proposed in earlier work to correct this.

We propose a trust region radius update mechanism that depends on the change in gradient direction between iterations and thus avoids restrictions on the step. To simplify the analysis, we limit ourselves to the unconstrained problem, and show first and second order global convergence. We also make the simplification that the range spaces are defined by canonical basis vectors.

**Key words.** trust region algorithm, partial separability, unconstrained, global convergence, structured problem, nonlinear programming, large-scale programming

## 1. Introduction.

Partial separability is a form of sparsity in systems that are composed of loosely-connected subsystems. Large optimization problems often have such a structure. A nonlinear function $f(x)$, $x \in \Re^n$ is defined to be **partially separable** if it can be written as

$$f(x) = \sum_{i=1}^{p} f_i(x),$$

where each $f_i(x),\ i = 1, \ldots, p$, is a nonlinear function with a large invariant subspace. The functions $f_i(x),\ i = 1, \ldots, p$, are known as **element functions**.

Trust region algorithms build a quadratic model of the objective function in each iteration and solve the model for an approximate minimum within a region known as the trust region. Typically, this region is a sphere defined by a radius, restricting the length of the step equally in all directions.

For a partially separable objective function, if a different radius defines a separate trust region for each of the various element functions, longer steps may be possible in the range spaces of the more accurately modeled element functions. A family of structured trust region algorithms have been proposed to investigate this possibility in [1].

The robustness of the basic trust region mechanism is affected by structuring. This well-known mechanism is: when the trust region size decreases, the function behavior is more accurately predicted by its exact quadratic model. If we look more carefully, we find this is not true for all directions, only for some such as the negative gradient direction, major directions of negative curvature, or the Newton direction, where the quadratic model dominates over higher order terms in the Taylor series expansion of the objective.

With a spherical trust region, only the good directions are ever taken, but a structured trust region shape may become skewed against these good directions, especially after a successful step is taken. For example, the trust region shape may allow only a relatively tiny step in the direction of the negative gradient, while allowing a long step in a direction orthogonal to it.

Conditions on the step were proposed in [1] in order to deal with this problem. Ideally, the trust region update mechanism should naturally take care of it by biasing the trust region shape towards the good directions. The method we propose here does just this, avoiding any conditions (beyond a conventional sufficient decrease condition) on the step. At the heart of the algorithm is an update mechanism that takes into account the change of gradient direction when an iterative step is taken.

The above problem does not affect the first order convergence results in [1] that show $\liminf \alpha_k = 0$, where $\alpha_k$ is a first order criticality measure. The extra conditions in [1] come into play only later, in the proof of $\lim_{k\to\infty} \alpha_k = 0$. (In [1], a sufficient decrease condition is used that the exact minimum of the model in the iterative subproblem cannot always attain. Thus, an additional restriction on the step is hidden within it.) This implies that in the limit the algorithms would move between points with differing gradient values. Expanding the trust region shape to favor some of the good directions described above, so as to compensate for the new slope conditions at the each iterate, we ensure moving away from noncritical points towards the critical points in the limit.

Thus, our proposed trust region update mechanism has two parts: a normal structured update based on model accuracy, on top of which there is an extra gradient-based increase in the elemental trust region radii in each iteration. We only have proofs for the unconstrained case so far, since the analysis here is complicated. The subproblem we use in the theory here has a 'box' trust region configuration rather than a more general structure, due to an assumption that each elemental trust region has a range space defined by canonical basis vectors.

We describe the algorithm in the next section. First and second order convergence results are obtained in the following two sections.

## 2. The gradient-dependent algorithm.

We begin with basic definitions and assumptions.

The unconstrained nonlinear minimization problem which we address is:

$$(P) \qquad \min f(x),$$

where $f : \Re^n \to \Re$ is partially separable.

The problem $(P)$ is solved iteratively, with $x_0$ as the given starting point. In each iteration, $f(x_k+s)-f(x_k), k = 0, 1, 2, \ldots$ is modeled in terms of its gradient and Hessian. The model, denoted by $m_k(s)$, is approximately minimized at $s_k$. The trust regions of the elements are then updated, and so is $x_k$. An iteration where $x_{k+1} = x_k + s_k$, is called a *successful iteration*. Otherwise, $x_{k+1} = x_k$, and the iteration is called *unsuccessful*.

The $l_2$-norm is used throughout this paper, unless otherwise specified.

The following basic assumptions are needed on $f$:

**Assumption 2.1.** *The function $f$ is bounded below on the set $\mathcal{L} := \{x \ : \ f(x) \leq f(x_0)\}$.*

**Assumption 2.2.** *Each $f_i$, $i = 1, \ldots, p$, and hence $f$, is twice continuously differentiable on an open set containing $\mathcal{L}$.*

**Assumption 2.3.** *There exists a positive constant $\chi_H \geq 1$ such that $\|\nabla^2 f(x)\| \leq \chi_H$ and $\|\nabla^2 f_i(x)\| \leq \chi_H$, $i = 1, \ldots, p$, on an open set containing $\mathcal{L}$.*

Define $g_k := \nabla f(x_k)$. Let $B_k$ be an approximation to the Hessian $\nabla^2 f(x_k)$.

DEFINITION 2.4. *The* **overall model** *$m_k(s)$ of $f(x_k + s) - f(x_k)$ is defined as:*

$$m_k(s) := g_k^T s + \frac{1}{2} s^T B_k s,$$

Each element function is modeled, in terms of the first three terms of its Taylor series. Let $g_{i,k} := \nabla f_i(x_k)$. Let $B_{i,k}$ be an approximation to $\nabla^2 f_i(x_k)$ such that $\sum_{i=1}^{p} B_{i,k} = B_k$.

DEFINITION 2.5. *The* **elemental model** *for $f_i(x_k + s) - f_i(x_k)$ is defined as follows:*

$$m_{i,k}(s) := g_{i,k}^T s + \frac{1}{2} s^T B_{i,k} s.$$

If an iteration is unsuccessful, so that $x_{k+1} = x_k$, then $B_{k+1} = B_k$ and $B_{i,k+1} = B_{i,k}$ for all $i$, so that the overall and elemental models are unchanged.

**Hessian approximation and Rayleigh quotient.** Several different assumptions related to the Hessian approximations $B_k$ have been used by earlier authors. We adopt the one used in [1],where it has been explained to be weaker than assuming that $B_k$ is uniformly bounded.

DEFINITION 2.6. *The* generalized Rayleigh quotient *of a function $f$ at $x$ along $s \neq 0$ is defined to be:*

$$\omega(f, x, s) := \frac{2}{\|s\|^2}[f(x + s) - f(x) - \nabla f(x)^T s].$$

Because of the assumption that $\nabla^2 f$ is bounded, we have $|\omega(f_i, x, s)| \leq L_h$ for all $i$ if $x$ and $x + s$ lie in $\mathcal{L}$, where $L_h \geq 1$ is a positive constant. We define a version of the generalized Rayleigh quotient of $m_k$:

DEFINITION 2.7.

$$\beta_k := 1 + \max_{q=1,\ldots,k} (\max(|\omega(m_q, 0, s_q)|, \max_{i=1,\ldots,p} (|\omega(m_{i,q}, 0, s_q)|))).c \tag{2.1}$$

Given these definitions, the assumption is:

**Assumption 2.8.** $\sum_{k=0}^{\infty} \frac{1}{\beta_k} = +\infty.$

**The trust region structure.** We define $\Delta_{i,k}, i = 1, \ldots, p$ to be the trust region radii for the $p$ element functions. These are updated in each iteration, and together define the overall trust region structure in the following manner:

DEFINITION 2.9. *The null space $N$ of a function $f(x)$ is defined to be the set $\{v \mid f(x + v) = f(x)\}$.*

DEFINITION 2.10. *The range space $R$ of a function $f(x)$ is defined to be the subspace orthogonal to $N$ in $R^n$.*

Let $R_i$ denote the range space of an element function $f_i, i = 1, \ldots, p$. Elemental models $m_{i,k}$ have the same range space $R_i$ as $f_i$, for all $i, k$. We simplify the trust region subproblem by making an assumption on the range spaces.

**Assumption 2.11.** *Each $R_i$ is a coordinate subspace, i.e., the span of some set $e_j, j \in q_i$, where $q_i \subset \{1, \ldots, n\}$ and $e_j$ denotes the $j$th unit vector, $j = 1, \ldots, n$.*

The partial separability assumption for $f$ implies that for the most part, each $e_j$ is contained in only a few of the $R_i$'s. We define:

$$p_j := \{i : e_j \in R_i\} \subset \{i = 1, \ldots, p\},$$

where $p_j, j = 1, \ldots, n$ maps a particular coordinate index $j$ onto the set of elements which use it. Notice that the reverse transformation from an element to a set of coordinates is given by $q_i$.

Now, each *elemental trust region* is given by the following constraint:

$$\|P_{R_i}(s_k)\|_\infty \leq \Delta_{i,k},$$

where $P_{R_i}(s)$ denotes the projection of a vector $s$ onto $R_i$. Because we use the $\infty$-norm, the elemental trust region constraints intersect to give upper and lower bounds on each coordinate, together defining a box-shaped *overall trust region.*

We can now fully state the subproblem $(SP)$ to be solved in each iteration:

$$(SP) \qquad \min m_k(s) = g_k^T s + \frac{1}{2} s^T B_k s$$
$$|s_j| \leq \Delta_k^j, \qquad j = 1, \ldots, n,$$

where $s_j, j = 1, \ldots, n$ is the $j$th component of the step $s$, and $\Delta_k^j := \min_{i \in p_j} \Delta_{i,k}$ is the *trust region radius for coordinate* $j$, and is defined as the minimum of the elemental trust region radii that affect that coordinate.

Having solved for $s_k$, the minimizing value of $s$, we update each elemental trust region radius twice. In the first update, *separation criteria* divide the elements into one of three classes depending on how accurately each element function $f_i$ is modeled. Also, depending on how accurately the overall function $f$ is modeled, the iteration is put into one of three classes. The combination is used to update the radii. The second update is to expand some of the elemental trust regions from the previous update, depending on the change in the gradient.

To check the accuracy of fit between functions and models, the following terms will be used:

$$\begin{aligned}
\delta f_k &= f(x_k) - f(x_k + s_k), \\
\delta f_{i,k} &= f_i(x_k) - f_i(x_k + s_k), \\
\delta m_k &= -m_k(s_k), \\
\delta m_{i,k} &= -m_{i,k}(s_k), \\
r_k &= \delta f_k / \delta m_k, \\
r_{i,k} &= \delta f_{i,k} / \delta m_{i,k}.
\end{aligned}$$

**Separation criteria.** The elemental trust region update is done in two stages, as described in the introduction. In the first stage, we update according to the following separation criterion (a way to classify the model accuracy for the element functions), and this stage is in common with the approach in [1], although they use a different separation criterion.

We assign each iteration a parameter $\tau_k$. If $r_k \geq \mu_2$ then $\tau_k = 2$, if $\mu_1 \leq r_k < \mu_2$ then $\tau_k = 1$, and otherwise $\tau_k = 0$.

2.12. **Parallel separation criterion.** *Let* $0 < \mu_1 \leq \mu_2 < 1$.
*If* $\delta f_{i,k} \geq \delta m_{i,k} - \frac{(1-\mu_2)}{p} \delta m_k$ *then* $\tau_{i,k}^1 = 2$,
*if* $\delta f_{i,k} < \delta m_{i,k} - \frac{(1-\mu_1)}{p} \delta m_k$ *then* $\tau_{i,k}^1 = 0$,
*else* $\tau_{i,k}^1 = 1$.

We calculate $\tau_{i,k} = \tau_k + \tau_{i,k}^1$, and update the elemental trust region radii as follows.
Let $0 < \gamma_1 \leq \gamma_2 < 1 \leq \gamma_3 \leq \gamma_4$.
If $\tau_{i,k} = 4$ then $\Delta_{i,k+1} = [1, \gamma_4]\Delta_{i,k}$,
if $\tau_{i,k} = 3$ then $\Delta_{i,k+1} = [1, \gamma_3]\Delta_{i,k}$,
if $\tau_{i,k} = 2$ then $\Delta_{i,k+1} = \Delta_{i,k}$,
if $\tau_{i,k} = 1$ then $\Delta_{i,k+1} = [\gamma_2, 1]$,
if $\tau_{i,k} = 0$ then $\Delta_{i,k+1} = [\gamma_1, \gamma_2]\Delta_{i,k}$.

The following result is of interest. It shows that in an unsuccessful iteration at least one elemental trust region radius will decrease, and that, less crucially for proving convergence, in a successful iteration at least one elemental trust region radius must increase.

LEMMA 2.13. *For the criteria described above, if all $\tau^1_{i,k} \geq 1$ then $\tau_k \geq 1$. And if all $\tau^1_{i,k} \leq 1$, then $\tau_k \leq 1$.*

*Proof.* We prove only the first assertion. The proof of the second assertion is similar.

If $\tau^1_{i,k} \geq 1$ for all $i$, then

$$\sum_{i=1}^{p} \delta f_{i,k} \geq \sum_{i=1}^{p} (\delta m_{i,k} - \frac{(1-\mu_1)}{p} \delta m_k).$$

So, $\delta f_k \geq \delta m_k - (1-\mu_1)\delta m_k$, and thus $\delta f_k \geq \mu_1 \delta m_k$. □

**The gradient expansion.** The gradient-based expansion of elemental trust region radii is a second stage of update applied to all the elemental trust region radii. As motivated in the introduction, we want to allow the algorithm to step away from non-critical points in the limit, by letting the trust region size expand towards the negative gradient direction at the new point.

We first define a function $g'_{i,k}$ of the overall gradient $g_k$, that is the *subvector*, or 'piece' of $g_k$ that corresponds to the coordinates associated with the $i$th element function.

DEFINITION 2.14. *The partial elemental gradient $g'_{i,k} \in \Re^{|q_i|}$, is defined to be the subvector of $g_k$ such that $g^j_k$ is a component of $g'_{i,k}$ if and only if $j \in q_i$.*

Let $I^*$ denote the set of elements which have $\|g'_{i,k}\|_\infty = \|g_k\|_\infty$.

Let $i^*$ denote an element $i$ in $I^*$ which has the smallest $\Delta_{i,k}$.

The update can now be stated:

$$\Delta_{i,k+1} = \max(\frac{\|g'_{i,k+1}\|_\infty}{\|g'_{i^*,k}\|_\infty} \Delta_{i^*,k}, \Delta_{i,k+1}), \quad i = 1, \ldots, p. \tag{2.2}$$

This update increases the trust region size of an element whose partial gradient is large in relation to $\|g'_{i^*,k}\|_\infty = \|g_k\|_\infty$. We used $I^*$ because the update $\Delta_{i,k+1} = \max(\frac{\|g'_{i,k+1}\|_\infty}{\|g'_{i,k}\|_\infty} \Delta_{i,k}, \Delta_{i,k+1})$ is not practical (since the denominator of the ratio of gradient subvectors may be equal to zero). The $\infty$-norm of the subvectors is needed for Lemma 3.3, which we could not prove with the $l_2$-norm of the subvectors.

**Sufficient decrease condition.** Trust region algorithms do not need an exact solution to the subproblem to converge. Approximate solutions, that guarantee sufficient decrease in the value of the model, do just as well. One such sufficient decrease condition on the approximate step, that is quite typical of the ones existing in the literature, is in [4]:

$$\delta m_k \geq \kappa \|g_k\| \min(\frac{\|g_k\|}{\|B_k\|}, \Delta_k),$$

where $\Delta_k$ is the single trust region radius.

The gradient-dependent algorithm requires the following sufficient decrease condition.

$$\delta m_k \geq \kappa \max_{j \in \{1,\ldots,n\}} (|g^j_k| \min(\frac{|g^j_k|}{|b^{jj}_k|}, \Delta^j_k)) \tag{2.3}$$

where, $\delta m_k = -m_k(s_k)$, $0 < \kappa < 1$, $g^j_k$ is the $j$th coordinate of $g_k$, and $b^{jj}_k$ is the $j$th diagonal element of the matrix $B_k$.

We now show that such a decrease is achievable.

LEMMA 2.15. *There exists a step $s_k$ for which (2.3) holds for any constant $0 < \kappa \leq \frac{1}{2}$.*

*Proof.* Consider first the decrease obtained when we minimize the model along the direction $d_{jk} = -\text{sgn}(g^j_k)e_j$ for a given $j$. Let $s^*_j$ be the minimizer of $m_k(s)$ over $s = td_{jk}$ such that $t \leq \Delta^j_k$. Then $m_k(s) = -\text{sgn}(g^j_k)tg^T_k e_j + \frac{t^2}{2}e^T_j B_k e_j = -t|g^j_k| + \frac{t^2}{2}b^{jj}_k$. When $b^{jj}_k > 0$, $t^* = \frac{|g^j_k|}{b^{jj}_k}$ minimizes $m_k(s)$.

If $b^{jj}_k > 0$ and $t^* \leq \Delta^j_k$, then $s^*_j = t^* d_{jk}$, and $m_k(s^*_j) = -\frac{|g^j_k|^2}{2b^{jj}_k}$. If $b^{jj}_k > 0$ and $t^* \geq \Delta^j_k$, then $s^*_j = \Delta^j_k d_{jk}$, and $m_k(s^*_j) \leq -|g^j_k|\Delta^j_k + \frac{1}{2}(\Delta^j_k)^2 \frac{|g^j_k|}{\Delta^j_k} = -\frac{1}{2}|g^j_k|\Delta^j_k$, since $\Delta^j_k \leq t^*$ implies that $b^{jj}_k \leq \frac{|g^j_k|}{\Delta^j_k}$.

Now if $b^{jj}_k \leq 0$, then $t^2 b^{jj}_k \leq 0$ and so $m_k(s) \leq -t|g^j_k|$. Choosing $t = \Delta^j_k$, we get $m_k(s^*_j) \leq -|g^j_k|\Delta^j_k$.

Thus, $m_k(s_j^*) \le \kappa(|g_k^j| \min(\frac{|g_k^j|}{|b_k^{jj}|}, \Delta_k^j))$. To obtain the result, we choose

$$s_k = \arg \min_{j\in\{1,\ldots,n\}} m_k(s_j^*). \tag{2.4}$$

□

COROLLARY 2.16. *There exists a step for which the following sufficient decrease condition holds for any* $0 < \kappa \le \frac{1}{2}$*:*

$$\delta m_k \ge \kappa \max_{j\in\{1,\ldots,n\}} (|g_k^j| \min(\frac{|g_k^j|}{\beta_k}, \Delta_k^j)). \tag{2.5}$$

*Proof.* Suppose we choose $s_k = s_j^*$ as in (2.4). Then $\beta_k \ge |b_k^{jj}|$, so that (2.3) implies (2.5). □

The above proofs show that the exact solution to $(SP)$ will always satisfy each of the two sufficient decrease conditions above. The condition (2.5) is used to prove the first order results. (It allows more flexibility in the choice of step than (2.3) allows.) In the third section we will use (2.3) to prove our second order results. The results that hold with (2.5) would continue to hold with (2.3), since the latter implies the former.

2.17. **The gradient-dependent algorithm.** *Given* $0 < \mu_1 \le \mu_2 < 1$*, a feasible* $x_0$*, and starting values for the trust region sizes such that* $\Delta_{i,0} \ge \|g_0\|$*, for all* $i = 1,\ldots,p$*, the kth iteration takes the following form:*

1. *Find an approximate solution* $s_k$ *to the subproblem* $(SP)$ *that satisfies a sufficient decrease condition.*
2. *Evaluate* $f(x_k + s_k)$ *and hence* $r_k$*.*
3. *Update the trust region radii according to one of the separation criteria, such as (2.12).*
4. *If* $r_k < \mu_1$ *then* $x_{k+1} = x_k$ *and the iteration ends here.*
   *Else* $x_{k+1} = x_k + s_k$*, calculate* $g_{k+1}, B_{k+1}$ *and go on to step 5.*
5. *Reset the trust region radii according to (2.2).*

We have now stated the assumptions, the structure of the trust region and how it is updated in each iteration, the sufficient decrease conditions we will use and finally, the algorithm. We now go on to our convergence results.

## 3. First Order Convergence.

We will show that every limit point of the sequence of $x_k$'s generated by the algorithm must be a critical point. The step must satisfy the sufficient decrease condition (2.5). The parallel separation criterion is used for the first trust region update stage.

We begin by a lemma that gives an upper bound to the difference between the change in the model and the change in the function for a given step.

LEMMA 3.1. *If Assumption 2.3 holds, then there exists a constant* $L \ge 1$ *such that for each* $k = 0, 1, 2, \ldots$*,* $|\delta f_k - \delta m_k| \le L\beta_k \|s_k\|^2$*, and for each* $i = 1, \ldots, p$*,* $|\delta f_{i,k} - \delta m_{i,k}| \le L\beta_k \|s_{i,k}\|^2$*, where* $s_{i,k}$ *is the projection of* $s_k$ *onto the range space* $R_i$*.*

*Proof.*

$$\begin{aligned}
|\delta f_k - \delta m_k| &= |f(x_k) - f_(x_k + s_k) - m(0) + m(s_k)| \\
&= |g_k^T s_k + \frac{1}{2}\|s_k\|^2 \omega(f, x_k, s_k) - g_k^T s_k - \frac{1}{2}\|s_k\|^2 \omega(m_k, 0, s_k)| \\
&\le \frac{1}{2}\|s_k\|^2 (|\omega(f, x_k, s_k)| + |\omega(m_k, 0, s_k)|) \\
&\le \frac{1}{2}(L_h + \beta_k)\|s_k\|^2 \\
&\le \frac{1}{2}(L_h + 1)\beta_k \|s_k\|^2 \\
&\le L\beta_k \|s_k\|^2,
\end{aligned}$$

where $L := \frac{1}{2}(L_h + 1)$ and $L_h \le \chi_H$ is an upper bound on the generalized Rayleigh quotient of $f$ for any choice of $x$ and $s$.

The proof for the elemental differences $|\delta f_{i,k} - \delta m_{i,k}| \le L\beta_k \|s_{i,k}\|^2$ is similar. □

LEMMA 3.2. *If, for any* $j = 1, \ldots, n$*,* $\Delta_k^j \le c_1 |g_k^j| / \beta_k$*, where* $c_1 := \frac{(1-\mu_2)\kappa\gamma_1^2}{pL}$*, and if* $\Delta_{i,k} \le \frac{\Delta_k^j}{\gamma_1}$ *for some* $i \in p_j$*, then* $\Delta_{i,k+1} \ge \Delta_{i,k}$*.*

*Proof.* Notice $c_1 \leq 1$ (all terms in the numerator are less than or equal to 1, all terms in the denominator are greater than or equal to 1). Hence $\Delta_k^j \leq c_1|g_k^j|/\beta_k$ implies that $\min(\frac{|g_k^j|}{\beta_k}, \Delta_k^j) = \Delta_k^j$. We substitute this into the sufficient decrease condition (2.5) to get $\delta m_k \geq \kappa |g_k^j| \Delta_k^j$.

Let $i \in p_j$ such that $\Delta_{i,k} \leq \frac{\Delta_k^j}{\gamma_1}$. From Lemma 3.1, $\delta m_{i,k} - \delta f_{i,k} \leq |\delta m_{i,k} - \delta f_{i,k}| \leq L\beta_k \Delta_{i,k}^2 \leq L\beta_k (\frac{\Delta_k^j}{\gamma_1})^2 \leq L\beta_k \frac{\Delta_k^j}{\gamma_1^2} \frac{\delta m_k}{\kappa |g_k^j|} \leq \frac{(1-\mu_2)}{p} \delta m_k$, substituting in $\Delta_k^j \leq c_1 |g_k^j|/\beta_k$ and the value of $c_1$. Thus, looking at steps 3 and 5 of the algorithm, by the parallel separation criterion, the $i$th element is not a candidate for reduction in its trust region size. So we have $\Delta_{i,k+1} \geq \Delta_{i,k}$. ◻

LEMMA 3.3. *For the sequence of iterates generated by the algorithm* $\Delta_k^j \geq c_1\gamma_1 |g_k^j|/\beta_k$, *for all* $j = 1, \ldots, n$.

*Proof.* This will be proved by induction. Notice that $\Delta_0^j$'s satisfy the lemma by our choice of $\Delta_{i,0}$ in Algorithm 2.17. We assume that this lemma holds for iteration $k$ and will now prove it will hold for iteration $k+1$.

If the iteration is unsuccessful ($x_{k+1} = x_k$) and $\Delta_k^j \geq c_1 |g_k^j|/\beta_k$ for some $j$, then for all $i \in p_j, \Delta_{i,k+1} \geq \gamma_1 \Delta_{i,k} \geq \gamma_1 \Delta_k^j \geq \gamma_1 c_1 |g_k^j|/\beta_k \geq \gamma_1 c_1 |g_{k+1}^j|/\beta_{k+1}$ since $|g_k^j| = |g_{k+1}^j|$ and $\beta_k$ is an increasing sequence, and the lemma holds.

If the iteration is unsuccessful and $\Delta_k^j \leq c_1 |g_k^j|/\beta_k$ for some $j$, then for $i \in p_j$, there are two possibilities. First, $\Delta_{i,k} \leq \frac{\Delta_k^j}{\gamma_1}$. Applying Lemma 3.2, $\Delta_{i,k+1} \geq \Delta_{i,k}$. Or second, $\Delta_{i,k} > \frac{\Delta_k^j}{\gamma_1}$, which implies that $\Delta_{i,k+1} \geq \gamma_1 \Delta_{i,k} > \Delta_k^j$. And so either way, $\Delta_{k+1}^j = \min_{i \in p_j} \Delta_{i,k+1} \geq \min_{i \in p_j} \Delta_{i,k} = \Delta_k^j$. Hence, $\Delta_{k+1}^j \geq \gamma_1 c_1 |g_{k+1}^j|/\beta_{k+1}$, as before.

Else, if the iteration is successful ($x_{k+1} = x_k + s_k$) then (2.2) we update the elemental trust regions so that $\Delta_{i,k+1} \geq \frac{\|g'_{i,k+1}\|_\infty}{\|g'_{i^*,k}\|_\infty} \Delta_{i^*,k}$. Note that for all $j \in q_i^*$, $\Delta_{i^*,k} \geq \Delta_k^j \geq c_1\gamma_1 |g_k^j|/\beta_k$ (by the induction hypothesis). Hence, $\Delta_{i^*,k} \geq c_1\gamma_1 \|g'_{i^*,k}\|_\infty/\beta_k$. From the update expression given above, $\Delta_{i,k+1} \geq c_1\gamma_1 \|g'_{i,k+1}\|_\infty/\beta_k \geq c_1\gamma_1 |g_{k+1}^j|/\beta_{k+1}$ for all $j \in q_i$ and $i = 1, \ldots, p$. So, $\Delta_{k+1}^j = \min_{i \in p_j} \Delta_{i,k+1} \geq c_1\gamma_1 |g_{k+1}^j|/\beta_{k+1}$. ◻

THEOREM 3.4. *If the algorithm has infinitely many successful iterations, then*

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

*Proof.* Suppose that $\limsup_{k \to \infty} \|g_k\| > 0$. Then there exists an infinite subsequence $\{l\}$ of successful iterates and a $j$ such that for some $\epsilon > 0$, $|g_l^j| > \epsilon$. For each successful iteration, $\delta f_l \geq \mu_1 \delta m_l \geq \mu_1 \kappa |g_l^j| \min(\frac{|g_l^j|}{\beta_k}, \Delta_l^j)$ from (2.5). Therefore by Lemma 3.3 we get $\delta f_l \geq \mu_1 \kappa \epsilon \min(\frac{\epsilon}{\beta_k}, \frac{c_1\gamma_1\epsilon}{\beta_k}) = \mu_1\kappa\gamma_1 c_1 \epsilon^2/\beta_k$. But Assumption 2.8 now implies that $f$ is unbounded below and we have arrived at a contradiction. Thus, $\limsup_{k\to\infty} \|g_k\| = 0$. ◻

THEOREM 3.5. *If the algorithm has finitely many successful iterations, then there is a single limit point* $x_*$, *such that* $g(x_*) = 0$.

*Proof.* From step 4 of the algorithm, a finite number of successful iterations means $x_k$ is unchanged for $k$ large enough, and that $x_* = x_l$, where $l-1$ is the index of the last successful iteration. Now we assume (for a proof by contradiction) $\|g(x_*)\| > 0$. Hence, there exists an $\epsilon > 0$ for which $|g_*^j| = \epsilon$ for some $j$. By Lemma 3.3, there must then exist $\epsilon_1 > 0$ satisfying $\Delta_k^j \geq \epsilon_1/\beta_k$ for all $k \geq l$, and therefore (from (2.5)), also an $\epsilon_2 > 0$ such that for all $k \geq l$

$$\delta m_k \geq \epsilon_2/\beta_k. \tag{3.1}$$

Since iteration $l-1$ is followed by only unsuccessful iterations and in each such iteration at least one $\Delta_{i,k}$ reduces (from Lemma 2.13), $\lim_{k\to\infty} \Delta_{i,k} = 0$ for at least one $i \in \{1, \ldots, p\}$. We now show that our assumption $\|g(x_*)\| > 0$ contradicts this.

First note that, because the overall and elemental models are unchanged after the $(l-1)$ iteration, all subsequent $\beta_k$'s are bounded: for $k \geq l$, $\beta_k \leq \max(\beta_l, \max_i(\|B_{i,l}\|+1), \|B_l\|+1)$.

Now let $k$ be the first iteration after the $(l-1)$st where $\Delta_{i,k} < c_2/\beta_k$, where $c_2 := \gamma_1(\frac{(1-\mu_2)\epsilon_2}{pL})^{1/2}$, for some $i$. Without loss of generality we can choose $\epsilon_2$ small enough so that $\Delta_{i,l} > c_2/\beta_l$ for all $i$ and so $k > l$. Thus, $\Delta_{i,k-1} < \Delta_{i,k}/\gamma_1 < \frac{c_2}{\beta_k\gamma_1}$. Now $\delta m_{i,k-1} - \delta f_{i,k-1} \leq L\beta_{k-1}\Delta_{i,k-1}^2$ (from Lemma 3.1) $\leq \frac{L\beta_{k-1}c_2^2}{\beta_k^2\gamma_1^2}$ (substituting for $\Delta_{i,k-1}$) $\leq \frac{(1-\mu_2)\epsilon_2}{p\beta_k}$ (from the definition of $c_2$) $\leq \frac{(1-\mu_2)}{p}\delta m_k$

(from (3.1)). Thus, $\Delta_{i,k} < c_2/\beta_k$ is not possible for any $i = 1, \ldots, p$ or $k \geq l$. But recall that $\beta_k$ is bounded. Hence $\lim_{k\to\infty} \Delta_{i,k} = 0$ cannot hold for any $i$, giving us a contradiction. □

**4. Second order convergence.** We begin with changes in our assumptions from the first order theory. For Algorithm 2.17, we find that only the weaker second order results from among the ones proved for unstructured trust region algorithms, as in [4], can be proved. We could not show that there always exists a step such that the sufficient decrease condition (2.3) for first order convergence, and (4.3) required for the stronger second order results, could be simultaneously satisfied. We then make a change in the algorithm to prove the stronger results.

All the assumptions (Assumptions 2.1–2.3, 2.8 and 2.11) in the first order theory stand. We must strengthen our assumptions on the second derivative as follows: We assume that the exact second derivative is available to us.

**Assumption 4.1.** *$B_k = \nabla^2 f(x_k)$ and $B_{i,k} = \nabla^2 f_i(x_k)$, $i = 1, \ldots, p$.*

Thus, Assumption 2.3 about the boundedness of the exact second derivatives now applies to $B_k$:

**Assumption 4.2.** *There exists a positive constant $\chi_B \geq 1$ such that $\|B_k\| \leq \chi_B$ and $\|B_{i,k}\| \leq \chi_B$, $i = 1, \ldots, p$, for all $k$.*

We also require:

**Assumption 4.3.** *$\nabla^2 f_{i,k}$ is Lipschitz continuous with a constant $L_c$ for all $i = 1, \ldots, p$.*

This allows us the following version of Lemma 3.1

LEMMA 4.4. *If Assumptions 2.3 and 4.2 hold, $|\delta f_k - \delta m_k| \leq L\|s_k\|^2$ and $|\delta f_{i,k} - \delta m_{i,k}| \leq L\|s_{i,k}\|^2$, for all $i = 1, \ldots, p$ and all $k$, where $L := \frac{1}{2}(L_h + \chi_B) \geq 1$, where $\chi_B$ is an upper bound on $\|B_k\|$ and $\|B_{i,k}\|$.*

The proof of this lemma is similar to that of Lemma 3.1.

We can guarantee (simply by choosing the best among the coordinate direction steps and a step $s_k$ such that $\|s_k\| = \Delta_{\min,k} = \min_{1\leq i\leq p} \Delta_{i,k}$ along a direction of sufficient negative curvature) that both (2.3) and the following condition are satisfied:

$$\delta m_k \geq -\kappa_1 \lambda_k \Delta_{\min,k}^2, \tag{4.1}$$

where $\kappa_1$ is a small positive constant and $\lambda_k$ is the minimum eigenvalue of the second derivative of $f$ at $x_k$. Further, if the step described above happens to lie along one of the coordinate directions and $\|s_k\| < \Delta_{\min,k}$, we reset $s_k$ to be the the above step plus an additional step along a negative curvature direction that is a non-ascent direction for the model at the intermediate point, until we reach the edge of the trust region, thereby ensuring that:

$$\|s_k\| \geq \Delta_{\min,k}. \tag{4.2}$$

Notice that, (4.2) also holds for the exact minimizer of the subproblem $(SP)$, whenever a direction of negative curvature exists. We now prove a theorem resembling standard results for trust region algorithms, in particular, as in [4].

THEOREM 4.5. *Suppose at each iteration $s_k$ satisfies conditions (2.3) and (4.1), with $\|s_k\| \geq \Delta_{\min,k}$ when $\lambda_k < 0$. With the parallel separation criterion, for the sequence $\{x_k\}$ generated by Algorithm 2.17:*

*(a) The sequence $\{g_k\}$ converges to zero.*

*(b) If $\{x_k\}$ is bounded then there is a limit point $x_*$ with $\nabla^2 f_*$ positive semidefinite.*

*Proof.*

(a) We need to show that the first order results still hold when the requirement on $s_k$ is the sufficient decrease condition (2.3) rather than (2.5).
We start by showing a slightly different Lemma 3.2: defining $c_1$ as before, we show that if $\Delta_k^j \leq c_1|g_k^j|$ and $\Delta_{i,k} \leq \frac{\Delta_k^j}{\gamma_1}$ for some $i \in p_j$, then $\Delta_{i,k+1} \geq \Delta_{i,k}$. The proof is similar to that of Lemma 3.2. Due to the uniform boundedness of $\nabla^2 f(x_k)$, Lemma 4.4 can be used now instead of Lemma 3.1. Since $Lc_1 < 1$ and $L > |b_k^{jj}|$, where $L = \frac{1}{2}(L_h + \chi_B)$ is as defined in Lemma 4.4, we have $\Delta_k^j \leq c_1|g_k^j| \leq Lc_1|g_k^j|/|b_k^{jj}| \leq |g_k^j|/|b_k^{jj}|$ which implies that $\min(\frac{|g_k^j|}{b_k^{jj}}, \Delta_k^j) = \Delta_k^j$. We substitute this into (2.3) to get $\delta m_k \geq \kappa|g_k^j|\Delta_k^j$. Using the result of Lemma 4.4, for $i \in p_j$ such that $\Delta_{i,k} \leq \frac{\Delta_k^j}{\gamma_1}$ we have $\delta m_{i,k} - \delta f_{i,k} \leq L\Delta_{i,k}^2 \leq L(\frac{\Delta_k^j}{\gamma_1})^2 \leq L\frac{c_1|g_k^j|}{\gamma_1^2}\frac{\delta m_k}{\kappa|g_k^j|} \leq \frac{(1-\mu_2)}{p}\delta m_k$ and the rest of the proof follows.

Again, we have a modified Lemma 3.3 where we show $\Delta_k^j \geq c_1\gamma_1|g_k^j|$ (instead of $\Delta_k^j \geq c_1\gamma_1|g_k^j|/\beta_k$, which we had before). The proof applies Lemma 3.2 as modified above, and is similar to the one given in Section 1, the only change being to replace each mention of $\beta_k$ or $\beta_{k+1}$ by 1.

The proof of Theorem 3.4 is changed by replacing $\beta_k$ with $|b_k^{jj}|$. But since $|b_k^{jj}|$ is bounded above by $\chi_H$, the lower bound $\mu_1\kappa\epsilon\min(\frac{\epsilon}{|b_k^{jj}|}, c_1\gamma_1\epsilon)$ on $\delta f_l$ (from the changed proof) implies that $f$ is unbounded below, giving us the contradiction we want.

Finally, the proof of Theorem 3.5 is the same as the one stated earlier except that every mention of $\beta_k$ or $\beta_{k-1}$ in it is replaced by 1.

(b) The proof is by contradiction. Assume that there exists $\epsilon_1 > 0$ such that for all $k$ large enough, say $k \geq k_0$, $-\lambda_k \geq \epsilon_1$. We will show that this contradicts the assumption that $f$ is bounded. We begin by showing that $\Delta_{\min,k} \geq c_2$ for all $k \geq k_1$ (also by contradiction), where $k_1 > k_0$, and $c_2 := \frac{(1-\mu_2)\gamma_1^3\kappa_1\epsilon_1}{pL_c}$. We choose $\epsilon_1$ to be small enough that $\Delta_{\min,k_1} \geq c_2$. Now suppose $\Delta_{\min,k} < c_2$ for the first time on the $k$th iteration, $k \geq k_1$. Consider the $i$th element, where $\Delta_{i,k} = \Delta_{\min,k}$. We have $\Delta_{i,k-1} \leq \Delta_{\min,k}/\gamma_1$ and $\Delta_{\min,k-1} \geq \Delta_{\min,k}$. From the mean-value theorem,

$$\begin{aligned}\frac{|\delta f_{i,k-1} - \delta m_{i,k-1}|}{\delta m_{k-1}} &\leq \frac{\|s_{i,k-1}\|^2 \max_{\xi\in[0,1]} \|\nabla^2 f_i(x_{k-1} + \xi s_{k-1}) - \nabla^2 f_i(x_{k-1})\|}{-\kappa_1\lambda_{k-1}\Delta_{\min,k-1}^2} \\ &\leq \frac{L_c\|s_{i,k-1}\|^3}{\kappa_1\epsilon_1\Delta_{\min,k-1}^2} \qquad \text{(by Lipschitz continuity and (4.1))} \\ &\leq \frac{L_c\Delta_{i,k-1}^3}{\kappa_1\epsilon_1\Delta_{\min,k-1}^2} \qquad \text{(from (4.2))} \\ &\leq \frac{L_c\Delta_{\min,k}}{\gamma_1^3\kappa_1\epsilon_1} \\ &\leq \frac{1-\mu_2}{p}.\end{aligned}$$

Therefore by the parallel separation criterion the $i$th element is not a candidate for reduction of its trust region, or $\Delta_{\min,k} < c_2$ is not possible. If we had only a finite number of steps where $r_k \geq \mu_1$, $\Delta_{\min,k}$ would converge to zero. Since it cannot, we must have an infinite number of successful steps, where $\Delta_{\min,k}$ and $-\lambda_k$ are bounded away from zero for all sufficiently large $k$. Thus, for all subsequent successful steps $\delta f_k \geq \mu_1\delta m_k \geq \mu_1\kappa_1\epsilon_1\Delta_{\min,k}^2$, which contradicts the boundedness of $f$.

□

We would have been able to prove the rest of the results that hold for unstructured trust region algorithms, as in [4], if there was a guarantee that an $s_k$ that satisfies both (2.3) and (4.3) below, exists:

$$\delta m_k \geq -\kappa_1\lambda_k\|s_k\|^2, \tag{4.3}$$

where $\kappa_1$ and $\lambda_k$ are as defined in (4.1) above. We have not been able to find a guarantee that (2.3) and (4.3) can be simultaneously satisfied.

However, instead of using (4.3), we can switch to the following algorithm which does allow us to prove some of the stronger second order results. In Algorithm 2.17 we had an extra expansion dependent on some gradient subvectors. In this one we make one more similar expansion: this one is dependent on the minimum eigenvalue $\lambda_k$ of the second derivative matrix at $x_k$.

4.6. **Gradient-dependent algorithm with a second order adjustment.** *Given $0 < \mu_1 \leq \mu_2 < 1$, a feasible $x_0$, and starting values for the trust region sizes such that $\Delta_{i,0} \geq \max(\|g_0\|, -\lambda_0)$, for all $i = 1, \ldots, p$, the kth iteration takes the following form:*

1. *Find an approximate solution $s_k$ to the subproblem $(SP)$ that satisfies the sufficient decrease conditions (2.3) and (4.1).*
2. *Evaluate $f(x_k + s_k)$, and hence $r_k$.*
3. *Update the trust region radii according to one of the separation criteria, such as (2.12).*
4. *If $r_k < \mu_1$ then $x_{k+1} = x_k$ and the iteration ends here. Else $x_{k+1} = x_k + s_k$, calculate $g_{k+1}$ and $\nabla^2 f_{k+1}$ and go to the next step.*
5. *Reset the trust region radii according to (2.2).*
6. *Once again, reset $\Delta_{i,k+1} = \max(-\kappa_2\lambda_{k+1}, \Delta_{i,k+1})$, where $\kappa_2 \geq \frac{(1-\mu_2)\kappa_1\gamma_1^2}{pL_c}$.*

The differences between Algorithm 2.17 and the above are: firstly, $\Delta_{i,0}$ must fulfill a different condition; and secondly, there is a new step 6 that involves expansion of elemental trust regions. Neither of these changes invalidate the first order convergence proved in the last section.

Define $\hat\lambda_k := \min(0, \lambda_k)$. Notice that $\hat\lambda_k$ is non-positive. We require the following result to prove the next two theorems.

LEMMA 4.7. *For the sequence of iterates generated by the algorithm,* $\Delta_{\min,k} \geq -c_3\hat\lambda_k$, *where* $c_3 := \frac{(1-\mu_2)\kappa_1\gamma_1^3}{pL_c}$.

*Proof.* The proof is by induction. We see that our choice of $\Delta_{i,0}$ satisfies this lemma. Now suppose that the lemma holds for $\Delta_{\min,k}$. We will prove it for the $(k+1)$th iteration by showing that $\Delta_{i,k+1} \geq -c_3\hat\lambda_{k+1}$ for all $i = 1, \ldots, p$. Note that there is nothing to prove if $\hat\lambda_{k+1} \geq 0$, so we assume $\hat\lambda_{k+1} < 0$.

Case 1. The $k$th iteration is unsuccessful. For all $i$ such that $\Delta_{i,k} \geq -c_3\hat\lambda_k/\gamma_1$ we have $\Delta_{i,k+1} \geq \gamma_1\Delta_{i,k} \geq -c_3\hat\lambda_k = -c_3\hat\lambda_{k+1}$.

For $i$ such that $\Delta_{i,k} < -c_3\hat\lambda_k/\gamma_1 \leq \Delta_{\min,k}/\gamma_1$,

$$\begin{aligned}
\frac{|\delta f_{i,k} - \delta m_{i,k}|}{\delta m_k} &\leq -\frac{L_c\|s_{i,k}\|^3}{\kappa_1\hat\lambda_k\Delta^2_{\min,k}} && \text{(by Lipschitz continuity)}\\
&\leq -\frac{L_c\|\Delta_{\min,k}\|}{\kappa_1\hat\lambda_k\gamma_1^3}\\
&\leq \frac{L_c c_3}{\kappa_1\gamma_1^3} && \text{(by our induction assumption)}\\
&\leq \frac{(1-\mu_2)}{p} && \text{(substituting for } c_3).
\end{aligned}$$

Therefore none of these elemental trust region radii would be reduced. Or, $\Delta_{i,k+1} \geq \Delta_{i,k} \geq -c_3\hat\lambda_k = -c_3\hat\lambda_{k+1}$.

Case 2. If the iteration is successful then from the new expansion of elemental trust region sizes in step 6, $\Delta_{i,k+1} \geq -\kappa_2\hat\lambda_{k+1} \geq -c_3\hat\lambda_{k+1}$. □

The stronger second order convergence results can now be stated and proved.

THEOREM 4.8. $\nabla^2 f_*$ *is positive semidefinite for all limits points* $x_*$ *of the sequence of iterates* $\{x_k\}$ *generated by Algorithm 4.6.*

*Proof.* From condition (4.1), for all successful iterations $k$ we have $\delta f_k \geq \mu_1\delta m_k \geq -\mu_1\kappa_1\hat\lambda_k\Delta^2_{\min,k} \geq -\mu_1\kappa_1 c_3^2\hat\lambda_k^3$ (applying Lemma 4.7). Thus by the boundedness of $f$, $\hat\lambda_k$ must converge to zero. This implies that $\nabla^2 f_*$ is positive semidefinite for all limit points $x_*$. □

The following theorem is the same as in the other algorithms:

THEOREM 4.9. *Let* $x_k$ *be the sequence generated by Algorithm 4.6 with* $s_k$ *satisfying the same conditions as in Theorem 4.5. If* $x_*$ *is a limit point of* $\{x_k\}$ *with* $\nabla^2 f(x_*)$ *positive definite then* $\{x_k\}$ *converges to* $x_*$, *all iterations are eventually successful, and* $\{\Delta_{\min,k}\}$ *is bounded away from zero.*

*Proof.* We first prove that $\{x_k\}$ converges to $x_*$. Choose $\epsilon > 0$ and $h > 0$ so that the minimum eigenvalue of $\nabla^2 f(x)$ is at least $\epsilon$ for $\|x - x_*\| \leq h$. Since the change in the value of the model $\delta m_k$ is nonnegative, we have $\|g_k\|\|s_k\| \geq -g_k^T s_k \geq \frac{1}{2}s_k^T\nabla^2 f_k s_k \geq \frac{1}{2}\lambda_k\|s_k\|^2$, where $\lambda_k$ is the minimum eigenvalue of $\nabla^2 f_k$. Thus $\|x_k - x_*\| \leq h$ implies that

$$\frac{1}{2}\epsilon\|s_k\| \leq \|g_k\|. \tag{4.4}$$

Theorem 3.4 guarantees that $\{g_k\}$ converges to zero, and thus there is an index $k_1$ for which $\|g_k\| \leq \frac{1}{4}\epsilon h$ for all $k \geq k_1$. Hence, (4.4) shows that if $\|x_k - x_*\| \leq \frac{1}{2}h$ for $k \geq k_1$, then $\|x_{k+1} - x_*\| \leq h$.

Since $g_* = 0$, from the Taylor series expansion of $f$ about $x_*$ we have

$$f(x) - f(x_*) = (x - x_*)^T\nabla^2 f(x_* + \xi x)(x - x_*)/2,$$

where $0 \leq \xi \leq 1$. This implies that for $\frac{1}{2}h < \|x - x_*\| \leq h$, $\nabla^2 f(x_* + \xi x)$ is positive definite and $f(x) - f(x_*) \geq \frac{1}{2}\epsilon\|x - x_*\|^2 > \frac{1}{8}\epsilon h^2$. Thus, there exists an index $k_2 > k_1$ such that $\|x_{k_2} - x_*\| \leq h/2$ and $f(x_{k_2}) \leq f(x_*) + \frac{1}{8}\epsilon h^2$. Applying (4.4) to $x_{k_2}$ and $x_{k_2+1}$, we get $\|x_{k_2+1} - x_{k_2}\| \leq h/2$. But then $\|x_{k_2+1} - x_*\| \leq h$. Now $f(x_*) + \frac{1}{2}\epsilon\|x_{k_2+1} - x_*\|^2 \leq f_(x_{k_2+1}) \leq f(x_{k_2}) \leq f(x_*) + \frac{1}{8}\epsilon h^2$, implying that $\|x_{k_2+1} - x_*\| \leq h/2$.

Hence, $\|x_k - x_*\| \le h/2$ for $k \ge k_2$. But since $h$ can be chosen arbitrarily small, $\{x_k\}$ converges to $x_*$.

For the rest,

$$\begin{aligned}
\delta m_k &\ge \kappa \max_{j\in\{1,\ldots,n\}} (|g_k^j| \min(\frac{|g_k^j|}{|b_k^{jj}|}, \Delta_k^j)) \quad \text{(from (2.3))} \\
&\ge \kappa \max_{j\in\{1,\ldots,n\}} (|g_k^j| \min(\frac{|g_k^j|}{|b_k^{jj}|}, c_1\gamma_1|g_k^j|)) \\
&\quad \text{(modified Lemma 3.3, as stated in the proof of Theorem 4.5(a))} \\
&\ge \kappa c_1 \gamma_1 \max_{j\in\{1,\ldots,n\}} (|g_k^j|^2) \\
&\quad \text{(since } L \ge |b_k^{jj}| \text{ implies that } c_1 < 1/|b_k^{jj}|) \\
&\ge \frac{\kappa c_1 \gamma_1}{n} \|g_k\|^2 \\
&\ge \frac{\kappa c_1 \gamma_1 \epsilon^2}{4n} \|s_k\|^2 \quad \text{(from (4.4) where } \epsilon \text{ is a positive constant).}
\end{aligned}$$

Now from Assumption 4.3, for sufficiently large $k$, $|r_k - 1| = \frac{|\delta f_k - \delta m_k|}{\delta m_k} \le \frac{L_c \|s_k\|}{\epsilon_1}$, where $\epsilon_1 = \frac{\kappa c_1 \gamma_1 \epsilon^2}{4n}$. Hence $|r_k - 1|$ converges to zero. Hence all iterations are eventually successful and $\{\Delta_{\min,k}\}$ is bounded away from zero. □

**5. Acknowledgement.** This work was done as part of my Ph.D. at the School of Operations Research and Industrial Engineering at Cornell University, Ithaca, New York. I thank Prof. Michael J. Todd, who was my Ph.D. advisor there.

REFERENCES

[1] A. R. Conn, N. Gould, A. Sartenaer, and Ph. L. Toint, *Convergence properties of minimization algorithms for convex constraints using a structured trust region*, SIAM J. Optim., 6 (1996), pp. 1059–1086.

[2] A. R. Conn, N. Gould, A. Sartenaer, and Ph. L. Toint, *Global convergence of a class of trust region algorithms for optimization using inexact projections on convex constraints*, SIAM J. Optim., 3 (1993), pp. 164–221.

[3] R. Fletcher, *Practical Methods of Optimization*, Second ed., Practical Methods of Optimization, John Wiley, New York, 1987.

[4] J. J. Moré, *Recent developments in algorithms and software for trust region methods*, in Mathematical Programming: The State of the Art, A. Bachem, M. Grötschel and B. Korte, eds., Springer-Verlag, 1982.

[5] J. J. Moré, and D. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.

[6] Ph. L. Toint, *Global convergence of a class of trust region methods for non-convex minimization in Hilbert space*, IMA J. Numer. Anal., 8 (1988), pp. 231–252.