On the Global Convergence of a Trust Region Method for Solving Nonlinear Constraints Infeasibility Problem

Choong Ming Chin*

**Department of Electrical Engineering, University of Malaya, Pantai Valley, Kuala Lumpur 50603, Malaysia.*
E-mail: `clp@tm.net.my`

Numerical Optimization Report, November 2003, revised January 2004

Abstract

A framework for proving global convergence for a class of nonlinear constraints infeasibility problem is presented without assuming that the Jacobian has full rank everywhere. The underlying method is based on the simple sufficient reduction criteria where trial points are accepted provided there is a sufficient decrease in the constraints violation function. The proposed methods solve a sequence of quadratic programming subproblems for which effective software is readily available, and instead of using line search strategies that could converge to singular non-stationary points, the methods utilize trust region techniques to induce global convergence.

The proof technique is presented in a fairly general context, allowing a range of specific algorithm choices associated with choosing the Hessian matrix representation and controlling the trust region radius.

Keywords nonlinear programming, global convergence, trust region, SQP, feasibility restoration phase.

1 Introduction

This paper concerns with the development of a trust region method algorithm for finding a feasible solution of a general system of nonlinear inequalities of the form

$$c_i(\mathbf{x}) \leq 0, \; i = 1, 2, \ldots, m. \tag{1.1}$$

The functions $c_i(\mathbf{x})$ are called constraints functions, where we assume $\mathbf{c} : \mathbb{R}^n \mapsto \mathbb{R}^m$ are twice differentiable. The main question we have to ask is given a set of problems of the form (1.1), does there exist any $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that $c_i(\hat{\mathbf{x}}) \leq 0, \; i = 1, 2, \ldots, m$? If the answer to (1.1) is yes, then the set of constraints $\{c_i(\mathbf{x})\}$ is said to be *feasible*; otherwise it is said to be *infeasible*. By expressing constraints in the above formulation we also allow nonlinear constraints $\mathbf{c}(\mathbf{x}) = \mathbf{0}$ to be included where we do so by expressing them as two separate inequality constraints $\mathbf{c}(\mathbf{x}) \leq \mathbf{0}$ and $-\mathbf{c}(\mathbf{x}) \leq \mathbf{0}$.

The set of inequalities (1.1) is a special case of nonlinear programming problem where it frequently occurs in practice especially in filter type problems of Chin and Fletcher [6], and Fletcher and Leyffer [8], when the linearized constraints of a subproblem are inconsistent with the trust region constraint, and hence the main algorithm has to temporarily exits to a feasibility restoration phase algorithm to generate iterates that approach the feasible region and also acceptable to the filter. The same scenario also occurs in line search based filter methods for nonlinear programming when the step size has reached a minimum threshold such that a trial iterate could not be accepted to the filter. Further details of such problems are given in Benson, Shanno and Vanderbei [2], Biegler and Wächer [3], Chin [5] and Ulbrich [18].

At present there exist many methods to solve problems of the form (1.1) and in the paper by Schnabel [17], the strategy is to minimize an exponential-type penalty function $w : \mathbb{R} \to \mathbb{R}$ of the form

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad \phi(\mathbf{x}, p) = \frac{1}{p}\sum_{i=1}^{m} w(p \cdot c_i(\mathbf{x}))$$

where p is a penalty parameter with monotonically increasing values. An example of the penalty function used is $w(z) = \mathrm{e}^z - 1$, $z \in \mathbb{R}$ which was first introduced by Hartman [12] and Kort and Bertsekas [13] for general nonlinear programming problem. The objective of Schnabel's strategy is to form a weighted sum of constraint values that penalizes constraint violation and rewards constraint satisfaction by increasing the penalty parameter p until either

(i) $\mathbf{x}^*(p)$ is feasible, or

(ii) $\phi^*(p) > 0$, that is infeasibility is established

where $\mathbf{x}^*(p) = \text{argmin } \phi(\mathbf{x}, p)$ and $\phi^*(p)$ is the local minimum value of $\phi(\mathbf{x}, p)$ evaluated at $\mathbf{x}^*(p)$. However the method suffers a drawback apart from $p \to +\infty$ that is the method could converge to a point $\mathbf{x}^*(p)$ such that $\phi^*(p) \leq 0$ and $\mathbf{x}^*(p)$ is infeasible, which can occur whether $\{c_i(\mathbf{x})\}$ is feasible or not.

Instead of using a penalty-type function to address (1.1) an alternative method is to solve a norm minimization problem of the form

$$H \left\{ \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad h(\mathbf{c}(\mathbf{x})) \right.$$

where $h(\mathbf{c}(\mathbf{x})) = \sum_{i=1}^{m}\{0, c_i(\mathbf{x})\}$. An equivalent smooth form of Problem H can be expressed as

$$P \left\{ \begin{array}{lll} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} & \displaystyle\sum_{i\in\mathcal{V}(\mathbf{x})} c_i(\mathbf{x}) & \\ \text{subject to} & c_i(\mathbf{x}) \leq 0 & i \in \mathcal{V}^{\perp}(\mathbf{x}), \end{array} \right.$$

where $\mathcal{V}(\mathbf{x}) = \{i \ : \ c_i(\mathbf{x}) > 0\}$ and $\mathcal{V}^{\perp}(\mathbf{x}) = \{i \ : \ c_i(\mathbf{x}) \leq 0\}$. As a result it can be shown that

$$\mathbf{x}^{\infty} \text{ minimizes Problem } H \text{ locally} \quad \Leftrightarrow \quad \mathbf{x}^{\infty} \text{ is a local solution of Problem } P.$$

At present there exist various strategies of solving Problem H (or P) efficiently and one of the methods is to solve a successive linearization of the constraints and use a line search technique on the merit function $h(\mathbf{c}(\mathbf{x}))$ to promote global convergence. Although the technique is well behaved where in most cases it makes a steady progress towards the solution, there are still potential difficulties associated with this approach namely convergence of the iterates to non-stationary points can occur (see Byrd, Marazzi and Nocedal [4] and Powell [16]). On the other hand, by using a trust region strategy of solving a sequence of ℓ_1LP subproblems (Sℓ_1LP method), the problem of iterates converging to singular non-stationary points can be circumvented (see Byrd, Marazzi and Nocedal [4]). However, as these methods rely only on first-order information alone, slow convergence of iterates to a solution point can occur in the neighbourhood of a non-zero local minimum of $h(\mathbf{c}(\mathbf{x}))$.

To promote fast local convergence of iterates, another technique to consider is to use Sℓ_1QP method (see Fletcher and Sainz de la Maza [9]), where the strategy is to use second-order information and solve the following subproblem at every iteration

$$\begin{array}{ll} \underset{\mathbf{d}\in\mathbb{R}^n}{\text{minimize}} & h(\mathbf{c}(\mathbf{x}_k) + \nabla\mathbf{c}(\mathbf{x}_k)^T\mathbf{d}) + \frac{1}{2}\mathbf{d}^T\mathbf{B}_k\mathbf{d} \\ \text{subject to} & \|\mathbf{d}\|_{\infty} \leq \rho \end{array}$$

where $\mathbf{x}_k$ denotes the k-th iterate, $\mathbf{d}$ is a displacement vector, $\mathbf{B}_k$ is an approximation of the Hessian of the Lagrangian and ρ is a trust region radius. Although both Sℓ_1LP and Sℓ_1QP methods are globally convergent, the most disadvantage aspects of these methods is the complicated form of the non-smooth subproblems where they need special purpose software to solve them. Even by converting the above non-smooth QP subproblem to a regular smooth QP subproblem of the form

$$\begin{array}{ll} \underset{\mathbf{d}\in\mathbb{R}^n,\mathbf{r}^+,\mathbf{r}^-\in\mathbb{R}^m}{\text{minimize}} & \sum_{i=1}^m r_i^+ + \frac{1}{2}\mathbf{d}^T\mathbf{B}_k\mathbf{d} \\ \text{subject to} & \mathbf{r}^+ - \mathbf{r}^- = \mathbf{c}(\mathbf{x}_k) + \nabla\mathbf{c}(\mathbf{x}_k)^T\mathbf{d} \\ & \mathbf{r}^+ \geq \mathbf{0},\, \mathbf{r}^- \geq \mathbf{0} \\ & \|\mathbf{d}\|_{\infty} \leq \rho \end{array}$$

where $r_i^+ = \max\{0, c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}\}$ and $r_i^- = -\min\{0, c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}\}$, $i = 1, 2, \ldots, m$, the strategy suffers with an increased $n+2m$ number of variables in addition to a complex data structure. Hence these approaches would not be a viable choice for solving large scale system of constraints infeasibility problems.

Recently Fletcher and Leyffer [7] proposed solving (1.1) as a bi-objective optimization problem, so that the idea of a "filter" (see Fletcher and Leyffer [8]) can be used in conjunction with a trust region SQP algorithm. Basically the idea is to divide the constraints into two sets, J and $J^{\perp}$ respectively, where $J^{\perp}$ denotes the complement of

$\{1, 2, \ldots, m\}\backslash J$. Constraints in J are those that are difficult to satisfy while the set $J^\perp$ represents constraints which are close to being satisfied. Hence the objective is then to minimize the sum of constraint violations in J subject to a system of inequalities given in the set $J^\perp$

$$\begin{array}{ll} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} & \sum_{i\in J} c_i(\mathbf{x}) \\ \text{subject to} & c_i(\mathbf{x}) \le 0 \quad i \in J^\perp. \end{array}$$

As the algorithm progresses, the definition of the sets J and $J^\perp$ is changed adaptively where the strategy is a two-pronged one, that is minimize the ℓ_1 norm of constraint violations for (i) the J constraints, and (ii) the $J^\perp$ constraints. To ensure the iterates are globally convergent, the filter is used as an alternative to a penalty function as a means to decide whether or not to accept a new point in the algorithm.

This paper has some ideas in common with the approaches of Sℓ_1QP method and the bi-objective optimization strategy of Fletcher and Leyffer [7]. However, the main difference of our algorithm from these two approaches is that trial points generated by a trust region SQP algorithm are accepted provided there is a sufficient decrease in the constraints violation function, $h(\mathbf{c}(\mathbf{x}))$. Thus with this approach there is no need for us to define a filter acceptability test for the trial iterates, and because our approach has some elements of Fletcher and Leyffer [7] strategy, we also do not need to solve a non-smooth optimization problem as discussed above. The algorithm we are considering is based on a trust region SQP method applied to Problem $\tilde{P}$

$$\tilde{P} \left\{ \begin{array}{lll} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} & \sum_{i\in\mathcal{J}(\mathbf{x})} c_i(\mathbf{x}) & \\ \text{subject to} & c_i(\mathbf{x}) \le 0 & i \in \mathcal{J}^\perp(\mathbf{x}) \cup \bar{\mathcal{J}}(\mathbf{x}), \end{array} \right.$$

where $\mathcal{J}(\mathbf{x}) = \mathcal{V}(\mathbf{x})$, $\mathcal{J}^\perp(\mathbf{x}) = \mathcal{V}^\perp(\mathbf{x})$ and $\bar{\mathcal{J}}(\mathbf{x}) \subseteq \mathcal{J}(\mathbf{x})$ is a set consists of violated constraints which are close to being satisfied, and the sets are also chosen adaptively as the algorithm proceeds. We will also show that under certain constraint qualifications that if the algorithm converges to a non-feasible point with respect to (1.1) then a local minimizer of $h(\mathbf{c}(\mathbf{x})) > 0$ is found, which is equivalent to a KKT point of Problem P.

Before presenting the algorithm and the convergence proof we first make a few definitions. We denote the Jacobian of the constraints by $\nabla\mathbf{c}(\mathbf{x})^T$ and the Lagrangian function by $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i\in\mathcal{J}(\mathbf{x})} c_i(\mathbf{x}) + \sum_{i\in\mathcal{J}^\perp(\mathbf{x})\cup\bar{\mathcal{J}}(\mathbf{x})} \lambda_i c_i(\mathbf{x})$ where $\boldsymbol{\lambda}$ is a vector of multiplier estimates corresponding to the nonlinear constraints in the set $\mathcal{J}^\perp(\mathbf{x}) \cup \bar{\mathcal{J}}(\mathbf{x})$. The Hessian is denoted by $\mathbf{W}$ in which $\mathbf{W}$ is some approximation of the Hessian of the Lagrangian. Subscript $_k$ refers to iteration indices and quantities relating to local solution of Problem P are superscripted with a $^\infty$.

2 Algorithmic Description

In this paper, at the current iterate $\mathbf{x}_k$ we consider using a trust region SQP algorithm to solve the following NLP problem

$$\tilde{P}(\mathbf{x}_k)\begin{cases} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} & \sum_{i\in\mathcal{J}(\mathbf{x}_k)} c_i(\mathbf{x}_k) \\ \text{subject to} & c_i(\mathbf{x}_k)\le 0 \qquad i\in\mathcal{J}^{\perp}(\mathbf{x}_k)\cup\bar{\mathcal{J}}(\mathbf{x}_k), \end{cases}$$

so that we can prove global convergence. It is also possible for the algorithm to terminate finitely, either if a feasible point of (1.1) is found, or a KKT point of Problem P is found. Otherwise we will show that for an infinite subsequence of iterates $\{\mathbf{x}_k\}$, there exists an accummulation point $\mathbf{x}^{\infty}$ such that either it is feasible with respect to (1.1) or is a KKT point for Problem P provided some constraints qualifications are satisfied.

At the current iterate $\mathbf{x}_k$, the QP subproblem in our algorithm is defined by

$$QP(\mathbf{x}_k,\mathcal{J}(\mathbf{x}_k),\rho)\begin{cases} \underset{\mathbf{d}\in\mathbb{R}^n}{\text{minimize}} & \sum_{i\in\mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T\mathbf{d}+\frac{1}{2}\mathbf{d}^T\mathbf{W}_k\mathbf{d} \\ \text{subject to} & \nabla c_i(\mathbf{x}_k)^T\mathbf{d}+c_i(\mathbf{x}_k)\le 0, \qquad i\in\mathcal{J}^{\perp}(\mathbf{x}_k)\cup\bar{\mathcal{J}}(\mathbf{x}_k) \\ & \|\mathbf{d}\|_\infty\le\rho \end{cases}$$

where $\mathcal{J}(\mathbf{x}_k)=\{i\ :\ c_i(\mathbf{x}_k)>0\}$, $\mathcal{J}^{\perp}(\mathbf{x}_k)=\{i\ :\ c_i(\mathbf{x}_k)\le 0\}$ and $\bar{\mathcal{J}}(\mathbf{x}_k)=\{i\ :\ c_i(\mathbf{x}_k)>0 \text{ and } \nabla c_i(\mathbf{x}_k)^T\mathbf{d}+c_i(\mathbf{x}_k)\le 0\}$, and we denote the solution, the search direction as $\mathbf{d}$.

As in all trust region based methods we define

$$\Delta h \;=\; h(\mathbf{c}(\mathbf{x}_k))-h(\mathbf{c}(\mathbf{x}_k+\mathbf{d}))$$

as the *actual reduction* in $h(\mathbf{c}(\mathbf{x}_k))$, and let

$$\begin{aligned}\Delta l &= l(\mathbf{0})-l(\mathbf{d})\\ &= -\sum_{i\in\mathcal{J}(\mathbf{x}_k)}\nabla c_i(\mathbf{x}_k)^T\mathbf{d}\end{aligned}$$

be the *linear predicted reduction* in $h(\mathbf{c}(\mathbf{x}_k))$ where

$$l(\mathbf{d})=\sum_{i\in\mathcal{J}(\mathbf{x}_k)}c_i(\mathbf{x}_k)+\sum_{i\in\mathcal{J}(\mathbf{x}_k)}\nabla c_i(\mathbf{x}_k)^T\mathbf{d}.$$

In addition, we also let

$$\begin{aligned}\Delta q &= q(\mathbf{0})-q(\mathbf{d})\\ &= -\sum_{i\in\mathcal{J}(\mathbf{x}_k)}\nabla c_i(\mathbf{x}_k)^T\mathbf{d}-\frac{1}{2}\mathbf{d}^T\mathbf{W}_k\mathbf{d}\end{aligned}$$

be the *quadratic predicted reduction* in $h(\mathbf{c}(\mathbf{x}_k))$ where we define

$$q(\mathbf{d}) = \sum_{i\in\mathcal{J}(\mathbf{x}_k)} c_i(\mathbf{x}_k) + \sum_{i\in\mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{W}_k\mathbf{d}.$$

In order for the trial step $\mathbf{x}_k + \mathbf{d}$ to be accepted by the algorithm, we require the trial step to satisfy the simple sufficient reduction condition

$$\Delta h \geq \sigma \Delta q$$

where $\Delta q > 0$ and $\sigma \in (0,1)$ is a pre-assigned parameter. If both $\Delta l < 0$ and $\Delta q < 0$, we then reject the step calculated from $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ and solve the following subproblem

$$\widetilde{QP}(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho) \begin{cases} \underset{\mathbf{d}\in\mathbb{R}^n}{\text{minimize}} & \displaystyle\sum_{i\in\mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{W}_k\mathbf{d} \\ \text{subject to} & \nabla c_i(\mathbf{x}_k)^T\mathbf{d} + c_i(\mathbf{x}_k) \leq 0, \qquad i \in \mathcal{J}^{\perp}(\mathbf{x}_k) \cup \bar{\mathcal{J}}(\mathbf{x}_k) \\ & \displaystyle\sum_{i\in\mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T\mathbf{d} \leq 0 \\ & \|\mathbf{d}\|_\infty \leq \rho \end{cases}$$

and we denote the solution as $\mathbf{d}$. Take note that for the above subproblem the set $\bar{\mathcal{J}}(\mathbf{x}_k)$ might not be the same as the set $\bar{\mathcal{J}}(\mathbf{x}_k)$ defined in $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ and by solving $\widetilde{QP}(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ we can ensure that $\Delta l > 0$, and we will show in Lemma 3 that the quadratic predicted reduction can be shown to satisfy the following condition

$$\Delta q \begin{cases} = \frac{1}{2}\Delta l^2/b & \text{if } \Delta l < b \\ \geq \frac{1}{2}\Delta l & \text{otherwise} \end{cases}$$

where $b = \mathbf{d}^T\mathbf{W}_k\mathbf{d}$.

Given all the necessary definitions, we are now in a position to state the trust region SQP algorithm by means of the following pseudo-code.

Feasibility Restoration Phase Algorithm

Given initial point $\mathbf{x}_0$, set $k := 0$, $\sigma \in (0,1)$ and $\rho \geq \rho_{\min}$. Set $\mathcal{V}(\mathbf{x}_k) = \{i \ : \ c_i(\mathbf{x}_k) > 0\}$ and $\mathcal{V}^{\perp}(\mathbf{x}_k) = \{i \ : \ c_i(\mathbf{x}_k) \leq 0\}$.

Step 1 If $\mathcal{V}(\mathbf{x}_k) = \emptyset$ **Then**

- STOP (all constraints are feasible)

Endif

Step 2 Find $\mathbf{d}$ by solving $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem.

Step 3 **If** *convergence criterion* is met **Then**

- STOP.

Else if $\Delta l < 0$ and $\Delta q < 0$ **Then**

- Goto Step 4.

Else

- Goto Step 5.

Endif

Step 4 Find $\mathbf{d}$ by solving $\widetilde{QP}(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$.

Step 5 **If** $\Delta h < \sigma \Delta q$ **Then**

- Goto Step 6.

Else

- Goto Step 7.

Endif

Step 6 Set $\rho := \frac{1}{2}\rho$ and return to Step 2.

Step 7 Set $\mathbf{d}_k = \mathbf{d}$, $\rho_k = \rho$, $\Delta h_k = \Delta h$, $\Delta q_k = \Delta q$.

Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$.

Set $\mathcal{V}(\mathbf{x}_{k+1}) = \{i \ : \ c_i(\mathbf{x}_{k+1}) > 0\}$ and $\mathcal{V}^{\perp}(\mathbf{x}_{k+1}) = \{i \ : \ c_i(\mathbf{x}_{k+1} \leq 0\}$.

Set $\rho = \begin{cases} \rho_k & \text{if } \|\mathbf{d}_k\|_\infty < \rho_k, \\ 2\rho_k & \text{otherwise.} \end{cases}$

Set $\rho \geq \rho_{\min}$ if $\rho < \rho_{\min}$.

Set $k := k + 1$.

Goto Step 1.

We begin with an initial guess $\mathbf{x}_0$ of the solution $\mathbf{x}^\infty$ and if $h(\mathbf{c}(\mathbf{x}_k)) = 0$, we would then terminate the algorithmic process. At every iteration k there is an inner loop (Steps 2 to 6) where decreasing values of ρ are generated until a trial point $\mathbf{x}_k + \mathbf{d}$ is acceptable by the algorithm. At the start of a new iteration, the trust region radius is always initialized with any value $\rho \geq \rho_{\min}$, where $\rho_{\min} > 0$ is a pre-set parameter. We will show in the next section that by employing such a strategy a trial step will be accepted by the algorithm in a finite number of iterations. For each new value for ρ in the inner loop, the set $\bar{\mathcal{J}}(\mathbf{x}_k) \subseteq \mathcal{J}(\mathbf{x}_k)$ is determined by means of the following steps

1. Set $\bar{\mathcal{J}}(\mathbf{x}_k) = \emptyset$ and solve $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ to find $\mathbf{d}$.
2. If $c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d} > 0$ for all $i \in \mathcal{J}(\mathbf{x}_k)$, then STOP.

3. Otherwise include indices $i \in \bar{\mathcal{J}}(\mathbf{x}_k)$ such that $c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d} \leq 0$ from $\mathcal{J}(\mathbf{x}_k)$.

4. Return to step 2 to find a new solution $\mathbf{d}$ of $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$.

Take note that as the number of constraints in $\mathcal{J}(\mathbf{x}_k)$ is finite, thus only a finite number of repeat steps are needed to determine the set $\bar{\mathcal{J}}(\mathbf{x}_k)$. In addition because the old $\mathbf{d}$ is feasible in the new $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem, hence the new QP subproblem must also be compatible.

After the set $\bar{\mathcal{J}}(\mathbf{x}_k)$ is found, the next step is to check whether $\Delta l > 0$ or $\Delta q > 0$, and if both of them are less than zero, we then proceed to Step 4 of the restoration phase algorithm to solve $\widetilde{QP}(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$. To ensure that there exists an optimal solution $\mathbf{d}$ in $\widetilde{QP}(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ such that $\Delta l > 0$, we can do so by using the same steps as discussed before with an additional constraint $\sum_{i \in \mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T \mathbf{d} \leq 0$ apart from $\|\mathbf{d}\|_\infty \leq \rho$ in the beginning of a new QP subproblem. By initializing $\bar{\mathcal{J}}(\mathbf{x}_k) = \emptyset$ at the beginning of $\widetilde{QP}(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem, we know the subproblem is compatible since $\mathbf{d} = \mathbf{0}$ also solves the problem.

In Step 5 of the restoration phase algorithm, a simple sufficient reduction test is performed and if the trial step is rejected, then ρ is halved and the inner loop (Steps 2 to 6) is repeated. Otherwise the inner loop terminates and the current values for ρ and $\mathbf{d}$ are denoted by ρ_k and $\mathbf{d}_k$ respectively. We then update $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$ and find the new sets $\mathcal{V}(\mathbf{x}_{k+1})$ and $\mathcal{V}^\perp(\mathbf{x}_{k+1})$, and proceed back to Step 1 of the main algorithm.

Note that in the restoration phase, the main objective is to generate iterates that improve the constraints infeasibility but there is always a possibility that the restoration phase might fail to terminate to a feasible point with respect to (1.1) and converge to an infeasible point instead. An example of this behaviour could happened if there exists a non-zero local minimum of $h(\mathbf{c}(\mathbf{x}))$ which indicates that the Problem P is locally incompatible. Note that in this paper we are not interested in making a global statement about the infeasibility of (1.1) which is not practical for any nonlinear systems of any size, as this would be akin to global minimization of Problem P. Thus in this paper we also allow the possibility that the restoration phase may fail to find a feasible point, and regard this scenario as an indication that the constraints of Problem P is locally incompatible.

3 Global Convergence

In this section we present the global convergence proof of the trust region based feasibility restoration phase algorithm. Before going further we make the following assumptions.

Standard Assumptions

(**A1**) Let $\{\mathbf{x}_k\}$ be generated by the feasibility restoration phase algorithm and suppose that $\{\mathbf{x}_k\}$ and $\{\mathbf{x}_k + \mathbf{d}\}$ are contained in a compact and convex set $\mathcal{S}$ of $\mathbb{R}^n$.

(**A2**) Assume $c_i(\mathbf{x})$, $i = 1, 2, \ldots, m$ are twice continuously differentiable on an open set containing $\mathcal{S}$.

(**A3**) Assume $\mathbf{W}_k$ is bounded for all k.

Remark A consequence of assumptions (**A2**) - (**A3**) is that there exists a constant $M > 0$, independent of $\mathbf{x}$ and k such that for all $\mathbf{x} \in \mathcal{S}$ and for all k, it follows that $\frac{1}{2}|\mathbf{s}^T\mathbf{W}_k\mathbf{s}| \leq M$, for all vectors $\mathbf{s}$ such that $\|\mathbf{s}\|_\infty = 1$. Without loss of generality, we may also assume $\|\nabla\mathbf{c}(\mathbf{x})\|_1 \leq M$.

We first present the following lemma which concerns about Lipschitz continuity of h.

Lemma 1 *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *where* $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ *and* $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ *then*

$$|h(\mathbf{x}) - h(\mathbf{y})| \leq \|\mathbf{x} - \mathbf{y}\|_1$$

where $h(\mathbf{x}) = \sum_{i=1}^{n} \max\{0, x_i\}$ *and* $h(\mathbf{y}) = \sum_{i=1}^{n} \max\{0, y_i\}$.

Proof We let $u_i = \max\{0, x_i\}$ and $v_i = \max\{0, y_i\}$ for $i = 1, 2, \ldots, n$. Hence

$$\begin{aligned} |h(\mathbf{x}) - h(\mathbf{y})| &= \left|\sum_{i=1}^{n} \max\{0, x_i\} - \sum_{i=1}^{n} \max\{0, y_i\}\right| \\ &= \left|\sum_{i=1}^{n}(u_i - v_i)\right| \\ &\leq \sum_{i=1}^{n} |u_i - v_i|. \end{aligned}$$

To prove the result we consider 4 cases

(a) If $x_i > 0$ and $y_i > 0$ then $u_i = \max\{0, x_i\} = x_i$ and $v_i = \max\{0, y_i\} = y_i$. Thus $|u_i - v_i| = |x_i - y_i|$.

(b) If $x_i > 0$ and $y_i \leq 0$ then $u_i = \max\{0, x_i\} = x_i$ and $v_i = \max\{0, y_i\} = 0$. Thus $|u_i - v_i| = |x_i| \leq |x_i - y_i|$.

(c) If $x_i \leq 0$ and $y_i \leq 0$ then $u_i = \max\{0, x_i\} = 0$ and $v_i = \max\{0, y_i\} = 0$. Thus $|u_i - v_i| = 0 \leq |x_i - y_i|$.

(d) If $x_i \leq 0$ and $y_i > 0$ then $u_i = \max\{0, x_i\} = 0$ and $v_i = \max\{0, y_i\} = y_i$. Thus $|u_i - v_i| = |y_i| \leq |x_i - y_i|$.

Therefore from case (a) to case (d)

$$\begin{aligned} |h(\mathbf{x}) - h(\mathbf{y})| &\leq \sum_{i=1}^{n} |u_i - v_i| \\ &\leq \sum_{i=1}^{n} |x_i - y_i| \\ &= \|\mathbf{x} - \mathbf{y}\|_1. \end{aligned}$$

q.e.d

The next lemma gives a result concerning the relationship between Δh and Δq.

Lemma 2 *Let the standard assumptions hold and consider the QP step* $\mathbf{d}$. *For any* $\mathbf{x}_k \in \mathcal{S}$, $\mathbf{d} \in \mathbb{R}^n$, *the line segment from* $\mathbf{x}_k$ *to* $\mathbf{x}_k + \mathbf{d}$ *is contained in* $\mathcal{S}$

$$\Delta h - \Delta q \geq \sum_{i \in \bar{\mathcal{J}}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}\} - \beta \rho^2$$

where $\beta > 0$

Proof From definition

$$\begin{aligned} \Delta h - \Delta q &= h(\mathbf{c}(\mathbf{x}_k)) - h(\mathbf{c}(\mathbf{x}_k + \mathbf{d})) + \sum_{i \in \mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T \mathbf{d} + \frac{1}{2} \mathbf{d}^T \mathbf{W}_k \mathbf{d} \\ &= h(\mathbf{c}(\mathbf{x}_k)) - h(\mathbf{c}(\mathbf{x}_k + \mathbf{d})) + \sum_{i \in \mathcal{J}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}\} - \\ &\quad \sum_{i \in \mathcal{J}(\mathbf{x}_k)} c_i(\mathbf{x}_k) + \frac{1}{2} \mathbf{d}^T \mathbf{W}_k \mathbf{d} \\ &= h(\mathbf{c}(\mathbf{x}_k)) - h(\mathbf{c}(\mathbf{x}_k + \mathbf{d})) + \sum_{i \in \mathcal{J}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}\} - \\ &\quad h(\mathbf{c}(\mathbf{x}_k)) + \frac{1}{2} \mathbf{d}^T \mathbf{W}_k \mathbf{d} \\ &= \sum_{i \in \mathcal{J}(\mathbf{x}_k) / \bar{\mathcal{J}}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}\} + \sum_{i \in \bar{\mathcal{J}}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}\} - \\ &\quad h(\mathbf{c}(\mathbf{x}_k + \mathbf{d})) + \frac{1}{2} \mathbf{d}^T \mathbf{W}_k \mathbf{d} \end{aligned}$$

$$\begin{aligned} &= \sum_{i\in\bar{\mathcal{J}}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}\} + h(\mathbf{c}(\mathbf{x}_k) + \nabla \mathbf{c}(\mathbf{x}_k)^T\mathbf{d}) - \\ &\quad h(\mathbf{c}(\mathbf{x}_k + \mathbf{d})) + \frac{1}{2}\mathbf{d}^T\mathbf{W}_k\mathbf{d} \\ &\geq \sum_{i\in\bar{\mathcal{J}}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}\} - |h(\mathbf{c}(\mathbf{x}_k) + \nabla \mathbf{c}(\mathbf{x}_k)^T\mathbf{d}) - \\ &\quad h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}))| - M\rho^2. \end{aligned} \tag{3.1}$$

By focussing on $|h(\mathbf{c}(\mathbf{x}_k) + \nabla \mathbf{c}(\mathbf{x}_k)^T\mathbf{d}) - h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}))|$ and using the results from Lemma 1 we have

$$\begin{aligned} |h(\mathbf{c}(\mathbf{x}_k) + \nabla \mathbf{c}(\mathbf{x}_k)^T\mathbf{d}) - h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}))| &= |h(\mathbf{c}(\mathbf{x}_k) + \nabla \mathbf{c}(\mathbf{x}_k)^T\mathbf{d}) - \\ &\quad h(\mathbf{c}(\mathbf{x}_k) + \int_0^1 \nabla \mathbf{c}(\mathbf{x}_k + t\mathbf{d})^T\mathbf{d}\ dt)| \\ &\leq \left|\nabla \mathbf{c}(\mathbf{x}_k)^T\mathbf{d} - \int_0^1 \nabla \mathbf{c}(\mathbf{x}_k + t\mathbf{d})^T\mathbf{d}\ dt\right| \\ &= \left|\int_0^1 \left\{\nabla \mathbf{c}(\mathbf{x}_k)^T\mathbf{d} - \nabla \mathbf{c}(\mathbf{x}_k + t\mathbf{d})^T\mathbf{d}\right\}\ dt\right| \\ &\leq \int_0^1 \|\nabla \mathbf{c}(\mathbf{x}_k) - \nabla \mathbf{c}(\mathbf{x}_k + t\mathbf{d})\|_1 \|\mathbf{d}\|_1\ dt \\ &\leq \int_0^1 M\|t\mathbf{d}\|_1\|\mathbf{d}\|_1\ dt \\ &= M\|\mathbf{d}\|_1^2 \int_0^1 t\ dt \\ &= \frac{1}{2}M\|\mathbf{d}\|_1^2. \end{aligned} \tag{3.2}$$

Therefore by subsituting (3.2) into (3.1) we have

$$\Delta h - \Delta q \ \geq \ \sum_{i\in\bar{\mathcal{J}}(\mathbf{x}_k)} \{c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}\} - \beta\rho^2$$

where $\beta = \frac{1}{2}Mn^2 + M$. *q.e.d*

We next give a result relating to Δq.

Lemma 3 *Let the standard assumptions hold and consider the step* $\mathbf{d}$ *such that* $\Delta l \geq \eta\rho > 0$ *where* $\eta > 0$ *is a constant. It follows that the quadratic predicted reduction* Δq *satisfies the condition*

$$\Delta q \ \geq \ \frac{1}{2}\min\left\{\eta\rho, \frac{\eta^2}{2M}\right\}. \tag{3.3}$$

In addition if

$$\rho \;\leq\; \frac{\eta}{2M} \tag{3.4}$$

then $\Delta q \geq \frac{1}{2}\eta\rho$.

Proof We consider minimizing a quadratic function $\phi(\alpha) = -\alpha\Delta l + \frac{1}{2}\alpha^2 b$ on the interval $\alpha \in [0,1]$ where $b = \mathbf{d}^T\mathbf{W}_k\mathbf{d} \leq 2M\rho^2$. Because $\Delta l > 0$, we need to consider $\alpha > 0$ only.

The unconstrained minimizer, if it exists, is at $\alpha = \Delta l/b$. Hence if $\Delta l < b$, the minimizer is $\alpha = \Delta l/b < 1$ and

$$\begin{aligned}\Delta q &= \frac{\Delta l^2}{b} - \frac{\Delta l^2}{2b}\\ &= \frac{\Delta l^2}{2b}\\ &\geq \frac{\eta^2}{4M}.\end{aligned}$$

Otherwise if $\Delta l \geq b$, the minimizer is $\alpha = 1$ and

$$\begin{aligned}\Delta q &= \Delta l - \frac{1}{2}b\\ &\geq \frac{1}{2}\eta\rho.\end{aligned}$$

Putting these two inequalities together gives

$$\Delta q \;\geq\; \frac{1}{2}\min\left\{\eta\rho, \frac{\eta^2}{2M}\right\}.$$

The second part of the result follows easily such that if the unconstrained minimizer $\alpha < 1$ then $\alpha = \Delta l/b$ and by substituting the upper bound for b and because $\Delta l \geq \eta\rho$ we have

$$\begin{aligned}\alpha &= \frac{\Delta l}{b}\\ &\geq \frac{\eta}{2M\rho}.\end{aligned}$$

Hence if $\rho \leq \eta/2M$ then $\alpha \geq 1$ which is a contradiction. Thus $\alpha = 1$ and hence $\Delta q \geq \frac{1}{2}\eta\rho$. *q.e.d*

We now proceed to analyze the feasibility restoration phase algorithm in detail by first showing that the inner loop of the algorithm terminates finitely. Here $\mathbf{x}_k$ is fixed and we consider what happens to the QP subproblem as ρ is reduced.

Theorem 1 *Let the standard assumptions hold, then the inner loop (Steps 2 to 6) terminates in a finite number of iterations.*

Proof To prove this result, we will only consider the case when $\mathbf{x}_k$ is not a feasible point that satisfies (1.1). Otherwise there is nothing to prove. If $\mathbf{x}_k$ is a KKT point of Problem P then $\mathcal{J}(\mathbf{x}_k) = \mathcal{V}(\mathbf{x}_k)$, $\mathcal{J}^{\perp}(\mathbf{x}_k) = \mathcal{V}^{\perp}(\mathbf{x}_k)$, $\bar{\mathcal{J}}(\mathbf{x}_k) = \emptyset$ and hence $\mathbf{d} = \mathbf{0}$ solves $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem. Thus the algorithm terminates.

On the other hand if $\mathbf{x}_k$ is not a KKT point, and if the inner loop does not terminate finitely then the rules for decreasing ρ ensures that $\rho \to 0$ and hence for all $i \in \mathcal{J}(\mathbf{x}_k)$, $\nabla c_i(\mathbf{x}_k)^T\mathbf{d} + c_i(\mathbf{x}_k) > 0$. Thus inactive constraints at $\mathbf{x}_k$ are inactive at any point for which $\|\mathbf{d}\|_\infty \le \rho$ for sufficiently small ρ. Because $\mathbf{x}_k$ is not a KKT point, then there exists a feasible descent direction $\mathbf{s}$, $\|\mathbf{s}\|_\infty = 1$ and an $\varepsilon > 0$ such that

$$\sum_{i\in\mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T\mathbf{s} \;\le\; -\varepsilon.$$

By denoting $\rho\mathbf{s}$ as the feasible descent direction with respect to $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem then from Lemma 3 by subsituting $\eta = \varepsilon$ in (3.4), if

$$\rho \;\le\; \frac{\varepsilon}{2M} \tag{3.5}$$

then $q(\mathbf{0}) - q(\rho\mathbf{s}) \ge \varepsilon\rho/3$. By optimality of $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ we therefore have

$$\begin{aligned} \Delta q &\ge q(\mathbf{0}) - q(\rho\mathbf{s}) && (3.6)\\ &\ge \frac{1}{3}\varepsilon\rho && (3.7) \end{aligned}$$

Furthermore because $\bar{\mathcal{J}}(\mathbf{x}_k) = \emptyset$ for sufficiently small ρ, if (3.5) holds such that $\Delta q \ge \varepsilon\rho/3$, then from Lemma 2 if

$$\rho \;\le\; \frac{(1-\sigma)\varepsilon}{2\beta}$$

we therefore have $\Delta h \ge \sigma\Delta q$. Hence if ρ lies in the range

$$\rho \;\le\; \min\left\{\frac{\varepsilon}{2M}, \frac{(1-\sigma)\varepsilon}{2\beta}\right\} \tag{3.8}$$

then $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ is consistent, $\Delta q \ge \varepsilon\rho/3$, $\Delta h \ge \sigma\Delta q$ and thus $\mathbf{x}_k + \mathbf{d}$ will be accepted by the algorithm. Hence the inner loop terminates for this case.

q.e.d

Our global convergence proof concerns with Karush-Kuhn-Tucker (KKT) necessary conditions under a strict Mangasarian-Fromowitz constraint qualification (SMFCQ) (see Kyparisis [14]). Let $\mathbf{x}^\infty$ be a feasible point of Problem P, then $\mathbf{x}^\infty$ satisfies SMFCQ if and only if

(1) $\nabla c_i(\mathbf{x}^\infty)$, $i \in \mathcal{V}^\perp_{\mathcal{E}_1}(\mathbf{x}^\infty)$ are linearly independent; and

(2) there exists a vector $\mathbf{s} \in \mathbb{R}^n$ such that

$$\begin{aligned} \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} &= 0, \quad i \in \mathcal{V}^\perp_{\mathcal{E}_1}(\mathbf{x}^\infty) \\ \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} &< 0, \quad i \in \mathcal{V}^\perp_{\mathcal{E}_2}(\mathbf{x}^\infty) \end{aligned}$$

where $\mathcal{E}(\mathbf{x}^\infty) = \{i : c_i(\mathbf{x}^\infty) = 0\}$, $\mathcal{V}^\perp_{\mathcal{E}_1}(\mathbf{x}^\infty) = \{i \in \mathcal{E}(\mathbf{x}^\infty) : \lambda_i^\infty > 0\}$ and $\mathcal{V}^\perp_{\mathcal{E}_2}(\mathbf{x}^\infty) = \{i \in \mathcal{E}(\mathbf{x}^\infty) : \lambda_i^\infty = 0\}$. Take note that in this paper, constraints that are in the set $\mathcal{V}^\perp_{\mathcal{E}_1}(\mathbf{x}^\infty)$ are considered as *strongly active* while constraints that are in the set $\mathcal{V}^\perp_{\mathcal{E}_2}(\mathbf{x}^\infty)$ are considered as *weakly active*.

We now state the KKT necessary conditions. Let $\mathbf{x}^\infty$ solves Problem P at which $\mathcal{V}(\mathbf{x}^\infty) \neq \emptyset$ and define the sets

$$\mathcal{G}_\mathcal{V} = \left\{ \mathbf{s} : \sum_{i \in \mathcal{V}(\mathbf{x}^\infty)} \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} < 0 \right\} \tag{3.9}$$

$$\mathcal{C}_{\mathcal{V}^\perp_{\mathcal{E}_1}} = \left\{ \mathbf{s} : \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} = 0, i \in \mathcal{V}^\perp_{\mathcal{E}_1}(\mathbf{x}^\infty) \right\} \tag{3.10}$$

$$\mathcal{C}_{\mathcal{V}^\perp_{\mathcal{E}_2}} = \left\{ \mathbf{s} : \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} < 0, i \in \mathcal{V}^\perp_{\mathcal{E}_2}(\mathbf{x}^\infty) \right\}. \tag{3.11}$$

Thus the KKT necessary conditions for $\mathbf{x}^\infty$ to solve Problem P are

(i) $\mathbf{x}^\infty$ is a feasible point; and

(ii) the set of directions $\mathcal{G}_\mathcal{V} \cap \mathcal{C}_{\mathcal{V}^\perp_{\mathcal{E}_1}} \cap \mathcal{C}_{\mathcal{V}^\perp_{\mathcal{E}_2}} = \emptyset$.

When both conditions (i) - (ii) are true, then we shall refer $\mathbf{x}^\infty$ as a KKT point. Furthermore the consequence of conditions (i) - (ii) implies the existence of unique multipliers (see Kyparisis [14]), and it can be shown (see Gauvin [11]) that the multiplier vector $\boldsymbol{\lambda}^\infty$ is bounded. Take note that if the strict complementary slackness conditions holds at $\mathbf{x}^\infty$, i.e. $\mathcal{C}_{\mathcal{V}^\perp_{\mathcal{E}_2}} = \emptyset$, then the SMFCQ is equivalent to the Cottle constraint qualification (see Bazaraa, Sherali and Shetty [1] and Mangasarian [15]).

In order to show global convergence, we note that if $\mathbf{x}^\infty$ is a feasible point of Problem P but not a KKT point then there exists an $\varepsilon > 0$ and a vector $\mathbf{s}$ such that $\|\mathbf{s}\|_\infty = 1$ for which

$$\sum_{i \in \mathcal{V}(\mathbf{x}^\infty)} \nabla c_i(\mathbf{x}_k)^T\mathbf{s} \leq -\varepsilon \tag{3.12}$$

$$\nabla c_i(\mathbf{x}_k)^T\mathbf{s} = 0 \quad i \in \mathcal{V}^\perp_{\mathcal{E}_1}(\mathbf{x}^\infty) \tag{3.13}$$

$$\nabla c_i(\mathbf{x}_k)^T\mathbf{s} \leq -\varepsilon \quad i \in \mathcal{V}^\perp_{\mathcal{E}_2}(\mathbf{x}^\infty) \tag{3.14}$$

for all $\mathbf{x}_k$ in some neighbourhood $\mathcal{N}^\infty$ of $\mathbf{x}^\infty$. The conditions (3.12) – (3.14) are a direct consequence of condition (ii) and the continuity of the vectors $\nabla c_i(\mathbf{x}_k)$, $i \in \mathcal{V}(\mathbf{x}^\infty) \cup$

$\mathcal{E}(\mathbf{x}^\infty)$. With the conditions (3.12) - (3.14), we are now in a position to prove the global convergence of our algorithm.

Before proving the global convergence theorem, we need to provide some results that describe the behaviour of QP subproblems in the neighbourhood of a feasible point $\bar{\mathbf{x}}$ with respect to Problem P at which $\mathcal{V}(\bar{\mathbf{x}}) \neq \emptyset$ and SMFCQ holds.

Lemma 4 *Let the standard assumptions hold, and let $\bar{\mathbf{x}} \in \mathcal{S}$ be a feasible point of Problem P at which $\mathcal{V}(\bar{\mathbf{x}}) \neq \emptyset$ and SMFCQ holds, but not a KKT point. Then there exists a neighbourhood $\mathcal{N}$ of $\bar{\mathbf{x}}$ and positive constants ε, μ and τ such that for all $\mathbf{x}_k \in \mathcal{N} \cap \mathcal{S}$ and if $\mathcal{J}(\mathbf{x}_k)$ is such that $\mathcal{V}(\bar{\mathbf{x}}) \subset \mathcal{J}(\mathbf{x}_k) \subset \mathcal{V}(\bar{\mathbf{x}}) \cup \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$, $\bar{\mathcal{J}}(\mathbf{x}_k) \subset \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$ and if*

$$\mu \|\mathbf{c}_k\|_2 \leq \rho \leq \tau \tag{3.15}$$

where $\|\mathbf{c}_k\|_2 = \sqrt{\sum_{i \in \mathcal{E}(\bar{\mathbf{x}})} \{c_i(\mathbf{x}_k)\}}$ then it follows that $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem is consistent, the quadratic predicted reduction $\Delta q \geq \frac{1}{3}\rho\varepsilon$ and the sufficient reduction $\Delta h \geq \sigma \Delta q$ holds.

Proof Suppose $\bar{\mathbf{x}}$ is a feasible point and not a KKT point of Problem P, then for $\mathbf{x}_k$ in some neighbourhood $\mathcal{N}$ of $\bar{\mathbf{x}}$, there exists a vector $\mathbf{s}^\circ$ such that $\|\mathbf{s}^\circ\|_\infty = 1$ for which

$$\begin{aligned} \sum_{i \in \mathcal{V}(\bar{\mathbf{x}})} \nabla c_i(\mathbf{x}_k)^T \mathbf{s}^\circ &< 0 \\ \nabla c_i(\mathbf{x}_k)^T \mathbf{s}^\circ &= 0, \quad i \in \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}}) \\ \nabla c_i(\mathbf{x}_k)^T \mathbf{s}^\circ &< 0, \quad i \in \mathcal{V}^{\perp}_{\mathcal{E}_2}(\bar{\mathbf{x}}). \end{aligned}$$

Using similar techniques as described in Chin [5], Fletcher and Leyffer [7] and, Fletcher, Leyffer and Toint [10], let the matrix $\mathbf{A}_k$ consists of column vectors $\nabla c_i(\mathbf{x}_k)$, $i \in \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$. By linear independence and continuity there exists a neighbourhood $\mathbf{x}_k \in \mathcal{N} \cap \mathcal{S}$ in which $\mathbf{A}_k$ has full column rank. Hence for the linear system

$$\mathbf{A}_k^T \mathbf{d} + \mathbf{b}_k = \mathbf{0}$$

where $\mathbf{b}_k$ is the corresponding column vector, we can write the range space vector $\mathbf{p}$ as

$$\mathbf{p} = -(\mathbf{A}_k^T)^+ \mathbf{b}_k$$

such that $(\mathbf{A}_k^T)^+$ is the pseudo-inverse of $\mathbf{A}_k^T$.

In addition since $\mathbf{A}_k$ has full column rank, the null space of $\mathbf{A}_k^T$ defines the tangent space to the strongly active inequality constraints, $i \in \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$ at $\mathbf{x}_k$. We can then write the projection onto this tangent space as

$$\mathbf{P}_k = \mathbf{I} - \mathbf{A}_k \left[\mathbf{A}_k^T \mathbf{A}_k\right]^{-1} \mathbf{A}_k^T$$

and we let

$$\mathbf{s} = \frac{\mathbf{P}_k\mathbf{s}^\circ}{\|\mathbf{P}_k\mathbf{s}^\circ\|_2}$$

which is the closest unit vector to $\mathbf{s}^\circ$ in the null space of $\mathbf{A}_k^T$. Hence by continuity there exists a (smaller) neighbourhood $\mathcal{N}$ and a constant $\varepsilon > 0$ such that

$$\begin{aligned} \sum_{i\in\mathcal{V}(\bar{\mathbf{x}})} \nabla c_i(\mathbf{x}_k)^T\mathbf{s} &\le -\varepsilon \\ \nabla c_i(\mathbf{x}_k)^T\mathbf{s} &= 0, \quad i \in \mathcal{V}_{\mathcal{E}_1}^{\perp}(\bar{\mathbf{x}}) \\ \nabla c_i(\mathbf{x}_k)^T\mathbf{s} &\le -\varepsilon, \quad i \in \mathcal{V}_{\mathcal{E}_2}^{\perp}(\bar{\mathbf{x}}) \end{aligned}$$

for any $\mathbf{x}_k \in \mathcal{N} \cap \mathcal{S}$. By definition the range space vector $\mathbf{p}$ satisfies

$$\begin{aligned} \|\mathbf{p}\|_2 &\le \|(\mathbf{A}_k^T)^+\|_2\|\mathbf{b}_k\|_2 \\ &\le M\|\mathbf{c}_k\|_2 \end{aligned}$$

where $M > 0$ and $\|\mathbf{c}_k\|_2 = \sqrt{\sum_{i\in\mathcal{E}(\bar{\mathbf{x}})} \{c_i(\mathbf{x}_k)\}^2}$. Furthermore if

$$\rho \ge M\|\mathbf{c}_k\|_2 \tag{3.16}$$

then $\|\mathbf{p}\|_\infty \le \|\mathbf{p}\|_2 \le \rho$.

We now consider the $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem and if (3.16) holds, we let the line segment be

$$\mathbf{d}^1 = \mathbf{p} + (\rho - \|\mathbf{p}\|_2)\mathbf{s}$$

where $\rho \ge \|\mathbf{p}\|_2$. Since the vectors $\mathbf{p}$ and $\mathbf{s}$ are orthogonal, it follows that $\|\mathbf{d}^1\|_\infty \le \|\mathbf{d}^1\|_2 \le \rho$, that is $\mathbf{d}^1$ satisfies the trust region constraint.

For strongly active constraints at $\bar{\mathbf{x}}$, $i \in \mathcal{V}_{\mathcal{E}_1}^{\perp}(\bar{\mathbf{x}})$

$$\begin{aligned} c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}^1 &= c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{p} + (\rho - \|\mathbf{p}\|_2)\nabla c_i(\mathbf{x}_k)^T\mathbf{s} \\ &= 0. \end{aligned} \tag{3.17}$$

On the other hand for weakly active constraints at $\bar{\mathbf{x}}$, $i \in \mathcal{V}_{\mathcal{E}_2}^{\perp}(\bar{\mathbf{x}})$

$$\begin{aligned} c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}^1 &= c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{p} + (\rho - \|\mathbf{p}\|_2)\nabla c_i(\mathbf{x}_k)^T\mathbf{s} \\ &\le \|\mathbf{c}_k\|_2 + M\|\mathbf{p}\|_2 - (\rho - \|\mathbf{p}\|_2)\varepsilon \\ &\le [1 + M(\varepsilon + M)]\,\|\mathbf{c}_k\|_2 - \varepsilon\rho \end{aligned} \tag{3.18}$$

and if

$$\rho \ge \frac{[1 + M(\varepsilon + M)]\,\|\mathbf{c}_k\|_2}{\varepsilon}$$

then the set of constraints (3.18) are consistent.

For non-active constraints $i \in \mathcal{V}^{\perp}(\bar{\mathbf{x}})/\mathcal{E}(\bar{\mathbf{x}})$, since $\mathbf{x}_k \in \mathcal{N} \cap \mathcal{S}$, then there exist $c_i(\mathbf{x}_k) \le -c$ and $\nabla c_i(\mathbf{x}_k)^T \mathbf{d}^1 \le a\rho$ where $a > 0$ and $c > 0$ are independent of $\mathbf{x}$ and k. It follows that

$$c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}^1 \;\; \le \;\; -c + a\rho \tag{3.19}$$

and if $\rho \le c/a$ then the set of constraints (3.19) are consistent.

For constraints $i \in \mathcal{V}(\bar{\mathbf{x}})$, there exist positive constants $\hat{c}$ and $\hat{a}$ independent of $\mathbf{x}$ and k such that $c_i(\mathbf{x}_k) > \hat{c}$ and $\nabla c_i(\mathbf{x}_k)^T \mathbf{d}^1 > -\hat{a}\rho$. Hence if $\rho \le \hat{c}/\hat{a}$, it follows that $c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}^1 > 0$, $i \in \mathcal{V}(\bar{\mathbf{x}})$. Hence it follows that if

$$\max\left\{ M\|\mathbf{c}_k\|_2, \frac{[1 + M(\varepsilon + M)]\,\|\mathbf{c}_k\|_2}{\varepsilon} \right\} \le \rho \le \min\left\{ \frac{c}{a}, \frac{\hat{c}}{\hat{a}} \right\} \tag{3.20}$$

then $\mathbf{d}^1 = \mathbf{p} + (\rho - \|\mathbf{p}\|_2)\mathbf{s}$ is a feasible point with respect to the linearized active inequality constraints, $\mathcal{V}(\bar{\mathbf{x}}) \subset \mathcal{J}(\mathbf{x}_k) \subset \mathcal{V}(\bar{\mathbf{x}}) \cup \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$ and the current $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem is compatible.

To show that $\mathbf{d}^1$ is a descent direction in the $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem, we note that for $\mathcal{V}(\bar{\mathbf{x}}) \subset \mathcal{J}(\mathbf{x}_k) \subset \mathcal{V}(\bar{\mathbf{x}}) \cup \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$ such that $\bar{\mathcal{J}}(\mathbf{x}_k) \subset \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$, we can therefore write

$$\begin{aligned}
\sum_{i \in \mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T \mathbf{d}^1 &= \sum_{i \in \mathcal{J}(\mathbf{x}_k)} \left\{ \nabla c_i(\mathbf{x}_k)^T \mathbf{p} + (\rho - \|\mathbf{p}\|_2) \nabla c_i(\mathbf{x}_k)^T \mathbf{s} \right\} \\
&\le M\|\mathbf{p}\|_2 + \varepsilon\|\mathbf{p}\|_2 - \varepsilon\rho \\
&\le M^2\|\mathbf{c}_k\|_2 + \varepsilon M\|\mathbf{c}_k\|_2 - \varepsilon\rho
\end{aligned}$$

since $\sum_{i \in \mathcal{V}(\bar{\mathbf{x}})} \nabla c_i(\mathbf{x}_k)^T \mathbf{s} \le -\varepsilon$ and $\sum_{i \in \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})} \nabla c_i(\mathbf{x}_k)^T \mathbf{s} = 0$. Hence if

$$\rho \;\; \ge \;\; \frac{3M(M + \varepsilon)\|\mathbf{c}_k\|_2}{\varepsilon} \tag{3.21}$$

then $\sum_{i \in \mathcal{J}(\mathbf{x}_k)} \nabla c_i(\mathbf{x}_k)^T \mathbf{d}^1 \le -\frac{2}{3}\varepsilon\rho$.

From the left-hand side of the inequality (3.20) and from (3.21), if we let

$$\max\left\{ M\|\mathbf{c}_k\|_2, \frac{[1 + M(\varepsilon + M)]\,\|\mathbf{c}_k\|_2}{\varepsilon}, \frac{3M(M + \varepsilon)\|\mathbf{c}_k\|_2}{\varepsilon} \right\} \;\; = \;\; \mu\|\mathbf{c}_k\|_2$$

where $\mu > 0$ and if

$$\mu\|\mathbf{c}_k\|_2 \le \rho \le \min\left\{ \frac{c}{a}, \frac{\hat{c}}{\hat{a}} \right\} \tag{3.22}$$

then $\mathbf{d}^1 = \mathbf{p} + (\rho - \|\mathbf{p}\|_2)\mathbf{s}$ is both a feasible and a descent direction for the $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem.

If (3.22) holds such that $l(\mathbf{0}) - l(\mathbf{d}^1) \geq \frac{2}{3}\varepsilon\rho$, then by subsituting $\eta = \frac{2}{3}\varepsilon$ in (3.4), if

$$\rho \leq \frac{\varepsilon}{3M}$$

then $q(\mathbf{0}) - q(\mathbf{d}^1) \geq \frac{1}{3}\varepsilon\rho$. Since $\mathbf{d}^1$ is both a feasible and descent direction, it follows by optimality of $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ that

$$\begin{aligned} \Delta q &\geq q(\mathbf{0}) - q(\mathbf{d}^1) \\ &\geq \frac{1}{3}\varepsilon\rho \end{aligned}$$

Hence for $\mathcal{V}(\bar{\mathbf{x}}) \subset \mathcal{J}(\mathbf{x}_k) \subset \mathcal{V}(\bar{\mathbf{x}}) \cup \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$ such that $\bar{\mathcal{J}}(\mathbf{x}_k) \subset \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$ and if

$$\mu\|\mathbf{c}_k\|_2 \leq \rho \leq \min\left\{\frac{c}{a}, \frac{\hat{c}}{\hat{a}}, \frac{\varepsilon}{3M}\right\} \tag{3.23}$$

then the $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem is consistent and $\Delta q \geq \frac{1}{3}\varepsilon\rho$.

To satisfy the sufficient reduction test, we note that if $\bar{\mathcal{J}}(\mathbf{x}_k) \subset \mathcal{V}^{\perp}_{\mathcal{E}_1}(\bar{\mathbf{x}})$ then from (3.17), by continuity the optimal QP step $\mathbf{d}$ would satisfy

$$\sum_{i \in \bar{\mathcal{J}}(\mathbf{x}_k)} \left\{\nabla c_i(\mathbf{x}_k)^T \mathbf{d} + c_i(\mathbf{x}_k)\right\} = 0.$$

Hence if the above condition and (3.23) holds, then from Lemma 2, $\Delta h \geq \Delta q - \beta\rho^2$ and if

$$\rho \leq \frac{(1-\sigma)\varepsilon}{3\beta} \tag{3.24}$$

then $\Delta h \geq \sigma\Delta q$.

From the right-hand side of the inequality (3.23) and (3.24), we let

$$\min\left\{\frac{c}{a}, \frac{\hat{c}}{\hat{a}}, \frac{\varepsilon}{3M}, \frac{(1-\sigma)\varepsilon}{3\beta}\right\} = \tau$$

where $\tau > 0$ is a constant and if

$$\mu\|\mathbf{c}_k\|_2 \leq \rho \leq \tau$$

then the $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem is consistent, $\Delta q \geq \frac{1}{3}\varepsilon\rho > 0$ and hence $\Delta h \geq \sigma\Delta q$.
q.e.d

In the next theorem we will show that the trust region based feasibility restoration phase algorithm does provide a global convergence proof for the class of problems that we are going to solve. Using similar strategy as in Fletcher and Leyffer [7] that when the algorithm is applied one of four different possible outcomes ($\mathbb{A}$ - $\mathbb{D}$) can occur.

($\mathbb{A}$) A feasible point $\mathbf{x}_k$ of (1.1) is found.

($\mathbb{B}$) A KKT point $\mathbf{x}_k$ of Problem P is found.

($\mathbb{C}$) There exists an accummulation point $\mathbf{x}^\infty$ which is a feasible point of (1.1).

($\mathbb{D}$) There exists an accummulation point $\mathbf{x}^\infty$ that either fails to satisfy SMFCQ conditions or is a KKT point of Problem P.

A consequence of Theorem 1 is that if the feasibility restoration phase algorithm does not terminate in either type ($\mathbb{A}$) or ($\mathbb{B}$) then an infinite subsequence of type ($\mathbb{C}$) or ($\mathbb{D}$) will occur. The next theorem completes the global convergence proof by showing that subsequences from type ($\mathbb{D}$) that satisfy SMFCQ conditions converge to a KKT point. But before we state the main results, we need to make an additional assumption. The discussion under which the condition might hold is further elaborated in Section 4.

Supplementary Assumptions

(**A4**) Under the conditions of Lemma 4, such that for $\mathbf{x}_k \in \mathcal{N} \cap \mathcal{S}$ and for all $\rho > \tau$, the linear predicted reduction satisfies $\Delta l \geq \frac{2}{3}\varepsilon\rho$ and hence from Lemma 3, $\Delta q \geq \frac{1}{3}\min\{\varepsilon\rho, 2\varepsilon^2/9M\}$.

Theorem 2 *Let the standard and the supplementary assumptions hold, then for the feasibility restoration phase algorithm either iterates of type* ($\mathbb{A}$) *or* ($\mathbb{B}$) *will occur, or there exists an infinite subsequence of type* ($\mathbb{C}$) *or* ($\mathbb{D}$). *If we denote* $\mathbf{x}^\infty$ *as an accumulation point for this subsequence then either* $\mathbf{x}^\infty$ *is a feasible point to (1.1), a KKT point or fails to satisfy SMFCQ conditions of Problem* P.

Proof We only need to consider the case in which ($\mathbb{A}$), ($\mathbb{B}$) or ($\mathbb{C}$) does not occur, and for case ($\mathbb{D}$), SMFCQ conditions are satisfied. Because we are not considering case ($\mathbb{C}$), it follows that $\mathcal{V}(\mathbf{x}^\infty) \neq \emptyset$. In order to prove this theorem, we first note that the inner loop of each iteration (Steps 2 to 6) is finite (Theorem 1), and the outer iteration sequence indexed by k is infinite. In addition from the standard assumptions we assume all iterates $\mathbf{x}_k \in \mathcal{S}$ which is bounded, hence it follows that the sequence or any infinite subsequence of iterates $\{\mathbf{x}_k\}$ has one or more accumulation points.

We let $\mathbf{x}^\infty$ be any accumulation point of the main sequence and examined the proposition (to be contradicted) that $\mathbf{x}^\infty$ is not a KKT point but a feasible point with respect to Problem P. It follows that there exists a neighbourhood $\mathcal{N}^\infty$ in which the set defined by (3.9), (3.10) and (3.11) is not empty, and hence Lemma 4 applies. We now consider any subsequence $k \in \mathcal{K}$ for which $\mathbf{x}_k \to \mathbf{x}^\infty$, and assume without loss of generality that $\mathbf{x}_k \in \mathcal{N}^\infty \cap \mathcal{S}$ and $k \geq K$ for all $k \in \mathcal{K}$.

From Lemma 4 we consider $\mathcal{V}(\mathbf{x}^\infty) \subset \mathcal{J}(\mathbf{x}_k) \subset \mathcal{V}(\mathbf{x}^\infty) \cup \mathcal{V}^{\perp}_{\mathcal{E}_1}(\mathbf{x}^\infty)$, $\bar{\mathcal{J}}(\mathbf{x}_k) \subset \mathcal{V}^{\perp}_{\mathcal{E}_1}(\mathbf{x}^\infty)$ and the interval of (3.15) where

$$\mu\|\mathbf{c}_k\|_2 \leq \rho \leq \tau \tag{3.25}$$

and it follows that for any value ρ in the interval $QP(\mathbf{x}_k, \mathcal{J}(\mathbf{x}_k), \rho)$ subproblem is consistent, the predicted quadratic reduction $\Delta q \geq \frac{1}{3}\rho\varepsilon$, the sufficient reduction $\Delta h \geq \sigma \Delta q$ holds and hence $\mathbf{x}_k + \mathbf{d}$ will be accepted by the algorithm.

The right hand side of (3.25) is a number, independent of k and $\mathbf{x}$, while the left hand side of (3.25) converges to zero. Thus there exists some iteration $\bar{K}$ such that $\mu\|\mathbf{c}_k\|_2 \leq \frac{1}{2}\tau$ for all $k \geq \bar{K} \geq K$. Thus, for any $k \geq \bar{K}$, as ρ is reduced in the inner loop, either it must eventually fall within the interval (3.25) or a value to the right of the interval is accepted. We can therefore guarantee that a value $\rho_k > \frac{1}{2}\tau$ will be chosen. If ρ falls within the interval such that $\rho \leq \tau$, then $\Delta q \geq \frac{1}{3}\varepsilon\rho$ and hence

$$\Delta h_k \;>\; \frac{1}{6}\sigma\varepsilon\tau.$$

If on the other hand if a value to the right of the interval (3.25) is accepted that is when $\rho > \tau$, then $\Delta h \geq \sigma\Delta q$ must hold and $\Delta l \geq \frac{2}{3}\tau\varepsilon$ holds (from the supplementary assumption) because Δl is a monotonically increasing function of ρ. We therefore deduce from Lemma 3 that by subsituting $\eta = \frac{2}{3}\varepsilon$

$$\Delta h_k \;\geq\; \frac{1}{3}\sigma \min\left\{\varepsilon\tau, \frac{2\varepsilon^2}{9M}\right\}.$$

Thus in either case, Δh_k is uniformly bounded away from zero, which, together with the monotonically decreasing h-values of the main sequence for $k \geq K$, contradicts the fact that the sequence $\{h(\mathbf{c}(\mathbf{x}_k))\}$ is bounded below. Thus any accumulation point is a KKT point.

q.e.d

4 Conclusion

A prototypical algorithm of applying simple sufficient reduction test in a trust region based SQP algorithm has been described demonstrating the fact that global convergence can also be achieved without resorting to a filter strategy or solving non-smooth optimization subproblems. Of course the feasibility restoration phase algorithm is incomplete in many areas and can only serve as a guide to what might be successfully implemented in practice. Central to this issue are the rule for adjusting ρ in the inner loop, the minimum threshold of a trust region size at the beginning of a new loop, $\rho_{\min}$ and also the strategy to express $\mathbf{W}_k$ used in the QP subproblems. One possibility is to use the Hessian of the Lagrangian calculated from the second derivatives of $\mathbf{c}$, and also using estimates of Lagrange multipliers. By doing so, we would expect fast local convergence of the iterates towards the local solution. However, the disadvantage of such a strategy is that $\mathbf{W}_k$ could be indefinite or the minimizer of the QP subproblem may not even exist, if the problem is unbounded. Hence the task of finding a global minimizer of the QP subproblem could be problematic. Another possibility is to utilize quasi-Newton methods to express $\mathbf{W}_k$ and update it at the start of a new iteration. This alternative strategy will help to ensure the positive-definiteness of $\mathbf{W}_k$ be maintained so that any KKT point of the QP subproblem is a global solution. If quasi-Newton updates to Hessians are used then there is an important question which update strategy to use without compromising the fast aymptotic behaviour of the iterates near the solution. Another important question to address also is whether the inclusion of second-order correction (SOC) steps in the algorithm would be advantageous to the local convergence rate of the iterates. It is hoped that with the full implementation of the algorithm, we can answer the above issues.

As for the validity of the supplementary assumption (**A4**), it is much more easier to guarantee than say, the additional assumption used in Fletcher and Leyffer [7]. This is due to fact that for $\Delta l \geq \frac{2}{3}\varepsilon\rho$ to hold, the only bounded criteria that we need is in the form of (3.21). As $\|\mathbf{c}_k\|_2$ converges to zero as $\mathbf{x}_k \to \mathbf{x}^\infty$, $k \in \mathcal{K}$, and by initializing $\rho \geq \rho_{\min}$ at the start of a new iteration, we can then guarantee the condition $\Delta l \geq \frac{2}{3}\varepsilon\rho$ to hold true. On the other hand, the only assumption that can cause difficulties is the case when the Hessian of the Lagrangian is used to express $\mathbf{W}_k$, and at some iterations the *a priori* bound for $\|\mathbf{W}_k\|$ is not available. This usually happens in the case when the multiplier estimates are not bounded, but as we have discussed earlier, if SMFCQ conditions hold then the multiplier estimates will be bounded. Hence it is very likely that $\|\mathbf{W}_k\|$ will be bounded as well for our case.

References

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, Nonlinear Programming Theory and Algorithms, 2nd Edition, John Wiley and Sons, New York, (1993).

[2] H.Y. Benson, D.F. Shanno and R.J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: filter methods and merit functions*, Technical Report ORFE-00-06, Operations Research and Financial Engineering, Princeton University, (2001).

[3] L.T. Biegler and A. Wächter, *Global and local convergence of line search filter methods for nonlinear programming*, CAPD Technical Report B-01-09, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, (2001).

[4] R.H. Byrd, M. Marazzi and J. Nocedal, *On the convergence of Newton iterations to non-stationary points*, Report OTC 2001/01, Optimization Technology Center, Northwestern University, Evanston, IL, (2001).

[5] C.M. Chin, *A global convergence theory of a filter line search method for nonlinear programming*, Numerical Optimization Report, Department of Statistics, University of Oxford, England, (2003)

[6] C.M. Chin and R. Fletcher, *On the global convergence of an SLP–filter algorithm that takes EQP steps*, Mathematical Programming, **96**(1), (2003), pp. 161 – 177.

[7] R. Fletcher and S. Leyffer, *Filter type algorithms for solving systems of algebraic equations and inequalities*, Numerical Analysis Report NA/204, Department of Mathematics, University of Dundee, Scotland, (2001).

[8] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Mathematical Programming, **91**(1), (2002), pp. 239 – 269.

[9] R. Fletcher and E. Sainz de la Maza, *Nonlinear programming and nonsmooth optimization by successive linear programming*, Mathematical Programming, **43**, (1989), pp. 235 – 256.

[10] R. Fletcher, S. Leyffer and Ph.L. Toint, *On the global convergence of a filter-SQP algorithm*, Numerical Analysis Report NA/197, Department of Mathematics, University of Dundee, (2000).

[11] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming, **12**, (1977), pp. 136 – 138.

[12] J.K. Hartman, *Iterative determination of parameters for an exact penalty function*, Journal of Optimization Theory and Applications, **16**, (1975), pp. 49 – 66.

[13] B.W. Kort and D.P. Bertsekas, *A new penalty function method for constrained minimization*, Proceedings of 1972 IEEE Conference on Decision and Control, (1972), pp. 162 – 166.

[14] J. Kyparisis, *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming*, Mathematical Programming, **32**, (1985), pp. 242 – 246.

[15] O.L. Mangasarian, Nonlinear Programming, McGraw–Hill Book Company, New York, (1969).

[16] M.J.D. Powell, *A hybrid method for nonlinear equations*, In P. Rabinowitz (editor), Numerical Methods for Nonlinear Algebraic Equations, Gordon and Breach, London, (1970), pp. 87 – 114.

[17] R.B. Schnabel, *Determining feasibility of a set of nonlinear inequality constraints*, Mathematical Programming Study, **16**, (1982), pp. 137 – 148.

[18] S. Ulbrich, *On the superlinear local convergence of a filter-SQP method*, Technical Report, Zentrum Mathematik, Technische Universität München, Germany, (2002)