Combining filter and non-monotone trust region algorithm for nonlinear systems of equalities and inequalities

Chao Gu * and Hua Wang

School of Math. and Info., Shanghai LiXin University of Commerce, P.R.China

Abstract   In this paper, we combine filter and non-monotone trust region algorithm for nonlinear systems of equalities and inequalities. The systems of equalities and inequalities are transformed into a continuous equality constrained optimization solved by the new algorithm. Filter method guarantees global convergence of the algorithm under appropriate assumptions. The second order correction step is used to overcome Maratos effect so that superlinearly local convergence is achieved. Preliminary numerical results are reported.

Keywords   Nonlinear systems; Complementarity; Trust region; Filter method; Non-monotone technique; Convergence analysis

AMS 2010 subject classifications 65K05, 90C30, 90C33

DOI: 10.19139/soic.v3i3.100

1. Introduction

Consider the following nonlinear systems of equalities and inequalities

$$c_i(x) = 0, \quad i \in \mathcal{E} \tag{1a}$$
$$c_i(x) \le 0, \quad i \in \mathcal{I} \tag{1b}$$

where $c_i(x)$: $\mathbb{R}^n \to \mathbb{R}$, $i \in \mathcal{E} \bigcup \mathcal{I}$, $\mathcal{E} \bigcap \mathcal{I} = \emptyset$. Systems of nonlinear equalities and inequalities appear in a wide variety of problems. These systems play a central role in the model formulation design and analysis of numerical techniques employed in solving problems arising in optimization, complementarity, and variational inequalities [12, 15, 18, 19, 20, 28, 30]. For example, nonlinear complementarity problem is a special case of (1).

Many researchers considered the problem, especially for the numerical methods such as Dennis, El-Alem and Williamson [4], Macconi, Morini and Porcelli [14], Yang, Chen and Tong [31], Zhang and Huang [34]. But none of them used filter idea. Filter methods were presented by Fletcher and Leyffer [6] for nonlinear programming, offering an alternative to merit functions, as a tool to guarantee global convergence of algorithms for nonlinear optimization. Filter methods were successfully applied to solve various optimization problems [2, 6, 7, 8, 9, 10, 11, 13, 21, 23, 24, 25, 29].

In this paper, we present a new non-monotone filter trust region algorithm which is different from [22, 26, 27]. The new algorithm is based on non-monotone line search technique presented by Zhang and Hager [33]. We transform the problem (1) into an equality constrained optimization problem solved by the new algorithm. The

second order correction step is used to overcome Maratos effect so that superlinearly local convergence is achieved. The paper is outlined as follows. In Section 2, we propose the new non-monotone filter trust region algorithm. Section 3 is devoted to the global convergence theory for the new algorithm. In Section 4, we study the local convergence. Finally, we report some numerical experiments in Section 5.

Notation. $\| \cdot \|$ is the ordinary Euclidean norm in the paper. Let $\nabla\Phi(x)$ denote the gradient of the function $\Phi(x)$, $\nabla c(x)$ denote the Jacobian of the constraint $c(x)$.

2. A new non-monotone filter trust region algorithm

Let the vector functions $C_{\mathcal{E}}(x)$ be the vector functions whose components are $c_i(x)$ for $i \in \mathcal{E}$ and $C_{\mathcal{I}}(x)$ be the vector functions whose components are $c_i(x)$ for $i \in \mathcal{I}$. Given a 0-1 diagonal indicator matrix $W(x) \in \mathbb{R}^{p \times p}$ whose diagonal entries are

$$w_i(x) = \begin{cases} 1, & i \in \mathcal{I} \text{ and } c_i(x) \geq 0, \\ 0, & i \in \mathcal{I} \text{ and } c_i(x) < 0. \end{cases} \tag{2}$$

Now, we define the function

$$\Phi(x) = \frac{1}{2}\|W(x)C_{\mathcal{I}}(x)\|^2, \tag{3}$$

and problem (1) can be written as

$$\min \quad \frac{1}{2}\|W(x)C_{\mathcal{I}}(x)\|^2 \tag{4a}$$

$$\text{subject to} \quad C_{\mathcal{E}}(x) = 0. \tag{4b}$$

In [4], the gradient of the function $\Phi(x)$ can be written as

$$\nabla\Phi(x) = \nabla C_{\mathcal{I}}(x)^T W(x) C_{\mathcal{I}}(x). \tag{5}$$

Following from Byrd [1], Omojokun [17], Dennis, El Alem and Maciel [5], we obtain the trial step $s_k = s_k^t + s_k^n$ at the current iterate x_k by computing a quasi-normal step s_k^n and a tangential step s_k^t. The purpose of the quasi-normal step s_k^n is to improve feasibility. It is obtained as approximate solution of the trust-region subproblem

$$\min \|C_{\mathcal{E}}(x_k) + \nabla C_{\mathcal{E}}(x_k)^T s^n\|^2, \quad s.t. \ \|s^n\| \leq \Delta_k, \tag{6}$$

where the $\Delta_k > 0$ denotes the trust region radius. The requirements on the step s_k^n are that there exist constants $K_1, K_2 > 0$, such that s_k^n admits the upper bound

$$\|s_k^n\| \leq \min\{K_1\|C_{\mathcal{E}}(x_k)\|, \Delta_k\}, \tag{7}$$

and satisfies the decrease condition

$$\|C_{\mathcal{E}}(x_k)\|^2 - \|C_{\mathcal{E}}(x_k) + \nabla C_{\mathcal{E}}(x_k)^T s_k^n\|^2 \geq K_2\|C_{\mathcal{E}}(x_k)\| \min\{\|C_{\mathcal{E}}(x_k)\|, \Delta_k\}. \tag{8}$$

We define a quadratic model

$$q_k(s) = \nabla\Phi_k^T s + \frac{1}{2}s^T H_k s = \nabla C_{\mathcal{I}}(x_k)^T W(x_k) C_{\mathcal{I}}(x_k)s + \frac{1}{2}s^T H_k s \tag{9}$$

of $\Phi(x)$ about the current point x_k, where, H_k is a symmetric approximation of the Hessian of Lagrangian function

$$\ell(x, \lambda) = \Phi(x) + \lambda^T C_{\mathcal{E}}(x) = \frac{1}{2}\|W(x)C_{\mathcal{I}}(x)\|^2 + \lambda^T C_{\mathcal{E}}(x), \tag{10}$$

where λ_k is the multiplier corresponding to equality. Based on the model, the tangential step s_k^t is computed as approximate solution of the trust region subproblem

$$\min q_k(s_k^n + s^t), \quad s.t. \ \nabla C_\mathcal{E}(x_k)^T s^t = 0, \ \|s^t\| \leq \Delta_k, \tag{11}$$

satisfying the decrease condition

$$q_k(s_k^n) - q_k(s_k^n + s_k^t) \geq K_3 \|Z_k^T \nabla q_k(s_k^n)\| \min\{\|Z_k^T \nabla q_k(s_k^n)\|, \Delta_k\} \tag{12}$$

with a constant $K_3 > 0$, where $Z(x)$ denotes a matrix whose columns form a basis of the null space of $\nabla C_\mathcal{E}(x)^T$. Set $D_0 = f(x_0)$, $Q_0 = 1$. To evaluate the descent properties of the step for the objective function, we use the predicted reduction of $\Phi(x)$

$$pred_k = q_k(0) - q_k(s_k)$$

and the non-monotone reduction of $\Phi(x)$

$$ared_k = D_k - \Phi(x_k + s_k),$$

where $\eta_k \in [\eta_{\min}, \eta_{\max}]$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $Q_{k+1} = \eta_k Q_k + 1$, $D_{k+1} = (\eta_k Q_k D_k + f(x_k))/Q_{k+1}$. Observe that D_{k+1} is a convex combination of D_k and $f(x_k)$. Since $D_0 = f(x_0)$, it follows that D_k is a convex combination of the function values $f(x_0), f(x_1), \cdots, f(x_k)$. The choice of η_k controls the degree of nonmonotonicity. If $\eta_k = 0$ for each k, the algorithm actual is a monotone filter trust region algorithm. If $\eta_k = 1$ for each k, we have that $D_k = A_k$, where $A_k = \frac{\sum_{i=0}^{k} f(x_i)}{k+1}$.

The first order necessary optimality conditions at a local solution x^* of (4) can be written as

$$C_\mathcal{E}(x^*) = 0, \quad Z(x^*)^T \nabla \Phi(x^*) = 0.$$

Define the constraint violation $\theta(x) = \|C_\mathcal{E}(x)\|_1$. Set $E_0 = \theta(x_0)$. A trial point $x_k + s_k$ is acceptable if

$$\theta(x_k + s_k) \leq (1 - \gamma_\theta)E_k \tag{13a}$$

or

$$\Phi(x_k + s_k) \leq D_k - \gamma_\Phi \theta(x_k) \tag{13b}$$

holds for $\gamma_\theta, \gamma_\Phi \in (0, 1)$, where

$$E_{k+1} = (\eta_k Q_k E_k + \theta(x_k))/Q_{k+1}.$$

When the condition

$$pred_k \geq \kappa_\theta \theta_k^\psi \tag{14}$$

holds, where $\kappa_\theta > 0$ and $\psi \in (0, \frac{1}{2})$, the trial point $x_k + s_k$ has to satisfy the condition

$$\rho_k = \frac{ared_k}{pred_k} \geq \eta, \tag{15}$$

where $0 < \eta < 1$.

For the sake of a simplified notation we define the filter in this paper not as a list but as a set $\mathcal{F}_k$ containing all (θ, Φ) that are prohibited in iteration k. We say that a trial point $x_k + s_k$ is acceptable to the filter if $(\theta(x_k + s_k), \Phi(x_k + s_k)) \notin \mathcal{F}_k$. At the beginning of the proposed algorithm, the filter is initialized to $\mathcal{F}_0 = \{(\theta, \Phi) \in \mathbb{R}^2 : \theta \geq 10^4, \emptyset\}$. If the accepted trial step size does not satisfy the condition (14), the filter is augmented for a new iteration using the non-monotone update formula

$$\mathcal{F}_{k+1} = \mathcal{F}_k \cup \left\{ (\theta, \Phi) \in \mathbb{R}^2 : \theta \geq (1 - \gamma_\theta)E_k, \ \Phi \geq D_k - \gamma_\Phi \theta(x_k) \right\}. \tag{16}$$

We do not augment the filter if (14) and (15) hold for the accepted step.

The algorithm is described as follows.

Algorithm 1

Given: Starting point x_0; $D_0 = f(x_0)$; $E_0 = \theta(x_0)$; $Q_0 = 1$; initial trust region radius Δ_0; symmetric matrix H_0; $\mathcal{F}_0 = \{(\theta, \Phi) \in \mathbb{R}^2 : (10^4, \emptyset)\}$; γ_θ, $\gamma_\Phi \in (0, 1)$; $\kappa_\theta > 0$; $\psi \in (0, \frac{1}{2})$; $0 < \gamma_0 < \gamma_1 \leq 1 \leq \gamma_2$; $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$; $0 < \eta < 1$; $\Delta^{\min} > 0$; $\epsilon > 0$.

1. Compute $\Phi(x_k), \nabla\Phi(x_k), C_\mathcal{E}(x_k), \theta(x_k)$. If $\|Z(x_k)^T \nabla\Phi(x_k)\| + \theta(x_k) \leq \epsilon$, stop.
2. Compute a step. Compute s_k^n and s_k^t, set $s_k = s_k^n + s_k^t$.
3. Tests to accept the trial step.

    3.1. Check acceptability to the filter. If $(\theta(x_k + s_k), \Phi(x_k + s_k)) \in \mathcal{F}_k$, reject the trial step, go to Step 3.3.
    3.2. Check sufficient decrease with respect to current iterate.

       3.2.1. Case 1. (14) holds: If (15) holds, go to Step 4. Otherwise, go to Step 3.3.
       3.2.2. Case 2. (14) is not satisfied: If (13) holds, go to Step 4. Otherwise, go to Step 3.3.

    3.3. Choose $\Delta_k \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k]$. If $\Delta_k < \Delta^{\min}$, go to Step 7, otherwise go to Step 3.1.

4. Accept trial point. Set $x_{k+1} = x_k + s_k$ and choose Δ_{k+1} such that $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$ if $\rho_k \geq \eta$ holds.
5. Augment filter. If (14) is not satisfied, augment the filter using (16). Otherwise set $\mathcal{F}_{k+1} = \mathcal{F}_k$.
6. Continue with next iteration. Choose $\eta \in [\eta_{\min}, \eta_{\max}]$, and set

$$Q_{k+1} = \eta_k Q_k + 1,$$

$$D_{k+1} = (\eta_k Q_k D_k + f(x_k))/Q_{k+1},$$

$$E_{k+1} = (\eta_k Q_k E_k + \theta(x_k))/Q_{k+1},$$

$k \leftarrow k + 1$, and go back to Step 1.

7. Feasibility restoration phase. Compute a new iterate x_{k+1} by decreasing the infeasibility measure θ so that x_{k+1} satisfies the sufficient decrease conditions (13) and is acceptable to filter. Go to Step 6.

Feasibility restoration phase

Consider the problem

$$\min \quad \theta(x) = \|C_\mathcal{E}(x)\|_1.$$

There exist some methods which can be successfully applied to solve it. For details, see Nie [16] and Yuan [32].

3. Global convergence

Assumptions G. In the following, we denote the set of indices of those iterations in which the filter has been augmented by $\mathcal{A}$.

(G1) $C_\mathcal{E}(x)$ and $C_\mathcal{I}(x)$ are continuously differentiable and bounded on $\mathbb{R}^n$.
(G2) The iterate $\{x_k\}$ remains in compact subset $\mathbb{S} \subset \mathbb{R}^n$.
(G3) $\nabla C_\mathcal{E}(x_k), \nabla C_\mathcal{I}(x_k), (Z(x_k)^T Z(x_k))^{-1}$ and H_k are bounded for all k.

Lemma 1. *Suppose Assumptions G hold. Then*

$$\min\{\theta : (\theta, \Phi) \in \mathcal{F}_k\} > 0. \tag{17}$$

Proof

The proof of (17) is by induction. Since $\mathcal{F}_0 = \{(\theta, \Phi) \in \mathbb{R}^2 : (10^4, \emptyset)\}$, the claim is valid for $k = 0$.

Suppose that the claim is true for k. If $\theta(x_k) > 0$ and the filter is augmented in iteration k, then $\min\{\theta : (\theta, \Phi) \in \mathcal{F}_{k+1}\} > 0$ from the update rule (16) and $E_k \geq \theta(x_k)$. If, on the other hand, $\theta(x_k) = 0$, we have $\|s_k^n\| = 0$ from (7). It follows from (12) that $pred_k = q_k(0) - q_k(s_k) = q_k(s_k^n) - q_k(s_k^n + s_k^t) \geq \kappa_\theta \theta_k^\psi$. So (15) must be satisfied when Δ_k is sufficiently small, i.e., $\mathcal{F}_{k+1} = \mathcal{F}_k$. $\qquad\square$

Lemma 2. *Suppose Assumptions* G *hold. Then algorithm 1 is well defined.*

Proof
The proof is by contradiction. Suppose that the inner loop does not terminate in a finite number of iterations. In this case the current filter or reduction condition will always rejected trial point that leads to $\Delta_k \to 0$. There are two cases.

If $\theta(x_k) = 0$, similar to the proof of lemma 1, $pred_k \geq \kappa_\theta \theta_k^\psi$ and $\|s_k^n\| = 0$ hold. Then the trial point will be accepted by (15) when Δ_k is sufficiently small. From $\theta(x_k) = 0$, $\|s_k^n\| = 0$, $\|\nabla C_{\mathcal{E}}(x_k)^T s_k^t\| = 0$ and the Taylor expansion of $C_{\mathcal{E}}(x)$ we get

$$\theta(x_k + s_k) \leq \xi_1 \|s_k\|^2 \leq 4\xi_1 \Delta_k^2,$$

where ξ_1 is a positive constant. When $\Delta_k \leq \sqrt{\frac{\min\{\theta:(\theta,\Phi)\in\mathcal{F}_k\}}{4\xi_1}}$, the trial size is acceptable to the filter which is a contradiction.

If $\theta(x_k) > 0$, then the algorithm go to the feasibility restoration phase when $\Delta_k < \Delta^{\min}$. Because $\min\{\theta : (\theta, \Phi) \in \mathcal{F}_k\} > 0$, the algorithm of feasibility restoration phase can find a new iterate x_{k+1} acceptable to the current filter and (13) which is a contradiction. $\qquad\square$

Theorem 1. *Suppose that Assumptions* G *hold and the filter is augmented infinitely, i.e., $|\mathcal{A}| = \infty$. Then there exists a subsequence $\{k_i\} \subseteq \mathcal{A}$ such that*

$$\lim_{i\to\infty} \|Z(x_{k_i})^T \nabla\Phi(x_{k_i})\| + \|C_{\mathcal{E}}(x_{k_i})\|_1 = 0. \tag{18}$$

Proof
Similar to the proof of Lemma 4 in [10], we get

$$\lim_{i\to\infty} \|C_{\mathcal{E}}(x_{k_i})\|_1 = 0, \quad k_i \in \mathcal{A}.$$

By Assumption G2 and $|\mathcal{A}| = \infty$, there exists an accumulation point $\bar{x}$, i.e., $\lim_{i\to\infty} x_{k_i} = \bar{x}$, $k_i \in \mathcal{A}$. If $\lim_{i\to\infty} \|Z(x_{k_i})^T \nabla\Phi(x_{k_i})\| = 0$, then $\bar{x}$ is a KKT point. Otherwise, there exist a subsequence $\{x_{k_{i_j}}\}$ of $\{x_{k_i}\}$ and a constant $\epsilon > 0$ so that for all k_{i_j}

$$\|Z(x_{k_{i_j}})^T \nabla\Phi(x_{k_{i_j}})\| \geq \epsilon.$$

By the choice of $\{x_{k_{i_j}}\}$, we have

$$k_{i_j} \in \mathcal{A} \text{ for all } k_{i_j}, \tag{19}$$

and

$$\begin{aligned}
\|Z_{k_{i_j}}^T \nabla q_{k_{i_j}}(s_{k_{i_j}}^n)\| &= \|Z_{k_{i_j}}^T (\nabla\Phi(x_{k_{i_j}}) + H_{k_{i_j}} s_{k_{i_j}}^n)\| \\
&\geq \|Z_{k_{i_j}}^T \nabla\Phi(x_{k_{i_j}})\| - \|Z_{k_{i_j}}^T H_{k_{i_j}} s_{k_{i_j}}^n\| \\
&\geq \epsilon - v_1 \|s_{k_{i_j}}^n\| \geq \epsilon - \tilde{v}_1 \|C_{\mathcal{E}}(x_{k_{i_j}})\|,
\end{aligned}$$

where v_1 and $\tilde{v}_1$ are positive constants. Since $\lim_{j\to\infty} \|C_{\mathcal{E}}(x_{k_{i_j}})\| = 0$, there exists $K_4 \in \mathbb{N}$ such that for $j \geq K_4$

$$\|Z_{k_{i_j}}^T \nabla q_{k_{i_j}}(s_{k_{i_j}}^n)\| \geq \frac{\epsilon}{2}. \tag{20}$$

By (9), (12) and (20), we obtain

$$
\begin{aligned}
& pred_{k_{i_j}} - \kappa_\theta \theta_{k_{i_j}}^\psi \\
= \ & q_{k_{i_j}}(0) - q_{k_{i_j}}(s_{k_{i_j}}) - \kappa_\theta \theta_{k_{i_j}}^\psi \\
= \ & q_{k_{i_j}}(0) - q_{k_{i_j}}(s_{k_{i_j}}^n) + q_{k_{i_j}}(s_{k_{i_j}}^n) - q_{k_{i_j}}(s_{k_{i_j}}) - \kappa_\theta \theta_{k_{i_j}}^\psi \\
\geq \ & \nabla \Phi_{k_{i_j}}^T s_{k_{i_j}}^n + \frac{1}{2}(s_{k_{i_j}}^n)^T H_{k_{i_j}} s_{k_{i_j}}^n + K_3 \|Z_{k_{i_j}}^T \nabla q_{k_{i_j}}(s_{k_{i_j}}^n)\| \min\{\|Z_{k_{i_j}}^T \nabla q_{k_{i_j}}(s_{k_{i_j}}^n)\|, \Delta_{k_{i_j}}\} \\
& - \kappa_\theta \theta_{k_{i_j}}^\psi \\
\geq \ & \frac{\epsilon}{2} K_3 \min\{\frac{\epsilon}{2}, \Delta^{\min}\} - v_2 \|s_{k_{i_j}}^n\| - \kappa_\theta \theta_{k_{i_j}}^\psi,
\end{aligned}
$$

where v_2 is a positive constant. When j is sufficiently large, $\|s_{k_{i_j}}^n\| \to 0$ and $\theta_{k_{i_j}} \to 0$ hold, therefore the condition (14) is satisfied for sufficiently large j. The reason for accepting the step s_k must have been that (15) is satisfied. Consequently, the filter is not augmented in iteration k_{i_j} which contradicts to (19). $\square$

Theorem 2. *Suppose that Assumptions* G *hold and the filter is augmented finitely, i.e.,* $|\mathcal{A}| < \infty$. *Then*

$$
\lim_{k \to \infty} \|Z(x_k)^T \nabla \Phi(x_k)\| + \|C_\mathcal{E}(x_k)\|_1 = 0.
$$

Proof
First, we will show

$$
\lim_{k \to \infty} \theta(x_k) = 0. \tag{21}
$$

From assumptions there exists $K \in \mathbb{N}$ so that for all iterations $k \geq K$ the filter is not augmented, i.e., $k \notin \mathcal{A}$ for all $k \geq K$. We then have that for all $k \geq K$ condition (15) is satisfied, i.e.,

$$
\Phi(x_{k+1}) \leq D_k - \eta pred_k.
$$

With above results we have

$$
\begin{aligned}
D_{k+1} & = \frac{\eta_k Q_k D_k + \Phi(x_{k+1})}{Q_{k+1}} \\
& \leq \frac{\eta_k Q_k D_k + D_k - \eta pred_k}{Q_{k+1}} \\
& = D_k - \frac{\eta pred_k}{Q_{k+1}} \leq D_k - \frac{\eta pred_k}{1 - \eta_{\max}}.
\end{aligned} \tag{22}
$$

It is similar to Lemma 1.1 in [33] that if $pred_k > 0$, we have $\Phi_k \leq D_k$. Since $\{\Phi_k\}$ is bounded below and $\Phi_k \leq D_k$ as $k \to \infty$, we get that D_k is bounded below. It follows from (22) that

$$
\sum_{r=K}^{k} pred_r < \infty.
$$

It follows from $pred_k \geq \kappa_\theta \theta_k^\psi$ that (21) holds, i.e.,

$$
\lim_{k \to \infty} \|C_\mathcal{E}(x_k)\|_1 = 0. \tag{23}
$$

Now let $\bar{x}$ be any limit point of $\{x_k\}$. If $\lim_{i \to \infty} \|Z(x_k)^T \nabla \Phi(x_k)\| = 0$, then $\bar{x}$ is a KKT point. Otherwise, there exist a subsequence $\{x_{k_i}\}$ of $\{x_k\}$ and a constant $\epsilon > 0$ so that for all k_i

$$
\|Z(x_{k_i})^T \nabla \Phi(x_{k_i})\| \geq \epsilon.
$$

As in the proof of Theorem 1, we obtain that

$$\|Z_{k_i}^T \nabla q_{k_i}(s_{k_i}^n)\| \geq \frac{\epsilon}{2}. \tag{24}$$

Notice that

$$pred_{k_i} = q_{k_i}(0) - q_{k_i}(s_{k_i}) = q_{k_i}(0) - q_{k_i}(s_{k_i}^n) + q_{k_i}(s_{k_i}^n) - q_{k_i}(s_{k_i}).$$

Moreover, $pred_{k_i} \to 0$, $\|s_{k_i}^n\| \to 0$ and

$$|q_{k_i}(0) - q_{k_i}(s_{k_i}^n)| \leq v_3\|s_{k_i}^n\| + v_4\|s_{k_i}^n\|^2,$$

which in turn yields that

$$\lim_{i \to \infty} \|q_{k_i}(s_{k_i}^n) - q_{k_i}(s_{k_i})\| = 0, \tag{25}$$

where v_3 and v_4 are positive constants. We also deduce from (12) and (24) that

$$q_{k_i}(s_{k_i}^n) - q_{k_i}(s_{k_i}) \geq \frac{\epsilon}{2}K_3 \min\{\frac{\epsilon}{2}, \Delta_{k_i}\}. \tag{26}$$

From (25) and (26), it is easy to see that $\lim_{i \to \infty} \Delta_{k_i} = 0$. Thus, there exists $K_5 \in \mathbb{N}$ such that $\Delta_{k_i} < \Delta^{\min}$ is satisfied for $i \geq K_5$, then the algorithm go to the feasibility restoration phase, i.e., $k_i \in \mathcal{A}$, which contradicts to the choice of k_i. $\qquad\square$

4. Local convergence

Fletcher and Leyffer comment that filter method, similar to penalty function method, can suffer from Maratos effect. This results in poor local convergence behavior. In order to overcome the Maratos effect, we modify the non-monotone filter trust region algorithm by introducing second order correction step, which is denoted by s_k^{soc}. Similar to Chin, Rashid and Nor [2], we calculate the second order correction step s_k^{soc} by solving a modified QP subproblem

$$\min \quad \nabla\Phi(x_k)^T(s_k + s_k^{soc}) + \frac{1}{2}(s_k + s_k^{soc})^T H_k(s_k + s_k^{soc}) \tag{27a}$$

$$\text{subject to} \quad c_i(x_k + s_k) + \nabla c_i(x_k)^T s_k^{soc} = -\|s_k\|^v, \ \ i \in \mathcal{E}. \tag{27b}$$

where $v \in (2, 3)$. Chin et al. prove that $\|s_k^{soc}\| = O(\|s_k\|^2)$. A second order correction step is tried whenever the full SQP step $x_k + s_k$ has not been accepted.

We now state the new non-monotone trust region algorithm with second order correction step.

Algorithm 2

Given: Starting point x_0; $D_0 = f(x_0)$; $E_0 = \theta(x_0)$; $Q_0 = 1$; initial trust region radius Δ_0; symmetric matrix H_0; $\mathcal{F}_0 = \{(\theta, \Phi) \in \mathbb{R}^2 : (10^4, \emptyset)\}$; $\gamma_\theta, \gamma_\Phi \in (0, 1)$; $\kappa_\theta > 0$; $\psi \in (0, \frac{1}{2})$; $0 < \gamma_0 < \gamma_1 \leq 1 \leq \gamma_2$; $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$; $0 < \eta < 1$; $\Delta^{\min} > 0$; $\epsilon > 0$.

1. Compute $\Phi(x_k)$, $\nabla\Phi(x_k)$, $C_{\mathcal{E}}(x_k)$, θ_k. If $\|Z(x_k)^T\nabla\Phi(x_k)\| + \theta(x_k) \leq \epsilon$, stop.
2. Compute a step. Compute s_k^n and s_k^t, set $s_k = s_k^n + s_k^t$.
3. Tests to accept the trial step.

   3.1. Check acceptability to the filter. If $x_k + s_k \in \mathcal{F}_k$, reject the trial step, go to Step 3.3.
   3.2. Check sufficient decrease with respect to current iterate.

      3.2.1. Case 1. (14) holds: If (15) holds, set $x_{k+1} = x_k + s_k$ and go to Step 4. Otherwise, go to Step 3.3.
      3.2.2. Case 2. (14) is not satisfied: If (13) holds, set $x_{k+1} = x_k + s_k$ and go to Step 4. Otherwise, go to Step 3.3.

   3.3. Compute second order correction step s_k^{soc} and $\hat{x}_{k+1} = x_k + s_k + s_k^{soc}$.

3.4. Check acceptability to the filter. If $\hat{x}_{k+1} \in \mathcal{F}_k$, reject s_k^{soc}, go to Step 3.6.

3.5. Check sufficient decrease with respect to current iterate.

3.5.1. Case 1. (14) holds: If

$$D_k - \Phi(\hat{x}_{k+1}) \geq \eta_1 pred_k \tag{28}$$

holds, set $x_{k+1} = \hat{x}_{k+1}$ and go to Step 4.

3.5.2. Case 2. (14) is not satisfied: If

$$\theta(\hat{x}_{k+1}) \leq (1 - \gamma_\theta)E_k$$

or

$$\Phi(\hat{x}_{k+1}) \leq D_k - \gamma_\Phi \theta_k$$

holds, set $x_{k+1} = \hat{x}_{k+1}$ and go to Step 4.

3.6. Choose $\Delta_k \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k]$. If $\Delta_k < \Delta^{\min}$, go to Step 7, otherwise go to Step 3.1.

4. Choose Δ_{k+1} such that $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$ if $\rho_k \geq \eta$ and (14) holds.

5. Augment filter. If (14) is not satisfied, augment the filter using (16). Otherwise set $\mathcal{F}_{k+1} = \mathcal{F}_k$.

6. Continue with next iteration. Choose $\eta \in [\eta_{\min}, \eta_{\max}]$, and set

$$Q_{k+1} = \eta_k Q_k + 1,$$

$$D_{k+1} = (\eta_k Q_k D_k + f(x_k))/Q_{k+1},$$

$$E_{k+1} = (\eta_k Q_k E_k + \theta(x_k))/Q_{k+1},$$

$k \leftarrow k + 1$, and go back to Step 1.

7. Feasibility restoration phase. Compute a new iterate x_{k+1} by decreasing the infeasibility measure θ so that x_{k+1} satisfies the sufficient decrease conditions (13) and is acceptable to filter. Go to Step 6.

The second order correction step does not affect the global convergence properties [2, 29]. In order to analyze the local convergence rate of the proposed algorithm, more assumptions are needed.

Assumptions H

(H1) x_* is a KKT point of problem (4), i.e.,

$$C_\mathcal{E}(x_*) = 0 \quad \text{and} \quad Z(x_*)^T \nabla \Phi(x_*) = 0.$$

(H2) There exists a constant $\tilde{\tau} > 0$ such that

$$p^T(Z_*^T \nabla_{xx}^2 \mathcal{L}(x_*, \lambda_*)Z_*)p \geq \tilde{\tau}\|p\|^2, \quad \forall p \in \mathbb{R}^t$$

i.e., the second order sufficient condition holds at (x_*, λ_*).

(H3) $x_k \rightarrow x_*$.

(H4) $(V_k - H_k)d_k = O(\|d_k\|^2)$, where $V_k = \nabla_{xx}^2 \mathcal{L}(x_*, \lambda_*)$.

We start convergence analysis by the penalty function

$$\phi_\rho(x) = \Phi(x) + \rho\|C_\mathcal{E}(x)\|_1, \tag{30}$$

and the following model of the penalty function $\phi_\rho(x)$:

$$h_\rho(x_k, d) = \Phi(x_k) + \nabla\Phi(x_k)^T s + \frac{1}{2}s^T H_k s + \rho\|\nabla C_\mathcal{E}(x)^T s + C_\mathcal{E}(x_k)\|_1. \tag{31}$$

Note that we employ the exact penalty function only as a technical device, but the algorithm never refers to it.

Lemma 3. *Suppose Assumptions* H *hold. Then whenever condition (14) holds, the condition (28) is satisfied for sufficiently large* k.

Proof
Since (14) holds and $\psi \in (0, \frac{1}{2})$, we have

$$\theta_k \overset{(14)}{<} \kappa_\theta^{-\frac{1}{\psi}}(pred_k)^{\frac{1}{\psi}} = \kappa_\theta^{-\frac{1}{\psi}}(q_k(0) - q_k(s_k))^{\frac{1}{\psi}} = \kappa_\theta^{-\frac{1}{\psi}}(\nabla\Phi_k^T s_k + \frac{1}{2}s_k^T H_k s_k)^{\frac{1}{\psi}} = o(\|s_k\|^2). \tag{32}$$

Before continuing, we recall the step decomposition from the paper [29].

$$s_k = q_k + p_k, \tag{33}$$
$$q_k = Y_k\bar{q}_k, \quad \text{and} \quad p_k = Z_k\bar{p}_k, \tag{34}$$
$$\bar{q}_k = -[\nabla C_\mathcal{E}(x_k)^T Y_k]^{-1} C_\mathcal{E}(x_k), \tag{35}$$
$$\bar{p}_k = -[Z_k^T H_k Z_k]^{-1} Z_k^T (\nabla\Phi_k + H_k q_k). \tag{36}$$

It follows from the orthonormality of $[\begin{array}{cc} Y_k & Z_k \end{array}]$ and (32) that

$$\|q_k\| = o(\|\bar{p}_k\|^2) \quad \text{and} \quad \|s_k\| = O(\|\bar{p}_k\|). \tag{37}$$

Similar to the proof of Lemma 4.4 in [29] we can get the following result with (32). Suppose Assumptions H hold, then for sufficiently large k

$$\Phi(x_k) - \Phi(x_k + s_k + s_k^{soc}) \geq -\left(\frac{1}{2} + \eta_\Phi\right)\left(\nabla\Phi(x_k)^T s_k + \frac{1}{2}s_k^T H_k s_k\right) + o(\|s_k\|^2) \tag{38}$$

holds with $\rho > \|\lambda_k^+\|_D$ and $\eta_\Phi \in (0, \frac{1}{2})$, where $\|\cdot\|_D$ is the dual norm to $\|\cdot\|$. From Assumption H2, we have for sufficiently large k

$$\bar{p}_k^T (Z_k^T H_k Z_k)\bar{p}_k \geq \frac{\tilde{\tau}}{2}\|\bar{p}_k\|^2.$$

Therefore

$$
\begin{aligned}
& D_k - \Phi(x_k + s_k + s_k^{soc}) - \eta pred_k \\
\geq\ & \Phi_k - \Phi(x_k + s_k + s_k^{soc}) - \eta pred_k \\
\overset{(37),(38)}{\geq}\ & -\left(\frac{1}{2} + \eta_\Phi + \eta\right)\left(\nabla\Phi(x_k)^T s_k + \frac{1}{2}s_k^T H_k s_k\right) + o(\|\bar{p}_k\|^2) \\
\overset{(33),(34)}{=}\ & -\left(\frac{1}{2} + \eta_\Phi + \eta\right)\nabla\Phi(x_k)^T Z_k\bar{p}_k - \frac{1}{2}\left(\frac{1}{2} + \eta_\Phi + \eta\right)\bar{p}_k^T Z_k^T H_k Z_k\bar{p}_k + o(\|\bar{p}_k\|^2) \\
\overset{(36)}{=}\ & \frac{1}{2}(\frac{1}{2} + \eta_\Phi + \eta)\bar{p}_k^T Z_k^T H_k Z_k\bar{p}_k + o(\|\bar{p}_k\|^2), \\
\geq\ & \frac{1}{4}(\frac{1}{2} + \eta_\Phi + \eta)\tilde{\tau}\|\bar{p}_k\|^2 + o(\|\bar{p}_k\|^2).
\end{aligned}
\tag{39}
$$

Since $\frac{1}{4}(\frac{1}{2} + \eta_\Phi + \eta)\tilde{\tau} > 0$ and $\|\bar{p}_k\| \to 0$ as $k \to \infty$, (28) holds for sufficiently large k. □

Lemma 4. *Suppose Assumptions* H *hold. Let* $\phi_\rho(x)$ *be the exact penalty function with* $\rho > \|\lambda_*\|_D$, *where* $\|\cdot\|_D$ *is the dual norm to* $\|\cdot\|$. *Then*

$$\lim_{k\to\infty} \frac{\phi_\rho(x_k) - \phi_\rho(x_k + s_k + s_k^{soc})}{h_\rho(x_k, 0) - h_\rho(x_k, s_k)} = 1. \tag{40}$$

Proof
The result follows from Theorem 15.3.7 in [3]. □

Theorem 3. *Suppose Assumptions* H *hold. Then the full SQP step $x_{k+1} = x_k + s_k$ or $x_{k+1} = x_k + s_k + s_k^{soc}$ is acceptable to Algorithm 2 for sufficiently large k.*

Proof

Suppose that x_k is acceptable to the filter and the full step $x_{k+1} = x_k + s_k$ is rejected by the filter. We will show that for sufficient large k, $x_{k+1} = x_k + s_k + s_k^{soc}$ is acceptable to the filter. The initialization of the filter and the update rule imply that for all k the filter has the following property

$$(\tilde{\Phi}, \tilde{h}) \notin \mathcal{F}_k \Rightarrow (\Phi, h) \notin \mathcal{F}_k \quad \text{if } \Phi \leq \tilde{\Phi} \text{ and } h \leq \tilde{h}. \tag{41}$$

Therefore, we only need to show that $\Phi(x_k + s_k + s_k^{soc}) \leq \Phi(x_k)$ and $h(x_k + s_k + s_k^{soc}) \leq h(x_k)$. By the construction of Algorithm we have $h(x_k) \to 0$. So we will show that $\Phi(x_k + s_k + s_k^{soc}) \leq \Phi(x_k) - \gamma_\Phi \theta(x_k)$. From (27b) we have

$$
\begin{aligned}
c_i(x_k + s_k + s_k^{soc}) &= c_i(x_k + s_k) + \nabla c_i(x_k)^T s_k^{soc} + \frac{1}{2} s_k^T \nabla^2 c_i(w_i) s_k^{soc} \\
&\quad + \frac{1}{2}(s_k^{soc})^T \nabla^2 c_i(z_i) s_k^{soc} \\
&= -\|s_k\|^v + O(\|s_k\|^3) = o(\|s_k\|^2),
\end{aligned}
$$

where w_i is between x_k and $x_k + s_k$, z_i is between $x_k + s_k$ and $x_k + s_k + s_k^{soc}$, $v \in (2,3)$. Thus, $\theta(x_k + s_k + s_k^{soc}) = o(\|s_k\|^2)$. It follows from Lemma 4, (30), (31), (33)-(36) that for sufficiently large k

$$
\begin{aligned}
&D_k - \Phi(x_k + s_k + s_k^{soc}) - \gamma_\Phi \theta(x_k) \\
\geq\ & \Phi(x_k) - \Phi(x_k + s_k + s_k^{soc}) - \gamma_\Phi \theta(x_k) \\
=\ & \phi_\rho(x_k) - \phi_\rho(x_k + s_k + s_k^{soc}) - \rho(\theta(x_k + s_k + s_k^{soc}) - \theta(x_k)) - \gamma_\Phi \theta(x_k) \\
\geq\ & (\frac{1}{2} + \eta_\Phi)(h_\rho(x_k, 0) - h_\rho(x_k, s_k)) + (\rho - \gamma)\|C_\mathcal{E}(x_k)\|_1 + o(\|s_k\|^2) \\
=\ & \left(\frac{1}{2} + \eta_\Phi\right)\left(\rho\|C_\mathcal{E}(x_k)\|_1 - \nabla\Phi(x_k)^T s_k - \frac{1}{2} s_k^T H_k s_k\right) + (\rho - \gamma)\|C_\mathcal{E}(x_k)\|_1 + o(\|s_k\|^2) \\
=\ & \frac{1}{2}\left(\frac{1}{2} + \eta_\Phi\right)\bar{p}_k^T Z_k^T H_k Z_k \bar{p}_k + \left(\left(\frac{3}{2} + \eta_\Phi\right)\rho - \gamma\right)\|C_\mathcal{E}(x_k)\|_1 + o(\|\bar{p}_k\|^2) + o(\|C_\mathcal{E}(x_k)\|_1) \\
\geq\ & \frac{1}{4}\left(\frac{1}{2} + \eta_\Phi\right)\tilde{\tau}\|\bar{p}_k\|^2 + \zeta\|C_\mathcal{E}(x_k)\|_1 + o(\|\bar{p}_k\|^2) + o(\|C_\mathcal{E}(x_k)\|_1) \\
\geq\ & 0, \tag{42}
\end{aligned}
$$

where $\zeta > 0$, $\eta_\Phi \in (0, \frac{1}{2})$. Thus, the trial point $x_k + s_k + s_k^{soc}$ is acceptable to the filter $\mathcal{F}_k$ and (13). This means that the full SQP step $x_{k+1} = x_k + s_k$ or $x_{k+1} = x_k + s_k + s_k^{soc}$ is acceptable to Algorithm 2 for sufficiently large k. Therefore the sequence x_k converges to x_* superlinearly [3]. □

5. Numerical experiments

In this section we present the numerical results of Algorithm 1 on HP i5 personal computer with 4G memory. The selected parameter values are: $\gamma_\theta = 0.5$, $\gamma_\Phi = 0.5$; $\kappa_\theta = 10^{-4}$; $\psi = 0.3$; $\gamma_0 = 0.1$; $\gamma_1 = 0.5$; $\gamma_2 = 2$; $\eta_{\min} = \frac{1}{4}$; $\eta_{\max} = \frac{3}{4}$; $\eta = 0.9$; $\Delta^{\min} = 10^{-3}$; $\epsilon = 10^{-6}$. The computation terminates when stopping criterion $\|Z(x_k)^T \nabla\Phi(x_k)\| + \theta(x_k) \leq \epsilon$ is satisfied. The test problems are selected from [34].

NIT, NΦ and NG stand for the numbers of iteration, function evaluations and gradient evaluations, respectively. Let Error=$\Phi(x_k) + \theta_k$. The results of the experiments are shown in Table 1. The preliminary numerical results demonstrate that the new algorithm discussed in this paper is effective for solving this class of problems.

Table 1: Numerical results

Ex.		Algorithm 1				
	x_0	NIT	NΦ	NG	Error	Approximate solution x^*
1	$\begin{pmatrix} 0 \\ 5 \\ 0 \end{pmatrix}$	6	6	6	8.9990e-007	$\begin{pmatrix} 3.6838e-005 \\ 1.1866e+000 \\ -4.0797e-001 \end{pmatrix}$
2	$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$	4	4	4	2.2584e-010	$\begin{pmatrix} -5.1065e-001 \\ -1.0267e-003 \\ 4.8875e-001 \end{pmatrix}$
3	$\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$	4	4	4	2.2457e-008	$\begin{pmatrix} 2.9053e+000 \\ 1.5687e+000 \\ -2.3413e-001 \end{pmatrix}$
4	$\begin{pmatrix} 0.5 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$	3	3	3	2.0653e-007	$\begin{pmatrix} 5.0040e-001 \\ 2.0535e+000 \\ 9.9980e-001 \\ 3.8152e-002 \\ 1.0085e-003 \end{pmatrix}$
5	$\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$	5	5	5	1.5642e-007	$\begin{pmatrix} -7.7668e-001 \\ -1.6889e+000 \\ 1.3461e+000 \end{pmatrix}$

6. Acknowledgement

REFERENCES

1. R.H.Byrd, *Robust trust region methods for constrained optimization*, in: Third SIAM Conference on Optimization, Houston, Texas, May, 1987.
2. C.M.Chin, A.H.A.Rashid, K.M.Nor, *Global and local convergence of a filter line search method for nonlinear programming*, Optimization Methods and Software. 22 (2006) 365-390.
3. A.R.Conn, N.I.M.Gould, Ph.L.Toint, *Trust Region Methods*, MPS/SIAM Ser. Optim. 1, SIAM, Philadelphia, 2000.
4. J.E.Dennis, M.El-Alem, K.Williamson, *A trust-region approach to nonlinear systems of equalities and inequalities*, SIAM J. Optim. 9 (1999) 291-315.
5. J.E.Dennis, M.EI-Alem, M.C.Maciel, *A global convergence theory for general trust region based algorithms for equality constrained optimization*, SIAM J. Optim. 7 (1997) 177-207.
6. R.Fletcher, S.Leyffer, *Nonlinear programming without a penalty function*, Math. Program. 91 (2002) 239-269.
7. R.Fletcher, S.Leyffer, P.L.Toint, *On the global convergence of a filter-SQP algorithm*, SIAM J. Optim. 13 (2002) 44-59 .
8. R.Fletcher, N.I.M.Gould, S.Leyffer, P.L.Toint, A.Wächter, *Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming*, SIAM J. Optim. 13 (2002) 635-659.
9. C.Gu, D.Zhu, *A filter interior-point algorithm with projected Hessian updating for nonlinear optimization*, J. Appl. Math. Comput. 29 (2009) 67-80.
10. C.Gu, D.Zhu, *A filter secant method with nonmonotone line search for equality constrained optimization*, J. Syst. Sci. Complex 23 (2010) 846-860.
11. C.Gu, D.Zhu, *A non-monotone line search multidimensional filter-SQP method for general nonlinear programming*, Numer. Algor. 56 (2011) 537-559.
12. B.He, H.Yang, C.Zhang, *A modified augmented Lagrangian method for a class of monotone variational inequalities*, European J. Oper. Res. 159 (2004) 35-51.
13. C.J.Li, W.Y.Sun, *On filter-successive linearization methods for nonlinear semidefinite programming*, Sci China Ser A 52 (2009) 2341-2361.

14. M.Macconi, B.Morini, M.Porcelli, *Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities*, Appl. Numer. Math. 59 (2009) 859-876.
15. A.Mohamed, *A weighted full-Newton step primal-dual interior point algorithm for convex quadratic optimization*, Statistics, Optimization and Information Computing 2 (2014) 21-32.
16. P.Y.Nie, C.F.Ma, *A trust region filter mehtod for general nonlinear programming*, Appl. Math. Comput. 172 (2006) 1000-1017.
17. E.O.Omojokun, *Trust region algorithms for optimization with nonlinear equality and inequality constraints*, Ph.D. Thesis, University of Colorado, Boulder Colorado, USA, 1989.
18. L.Qi, D.Sun, G.Zhou, *A new look at smoothing newton methods for nonlinear complementarity problems and box constrained variational inequality problems*, Math. Program. 87 (2000) 1-35.
19. L.Qi, Y.Yang, *NCP functions applied to Lagrangian globalization for the nonlinear complementarity problem*, Journal of Global Optimization 24 (2) (2002) 261-283.
20. S.P.Rui, C.X.Xu, *A smoothing inexact Newton method for nonlinear complementarity problems*, J. Comput. Appl. Math. 233 (2010) 2332-2338.
21. W.Sun, Y.Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer, New York, 2006.
22. W.Sun, *Non-monotone trust region method for optimization*, Appl. Math. Comput. 156 (2004) 159-174.
23. W.Sun, D.Xu, *A filter trust-region method based on conic model for unconstrained optimization*, Science China Mathematics, 55 (2012) 527-543.
24. Y,Chen, W.Sun, *A dwindling filter line search method for unconstrained optimization*, Mathematics of Compuation, 84 (2015) 187-208.
25. K.Su, Z.Yu, *A modified SQP method with nonmonotone technique and its global convergence*, Comput. Math. Appl. 57 (2009) 240-247.
26. K.Su, D.G.Pu, *A nonmonotone filter trust region method for nonlinear constrained optimization*, J. Comput. Appl. Math. 223 (2009) 230-239.
27. H.Wang, D.G.Pu, *A nonmonotone filter trust region method for the system of nonlinear equations*, Appl. Math. Model. 37 (2013) 498-506.
28. Y.J.Wang, D.Zhu, *An affine scaling interior trust region method via optimal path for solving monotone variational inequality problem with linear constraints*, Chin. Ann. Math. 29B (2008) 273-290.
29. A.Wächter, L.T.Biegler, *Line search filter methods for nonlinear programming: Local convergence*, SIAM J. Optim. 6: 32-48 (2005).
30. J.Wu, G.Yu, *On the Convergence and O (1/N) Complexity of a Class of Nonlinear Proximal Point Algorithms for Monotonic Variational Inequalities*, Statistics, Optimization and Information Computing, 2 (2014) 105-113.
31. L.Yang, Y.Chen, X.Tong, *Smoothing newton-like method for the solution of nonlinear systems of equalities and inequalities*, Numer. Math. Theor. Meth. Appl. 2 (2009) 224-236.
32. Y.Yuan, *Trust region algorithms for nonlinear programming*, in Computational Mathematics in China, Contemp. Math. 163, Z. C. Shi, ed., AMS, Providence, RI, 1994, 205-225.
33. H.Zhang, W.W.Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim. 14 (2004) 1043-1056.
34. Y.Zhang, Z.Huang, *A nonmonotone smoothing-type algorithm for solving a system of equalities and inequalities*, J. Comput. Appl. Math. 233 (2010) 2312-2321.