A Class of General Trust-Region Multilevel Algorithms for Nonlinear Constrained Optimization: Global Convergence Analysis

Natalia M. Alexandrov and J.E. Dennis, Jr.

CRPC-TR99786-S
March 1999

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

Submitted March 1999

A CLASS OF GENERAL TRUST-REGION MULTILEVEL ALGORITHMS FOR NONLINEAR CONSTRAINED OPTIMIZATION: GLOBAL CONVERGENCE ANALYSIS

NATALIA M. ALEXANDROV* AND J. E. DENNIS, JR.†

Abstract. This paper presents a broad class of trust-region multilevel algorithms for solving large, nonlinear, equality constrained optimization problems, as well as a global convergence analysis of the class. The work is motivated by engineering optimization problems with naturally occurring, densely or fully-coupled subproblem structure.

The constraints are partitioned into blocks, the number and composition of which are determined by the application. At every iteration, a multilevel algorithm minimizes models of the reduced constraint blocks, followed by a reduced model of the objective function, in a sequence of subproblems, each of which yields a substep. The trial step is the sum of these substeps. The salient feature of the multilevel class is that there is no prescription on how the substeps must be computed. Instead, each substep is required to satisfy mild sufficient decrease and boundedness conditions on the restricted model that it minimizes. Within a single trial step computation, all substeps can be computed by different methods appropriate to the nature of each subproblem. This feature is important for the applications of interest in that it allows for a wide variety of step-choice rules.

The trial step is evaluated via one of two merit functions that take into account the autonomy of subproblem processing.

The multilevel procedure presented in this work is sequential. If a problem exhibits full or partial separability, or if separability is induced by introducing auxiliary variables, then the multilevel algorithms can easily be stated in parallel form. However, since this work is devoted to analysis, we consider the most general case—that of a fully coupled problem.

Key words. Constrained optimization, nonlinear programming, multilevel algorithms, global convergence, trust region, equality constrained, multidisciplinary design optimization

AMS subject classifications. 65K05, 49D37

1. Introduction.

This work is concerned with establishing global convergence theory for a new, broad class of optimization methods, called multilevel algorithms for large-scale trust-region optimization, abbreviated as MAESTRO. The class will be presented in application to the smooth nonlinear equality constrained problem:

$$\textbf{Problem EQC:} \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & C(x) = 0, \end{cases}$$

where $f : \Re^n \to \Re$ and $C : \Re^n \to \Re^m, m \leq n$. A subclass of multilevel algorithms is applicable to solving square and underdetermined systems of nonlinear equations [1, 2].

Multilevel methods can be used to solve any general constrained optimization problem or a general system of nonlinear equations, but their development has been motivated by engineering design problems that give rise to large-scale optimization formulations with naturally occurring, tightly or fully-coupled block subproblems, such as problems that arise in the engineering multidisciplinary design optimization (MDO) environment. Overviews of MDO may be found in, e.g., Alexandrov and Hussaini [5] and Sobieszczanski-Sobieski and Haftka [30]. One of the main requirements for any prospective optimization algorithm intended for solving engineering design

*Multidisciplinary Optimization Branch, NASA Langley Research Center, Mail Stop 159, Hampton, Virginia 23681-2199 (`n.alexandrov@larc.nasa.gov`).

†Department of Computational and Applied Mathematics, Rice University, P. O. Box 1892, Houston, Texas 77251.

problems is the capability to solve parts of the problem autonomously. Multilevel methods satisfy this requirement.

The proposed algorithms were inspired by the local class of Brown-Brent methods for nonlinear equations (e.g., Brown [8, 10], Brent [6], Gay [19], Dennis [20], Martinez [24]). Subclasses of MAESTRO may be viewed as a globalization of the Brown-Brent methods and their generalization to constrained optimization.

The multilevel class is characterized by the following features:

- The constraints or equations of the problem are partitioned into blocks in a manner suitable to an application. No assumptions are made on the structure of the problem or the strength of coupling among its components. Once the problem is partitioned, at each iteration the algorithm solves a sequence of progressively smaller dimensional subproblems, each of which yields a substep. The trial step is the sum of the substeps.
  Separability or partial separability can be used to extract parallelism. They may be introduced into problems where they do not occur naturally. However, since this paper is devoted to analysis, we will consider the most general case—that of a completely non-separable (fully coupled) problem.
- Instead of specifying a method for solving the algorithm's subproblems, we require that substeps comprised by the trial step satisfy mild decrease and boundedness conditions with respect to model subproblems. Both the boundedness and the decrease requirements are satisfied by most methods of interest for solving the subproblems, allowing for a wide variety of substeps.
  This feature is significant because in applications such as MDO, constraint blocks originate from different disciplines and almost certainly require different approaches to solving the subproblems.
  Specific ways to compute the substeps give rise to members of the MAESTRO class.
- The globalization strategy for the multilevel class is the model trust-region approach. Two variants of the ℓ_2 penalty function are used as merit functions. The procedure for updating the penalty parameters accounts for autonomous processing of the subproblems. The updates are based on those introduced by El-Alem in [14]. The approximation to the Hessian of the objective function is only assumed to be bounded.
- The multilevel methods belong to the class of out-of-core methods. In principle, there is no limit to the size of the problems they can solve.
- When they are applied to linear systems of equations with subsystems solved by iterative methods, multilevel algorithms are block-approximate direct solvers. For example, one choice of the trial step yields the block approximate Gaussian elimination.

In the next section we discuss briefly the notion of sufficient decrease in the context of the trust-region approach to globalization. Section 3 describes the multilevel class. In Section 4, we present the global convergence theory for the first variant of the merit function. Section 5 discusses the simpler global convergence analysis for the second variant of the merit function. Section 6 concludes with a brief summary.

2. Sufficient predicted decrease. The concept of sufficient decrease plays an important role in the trust-region strategy for improving the global behavior of local model-based algorithms. Given x_i, the current approximation to the solution, a trust-region algorithm for minimizing an unconstrained function $f(x)$ finds a trial step, s_i,

by approximately solving the *unconstrained trust-region subproblem*

$$\min\{\phi_i(s) \equiv f(x_i) + \nabla f(x_i)^T s + \frac{1}{2} s^T H_i s \;:\; ||s|| \leq \delta_i\}, \tag{2.1}$$

where x_i, ∇f, $s \in \Re^n$; $H_i \in \Re^{n \times n}$ approximates $\nabla^2 f(x_i)$; and $\delta_i > 0$ is the trust-region radius.

Global convergence analysis requires the trial step to satisfy a *fraction of Cauchy decrease* (FCD) condition. That is, s_i must predict at least a fraction of the improvement in f predicted by the steepest descent step within the trust region (the Cauchy step). Specifically, we must have, for some positive constant σ not dependent on i,

$$\phi_i(0) - \phi_i(s_i) \geq \sigma[\phi_i(0) - \phi_i(s_i^{CP})], \tag{2.2}$$

where

$$s_i^{CP} = -\alpha_i^{CP} \nabla f(x_i),$$

$$\alpha_i^{CP} = \begin{cases} \frac{||\nabla f(x_i)||^2}{\nabla f(x_i)^T H_i \nabla f(x_i)} & \text{if } \frac{||\nabla f(x_i)||^3}{\nabla f(x_i)^T H_i \nabla f(x_i)} \leq \delta_i \text{ and } \nabla f(x_i)^T H_i \nabla f(x_i) > 0 \\ \frac{\delta_i}{||\nabla f(x_i)||} & \text{otherwise.} \end{cases}$$

FCD is easily satisfied computationally, and it implies a weaker condition in a convenient form frequently used in analysis of trust-region algorithms (e.g., Powell [23]; Moré [25]):

LEMMA 2.1. *Let s_i satisfy the fraction of Cauchy decrease condition. Then*

$$\phi_i(0) - \phi_i(s_i) \geq \frac{\sigma}{2} ||\nabla f(x_i)|| \min\left\{\frac{||\nabla f(x_i)||}{||H_i||}, \delta_i\right\}. \tag{2.3}$$

Powell [23] established a global convergence result under a number of very mild assumptions. Namely, if f is continuously differentiable and bounded below on the level set $\{x \in \Re^n : f(x) \leq f(x_0)\}$, if the Hessian approximations $\{H_i\}$ are uniformly bounded, and if the iterates $\{x_i\}$ satisfy either (2.2) or (2.3), then weak first-order stationary convergence holds; that is,

$$\liminf_{i \to \infty} ||\nabla f(x_i)|| = 0.$$

Surveys of the trust-region approach to unconstrained optimization and nonlinear equations can be found in Dennis and Schnabel [21] and Moré [25]. Detailed treatment of the corresponding convergence theory can be found in Powell [23], Moré [25], Moré and Sorensen [26], Shultz et al. [29], and Sorensen [32].

Note that the model Hessians in [23] do not have to be symmetric or positive definite or have to satisfy any condition of accuracy (compatibility) as approximations of $\nabla^2 f(x_i)$. Thus, instead of describing a specific, detailed algorithm, [23] provides a general globally convergent framework for unconstrained optimization.

Some recent work on constrained optimization (e.g., [1, 4, 13, 15, 16]), including the present research, has similar intent. Namely, instead of proposing and analyzing a specific algorithm for constrained optimization, a framework is proposed that allows for a wide variety of step choices under the weakest possible requirements placed on the step. Such frameworks are particularly interesting for realistic design problems, because engineering optimization frequently gives rise to formulations that include computational components produced under widely varying conditions.

3. A class of general trust-region multilevel algorithms for equality constrained optimization. This section starts with an overview of the MAESTRO class applied to problem EQC, followed by a summary of notation and the statement of assumptions. We then discuss the trial step computation and other components of the algorithm—the merit function and its model, the stopping criterion, the trial step evaluation, updating the penalty parameters, the trust-region radii, and the iterate. The section ends with a formal statement of the algorithm.

3.1. Overview of the class. The MAESTRO class applied to problem EQC can be conceptually described as follows. The constraint system is partitioned into M blocks $C_1(x), \ldots, C_M(x)$, with each $C_k : \Re^n \to \Re^{m_k}$, $\sum_{k=1}^{M} m_k = m$. (Individual components of the system, as opposed to blocks, will be denoted by $c_j(x), j = 1, \ldots, m$.) In practice, this block decomposition may be obvious or it may itself be a topic of research.

At the current approximation x_i to a solution x_* we set $y_0 = x_i$ and compute the trial step $\hat{s}_i$ as follows.

A substep s_1 approximately minimizes $||C_1(y_0)+\nabla C_1^T(y_0)s||^2$, the Gauss-Newton model of the first constraint block about y_0 in a trust region of radius δ_1. It yields $y_1 = y_0 + s_1$.

The substep s_1 defines a hyperplane $\nabla C_1^T(y_0)(s - s_1) = 0$ parallel to the null space of the Jacobian of the first constraint block. All further substeps of trial step $\hat{s}_i$ are restricted to this hyperplane. Thus the substeps s_2, ..., s_{M+1} do not change the predicted improvement in the model of $||C_1(x)||^2$.

A substep s_2 approximately minimizes a model of $||C_2(x)||^2$ about y_1, restricted to the hyperplane defined by s_1, in a trust region of radius δ_2. The substep s_2 is taken to the point y_2.

The process continues to compute substeps that approximately minimize restricted models of progressively smaller dimensions. When all the constraint blocks have been processed, $n - m$ degrees of freedom still remain. They are used in building a quadratic model of the objective function about y_M. The final substep s_{M+1} approximately minimizes this model restricted to the intersection of the hyperplanes parallel to the null spaces of the Jacobians of all constraint blocks. The final substep yields the next approximation to a solution of problem EQC. The total trial step from $x_i = y_0$ is $\hat{s}_i = \sum_{k=1}^{M+1} s_k$.

Unless the convergence criteria are met, the trial step is evaluated and, upon updating all parameters of the algorithm, the procedure repeats in trust regions of the size determined by the success or failure of the previous trial step.

It should be noted that although the MAESTRO procedure is consistent with the Gauss-Seidel principle of using new information as soon as it is available, it is *not* the Gauss-Seidel algorithm. The latter, applied to a linear system of equations, is an iterative procedure; MAESTRO solves a linear system in one iteration.

3.2. Summary of notation. Given the block structure of the algorithm, notation becomes cumbersome. We abbreviate the notation whenever possible while attempting to retain clarity.

The iterates and the trial steps are denoted by x_i and $\hat{s}_i$, respectively. Computing one trial step $\hat{s}_i$ from the point x_i involves solving a "sweep" of $M+1$ subproblems. In the course of the algorithm description, it will become evident that the subproblems serve to compute a basis for the null space of the constraint system Jacobian. This computation, however, is accomplished step by step, using information at different points, in contrast with, say, the SQP methods, which use information at the current

iterate and the entire block of constraints even when the constraint system is solved iteratively.

The subproblems of a single trial step computation and the components within a single sweep of subproblems are indexed by k. Unless stated otherwise, we will always refer to one generic trial step computation and omit the reference to that step while referring to the subproblem entities. For instance, the substep produced by solving the k-th subproblem of a sweep is denoted by s_k $(k = 1, \ldots, M+1)$, while the points generated within each sweep by taking these substeps are denoted by y_k $(k = 0, \ldots, M+1)$ with $y_0 = x_i$ and $y_{M+1} = x_{i+1}$. The total trail step $\hat{s}_i = \sum_{k=1}^{M+1} s_k$.

If one has to distinguish between the subproblem components that belong to different trial steps, superscripts make the notation explicit. For instance, $\hat{s}_i = \sum_{k=1}^{M+1} s_k^i$, while $\hat{s}_{i+1} = \sum_{k=1}^{M+1} s_k^{i+1}$.

The trust-region radius for subproblem k, centered at y_{k-1}, is denoted by δ_k, while the radius of the total trust region, centered at $x_i = y_0$, is $\hat{\delta}_i$. P_k denotes the generic reducer matrix whose columns form a basis for the intersection of the null spaces of $\nabla C_1^T(y_0), \ldots, \nabla C_k^T(y_{k-1})$. H_M denotes the approximation to Hessian of f at y_M.

The components of C_1 will be numbered from $n_1 = 1$ to $n_2 - 1$, the components of C_2 will be numbered from n_2 to $n_3 - 1$, and so on. Finally, the components of C_M will be numbered from n_M to m.

All norms are ℓ_2 norms.

3.3. Assumptions. Let $\{x_i\}$ be the sequence of iterates generated by a member of the MAESTRO class with the corresponding subproblem iterates $\{y_k^i\}, k = 1, \ldots, M+1$. We make the following assumptions:

1. All x_i and y_k^i, $k = 1, \ldots, M+1, i = 1, \ldots$ lie in a convex set $\Omega \subseteq \Re^n$.
2. $f, C \in C^2(\Omega)$.
3. $f(x)$, $\nabla f(x)$, $\nabla^2 f(x)$, H_M^i, $C(x)$, $\nabla C(x)$ [all $C_k(x)$, $\nabla C_k(x)$], $\nabla^2 c_j(x)$, $j = 1, \ldots, m$, $(\nabla C_k^T(x) \nabla C_k(x))^{-1}$, P_k^i, $k = 0, \ldots, M$ $(P_0^i = I)$, are uniformly bounded in Ω. That is, there exist positive constants $\sigma_1, \ldots, \sigma_9$, such that $\|f(x)\| \leq \sigma_1$, $\|\nabla f(x)\| \leq \sigma_2$, $\|\nabla^2 f(x)\| \leq \sigma_3$, $\|H_M^i\| \leq \sigma_4$, $\|C(x)\| \leq \sigma_5$, $\|\nabla C(x)\| \leq \sigma_6$ $\|\nabla^2 c_j(x)\| \leq \sigma_7$, $j = 1, \ldots, m$, $\|(\nabla C_k^T(x) \nabla C_k(x))^{-1}\| \leq \sigma_8$, $\|P_k^i\| \leq \sigma_9$, $k = 0, \ldots, M$ for all $x \in \Omega$.
4. $\nabla C_k(x)$, $k = 1, \ldots, M$, have full rank.
5. If the substeps of the algorithm are block-linearly feasible (see §3.4.1), then the inverses $\{B_k^{-1}\}$ are assumed to be bounded.

These assumptions are conventional (see, for instance, [12, 13, 15, 16, 28]), although some are stated in block form.

The analysis will be done under these assumptions and so they will not be repeated in the statements of lemmas and theorems.

3.4. The trial steps. During the constraint elimination stage of computing a trial step $\hat{s}_i$, the substeps approximately solve the following subproblems:

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}\|C_k(y_{k-1}) + \nabla C_k^T(y_{k-1}) s\|^2 \\
\text{subject to} & \nabla C_j^T(y_{j-1}) s = 0, j = 1, \ldots, k-1, \\
& \|s\| \leq \delta_k,
\end{array}
\tag{3.1}
$$

$k = 1, \ldots, M$ (there is no null space constraint for $k = 1$). The objective function subproblem is

$$
\text{(3.2)} \qquad \begin{array}{ll} \text{minimize} & \phi_M(s) \equiv f(y_M) + \nabla f^T(y_M)s + \frac{1}{2}s^T H_M s \\ \text{subject to} & \nabla C_j^T(y_{j-1})s = 0, j = 1, \ldots, M \\ & \|s\| \leq \delta_{M+1}. \end{array}
$$

A change of variables $s = P_{k-1}v$, where the columns of the reducer matrix P_{k-1} form a basis for $\bigcap_{j=1}^{k-1}[\mathcal{N}(\nabla C_j^T(y_{j-1}))]$, converts subproblem k with null space constraints into an unconstrained trust-region subproblem. Reducer matrices and substep computation will be discussed presently.

3.4.1. Some choices for reducer matrices and substeps. We describe two explicit choices for the reducer matrix P_k and substeps.

Orthogonal substeps. For relatively small problems, P_1 can be computed via the QR decomposition of $\nabla C_1^T(y_0)$ and updated for subsequent subproblems to bases of progressively smaller dimensional intersections. All subproblems are explicit unconstrained trust-region minimization problems.

Orthogonal substeps are appealing but can be expensive to compute. It is easy to observe the relation between the unconstrained minimum norm substeps and solutions of the subproblems. Direct examination establishes the following:

PROPOSITION 3.1. *Let $s_k^{\perp}$ be the orthogonal substep from y_{k-1} to y_k with the associated trust-region radius δ_k. Then $s_k^{\perp}$ is equal to the unconstrained minimum norm substep on the shifted equation $C_k(y) - [C_k(y_{k-1}) + \nabla C_k^T(y_{k-1})s_k^{\perp}] = 0$.*

Readers familiar with Brent's algorithm for nonlinear equations [7] will notice that the unconstrained minimum norm step is the Brent step. Thus, for orthogonal substeps, the multilevel class is a direct extension of Brent's method to constrained optimization and, in the case of nonlinear equations, a trust-region globalization of Brent's method. Details can be found in [2].

Another way to express the generalization of the Brent substep is to compute the minimum norm substep and to truncate it to lie within the trust region if its length exceeds the trust-region radius. Such a substep will be shown to satisfy the conditions imposed on multilevel substeps.

Extension of the general block-linearly feasible substep. For large problems, the use of orthogonal substeps becomes prohibitively expensive. Extending the well-known reduced-basis technique for computing the reducer matrices and the substeps is an approach particularly well suited for the applications we have in mind. Experience in the context of single-level, large optimization problems with constraint systems formed by discretizations of partial differential equations [22] provides numerical reinforcement for this idea.

In the block-linearly feasible approach, only the last subproblem of the sweep needs to be an explicit optimization problem. The subproblems that eliminate the constraints are solved implicitly.

Consider an example. Let the constraint system $C(x) = 0$ be partitioned into three blocks, $C_k(x) = 0$, $C_k : \Re^n \to \Re^{m_k}$, $k = 1, \ldots, 3$.

Let us partition $\nabla C_1^T(y_0)$ into $[B_1|N_1]$, where B_1 is an $m_1 \times m_1$ nonsingular matrix. Column pivoting may be required to make B_1 nonsingular or to obtain the most advantageous B_1. Now

$$
\nabla C_1^T(y_0)s = B_1 s_{B_1} + N_1 s_{N_1} = -C_1(y_0).
$$

Choosing $s_{N_1} = 0$ yields $s_{B_1} = -(B_1)^{-1}C_1(y_0)$ and a linearly feasible point

$$s_1^{lin} = (s_{B_1}, s_{N_1})^T = (-B_1^{-1}C_1(y_0), 0)^T.$$

If s_1^{lin} lies within δ_1, set $s_1 = s_1^{lin}$. Otherwise, truncate the substep to lie within the trust region; i.e.,

$$s_1 = \frac{\delta_1 s_k^{lin}}{||s_k^{lin}||}.$$

The substep is taken to obtain $y_1 = y_0 + s_1$. Now we compute the substep on the second block of constraints, restricted to the null space of the Jacobian of the first block. We partition $\nabla C_2^T(y_1)Z_1$ into $[B_2|N_2]$, where

$$Z_1 = \left[\begin{array}{c} -B_1^{-1}N_1 \\ I_{n-m_1} \end{array} \right]$$

and B_2 is a nonsingular $m_2 \times m_2$ matrix, possibly obtained by column pivoting. Similarly to s_1^{lin}, we obtain

$$s_2^{lin} = (s_{B_2}, s_{N_2})^T = (-B_2^{-1}C_2(y_1), 0)^T.$$

The substep s_2 is truncated if it lies outside δ_2 to obtain $y_2 = y_1 + s_2$. For the final block of constraints, we partition $\nabla C_3^T(y_2)Z_2Z_1$ into $[B_3|N_3]$, where

$$Z_2 = \left[\begin{array}{c} -B_2^{-1}N_2 \\ I_{n-m_1-m_2} \end{array} \right],$$

and proceed as above to obtain s_3 and

$$Z_3 = \left[\begin{array}{c} -B_3^{-1}N_3 \\ I_{n-m_1-m_2-m_3} \end{array} \right].$$

Finally, we solve the unconstrained trust-region subproblem:

$$\begin{array}{ll} \text{minimize} & f(y_M) + \nabla f^T(y_M)Z_3Z_2Z_1v + \frac{1}{2}(Z_3Z_2Z_1v)^T H_M (Z_3Z_2Z_1v) \\ \text{subject to} & ||Z_3Z_2Z_1v|| \le \delta_4 \end{array}$$

for v to obtain $s_4 = Z_3Z_2Z_1v$.

This procedure causes the substeps to be parallel to hyperplanes formed by a number of coordinate directions. Readers familiar with Brown's method [9, 11] for nonlinear equations will recognize that this way of computing the substeps includes, as a special case, an extension of Brown's method to constrained optimization, while in the case of nonlinear equations it provides a trust-region globalization strategy for the local Brown's method.

As there is a Brown-Brent analog for any matrix decomposition in the linear case, the Brown-Brent class may be viewed as an extension to the nonlinear case of the direct linear solvers. Therefore, based on the choice of partitioning and pivoting for the Jacobian of the constraints, MAESTRO class includes, as a special case, the extension of the Brown-Brent class to constrained optimization and the trust-region globalization of the class for nonlinear equations [2].

Methods for solving reduced subproblems. The block-linearly feasible approach leaves only one explicit reduced unconstrained trust-region subproblem to

solve: the objective function subproblem. In general, every substep computation may be an unconstrained trust-region subproblem. These can be solved by any method suitable to the application, as long as they satisfy the two requirements described in the next subsection. Because the MAESTRO class is intended primarily for large-scale problems, the algorithms of Steihaug-Toint [34, 35] and Sorensen [31, 33] are promising.

3.4.2. The general substep requirements. In the previous section, we described some reasonable methods for computing the substeps, but the MAESTRO class assumes that the user will select the methods most fitting the properties of a specific block. The substeps are only required to satisfy two mild conditions.

Boundedness. Global convergence for multilevel algorithms with orthogonal substeps was established by Alexandrov [1]. However, the expense of computing orthogonal steps prompted us to propose using nonorthogonal substeps (Alexandrov and Dennis [4]). Here we complete the analysis of convergence for a broader class whose substeps s_k are required to satisfy a weaker condition:

$$||s_k^i|| \leq \Gamma_1 ||C_k(y_{k-1}^i)||, \;\; k = 1, \ldots, M, \tag{3.3}$$

for some positive constant Γ_1 independent of i. The constant need not be the same for all blocks of constraints, but we take Γ_1 to be the maximum of the individual constants. Note that it is also true that

$$||C_k(y_{k-1}^i)|| \leq \Gamma_2 ||s_k^i||, \;\; k = 1, \ldots, M, \tag{3.4}$$

for some constant $\Gamma_2 > 0$. This follows directly from the definition of the orthogonal or a block-linearly feasible step, and it follows from the sufficient decrease condition (described below) in the case of a substep that solves the k-th subproblem explicitly.

Let $N \leq M+1$ be the index of one of the substeps in the step $\hat{s}_i = \sum_{k=1}^{M+1} s_k$. If the substeps are orthogonal to each other, the Pythagorean theorem yields

$$||\sum_{k=j}^{N} s_k^{\perp}|| \leq ||\hat{s}_i||$$

for any j between 1 and N; i.e., the length of any intermediate sum of the substeps is bounded by the length of the total trial step. We obtain the following relaxation for nonorthogonal substeps:

LEMMA 3.2. *Let $\hat{s}_i = s_1 + \ldots s_{M+1}$ be a trial step generated by a member of the multilevel class from x_i. Then for $1 \leq N \leq M+1$, the following inequality holds:*

$$||\sum_{k=j}^{N} s_k|| \leq \Gamma_3 ||\hat{s}_i||$$

for all $j \in [1, N]$ and for some constant $\Gamma_3 > 0$ fixed across all iterations.

Proof. For $k = 1$,

$$||\nabla C_1^T(y_0)s_1|| = ||\nabla C_1^T(y_0)(s_1 + \ldots + s_{M+1})|| \equiv ||\nabla C_1^T(y_0)\hat{s}_i|| \leq ||\nabla C_1|| ||\hat{s}_i||.$$

On the other hand,

$$||\nabla C_1^T(y_0)s_1|| \geq \frac{1}{||[\nabla C_1^T(y_0)]^{\dagger}||} ||s_1||.$$

The problem assumptions yield the constant. A simple inductive argument leads to similar inequalities for other substeps, with Γ_3 taken to be the maximum of all block constants. $\square$

Sufficient decrease. Each substep s_k is required to satisfy FCD on the reduced quadratic model of the subproblem that it solves. The following lemma expresses FCD for the subproblems in a convenient form.

LEMMA 3.3. *Let $\hat{s}_i = s_1, \ldots, s_{M+1}$ be a trial step generated from $x_i = y_0$. Then there exist positive constants Γ_4, Γ_5, and Γ_6, not dependent on i, such that*

$$||C_k(y_{k-1})||^2 - ||C_k(y_{k-1}) + \nabla C_k^T(y_{k-1})s_k||^2 \geq \Gamma_4 ||C_k(y_{k-1})|| \min\{\Gamma_5 ||C_k(y_{k-1})||, \delta_k\},$$

where $k = 1, \ldots, M$, and

$$\phi_M(0) - \phi_M(s_{M+1}) \geq \frac{\sigma}{2} ||P_M^T \nabla f(y_M)|| \min\{\Gamma_6 ||P_M^T \nabla f(y_M)||, \delta_{M+1}\},$$

where P_M forms a basis for $\bigcap_{j=1}^{M} [\mathcal{N}(\nabla C_j^T(y_{j-1}))]$.

Proof. The proof is a straightforward application of Lemma (2.1) and the problem assumptions to subproblems $k = 1, \ldots, M+1$ to obtain the values of the constants. $\square$

Properties of some substeps. The following theorem shows that a wide class of methods for computing the substeps satisfies both requirements.

THEOREM 3.4. *Let $s_k^\perp$, $1 \leq k \leq M$, be an orthogonal substep generated by subproblem k of any trial step computation. Further, let s_k^{lin} be any block-linearly feasible substep from y_{k-1} to the hyperplane defined by the linearization of the reduced block C_k about y_{k-1}. Let δ_k be the associated trust-region radius. If $||s_k^{lin}|| \leq \delta_k$, then let $s_k = s_k^{lin}$. Otherwise let*

$$s_k = \frac{\delta_k s_k^{lin}}{||s_k^{lin}||}.$$

Then there exists a positive constant Γ_1, not dependent on i, such that

$$||s_k^\perp|| \leq \Gamma_1 ||C_k(y_{k-1})||$$

and

$$||s_k|| \leq \Gamma_1 ||C_k(y_{k-1})||.$$

Moreover, $s_k^\perp$ and s_k satisfy FCD on the reduced Gauss-Newton model of C_k about y_{k-1}.

Proof. First, consider the boundedness condition. If $C_k(y_{k-1}) = 0$, then $s_k = 0$ and the result holds. If $C_k(y_{k-1}) \neq 0$, then

$$||s_k^\perp|| \leq ||P_{k-1}^T \nabla C_k(y_{k-1}) \{[P_{k-1}^T \nabla C_k(y_{k-1})]^T [P_{k-1}^T \nabla C_k(y_{k-1})]\}^{-1}|| \, ||C_k(y_{k-1})||,$$

and the definition of the constant follows from the problem assumptions.

In the case of nonorthogonal substeps, our assumptions guarantee that we can always find, possibly with pivoting, an invertible B_k in a reduced block-Jacobian with B_k^{-1} uniformly bounded. Therefore, $P_k \equiv Z_k, k = 1, \ldots, M$, are uniformly bounded as well. It follows that the block-linearly feasible substeps, s_k^{lin}, and the truncated block-linearly feasible steps, s_k, are bounded in terms of the constraint block norms.

The problem assumptions provide the definition of the constant, and the constant Γ_1 is the maximum of those for $s_k^{\perp}$ and s_k.

Now consider the sufficient decrease condition. The conclusion is obvious for $s_k^{\perp}$ and for $s_k = s_k^{lin}$. Therefore, let us assume that $0 < \zeta_k \equiv \frac{\delta_k}{||s_k^{lin}||} < 1$.

We have

$$\begin{aligned}
&||C_k(y_{k-1})||^2 - ||C_k(y_{k-1}) + \nabla C_k^T(y_{k-1})s_k||^2 \\
&= ||C_k(y_{k-1})||^2 - ||C_k(y_{k-1}) + \zeta_k \nabla C_k^T(y_{k-1})s_k^{lin}||^2 \\
&\geq ||C_k(y_{k-1})||^2 - [(1-\zeta_k)||C_k(y_{k-1})|| + \zeta_k||C_k(y_{k-1}) + \nabla C_k^T(y_{k-1})s_k^{lin}||]^2 \\
&= [1-(1-\zeta_k)^2]||C_k(y_{k-1})||^2.
\end{aligned}$$

The definition of s_k^{lin} and the boundedness of the reducer matrix yield the conclusion. $\square$

3.5. Measuring progress toward solution. In this work, we focus on modified ℓ_2 penalty functions in keeping with using weak assumptions sufficient for establishing a global convergence analysis. It would also be easy to include a modified augmented Lagrangian as a merit function, by imposing the boundedness assumptions on the Lagrange multipliers. However, we relegate this line of research to the study of the local convergence properties of the class.

The substep s_k predicts decrease for $||C_k||^2$ from y_{k-1} to y_k. Because the model of $||C_k||^2$ is restricted to the intersection of the null spaces of Jacobians numbered $1, \ldots, k-1$, s_k conserves predicted improvement for all $||C_j||^2$, $j = 1, \ldots, k-1$. However, there is no prediction at all about how $s_1 + \ldots + s_k$ changes, and possibly increases, $||C_j||^2$, $j = k+1, \ldots, M$. Neither does any substep, except s_{M+1}, predict the behavior of the objective function.

Components of a progress measuring scheme include a merit function and its model, the definitions of the actual and predicted reductions in the merit function, and procedures for updating the parameters of the merit function. To account for the autonomous processing of subproblems, we consider two alternatives for such a scheme. Detailed global convergence analysis will be presented for Variant I. We will then point out the modifications to accommodate the simpler Variant II.

3.5.1. Variant I. Let $\rho_k, k = 1, \ldots, M$, be positive scalars greater than or equal to 1. To simplify notation, let $\boldsymbol{\rho}_k$ be the vector of the first k such scalars; i.e., $\boldsymbol{\rho}_k \equiv (\rho_1, \ldots, \rho_k)$. Let $S_k \equiv \{s_1, \ldots, s_k\}$ be the set of the first k substeps.

To measure the progress of the algorithm toward a solution, we introduce a new merit function—a modified ℓ_2 penalty function:

$$\bar{\mathcal{P}}(x; \boldsymbol{\rho}_M) \equiv f(x) + \sum_{k=1}^{M}(\prod_{j=k}^{M} \rho_j)||C_k(x)||^2.$$

For instance, in the case of two blocks of constraints, we have

$$\bar{\mathcal{P}}(x; \boldsymbol{\rho}_M) \equiv f(x) + \rho_2[||C_2(x)||^2 + \rho_1||C_1(x)||^2].$$

The initial choice, $\rho_k = 1$, is arbitrary and scale-dependent. The only requirement for our analysis is that $\rho_k \geq 1$. Of course, in practice, more care would likely be needed to choose initial penalty constants.

Because $\nabla C_k^T(y_{k-1})s_j = 0$ for all $j = k+1, \ldots, M+1$, we model each $||C_k(x_i + \hat{s}_i)||^2$ at $y_{M+1} = x_i + \hat{s}_i$ by $||C_k(y_{k-1}) + \nabla C_k^T(y_{k-1})s_k||^2$, and so we model the merit function at $x_i + \hat{s}_i$ by

$$\mathcal{M}(S_{M+1}; \boldsymbol{\rho}_M) \equiv \phi_M(s_{M+1}) + \sum_{k=1}^{M}(\prod_{j=k}^{M} \rho_j)||C_k(y_{k-1}) + \nabla C_k^T(y_{k-1})s_k||^2.$$

We define the actual reduction in the merit function by

$$ared_i \equiv \bar{\mathcal{P}}(x_i; \boldsymbol{\rho}_M^i) - \bar{\mathcal{P}}(x_i + \hat{s}_i; \boldsymbol{\rho}_M^{i+1}) \equiv \bar{\mathcal{P}}(y_0; \boldsymbol{\rho}_M^i) - \bar{\mathcal{P}}(y_{M+1}; \boldsymbol{\rho}_M^{i+1}). \tag{3.5}$$

The predicted reduction models the actual reduction and is defined as

$$pred_i \equiv \bar{\mathcal{P}}(x_i; \boldsymbol{\rho}_M^i) - \mathcal{M}(S_{M+1}; \boldsymbol{\rho}_M^{i+1}). \tag{3.6}$$

This function penalizes for the possible predicted increase in the constraint blocks $k, \ldots, M$, and in the objective function that may have occurred while computing the substeps numbered $1, \ldots, k-1$.

The arguments in the definition of $ared_i$ and $pred_i$ are evident, but we shall not use them explicitly—throughout throughout the remainder of the paper. The same convention holds for the partial predicted reductions. Thus

$$Cpred_1 \equiv ||C_1(y_0)||^2 - ||C_1(y_0) + \nabla C_1^T(y_0)s_1||^2$$

denotes the partial predicted reduction in the first block of constraints and

$$Cpred_k \equiv ||C_k(y_0)||^2 - ||C_k(y_{k-1}) + \nabla C_k(y_{k-1})^T s_k||^2 + \rho_{k-1} Cpred_{k-1}$$

denotes the partial predicted reduction in the $k-th$ block of constraints, $k = 2, \ldots, M$.

3.5.2. Variant II. This is a simplified merit function based on [2]. It replies on the fact that the norms of the substeps are bounded by the norms of the constraint values, which "pulls" the subiterates to a single point.

The merit function is defined simply as

$$\bar{\mathcal{P}}(x; \rho) \equiv f(x) + \rho \sum_{j=1}^{M} ||C_j(x)||^2, \tag{3.7}$$

with $ared_i$ defined as for the Variant I. But now the predicted reduction will be defined as

$$pred_i \equiv f(y_0) - \phi_M(s_{M+1}) + \rho_i \sum_{j=1}^{M} [||C_j(y_{j-1})||^2 - ||C_j(y_{j-1}) + \nabla C_j^T(y_{j-1})s_j||^2].$$

Note that this merit function has a single penalty parameter. While this variant is simpler than the first one, the relative efficiency of the two variants remains to be determined by numerical experimentation.

3.6. Updating the penalty parameters. The penalty parameters are updated upon completion of each trial step, before the step is evaluated. The updating scheme for Variant II is simply that proposed by El-Alem [15], with the distinction that the predicted reduction contains information about M constraint blocks instead of the single block.

The updating scheme for Variant II extends the one proposed by El-Alem [15, 16] to account for cumulative predicted reduction. It ensures that, unless an optimum is reached, the total trial step predicts at least a portion of FCD predicted in the model of the first block of constraints.

ALGORITHM 3.1. **Updating penalty parameters**

Set $\rho_1^1 = \ldots = \rho_M^1 = 1$ and choose $\beta \in (0,1)$.
do $k = 1, M-1$
    if at least one of $s_j \neq 0, j = 1, \ldots, k$; **then** update ρ_k:
        if $k = 1$; **then**
            Compute $Cpred_1$ and $Cpred_2$
        else
            Compute $Cpred_{k+1}$
        end if
        if $Cpred_{k+1} \geq \frac{\rho_k^i}{2} Cpred_k$; **then**
            $\rho_k^{i+1} = \rho_k^i$
        else
            $\rho_k^{i+1} = \bar{\rho_k} + \beta$,
            where $\bar{\rho_k} = \frac{2[||C_{k+1}(y_k) + \nabla C_{k+1}^T(y_k) s_{k+1}||^2 - ||C_{k+1}(y_0)||^2]}{Cpred_k}$
            Recompute $Cpred_{k+1}$
        end if
    else
        $\rho_k^{i+1} = \rho_k^i$
    end if
end do
if at least one of $s_j \neq 0, j = 1, \ldots, M+1$; **then** update ρ_M^i:
    Compute $pred_i$
    if $pred_i \geq \frac{\rho_M^i}{2} Cpred_M$; **then**
        $\rho_M^{i+1} = \rho_M^i$
    else
        $\rho_M^{i+1} = \bar{\rho_M} + \beta$,
        where $\bar{\rho_M} = \frac{2[\phi_M(s_{M+1}) - f(y_0)]}{Cpred_M}$
        Recompute $pred_i$
    end if
else
    $\rho_M^{i+1} = \rho_M^i$
end if

We choose the penalty parameters so that for each substep s_k the predicted reduction accumulated by $s_1 + \ldots + s_k$ is at least a fraction of the predicted decrease accumulated by $s_1 + \ldots + s_{k-1}$.

This scheme increases the penalty parameters but does not do so excessively, which helps to alleviate the numerical problems caused by a possibly too rapid growth of the penalty parameters. Also note that taking substep k restricted to previous null spaces preserves the decrease prediction for the blocks already processed before the k-th block and, thus, should weaken the rate of growth of the penalty parameters as well. However, if the penalty parameters do become large, they can be "re-started" from smaller values. The subject of nonmonotone penalty parameters is studied in El-Alem [17]. For the purposes of our discussion, we will consider only nondecreasing

penalty parameters.

The following lemma summarizes the properties of the penalty parameters computed in this scheme.

LEMMA 3.5. *Let ρ_k, $k = 1, \ldots, M$ be the penalty parameters generated by Algorithm (3.1). Then*

1. *The sequences $\{\rho_k^i\}$, $k = 1, \ldots, M$ are nondecreasing.*
2. *The partial predicted reductions satisfy*

$$Cpred_{k+1} \geq \frac{\rho_k}{2} Cpred_k \geq \frac{\prod_{j=1}^{k} \rho_j}{2^k} Cpred_1, \tag{3.8}$$

and the total predicted reduction satisfies

$$pred_i \geq \frac{\rho_M}{2} Cpred_M \geq \frac{\prod_{j=1}^{M} \rho_j}{2^M} Cpred_1. \tag{3.9}$$

3. *If a ρ_k is increased, it is increased by at least β.*

Proof. Direct examination of the updating scheme yields a straightforward proof. Note that inequalities (3.8) and (3.9) are satisfied whether or not the penalty parameters are increased. If they are not increased, the inequalities are satisfied with the previous values. Otherwise, they are satisfied with the updated values. ☐

In summary, the only substep that is certain to predict an improvement from $x_i \equiv y_0$ in $||C_1(x)||^2$ is s_1, which is the reason for the first block being the most heavily weighted one. Equivalently, without adjusting the penalty parameters, the entire step $\hat{s}_i$ predicts decrease for the first block only. An alternative scheme would be to place the penalty parameter only on the first block of constraints, i.e., to have a merit function of the form

$$\tilde{\mathcal{P}}(x) = f(x) + \sum_{j=2}^{M} ||C_j(x)||^2 + \rho ||C_1(x)||^2$$

with an appropriate scheme for updating ρ. While theoretically this scheme would not be fundamentally different from the one we adopted, in practice it is expected to have more numerical difficulties.

Finally, we emphasize that the step computation is independent of the penalty parameter computation.

3.7. Evaluating the step and updating the radii and the iterates. Although there are various interesting schemes for evaluating the trial step and updating the trust-region radii, for ease of exposition, we adopt the following strategy: the total trial step is evaluated, and all individual trust-region radii are equal and are updated simultaneously by the same factor. Thus, for every k, $\hat{\delta}_i \leq (M+1)\delta_k$. The simultaneous expansion or contraction of the trust-region radii is not a technical requirement. The algorithm for evaluating the step and updating the trust-region radii follows.

Let $ared_i$ and $pred_i$ be defined as in expressions (3.5) and (3.6). Let $r = \frac{ared_i}{pred_i}$.

ALGORITHM 3.2. **Step evaluation / trust-region update**

Given $0 < \delta_{min} \leq \delta_k \leq \delta_{max}$, $k = 1, \ldots, \delta_{M+1}$, $0 < \eta_1 < \eta_2 < 1$, $\gamma_1 \in (0, 1]$, $\gamma_2 > 1$, r,

if $r < \eta_1$ **then** (step not accepted; δ_k decreased)
  $\delta_k = \gamma_1 ||s_k||$, for $k = 1, \ldots, M+1$
  $x_{i+1} = x_i$

else if $r \geq \eta_2$ **then** (step accepted; δ_k increased)
$\delta_k = \min\{\delta_{max}, \max\{\delta_{min}, \gamma_2\delta_k\}\}$ for $k = 1, \ldots, M+1$
$x_{i+1} = x_i + \hat{s}_i$
else (step accepted; δ_k unchanged)
$\delta_k = \max\{\delta_{min}, \delta_k\}$ for $k = 1, \ldots, M+1$
$x_{i+1} = x_i + \hat{s}_i$
end if

Typical values of the constants can be found in [13], for instance.

We note that if the step is not accepted, the trust-region radii are decreased without any restrictions. However, if the step is accepted, the next trust-region radius is set to be no smaller than a predetermined positive number, δ_{min}. This technique is used by El-Hallabi and Tapia [18], for instance. Its role in the global convergence analysis is to ensure that the trust-region radii are bounded away from zero.

3.8. The stopping criteria. To terminate the algorithm, we require that

$$||P_M^T \nabla f(y_M)|| + \sum_{k=1}^{M} ||C_k(y_{k-1})|| \leq \epsilon_{tol} \tag{3.10}$$

holds for some small $\epsilon_{tol} > 0$.

While this is a conventional termination criterion useful for analysis, in practice one would impose a distinct termination criterion for each block of constraints followed by the projected gradient of the objective. The more complex criterion is not used here, as it complicates the sufficient decrease condition for the subproblems and is better addressed in a work on implementation.

3.9. The statement of the algorithm. We can now provide a complete formal statement of the MAESTRO class.

ALGORITHM 3.3. **MAESTRO class for problem EQC**
Given $x_0 \in \Re^n$, $\delta_k > 0, k = 1, \ldots, M+1$,
do
Set $y_0 = x_c$
compute the trial step $\hat{s}_i$:
do $k = 1, M$
if $||C_k(y_{k-1})|| > 0$ **then**
compute s_k that satisfies FCD on
minimize $\quad ||C_k(y_{k-1}) + \nabla C_k^T(y_{k-1})s||^2$
subject to $\quad \nabla C_j^T(y_{j-1})s = 0, j = 1, \ldots, k-1,$
$\quad ||s|| \leq \delta_k$
else
$s_k = 0$
end if
$y_k = y_{k-1} + s_k$
end do
if $||P_M^T \nabla f(y_M)|| + \sum_{j=1}^{M} ||C_j(y_{j-1})|| \leq \epsilon_{tol}$ **then terminate**
if $||P_M^T \nabla f(y_M)|| > 0$ **then**
compute s_{M+1} that satisfies FCD on
minimize $\quad f(y_M) + \nabla f^T(y_M)s + \frac{1}{2}s^T H_M s$
subject to $\quad \nabla C_j^T(y_{j-1})s = 0, j = 1, \ldots, M$
$\quad ||s|| \leq \delta_{M+1}$

else

$s_{M+1} = 0$

end if

$y_{M+1} = y_M + s_{M+1}$

the trial step is $\hat{s}_i = \sum_{i=1}^{M+1} s_k$

update the penalty parameters using Algorithm (3.1)

update x_i **and** $\delta_k, k = 1, \ldots, M+1$ **using Algorithm (3.2)**

end do

Notice that every trial step $\hat{s}_i$ results in an increase in the iteration count whether or not it is an acceptable step. Recall also that after each acceptable step $x_{i+1} \neq x_i$, $\delta_k \geq \delta_{min}$.

We should note that there is an option to eliminate only a subset of constraints via the described procedure. In that case, the rest of the constraints and the objective function would be restricted to the intersection of the null spaces of the Jacobians of the processed constraints, and the resulting reduced optimization problem would be solved by a chosen method. This approach has bearing on the computational structure of some problems in engineering optimization, and we relegate it to future work.

4. Global convergence analysis for Variant I. Global convergence analysis for the MAESTRO class is composed of the following main ingredients. First, we show that the total trial step satisfies a sufficient decrease condition, given that the substeps satisfy FCD on the reduced subproblems. Second, the difference between the actual and predicted reduction is shown to be bounded above by appropriate powers of the total trial step norm. Third, the algorithm is shown to be well-defined; that is, a successful step can be found after a finite number of trial steps. Fourth, the penalty parameters are shown to be bounded under the nontermination hypothesis. The trust-region radii are shown bounded away from zero. Finally, we establish weak first-order convergence. The analysis follows that in El-Alem [15].

4.1. Accounting for autonomous processing of the subproblems. The following lemma accounts for the effect of each substep s_k on the constraint blocks numbered $k+1, \ldots, M$ and on the objective function.

LEMMA 4.1. *There exist positive constants $\mu_1, \ldots, \mu_M$ independent of i, such that*

$$||C_k(y_0)||^2 - ||C_k(y_{k-1})||^2 \geq -\mu_{k-1}\hat{\delta}_i \sum_{j=1}^{k-1} ||C_j(y_{j-1})||, \quad k = 2, \ldots, M, \text{ and} \tag{4.1}$$

$$f(y_0) - f(y_M) \geq -\mu_M \sum_{j=1}^{M} ||C_j(y_{j-1})||. \tag{4.2}$$

Proof. We have, for some $z_k \in (y_0, y_{k-1})$ and $k = 2, \ldots, M$,

$$||C_k(y_0)||^2 - ||C_k(y_{k-1})||^2 = -2[\nabla C_k(y_0) C_k(y_0)]^T \sum_{j=1}^{k-1} s_j$$

$$- \frac{1}{2}(\sum_{j=1}^{k-1} s_j)^T [\nabla C_k(z_k) \nabla C_k(z_k)^T + \sum_{l=n_k}^{n_{k+1}-1} c_l(z_k) \nabla^2 c_l(z_k)] \sum_{j=1}^{k-1} s_j$$

$$\geq -2\|\nabla C_k(y_0)\|\|C_k(y_0)\|\sum_{j=1}^{k-1}\|s_j\| - \frac{1}{2}[\|\nabla C_k(y_0)\|^2$$
$$+ \sum_{l=n_k}^{n_{k+1}-1}\|c_l(z_k)\|\|\nabla^2 c_l(z_k)\|](\sum_{j=1}^{k-1}\|s_j\|)^2.$$

Taking into account that

$$\|C_k(y_0)\| - \|C_k(y_{k-1})\| \leq \|C_k(y_0) - C_k(y_{k-1})\| \leq \sigma_6\|\sum_{j=1}^{k-1} s_j\| \leq \sigma_6 \sum_{j=1}^{k-1}\|s_j\|,$$

we have

$$\|C_k(y_0)\| \leq \sigma_6 \sum_{j=1}^{k-1}\|s_j\| + \|C_k(y_{k-1})\| \leq \sigma_6 \sum_{j=1}^{k-1}\|s_j\| + \Gamma_2\|s_k\|$$
$$\leq (\sigma_6 + \Gamma_2\|s_k\|)\sum_{j=1}^{k-1}\|s_j\| \leq (\sigma_6 + \Gamma_2\delta_{max})\sum_{j=1}^{k-1}\|s_j\|.$$

Then

$$\|C_k(y_0)\|^2 - \|C_k(y_{k-1})\|^2$$
$$\geq -\{2\sigma_6(\sigma_6 + \Gamma_2\delta_{max}) + \frac{1}{2}[\sigma_6^2 + (n_{k+1} - n_k)\sigma_5\sigma_7]\}(\sum_{j=1}^{k-1}\|s_j\|)^2$$
$$\geq -\{2\sigma_6(\sigma_6 + \Gamma_2\delta_{max}) + \frac{1}{2}[\sigma_6^2 + (n_{k+1} - n_k)\sigma_5\sigma_7]\}\Gamma_1\Gamma_3\hat{\delta}_i \sum_{j=1}^{k-1}\|C_j(y_{j-1})\|$$
$$\equiv -\mu_{k-1}\hat{\delta}_i \sum_{j=1}^{k-1}\|C_j(y_{j-1})\|.$$

Similarly, for some $z_M \in (y_0, y_M)$, we have

$$f(y_0) - f(y_M) = -\nabla f(z_M)^T(s_1 + \ldots + s_M)$$
$$\geq -\sigma_2\sum_{j=1}^{M}\|s_j\| \geq -\sigma_2\Gamma_1\sum_{j=1}^{M}\|C_j(y_{j-1})\|$$
$$\equiv -\mu_M \sum_{j=1}^{M}\|C_j(y_{j-1})\|,$$

which concludes the proof. □

4.2. The behavior of the model. The following lemma provides an estimate for the cumulative predicted reductions.

LEMMA 4.2. *Let $s_1, \ldots, s_{M+1}$ be the substeps generated at the current iterate $x_i = y_0$, and let $\rho_1, \ldots, \rho_M$ be the penalty parameters. Then the partial and the total predicted reductions satisfy the following estimates:*

$$Cpred_k \geq \Gamma_4\|C_k(y_{k-1})\|\min\{\Gamma_5\|C_k(y_{k-1})\|, \delta_k\} - \mu_{k-1}\hat{\delta}_i\sum_{j=1}^{k-1}\|C_j(y_{j-1})\| + \rho_{k-1}Cpred_{k-1} \tag{4.3}$$

and

$$\text{(4.4)} \qquad pred_i \geq \frac{\sigma}{2}||P_M^T \nabla f(y_M)|| \min\{\Gamma_6 ||P_M^T \nabla f(y_M)||, \delta_{M+1}\}$$
$$- \mu_M \sum_{j=1}^{M} ||C_j(y_{j-1})|| + \rho_M Cpred_M,$$

where $\Gamma_4, \Gamma_5, \Gamma_6$ *are as in Lemma (3.3) and* $\mu_1, \ldots, \mu_M$ *are as in Lemma (4.1).*

Proof. Consider

$$Cpred_k = ||C_k(y_0)||^2 - ||C_k(y_{k-1}) + \nabla C_k^T(y_{k-1}) s_k||^2 + \rho_{k-1} Cpred_{k-1}$$
$$\pm ||C_k(y_{k-1})||^2.$$

Applying Lemmas (3.3) and (4.1) to the right-hand side yields

$$Cpred_k \geq \Gamma_4 ||C_k(y_{k-1})|| \min\{\Gamma_5 ||C_k(y_{k-1})||, \delta_k\}$$
$$- \mu_{k-1} \hat{\delta}_i \sum_{j=1}^{k-1} ||C_j(y_{j-1})|| + \rho_{k-1} Cpred_{k-1}.$$

Similarly, for the total predicted reduction we have

$$\begin{aligned} pred_i &= f(y_0) - \phi_M(s_{M+1}) + \rho_M Cpred_M \pm f(y_M) \\ &\geq \frac{\sigma}{2}||P_M^T \nabla f(y_M)|| \min\{\Gamma_6 ||P_M^T \nabla f(y_M)||, \delta_{M+1}\} \\ &\quad - \mu_M \sum_{j=1}^{M} ||C_j(y_{j-1})|| + \rho_M Cpred_M, \end{aligned}$$

which concludes the proof. $\square$

The following lemma provides an upper bound on the difference between the actual reduction and the predicted reduction. In particular, if $\{\rho_k^i\}$ are bounded, the estimate indicates that the predicted reduction approximates the actual reduction with $O(||\hat{s}_i||^2)$ accuracy.

LEMMA 4.3. *There exist positive constants* $\Gamma_7, \Gamma_8, \Gamma_9$, *and* $\nu_k, k = 1, \ldots, M$, *independent of the iterates, such that*

$$\text{(4.5)} \quad |ared_i - pred_i| \leq \Gamma_7 ||\hat{s}_i||^2 + \Gamma_8 (\prod_{j=1}^{M} \rho_j) ||\hat{s}_i||^3 + \sum_{k=1}^{M} (\prod_{j=k}^{M} \rho_j) \nu_k ||C_k(y_{k-1})|| ||\hat{s}_i||^2$$

and

$$\text{(4.6)} \qquad |ared_i - pred_i| \leq \Gamma_9 (\prod_{j=1}^{M} \rho_j) ||\hat{s}_i||^2.$$

Proof. The proof of inequality (4.5) is a straightforward, repeated application of the mean-value theorem, the Cauchy-Schwartz inequality, Lemma (3.2), and the problem assumptions. Boundedness of $||s_k||$ and $||C_k||$, together with (4.5), yields inequality (4.6). Detailed proof follows the work of El-Alem [14] and can be found in [1]. $\square$

If a particular subiterate y_{k-1} is cumulatively block-feasible, i.e., if $||C_j(y_{j-1})|| = 0$, $j = 1, \ldots, k-1$, or if $||C_k(y_{k-1})|| = 0$, the corresponding penalty parameter ρ_{k-1} is not increased. It is necessary to show, in addition, that the penalty parameters will not be increased if the iterates are sufficiently close to cumulative block-feasibility. Note that because of the autonomous processing of the constraint blocks, one has to be concerned not only about the overall feasibility, but also about the relative feasibility of one block with respect to another.

LEMMA 4.4. *Let $\omega_1 < \ldots < \omega_M$ be positive constants defined recursively as*

$$\omega_M \leq \min\{\frac{\epsilon_{tol}}{2\delta_{max}}, \frac{\sigma\epsilon_{tol}}{8\mu_M}\min\{\frac{\Gamma_6\epsilon_{tol}}{2\delta_{max}}, \frac{1}{M+1}\}\} \text{ and}$$

$$\omega_{k-1} \leq \min\{\frac{\omega_k}{2}, \frac{\Gamma_4\omega_k}{4\mu_k}\min\{\frac{\Gamma_5\omega_k}{2}, \frac{1}{M+1}\}\} \text{ for } k = 2, \ldots, M.$$

If $||P_M^T\nabla f(y_M)|| + \sum_{j=1}^M ||C_j(y_{j-1})|| > \epsilon_{tol}$ and $\sum_{j=1}^M ||C_j(y_{j-1})|| \leq \omega_M\hat{\delta}_i$, then

$$pred_i \geq \frac{\sigma}{4}||P_M^T\nabla f(y_M)||\min\{\Gamma_6||P_M^T\nabla f(y_M)||, \delta_{M+1}\} + \rho_M Cpred_M \tag{4.7}$$

for any $\rho_M > 0$, and ρ_M is not increased in Algorithm (3.1).

If $\sum_{j=1}^k ||C_j(y_{j-1})|| > \omega_k\hat{\delta}_i$ for some $2 \leq k \leq M$ and $\sum_{j=1}^{k-1} ||C_j(y_{j-1})|| \leq \omega_{k-1}\hat{\delta}_i$, then

$$Cpred_k \geq \frac{\Gamma_4}{2}||C_k(y_{k-1})||\min\{\Gamma_5||C_k(y_{k-1})||, \delta_k\} + \rho_{k-1}Cpred_{k-1} \tag{4.8}$$

for any $\rho_{k-1} > 0$, and ρ_{k-1} is not increased in Algorithm (3.1).

Proof. Note that $\omega_1 < \omega_2 < \ldots < \omega_M$ by construction.

By Lemma (4.2),

$$\begin{aligned} pred_i \geq\ & \frac{\sigma}{2}||P_M^T\nabla f(y_M)||\min\{\Gamma_6||P_M^T\nabla f(y_M)||, \delta_{M+1}\} \\ & - \mu_M\sum_{j=1}^M ||C_j(y_{j-1})|| + \rho_M Cpred_M. \end{aligned}$$

If $||P_M^T\nabla f(y_M)|| + \sum_{j=1}^M ||C_j(y_{j-1})|| > \epsilon_{tol}$ and $\sum_{j=1}^M ||C_j(y_{j-1})|| \leq \omega_M\hat{\delta}_i$, then $||P_M^T\nabla f(y_M)|| > \epsilon_{tol} - \omega_M\hat{\delta}_i \geq \frac{\epsilon_{tol}}{2}$, where the last inequality follows from the definition of ω_M. Taking into account that $\delta_{M+1} \geq \frac{\hat{\delta}_i}{M+1}$, we have

$$\begin{aligned} pred_i \geq\ & \frac{\sigma}{4}||P_M^T\nabla f(y_M)||\min\{\Gamma_6||P_M^T\nabla f(y_M)||, \delta_{M+1}\} \\ & + \frac{\sigma\epsilon_{tol}\hat{\delta}_i}{8}\min\{\frac{\Gamma_6\epsilon_{tol}}{2\delta_{max}}, \frac{1}{M+1}\} - \mu_M\omega_M\hat{\delta}_i \\ & + \rho_M Cpred_M. \end{aligned}$$

Because $\omega_M \leq \frac{\sigma\epsilon_{tol}\hat{\delta}_i}{8\mu_M}\min\{\frac{\Gamma_6\epsilon_{tol}}{2\delta_{max}}, \frac{1}{M+1}\}$,

$$\frac{\sigma\epsilon_{tol}\hat{\delta}_i}{8}\min\{\frac{\Gamma_6\epsilon_{tol}}{2\delta_{max}}, \frac{1}{M+1}\} - \mu_M\omega_M\hat{\delta}_i \geq 1$$

and (4.7) holds. Hence, ρ_M is not increased in Algorithm (3.1).

To account for $\rho_1, \ldots, \rho_{M-1}$ Lemma (4.2) gives us for $2 \le k \le M$,

$$Cpred_k \ge \Gamma_4 \|C_k(y_{k-1})\| \min\{\Gamma_5 \|C_k(y_{k-1})\|, \delta_k\} - \mu_{k-1}\hat{\delta}_i \sum_{j=1}^{k-1} \|C_j(y_{j-1})\| + \rho_{k-1} Cpred_{k-1}.$$

If $\sum_{j=1}^{k} \|C_j(y_{j-1})\| > \omega_k \hat{\delta}_i$ and $\sum_{j=1}^{k-1} \|C_j(y_{j-1})\| \le \omega_{k-1}\hat{\delta}_i$, then $\|C_k(y_{k-1})\| > (\omega_k - \omega_{k-1}\hat{\delta}_i \ge \frac{\omega_k}{2}\hat{\delta}_i$, because $\omega_{k-1} \le \frac{\omega_k}{2}$. Taking into account that $\delta_k \ge \frac{\hat{\delta}_i}{M+1}$, we have

$$\begin{aligned} Cpred_k \ge\; & \frac{\Gamma_4}{2} \|C_k(y_{k-1})\| \min\{\Gamma_5 \|C_k(y_{k-1})\|, \delta_k\} \\ & + \frac{\Gamma_4 \omega_k \hat{\delta}_i^2}{4} \min\{\frac{\Gamma_5 \omega_k}{2}, \frac{1}{M+1}\} - \omega_{k-1}\mu_{k-1}\hat{\delta}_i^2 \\ & + \rho_{k-1} Cpred_{k-1}. \end{aligned}$$

Because $\omega_{k-1} \le \frac{\Gamma_4 \omega_k}{4\mu_{k-1}} \min\{\frac{\Gamma_5\omega_k}{2}, \frac{1}{M+1}\}$,

$$\frac{\Gamma_4 \omega_k \hat{\delta}_i^2}{4} \min\{\frac{\Gamma_5 \omega_k}{2}, \frac{1}{M+1}\} - \omega_{k-1}\mu_{k-1}\hat{\delta}_i^2 \ge 0,$$

and inequality (4.8) holds. Therefore, ρ_{k-1} is not increased in Algorithm (3.1). ☐

The following lemma provides a lower bound on the predicted reduction if the iterate is sufficiently close to feasibility.

Lemma 4.5. *Let* $\|P_M^T \nabla f(y_M)\| + \sum_{j=1}^{M} \|C_j(y_{j-1})\| > \epsilon_{tol}$ *and* $\sum_{j=1}^{M} \|C_j(y_{j-1})\| \le \omega_M \hat{\delta}_i$, *with* ω_M *defined in Lemma (4.4). Then there exists a positive constant* τ_M *that depends on* ϵ_{tol} *but not on* ρ_M *or* i, *such that*

$$pred_i \ge \tau_M \hat{\delta}_i. \tag{4.9}$$

If $\sum_{j=1}^{k} \|C_j(y_{j-1})\| > \omega_k \hat{\delta}_i$ *and* $\sum_{j=1}^{k-1} \|C_j(y_{j-1})\| \le \omega_{k-1}\hat{\delta}_i$ *for some* $2 \le k \le M$, *where* ω_{k-1} *is defined in Lemma (4.4), then there exists a positive constant* τ_{k-1}, *that depends on* ϵ_{tol} *but not on* ρ_k *or* i, *such that*

$$Cpred_k \ge \tau_{k-1} \hat{\delta}_i^2. \tag{4.10}$$

Proof. If $\|P_M^T \nabla f(y_M)\| + \sum_{j=1}^{M} \|C_j(y_{j-1})\| > \epsilon_{tol}$ and $\sum_{j=1}^{M} \|C_j(y_{j-1})\| \le \omega_M \hat{\delta}_i$, then by Lemma (4.4),

$$\begin{aligned} pred_i &\ge \frac{\sigma}{4} \|P_M^T \nabla f(y_M)\| \min\{\Gamma_6 \|P_M^T \nabla f(y_M)\|, \delta_{M+1}\} + \rho_M Cpred_M \\ &\ge \frac{\sigma}{4} \|P_M^T \nabla f(y_M)\| \min\{\Gamma_6 \|P_M^T \nabla f(y_M)\|, \delta_{M+1}\} \\ &\ge \frac{\sigma \epsilon_{tol}}{8} \min\{\frac{\Gamma_6 \epsilon_{tol}}{2}, \delta_{M+1}\} \ge \frac{\sigma \epsilon_{tol} \hat{\delta}_i}{8} \min\{\frac{\Gamma_6 \epsilon_{tol}}{\delta_{max}}, \frac{1}{M+1}\} \equiv \tau_M \hat{\delta}_i. \end{aligned}$$

Similarly, if $\sum_{j=1}^{k} \|C_j(y_{j-1})\| > \omega_k \hat{\delta}_i$ and $\sum_{j=1}^{k-1} \|C_j(y_{j-1})\| \le \omega_{k-1}\hat{\delta}_i$ for some $2 \le k \le M$, then by Lemma (4.4),

$$Cpred_k \ge \frac{1}{2}\Gamma_4 \|C_k(y_{k-1})\| \min\{\Gamma_5 \|C_k(y_{k-1})\|, \delta_k\} + \rho_{k-1} Cpred_{k-1}$$

$$\geq \frac{1}{2}\Gamma_4\|C_k(y_{k-1})\| \min\{\Gamma_5\|C_k(y_{k-1})\|, \delta_k\}$$
$$\geq \frac{\Gamma_4\omega_k\hat{\delta}_i^2}{4}\min\{\frac{\Gamma_5\omega_k}{2}, \frac{1}{M+1}\} \equiv \tau_{k-1}\hat{\delta}_i^2,$$

which completes the proof. $\square$

4.3. Behavior of the penalty parameters. The penalty parameter sequences are shown to be nondecreasing, which, together with their boundedness from above, will allow us to conclude that the penalty parameters tend to a limit and, moreover, stay constant after a finite number of outer iterations. The following lemma establishes a relation between the trust-region radii and the penalty parameters.

LEMMA 4.6. *For each $k = 1, \ldots, M$, there exists a constant Θ_k, independent of i, such that if ρ_k is increased at some iteration i, then*

$$\rho_k\,\delta_k \;\leq\; \Theta_k. \tag{4.11}$$

Proof. Suppose ρ_M is increased; that is,

$$\rho_M = \frac{2\,\overbrace{[\phi_M(s_{M+1}) - f(y_0)]}^{A}}{Cpred_M} + \beta, \quad \text{with } \beta \in (0,1).$$

Then for some $z_1, z_2 \in (y_0, y_{M+1})$, we have

$$\begin{aligned}|A| &= |\phi_M(s_{M+1}) - f(y_0) \pm f(y_{M+1})| \\ &\leq \frac{1}{2}|(s_{M+1})^T[H_M - \nabla^2 f(z_1)]s_{M+1}| + \|\nabla f(z_2)\|\|y_{M+1} - y_0\| \\ &\leq \frac{1}{2}(\sigma_3+\sigma_4)\Gamma_3\|\hat{s}_i\|^2 + \sigma_2\|\hat{s}_i\| \quad \text{and}\end{aligned}$$

$$\rho_M Cpred_M \leq \underbrace{2\sigma_2\|\hat{s}_i\| + (\sigma_3+\sigma_4)\Gamma_3\|\hat{s}_i\|^2}_{B} + \beta Cpred_M. \tag{4.12}$$

Observe that because ρ_M is increased, the algorithm does not terminate at x_i; i.e., $\|P_M^T\nabla f(y_M)\| + \sum_{j=1}^M \|C_j(y_{j-1}\| > \epsilon_{tol}$. Lemma (4.4) applies and $\sum_{j=1}^M \|C_j(y_{j-1}\| > \omega_M\hat{\delta}_i$. Now, either $\sum_{j=1}^{M-1}\|C_j(y_{j-1})\| \leq \omega_{M-1}\hat{\delta}_i$ or $\sum_{j=1}^{M-1}\|C_j(y_{j-1})\| > \omega_{M-1}\hat{\delta}_i$.

Case 1: $\sum_{j=1}^{M-1}\|C_j(y_{j-1})\| \leq \omega_{M-1}\hat{\delta}_i$. Applying Lemma (4.4) to the left hand side of inequality (4.12) yields

$$\begin{aligned}&\frac{\rho_M}{2}[\Gamma_4\|C_M(y_{M-1})\|\min\{\Gamma_5\|C_M(y_{M-1})\|, \delta_M\} + \rho_{M-1}Cpred_{M-1}] \\ &\leq B + \beta[\|C_M(y_0)\|^2 - \|C_M(y_{M-1}) + \nabla C_M^T(y_{M-1})s_M\|^2 + \rho_{M-1}Cpred_{M-1}]\end{aligned}$$

or, because $\rho_j \geq 1, j = 1, \ldots, M$,

$$\begin{aligned}&\frac{\rho_M}{2}\Gamma_4\|C_M(y_{M-1})\|\min\{\Gamma_5\|C_M(y_{M-1})\|, \delta_M\} \\ &\leq B + \beta\,\underbrace{[\|C_M(y_0)\|^2 - \|C_M(y_{M-1}) + \nabla C_M^T(y_{M-1})s_M\|^2]}_{D} \\ &+\rho_{M-1}(\beta-1)Cpred_{M-1} \leq B + \beta D,\end{aligned} \tag{4.13}$$

where the last inequality is due to $\beta - 1 < 0$.

Now for some $z_3, z_4 \in (y_{M-1}, y_M)$,

$$\begin{aligned}
|D| &\leq |||C_M(y_0)||^2 - ||C_M(y_M)||^2| + |||C_M(y_M)||^2 - ||C_M(y_{M-1}) + \nabla C_M^T(y_{M-1})s_M||^2| \\
&\leq 2||\nabla C_M(z_3)||||C_M(z_3)|| \sum_{j=1}^{M} ||s_j|| \\
&+ |s_M^T[\nabla C_M(y_{M-1})\nabla C_M(y_{M-1})^T - \frac{1}{2}\nabla C_M(z_4)\nabla C_M(z_4)^T - \frac{1}{2}\sum_{l=n_M}^{m} c_l(z_4)\nabla^2 c_l(z_4)]s_M| \\
&\leq 2\sigma_5\sigma_6\Gamma_3||\hat{s}_i|| + [\frac{3}{2}\sigma_6^2 + \frac{1}{2}(m - n_m + 1)\sigma_5\sigma_7]\Gamma_3||\hat{s}_i||^2.
\end{aligned}$$

Substituting expressions for B and D into inequality (4.13) yields

$$\begin{aligned}
&\frac{\rho_M}{2}\Gamma_4||C_M(y_{M-1})|| \min\{\Gamma_5||C_M(y_{M-1})||, \delta_M\} \\
&\leq 2(\sigma_2 + \beta\sigma_5\sigma_6\Gamma_3)||\hat{s}_i|| + \Gamma_3\{\sigma_3 + \sigma_4 + \beta[\frac{3}{2}\sigma_6^2 + \frac{1}{2}(m - n_M + 1)\sigma_5\sigma_7)]\}||\hat{s}_i||^2 \\
&\leq \underbrace{\{2(\sigma_2 + \beta\sigma_5\sigma_6\Gamma_3) + \Gamma_3[\sigma_3 + \sigma_4 + \beta(\frac{3}{2}\sigma_6^2 + \frac{1}{2}(m - n_M + 1)\sigma_5\sigma_7)]\delta_{max}\}}_{E}\hat{\delta}_i.
\end{aligned}$$

Because $\sum_{j=1}^{M}||C_j(y_{j-1})|| > \omega_M\hat{\delta}_i$ and $\sum_{j=1}^{M-1}||C_j(y_{j-1})|| \leq \omega_{M-1}\hat{\delta}_i$,

$$||C_M(y_{M-1})|| > (\omega_M - \omega_{M-1})\hat{\delta}_i > 0, \text{ (since } \omega_M > \omega_{M+1}).$$

Hence

$$\frac{\rho_M}{2}\Gamma_4(\omega_M - \omega_{M-1})\hat{\delta}_i^2 \min\left\{\Gamma_5(\omega_M - \omega_{M-1}), \frac{1}{M+1}\right\} \leq E\hat{\delta}_i.$$

Denoting $E/\left[\frac{\Gamma_4}{2}(\omega_M - \omega_{M-1})\min\left\{\Gamma_5(\omega_M - \omega_{M-1}), \frac{1}{M+1}\right\}\right]$ by Θ_{M_1} yields $\rho_M\hat{\delta}_i \leq \Theta_{M_1}$.

$\boxed{\textbf{Case 2:}}$ $\sum_{j=1}^{M-1}||C_j(y_{j-1})|| > \omega_{M-1}\hat{\delta}_i$. Again consider: $\sum_{j=1}^{M-2}||C_j(y_{j-1})|| \leq \omega_{M-2}\hat{\delta}_i$ and $\sum_{j=1}^{M-2}||C_j(y_{j-1})|| > \omega_{M-2}\hat{\delta}_i$.

$\boxed{\textbf{Case 2.1:}}$ $\sum_{j=1}^{M-2}||C_j(y_{j-1})|| \leq \omega_{M-2}\hat{\delta}_i$. Applying Lemma (3.5) once to the left hand side of inequality (4.12) results in

$$\begin{aligned}
\frac{\rho_M\rho_{M-1}}{2}Cpred_{M-1} &\leq B + \beta[||C_M(y_0)||^2 - ||C_M(y_{M-1}) + \nabla C_M^T(y_{M-1})s_M||^2] \\
&+ \beta\rho_{M-1}Cpred_{M-1},
\end{aligned}$$

and because $\rho_j \geq 1$ for all j,

$$\frac{\rho_M}{2}Cpred_{M-1} \leq B + D\hat{\delta}_i + \beta\rho_{M-1}Cpred_{M-1}.$$

And now, following an argument analogous to that of Case 1 leads to $\rho_M\hat{\delta}_i \leq \Theta_{M_2}$.

$\boxed{\textbf{Case 2.2:}}$ $\sum_{j=1}^{M-2}||C_j(y_{j-1})|| > \omega_{M-2}\hat{\delta}_i$. And again we consider two cases. $\sum_{j=1}^{M-3}||C_j(y_{j-1})|| \leq \omega_{M-3}\hat{\delta}_i$ leads to $\rho_M\hat{\delta}_i \leq \Theta_{M_3}$, while $\sum_{j=1}^{M-3}||C_j(y_{j-1})|| > \omega_{M-3}\hat{\delta}_i$ splits into two other cases.

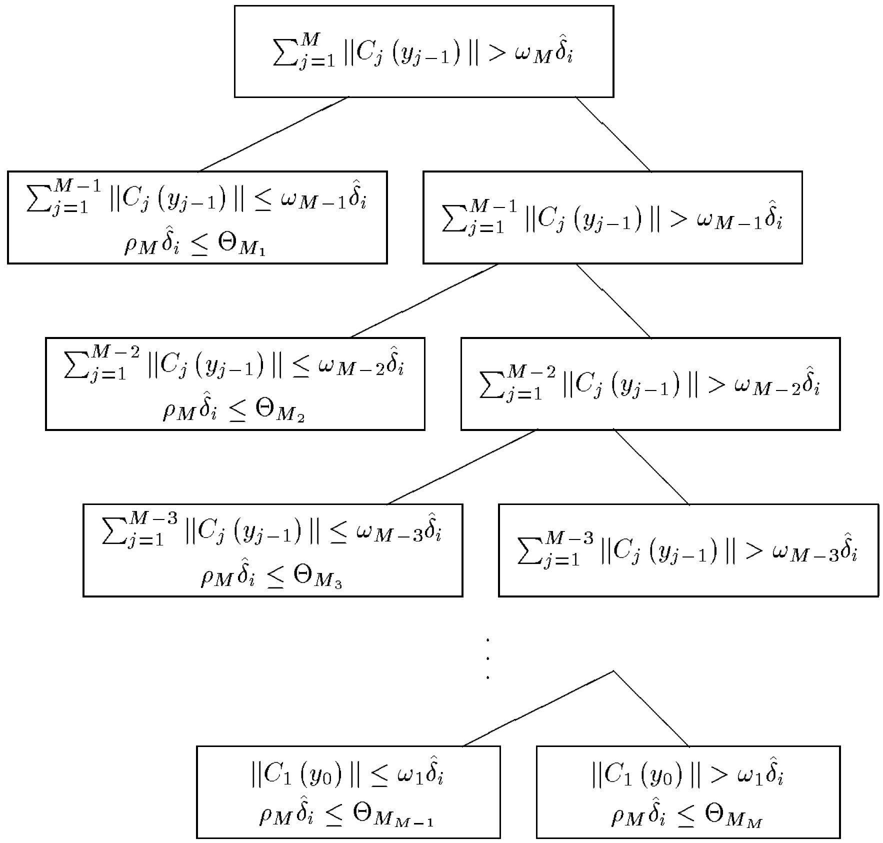

FIG. 4.1. *Flow chart for the proof of Lemma (4.6).*

The branching argument—illustrated in Figure 4.1—continues to obtain upper bounds on $\rho_M \hat{\delta}_i$ until we reach the branch on $\|C_1(y_0)\|$. For $\|C_1(y_0)\| \le \omega_1 \hat{\delta}_i$, the reasoning is analogous to that of the previous branches, resulting in $\rho_M \hat{\delta}_i \le \Theta_{M_{M-1}}$. Finally, for $C_1(y_0) > \omega_1 \hat{\delta}_i$, the application of Lemma (3.5) to inequality (4.12) yields

$$\begin{aligned}
\frac{\rho_1 \rho_2 \cdots \rho_M}{2^M} Cpred_1 &\le B + \beta Cpred_M \\
&= B + \beta[\|C_M(y_0)\|^2 - \|C_M(y_{M-1}) + \nabla C_M^T(y_{M-1}) s_M\|^2] \\
&+ \beta \sum_{k=1}^{M-1} \Big(\prod_{j=k}^{M-1} \rho_j \Big) [\|C_k(y_0)\|^2 - \|C_k(y_{k-1}) + \nabla C_k^T(y_{k-1}) s_k\|^2].
\end{aligned}$$

Taking into account $\rho_j \ge 1$ for all j and applying Lemma (3.3) to the left hand side, we obtain

$$\begin{aligned}
&\frac{\rho_M}{2^M} \Gamma_4 \|C_1(y_0)\| \min\{\Gamma_5 \|C_1(y_0)\|, \delta_1\} \\
&\le B + \beta \sum_{k=1}^{M} [\|C_k(y_0)\|^2 - \|C_k(y_{k-1}) + \nabla C_k^T(y_{k-1}) s_k\|^2].
\end{aligned}$$

Because $||C_1(y_0)|| > \omega_1 \hat{\delta}_i$, we have

$$
\begin{aligned}
&\frac{\rho_M}{2^M}\Gamma_4\omega_1\hat{\delta}_i \min\{\Gamma_5\omega_1\hat{\delta}_i, \frac{\hat{\delta}_i}{M+1}\} \\
&\leq B + \beta \sum_{k=1}^{M} [||C_k(y_0)||^2 - ||C_k(y_{k-1})^T s_k||^2 \pm ||C_k(y_{k-1})||^2] \\
&\leq B + \beta M \Gamma_3 [2\sigma_5\sigma_6 ||\hat{s}_i|| + (\frac{3}{2}\sigma_6^2 + \frac{1}{2}\sigma_5\sigma_7)||\hat{s}_i||^2]
\end{aligned}
$$

or, replacing B with its value,

$$
\begin{aligned}
&\frac{\rho_M}{2}\Gamma_4\omega_1\hat{\delta}_i^2 \min\{\Gamma_5\omega_1, \frac{1}{M+1}\} \\
&\leq \{2\sigma_2 + \beta M\sigma_5\sigma_6\Gamma_3 + \Gamma_3[\sigma_3 + \sigma_4 + \beta M(\frac{3}{2}\sigma_6^2 + \frac{1}{2}M\sigma_5\sigma_7)](M+1)\,\delta_{max}\}\hat{\delta}_i,
\end{aligned}
$$

which yields $\rho_M \hat{\delta}_i \leq \Theta_{M_M}$ with

$$
\Theta_{M_M} \equiv \frac{\{2(\sigma_2 + \beta M\sigma_5\sigma_6\Gamma_3) + \Gamma_3[\sigma_3 + \sigma_4 + \beta M(\frac{3}{2}\sigma_6^2 + \frac{1}{2}\sigma_5\sigma_7)]\}\delta_{max}}{\frac{\Gamma_4\omega_1}{2^M}\min\{\Gamma_5\omega_1, \frac{1}{M+1}\}}.
$$

Now setting $\Theta_M = \max\{\Theta_{M_1}, \ldots, \Theta_{M_M}\}$, we have

$$
\rho_M \hat{\delta}_i \leq \Theta_M.
$$

For $\rho_k, k = 1, \ldots, M-1$, the proof is analogous but it starts at a lower "branch" of the argument illustrated in Figure 4.1. □

The following lemma shows that the trust-region radius is bounded below and the penalty parameter is bounded above at an iterate where the algorithm does not terminate and any penalty parameter is increased. The two results are shown pairwise for ρ_1 before they can be proven for ρ_2 and the subsequent penalty parameters.

LEMMA 4.7. *Assume that the algorithm does not terminate at an iterate x_i and that any of the penalty parameters are increased. Then*

1. *there exists a positive constant $\bar{\delta}$ such that*

$$
\hat{\delta}_i \geq \bar{\delta} \quad \textit{and} \tag{4.14}
$$

2. *for $k = 1, \ldots, M$, $\{\rho_k^i\}$ converges to a limit ρ_k^*. Moreover, there exist positive integers $i_{\rho_k}, k = 1, \ldots, M$, such that $\rho_k^i = \rho_k^*$ for all $i \geq i_{\rho_k}$;*
3. *finally, the merit function $\bar{\mathcal{P}}$ is bounded on Ω.*

Proof. The general idea is to use the two upper bounds from Lemma (4.3) and the lower bound on the predicted reduction from Lemma (3.5) to obtain an upper bound on the ratio of the difference of the reductions to the predicted reduction in terms of the step norm.

Let $\hat{s}_i = s_1^i + \ldots + s_{M+1}^i$ be the total trial step generated form x_i and let $\hat{s}_-$ be the last acceptable step. Let the steps between $\hat{s}_-$ and $\hat{s}_i$ be numbered

$$
\hat{s}_-, \hat{s}_{i_1}, \hat{s}_{i_2}, \ldots, \hat{s}_{i_L}, \hat{s}_{i_{L+1}} = \hat{s}_i.
$$

The current trial step $\hat{s}_i$ can be either acceptable or unacceptable.

Let us first assume that ρ_1^i is increased. Without loss of generality, we can assume that $\sum_{j=1}^{2} ||C_j(y_{j-1}^i)|| > \omega_2\hat{\delta}_i$. By Lemma (4.4), $||C_1(y_0^i)|| > \omega_1\hat{\delta}_i$.

If there are no unacceptable steps between $\hat{s}_-$ and $\hat{s}_i$, then the method of updating the trust-region radius ensures that

$$\hat{\delta}_i \geq \max\{\hat{\delta}_-, \delta_{min}\} \geq \delta_{min}. \tag{4.15}$$

Thus, let us assume that the set of rejected steps between $\hat{s}_-$ and $\hat{s}_i$ is nonempty. Now, either $\|C_1(y_0^i)\| > \omega_1 \hat{\delta}_{i_l}$ for all $l = 1, \dots, L+1$ or only for some.

Suppose, first, that $\|C_1(y_0^i)\| > \omega_1 \hat{\delta}_{i_l}$ for all $l = 1, \dots, L+1$. By Lemma (4.3),

$$|ared_{i_l} - pred_{i_l}| \leq \Gamma_9 (\prod_{j=1}^{M} \rho_j^{i_l}) \|\hat{s}_{i_l}\|^2.$$

By Lemmas (3.5) and (3.3), we have

$$\begin{aligned} pred_{i_l} &\geq \frac{1}{2^M} (\prod_{j=1}^{M} \rho_j^{i_l}) [\|C_1(y_0^i)\|^2 - \|C_1(y_0^i) + \nabla C_1(y_0^i)^T s_1^{i_l}\|^2 \\ &\geq \frac{1}{2^M} (\prod_{j=1}^{M} \rho_j^{i_l}) \Gamma_4 \|C_1(y_0^i)\| \min\{\Gamma_5 \|C_1(y_0^i)\|, \delta_1^{i_l}\}. \end{aligned}$$

Because $\|C_1(y_0^i)\| > \omega_1 \hat{\delta}_{i_l}$ and $\hat{\delta}_{i_l} \leq (M+1)\, \delta_1^{i_l}$,

$$pred_{i_l} \geq \frac{1}{2^M} (\prod_{j=1}^{M} \rho_j^{i_l}) \Gamma_4 \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\} \hat{\delta}_{i_l}.$$

Therefore, because $\|\hat{s}_{i_l}\| \leq \hat{\delta}_{i_l}$,

$$\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} \leq \frac{2^M \Gamma_9 \|\hat{s}_{i_l}\|}{\Gamma_4 \|C_1(y_0^i)\| \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}}.$$

Because all the steps between $\hat{s}_-$ and $\hat{s}_i$ are rejected, $\frac{ared_{i_l}}{pred_{i_l}} < \eta_1$ where $\eta_1 \in (0,1)$ is defined in Algorithm 3.2 or $1 - \frac{ared_{i_l}}{pred_{i_l}} > 1 - \eta_1$. Hence

$$\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} > 1 - \eta_1$$

and

$$\frac{2^M \Gamma_9 \|\hat{s}_{i_l}\|}{\Gamma_4 \|C_1(y_0^i)\| \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}} > 1 - \eta_1,$$

which yields

$$\|\hat{s}_{i_l}\| \geq \frac{(1-\eta_1)\Gamma_4 \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}}{2^M \Gamma_9} \|C_1(y_0^i)\|, \quad l = 1, \dots, L.$$

We have $\hat{\delta}_i \equiv \hat{\delta}_{i_{L+1}} \geq \alpha_1 \|\hat{s}_{i_L}\|$, where α_1 is the trust-region radius reduction factor in the case of a trial step rejection. Because $\|C_1(y_0^i)\| > \omega_1 \hat{\delta}_{i_1}$, it follows that

$$\hat{\delta}_i \geq \alpha_1 \|\hat{s}_{i_L}\| \geq \alpha_1 \left[\frac{(1-\eta_1)\Gamma_4 \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}}{2^M \Gamma_9} \right] \omega_1 \hat{\delta}_{i_1}. \tag{4.16}$$

But because $\hat{\delta}_{i_1}$ is the first step after a successful step,

$$\hat{\delta}_{i_1} = \max\{\hat{\delta}_-, \delta_{min}\} \geq \delta_{min}.$$

Therefore

$$\hat{\delta}_i \geq \frac{\omega_1 \delta_{min} \alpha_1 (1-\eta_1) \Gamma_4 \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}}{2^M \Gamma_9} \equiv \Gamma_{10}.$$

Now suppose that $||C_1(y_0^i)|| > \omega_1 \hat{\delta}_{i_l}$ holds only for some of the rejected total steps $l = 1, \ldots, L+1$.

Let J be the largest index such that $||C_1(y_0^i)|| \leq \omega_1 \hat{\delta}_{i_l}$. Because after each step is rejected, the trust-region radius is decreased the following situation holds:

$$\hat{s}_-, \underbrace{\overbrace{\hat{s}_{i_1}, \hat{s}_{i_2}, \ldots, \hat{s}_{i_J}}^{||C_1(y_0^i)|| \leq \omega_1 \hat{\delta}_{i_l}}, \overbrace{\hat{s}_{i_{J+1}}, \ldots, \hat{s}_{i_L}}^{||C_1(y_0^i)|| > \omega_1 \hat{\delta}_{i_l}}}_{\text{rejected}}, \hat{s}_{i_{L+1}} = \hat{s}_i;$$

i.e., once $C_1(y_0^i) > \omega_1 \hat{\delta}_{i_J}$, then $||C_1(y_0^i)|| > \omega_1 \hat{\delta}_{i_l}$ holds for all $l = J+1, \ldots, L+1$.

The case $\hat{s}_{i_{J+1}} = \hat{s}_i$ is included in the previous case. Therefore we see that

$$\hat{\delta}_i \geq \alpha_1 ||\hat{s}_{i_J}||.$$

If $\hat{s}_{i_{J+1}} \neq \hat{s}_i$, then the hypothesis for inequality (4.16) holds, and using inequality (4.16), we have

$$\hat{\delta}_i \geq \frac{(1-\eta_1)\omega_1 \alpha_1 \Gamma_4 \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}}{2^M \Gamma_9} ||\hat{s}_{i_{J+1}}||$$

because $||C_1(y_0)|| > \omega_1 \hat{\delta}_{i_l}$ for all $l = J+1, \ldots, L+1$. Letting

$$\Gamma_{11} \equiv \min\{\frac{(1-\eta_1)\omega_1 \alpha_1 \Gamma_4 \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}}{2^M \Gamma_9}, \alpha_1\},$$

we have

$$\hat{\delta}_i \geq \Gamma_{11} ||\hat{s}_{i_J}||.$$

Now, keeping the same notation for the actual and predicted reductions as in the previous case, we have by Lemma (4.3)

$$|ared_{i_l} - pred_{i_l}| \leq \Gamma_7 ||\hat{s}_{i_l}||^2 + \Gamma_8 (\prod_{j=1}^{M} \rho_j) ||\hat{s}_{i_l}||^3 + \sum_{k=1}^{M} (\prod_{j=k}^{M} \rho_j) \nu_k ||C_k(y_{k-1})|| ||\hat{s}_{i_l}||^2.$$

By Lemmas (3.5) and (4.5),

$$pred_{i_l} \geq \frac{\rho_2 \ldots \rho_M}{2^{M-1}} Cpred_2^{i_l} \geq \frac{\rho_2 \ldots \rho_M}{2^{M-1}} \tau_1 \hat{\delta}_{i_l},$$

And because $\rho_j \geq 1$ for all j, we have

$$\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} \leq$$

$$\frac{2^{M-1}[\Gamma_7 ||\hat{s}_{i_l}||^2 + \Gamma_8 \rho_1 ||\hat{s}_{i_l}||^3 + \sum_{k=2}^{M} \nu_k ||\hat{s}_{i_l}||^2 ||C_k(y_{k-1})|| + \rho_1 \nu_1 ||\hat{s}_{i_l}||^2 ||C_1(y_0)||]}{\tau_1 \hat{\delta}_{i_l}}$$

$$\leq \frac{2^{M-1}\{\Gamma_7 + \sigma_5 (M-1)\nu_k + \Gamma_8 \rho_1 ||\hat{s}_{i_l}|| + \rho_1 \nu_1 ||C_1(y_0)||\} ||\hat{s}_{i_l}||^2}{\tau_1 \hat{\delta}_{i_l}}.$$

But $\|C_1(y_0)\| \leq \omega_1 \hat{\delta}_{i_l}$ and $\|\hat{s}_{i_l}\| \leq \hat{\delta}_{i_l}$. Hence

$$\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} \leq \frac{2^{M-1}[\Gamma_7 + \sigma_5(M-1)\nu_k + \Gamma_8\rho_1\hat{\delta}_{i_l} + \rho_1\nu_1\omega_1\hat{\delta}_{i_l}]\|\hat{s}_{i_l}\|^2}{\tau_1\hat{\delta}_{i_l}}.$$

By Lemma (4.6), $\rho_1\hat{\delta}_{i_l} \leq \Theta_1$. Therefore,

$$\begin{aligned}\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} &\leq \frac{2^{M-1}[\Gamma_7 + \sigma_5(M-1)\nu_k + \Gamma_8\Theta_1 + \Theta_1\nu_1\omega_1]\|\hat{s}_{i_l}\|^2}{\tau_1\hat{\delta}_{i_l}} \\ &\leq \underbrace{\frac{2^{M-1}[\Gamma_7 + \sigma_5(M-1)\nu_k + \Gamma_8\Theta_1 + \Theta_1\nu_1\omega_1]}{\tau_1}}_{\Gamma_{12}}\|\hat{s}_{i_l}\|.\end{aligned}$$

Because $\hat{s}_{i_l}$ is rejected,

$$(1-\eta_1) \leq \frac{2^{M-1}\Gamma_{12}}{\tau_1}\|\hat{s}_{i_l}\|.$$

Therefore,

$$\|\hat{s}_{i_l}\| \geq \frac{(1-\eta_1)\tau_1}{2^{M-1}\Gamma_{12}} \tag{4.17}$$

and so

$$\hat{\delta}_i \geq \frac{\Gamma_{11}(1-\eta_1)\tau_1}{2^{M-1}\Gamma_{12}}.$$

Defining

$$\bar{\delta}_1 = \min\left\{\delta_{min}, \Gamma_{10}, \frac{\Gamma_{11}(1-\eta_1)\tau_1}{2^{M-1}\Gamma_{12}}\right\},$$

we have

$$\hat{\delta}_i \geq \bar{\delta}_1,$$

which proves the first part of the lemma for $k = 1$.

For the second part of the lemma for $k = 1$, considering that $\{\rho_1^i\}$ forms a nondecreasing sequence, it remains to be shown that it is bounded above. If ρ_1^i is increased, then by Lemma (4.6) and the first part of the present lemma, we have

$$\rho_1^i \leq \frac{\Theta_1}{\hat{\delta}_i} \leq \frac{\Theta_1}{\bar{\delta}_1}.$$

Hence $\{\rho_1^i\}$ is a bounded, nondecreasing sequence, and therefore

$$\lim_{i\to\infty} \rho_1^i = \rho_1^* < \infty.$$

By Lemma (3.5), if ρ_1^i is increased, it is increased by at least $\beta > 0$. Therefore, because $\{\rho_1^i\}$ converges to a finite number, the number of increases has to be finite; i.e., $\rho_1^i = \rho_1^{i_{\rho_1}}$ for some index i_{ρ_1} and all $i \geq i_{\rho_1}$. Hence both results of the lemma are established for ρ_1^i.

To prove the result for the case $k = 2$, we assume that ρ_2^i is increased and show that $\hat{\delta}_i$ is bounded from below.

The argument for ρ_2^i is nearly identical to that for ρ_1^i. We assume that there are some rejected steps between between $\hat{s}_-$ and $\hat{s}_i$, for otherwise δ_{min} provides the lower bound on the trust-region radius. Again, two cases are considered. If $||C_2(y_1^i)|| > (\omega_2 - \omega_1)\hat{\delta}_{i_l}$ for all $l = 1, \ldots, L+1$, we use

$$pred_{i_l} \geq \frac{\rho_2 \ldots \rho_M}{2^{M-1}} Cpred_2$$

to arrive at the estimate

$$\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} \leq \frac{2^M \Gamma_9 \rho_1 ||\hat{s}_{i_l}||}{\Gamma_4 ||C_2(y_1^c)|| \min\{\Gamma_5(\omega_2 - \omega_1), \frac{1}{M+1}\}}.$$

However, we proved that $\rho_1 \leq \rho_1^*$, so

$$\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} \leq \frac{2^M \Gamma_9 \rho_1^*}{\Gamma_4 ||C_2(y_1^i)|| \min\{\Gamma_5(\omega_2 - \omega_1), \frac{1}{M+1}\}} ||\hat{s}_{i_l}||,$$

and the rest of the argument is identical to that of the case $k = 1$.

If $||C_2(y_1^i)|| > (\omega_2 - \omega_1)\hat{\delta}_{i_l}$ only for some of $l = 1, \ldots, L+1$, $\rho_1 \leq \rho_1^*$ is used again to remove the dependence of the estimates on the penalty parameters, and the argument proceeds identically to the case $k = 1$ to yield

$$\hat{\delta}_i \geq \bar{\delta}_2.$$

Again, if ρ_2 is increased, we have

$$\rho_2^i \leq \frac{\Theta_2}{\hat{\delta}_i} \leq \frac{\Theta_2}{\bar{\delta}_2}$$

and we obtain that $\rho_2^i \to \rho_2^*$ and the existence of an index i_{ρ_2} such that

$$\rho_2^i = \rho_2^{i_{\rho_2}} \quad \text{for all } i \geq i_{\rho_2}.$$

Continuing this procedure, for ρ_k^i we have the estimate

$$\frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} \leq \frac{2^k \Gamma_9 \rho_1 \ldots \rho_{k-1}}{\Gamma_4 ||C_k(y_{k-1})|| \min\{\Gamma_5(\omega_k - \omega_{k-1}), \frac{1}{M+1}\}} ||\hat{s}_{i_l}||.$$

But at this point, $\rho_1, \ldots, \rho_{k-1}$ have been shown to be bounded by $\rho_1^*, \ldots, \rho_{k-1}^*$, respectively, thus eliminating the dependence of the estimates on the penalty parameters. The rest of the argument proceeds identically as in the case $k = 1$ to show that $\hat{\delta}_i$ is bounded, and this result is then used to show that the sequence $\{\rho_k^i\}$ is bounded from above by ρ_k^*. Setting $\bar{\delta} = \min\{\bar{\delta}_1, \ldots, \bar{\delta}_M\}$ completes the proof for the first and second parts of the lemma.

Finally, the boundedness of the merit function $\bar{\mathcal{P}}$ follows directly from the problem assumptions and the boundedness of the penalty parameters. $\square$

4.4. Boundedness of the trust-region radius. We have shown that the total trust-region radius $\hat{\delta}_i$ is bounded away from zero if any of the penalty parameters are increased. We shall now show that $\hat{\delta}_i$ is always bounded away from zero. The trust-region updating strategy ensures that $\hat{\delta}_i$ is bounded from above.

THEOREM 4.8. *If the algorithm does not terminate at an iteration i, there exists a constant $\hat{\delta_*} > 0$, independent of the iterates, such that*

$$\hat{\delta}_i \geq \hat{\delta_*} \quad \text{for all } i. \tag{4.18}$$

Proof. The proof follows along the lines of Lemma (4.7), but we do not have to consider separate cases based on which penalty parameter is increased.

If there are no rejected steps between $\hat{s}_-$ and $\hat{s}_i$, or if $||C_1(y_0^i)|| > \omega_1 \hat{\delta}_{i_l}$ for all rejected steps $l = 1, \ldots, L+1$, the reasoning is identical to that of Lemma (4.7). If $||C_1(y_0^i)|| > \omega_1 \hat{\delta}_{i_l}$ only for some of $l = 1, \ldots, L+1$, letting J be the largest index such that $||C_1(y_0^i)|| \leq \omega_1 \hat{\delta}_{i_l}$ for all $i = i_1, \ldots, i_J$, we have

$$\hat{\delta}_i \geq \Gamma_{11} ||\hat{s}_{i_J}||. \tag{4.19}$$

By Lemma (4.3), $|ared_{i_l} - pred_{i_l}| \leq \Gamma_9 (\prod_{j=1}^{M} \rho_j) ||\hat{s}_{i_l}||^2$. By Lemma (4.5), $Cpred_2 \geq \tau_1 \hat{\delta}_{i_l}$, and thus by Lemma (3.5),

$$pred_{i_l} \geq (\frac{\prod_{j=2}^{M}}{2^{M-1}} \rho_j^{i_l}) Cpred_2 \geq (\frac{\prod_{j=2}^{M}}{2^{M-1}} \rho_j)^{i_l} \tau_1 \hat{\delta}_{i_l}.$$

Therefore, because $\hat{s}_{i_l}$ is rejected,

$$(1-\eta_1) \leq \frac{|ared_{i_l} - pred_{i_l}|}{pred_{i_l}} \leq \frac{2^{M-1} \Gamma_9 \rho_1^{i_l} ||\hat{s}_{i_l}||^2}{\tau_1 \hat{\delta}_{i_l}} \leq \frac{2^{M-1} \Gamma_9 \rho_1^* ||\hat{s}_{i_l}||}{\tau_1}.$$

Hence

$$||\hat{s}_{i_l}|| \geq \frac{(1-\eta_1)\tau_1}{\Gamma_9 \rho_1^*}.$$

Then by inequality (4.19), we have

$$\hat{\delta}_i \geq \frac{(1-\eta_1)\tau_1 \Gamma_{11}}{\Gamma_9 \rho_1^*}.$$

Letting $\hat{\delta_*} = \min\left\{\frac{(1-\eta_1)\tau_1\Gamma_{11}}{\Gamma_9 \rho_1^*}, \delta_{min}\right\}$, we have $\hat{\delta}_i \geq \hat{\delta_*}$, which concludes the proof. ☐

4.5. The algorithm is well-defined. The following theorem guarantees that the algorithm is well-defined, i.e., that after a finite number of iterations an acceptable total trial step $\hat{s}_i$ with $\frac{ared_i}{pred_i} \geq \eta_1$ will be found.

THEOREM 4.9. *Unless the current iterate x_i satisfies the termination criterion of the algorithm, an acceptable step $\hat{s}_i$ from x_i will be found after a finite number of trials.*

Proof. It suffices to consider two cases.

If $||C_1(y_0)|| > \omega_1 \hat{\delta}_i$, we have by Lemmas (3.5) and (3.3)

$$pred_i \geq \frac{1}{2^M}(\prod_{j=1}^{M} \rho_j) Cpred_1 \geq \frac{1}{2^M}(\prod_{j=1}^{M} \rho_j)\Gamma_4 ||C_1(y_0)|| \min\{\Gamma_5 ||C_1(y_0)||, \frac{\hat{\delta}_i}{M+1}\}$$
$$\geq \frac{1}{2^M}(\prod_{j=1}^{M} \rho_j)\Gamma_4 ||C_1(y_0)|| \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}\hat{\delta}_i.$$

By Lemma (4.3), the last inequality gives us

$$\frac{|ared_i - pred_i|}{pred_i} \leq \frac{2^M \Gamma_9 ||\hat{s}_i||^2}{\Gamma_4 ||C_1(y_0)|| \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}\hat{\delta}_i} \leq \frac{2^M \Gamma_9}{\Gamma_4 ||C_1(y_0)|| \min\{\Gamma_5 \omega_1, \frac{1}{M+1}\}}\hat{\delta}_i.$$

The last line indicates that $\left|\frac{ared_i}{pred_i} - 1\right|$ approaches 0 as $\hat{\delta}_i$ becomes smaller. Therefore the criterion

$$\frac{ared_i}{pred_i} \geq \eta_1 \in (0,1)$$

will be satisfied after a finite number of iterations.

If $||C_1(y_0)|| \leq \omega_1 \hat{\delta}_i$, by Lemmas (3.5) and (4.5)

$$pred_i \geq \frac{1}{2^{M-1}}(\prod_{k=2}^{M} \rho_k) Cpred_2 \geq \frac{1}{2^{M-1}}(\prod_{k=2}^{M} \rho_k)\tau_1 \hat{\delta}_i.$$

Hence

$$\frac{|ared_i - pred_i|}{pred_i} \leq \frac{2^{M-1}\Gamma_9 \rho_1 \hat{\delta}_i^2}{\tau_1 \hat{\delta}_i} \leq \frac{2^{M-1}\Gamma_9 \rho_1^*}{\tau_1}\hat{\delta}_i.$$

Again the ratio goes to 0 with decreasing $\hat{\delta}_i$, so the acceptance criterion will be satisfied after finitely many trials. $\square$

4.6. Global convergence. We now show that the sequence of iterates generated by Algorithm 3.3 has a subsequence convergent to a stationary point of the equality constrained minimization problem.

THEOREM 4.10. *If* $||P_M^T \nabla f(y_M^i)|| + \sum_{j=1}^{M} ||C_j(y_{j-1}^i)|| > \epsilon_{tol}$ *for all* i *and some* $\epsilon_{tol} > 0$, *then*

$$\lim_{i \to \infty} \sum_{j=1}^{M} ||C_j(y_{j-1}^i)|| = 0. \tag{4.20}$$

Proof. The proof by contradiction is analogous to that in [13].

Suppose the result does not hold, that is, that there exists an infinite sequence of indices $\{i_l\}$ such that $\sum_{j=1}^{M} ||C_j(y_{j-1}^i)|| > \theta$ for some $\theta > 0$ and for all $i \in \{i_l\}$. In particular, without loss of generality, it suffices to make this assumption about $||C_1(y_0^i)||$. Analogous reasoning will apply to the rest of the constraint blocks.

By Lemmas (4.7), (3.5), and (3.3), for each $i_l \geq \max\{i_{\rho_1}, \ldots, i_{\rho_M}\}$,

$$
\begin{aligned}
pred_{i_l} &\geq \frac{\prod_{j=1}^{M} \rho_j^{i_l}}{2^M} Cpred_1^{i_l} \\
&= \frac{\prod_{j=1}^{M} \rho_j^{i_l}}{2^M} [||C_1(y_0^{i_l})||^2 - ||C_1(y_0^{i_l}) + \nabla C_1^T(y_0^{i_l}) s_1^{i_l}||^2] \\
&\geq \frac{\prod_{j=1}^{M} \rho_j^{*}}{2^M} \Gamma_4 ||C_1(y_0^{i_l})|| \min\{\Gamma_5 ||C_1(y_0^{i_l})||, \frac{\hat{\delta}_i}{M+1}\} \\
&\geq \frac{\prod_{j=1}^{M} \rho_j^{*}}{2^M} \Gamma_4 \theta \min\{\Gamma_5 \theta, \frac{\hat{\delta}_*}{M+1}\} \equiv \Gamma_{13} > 0.
\end{aligned}
$$

Because due to Theorem (4.9) we need to consider only accepted steps, we have

$$
\bar{\mathcal{P}}_{i_l} - \bar{\mathcal{P}}_{i_l+1} = ared_{i_l} \geq \eta_1 pred_{i_l} \geq \eta_1 \Gamma_{13} > 0. \tag{4.21}
$$

This inequality leads to a contradiction if $i_l \to \infty$, because $\bar{\mathcal{P}}$ is bounded below. □

THEOREM 4.11. *Given any criterion $\epsilon_{tol} > 0$, the algorithm will terminate because*

$$
||P_M^T \nabla f(y_M)|| + \sum_{j=1}^{M} ||C_j(y_{j-1})|| < \epsilon_{tol}.
$$

Proof. If the algorithm does not terminate and we assume that a subsequence of $\{||P_M^T \nabla f(y_M)||\}$ converges to zero, then Theorem (4.10) produces a contradiction. Therefore, let us assume that $||P_M^T \nabla f(y_M)|| > \nu$ for some positive constant ν. By Theorem (4.10), $\sum_{j=1}^{M} ||C_j(y_{j-1})|| \to 0$ and the sequence $\{\hat{\delta}_i\}$ is bounded below by $\hat{\delta}_*$. Hence there exists an index $I > i_{\rho_M}$ such that for all $i > I$,

$$
\sum_{j=1}^{M} ||C_j(y_{j-1})|| \leq \omega_M \hat{\delta}_* \leq \omega_M \hat{\delta}_i.
$$

Thus, Lemma (4.5) and Theorem (4.8) yield an infinite sequence of steps with the actual decrease in $\bar{\mathcal{P}}$ of at least $\eta_1 \tau_m \hat{\delta}_*$, which contradicts the boundedness of $\bar{\mathcal{P}}$. □

By assumption, the norms of the substeps are bounded from above by constant multiples of the norms of the constraint blocks. Therefore, as the norm of each block goes to zero, so does the norm of the substep, and the asymptotic convergence is at a single point.

5. Global convergence for Variant II. Global convergence analysis for Variant II is a simplified version of that for Variant I, because the merit function contains only one penalty parameter. In particular, there is no need to make a distinction between feasibility and the relative feasibility among the constraint blocks as it is done in Lemma (4.4). One no longer needs to consider the branches on the cumulative behavior of the constraint blocks. In fact, the global convergence analysis of Dennis et al. [13] is almost directly applicable to Variant II. Because the norms of the substeps are bounded by the norms of the constraint blocks, again, asymptotic convergence occurs at a single point.

6. Concluding remarks. We have presented a class of multilevel algorithms for solving the equality constrained optimization problem together with a global convergence analysis. The algorithms are applicable to solving nonlinear equations as well.

The main practical appeal of the multilevel algorithms is that they allow the user to partition the constraint system according to the needs of the application and to process the blocks of constraints autonomously. In addition, the trial steps are only required to obey mild conditions, satisfied by many methods for solving the quadratic subproblems. In fact, the substeps comprised by the trial step can be obtained by different methods within a single solution procedure, as long as the substeps satisfy the required conditions of boundedness and sufficient decrease.

The proposed algorithms are expected to be of use in applications that exhibit natural block structure. The design of complex engineering systems is characterized by such a structure. While the multilevel class can take advantage of natural or induced separability to process the subproblems concurrently, it provides a way to solve fully or densely coupled problems sequentially, but with a degree of autonomy.

Given the problem assumptions, a well-defined partition of the constraint system exists, although pivoting may be required to obtain it. If the application does not suggest a natural partition, an advantageous partitioning may be obtained following the simple rules applicable to the local Brown-Brent methods for nonlinear equations. Because the last subproblems of the sequence will have the smallest dimension, the most computationally intensive block should be placed last, while the linear, or most linear, block should be placed first. On the other hand, because the first block of constraints will be satisfied more quickly and directly, the most important block of information should be placed first.

Current development in multilevel methods is proceeding along several directions, including extensions to general multilevel optimization with arbitrary objectives and applications to multicriteria optimization. In addition, in the preliminary implementation, the inequality constraints have been handled with the help of squared slack variables. However, other promising directions are under investigation.

To date, the class has been tested on small nonlinear equations and equality constrained problems [27] with promising results. The codes are now being tested on increasingly realistic problems of interest to engineering design applications. In particular, the approximation model management capability has been incorporated into MAESTRO [3] and is undergoing computational testing. Among the implementation issues considered are various strategies for evaluating the step and for updating the individual trust-region radii.

7. Acknowledgments. The authors wish to thank R.M. Lewis of ICASE for pointing out the expense of using orthogonal steps and for encouraging the use of reduced-basis steps.

REFERENCES

[1] N. Alexandrov, *Multilevel algorithms for nonlinear equations and equality constrained optimization*, PhD thesis, Rice University, Department of Computational and Applied Mathematics, May 1993. Also available as Technical Report TR93–20.

[2] ———, *On a class of trust-region multilevel algorithms for nonlinear equations: a globalization of brown-brent methods*, SIAM Journal on Optimization, (1998). submitted.

[3] ———, *On managing the use of surrogates in general nonlinear optimization*, in Proceedings of the Seventh AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA, 1998. AIAA paper AIAA-98-4798.

[4] N. Alexandrov and J. E. Dennis, Jr., *Multilevel algorithms for nonlinear optimization*, in Optimal Design and Control, J. Borggaard, J. Burkardt, M. Gunzburger, and J. Peterson, eds., Birkhäuser, 1994, pp. 1–22.

[5] N. Alexandrov and M. Y. Hussaini, eds., *Multidisciplinary Design Optimization: State of the Art*, Philadelphia, 1997, Society for Industrial and Applied Mathematics.

[6] R. P. Brent, *Some efficient algorithms for solving systems of nonlinear equations*, SIAM Journal on Numerical Analysis, 10 (1973), pp. 327–344.

[7] ———, *Some efficient algorithms for solving systems of nonlinear equations*, SIAM J. Numerical Analysis, 10 (1973), pp. 327—344.

[8] K. M. Brown, *A Quadratically Convergent Method for Solving Simultaneous Non-linear Equations*, PhD thesis, Purdue University, 1966.

[9] ———, *A Quadratically Convergent Method for Solving Simultaneous Nonlinear Equations*, PhD thesis, Purdue University, 1966.

[10] ———, *Algorithm 316. Solution of simultaneous nonlinear equations*, Communications of the Association for Computing Machinery (ACM), 10 (1967), pp. 728–729.

[11] ———, *A quadratically convergent Newton-like method based on Gaussian elimination*, SIAM J. Numerical Analysis, 6 (1969), pp. 560—569.

[12] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, *A trust region algorithm for nonlinearly constrained optimization*, SIAM Journal on Numerical Analysis, 24 (1987), pp. 1152–1170.

[13] J. E. Dennis, Jr., M. El-Alem, and M. C. Maciel, *A global convergence theory for general trust-region-based algorithms for equality constrained optimization*, Technical Report TR92-28, Rice University, Department of Computational and Applied Mathematics, Houston, Texas 77251, September 1992.

[14] M. El-Alem, *A Global Convergence Theory for a Class of Trust Region Algorithms for Constrained Optimization*, PhD thesis, Rice University, Department of Mathematical Sciences, 1988.

[15] M. M. El-Alem, *A Global Convergence Theory for a Class of Trust Region Algorithms for Constrained Optimization*, PhD thesis, Rice University, Department of Mathematical Sciences, 1988.

[16] ———, *A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization*, SIAM J. Numerical Analysis, 28 (1991), pp. 266—290.

[17] ———, *Robust trust-region algorithm with non-monotonic penalty parameter scheme for constrained optimization*, Technical Report TR92-30, Rice University, Department of Computational and Applied Mathematics, Houston, Texas 77251, September 1992.

[18] M. El Hallabi and R. A. Tapia, *A global convergence theory for arbitrary norm trust region methods for nonlinear equations*, Tech. Report TR 87–25, Rice University, Department of Computational and Applied Mathematics, 1987.

[19] D. M. Gay, *Brown's Method and Some Generalizations, with Applications to Minimization Problems*, PhD thesis, Cornell University, 1975.

[20] J. E. Dennis, Jr., *Non-linear least squares and equations*, in The State-of-the-Art in Numerical Analysis, 1976.

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

[22] R. M. Lewis. Oral communication.

[23] M. J. D. Powell, *Convergence properties of a class of minimization algorithms*, in Nonlinear Programming 2, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, 1975.

[24] J. M. Martínez, *Generalization of the methods of Brent and Brown for solving nonlinear simultaneous equations*, SIAM J. Numerical Analysis, 1 (1979), pp. 434—448.

[25] 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, 1983.

[26] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Stat. Comput., 4 (1983), pp. 553—572.

[27] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, *Testing unconstrained optimization software*, ACM Transactions on Mathematical Software, 7 (1981), pp. 17—41.

[28] M. J. D. Powell and Y.-X. Yuan, *A trust region algorithm for equality constrained optimization*, Mathematical Programming, 49 (1991), pp. 189–211.

[29] G. A. Shultz, R. B. Schnabel, and R. H. Byrd, *A family of trust-region-based algorithms for*

unconstrained minimization with strong global convergence properties, SIAM J. Numerical Analysis, 22 (1985), pp. 47—67.

[30] J. Sobieszczanski-Sobieski and R. T. Haftka, *Multidisciplinary aerospace design optimization: survey of recent developments*, Structural Optimization, 14 (1997), pp. 1–23.

[31] D. Sorensen, *Minimization of a large-scale quadratic function subject to a spherical constraint*, SIAM Journal on Optimization, 7 (1997), pp. 141–161.

[32] D. C. Sorensen, *Newton's method with a model trust region modification*, SIAM J. Numerical Analysis, 19 (1982), pp. 409—426.

[33] ———, *Minimization of a large scale quadratic function subject to an ellipsoidal constraint*, Tech. Report TR94-27, Department of Computational and Applied Mathematics, Rice University, August 1994.

[34] T. Steihaug, *The conjugate gradient method and trust regions in large scale optimization*, SIAM Journal on Numerical Analysis, 20 (1983), pp. 626–637.

[35] P. L. Toint, *Towards an efficient sparsity exploiting newton method for minimization*, Technical Report 80/4, Departément de Mathématique, Facultés Universitaires de Namur, Belgium, 1980.