SCIENCE CHINA
Mathematics

• ARTICLES •

June 2014 Vol. 57 No. 6: 1287–1302
doi: 10.1007/s11425-013-4677-y

On convergence analysis of a derivative-free trust region algorithm for constrained optimization with separable structure

XUE Dan & SUN WenYu*

School of Mathematical Sciences, Jiangsu Key Labratory for NSLSCS, Nanjing Normal University, Nanjing 210046*, China*
Email: wtxuedan@126.com, wysun@njnu.edu.cn

Received May 14, 2013; accepted June 15, 2013; published online June 25, 2013

Abstract In this paper, we propose a derivative-free trust region algorithm for constrained minimization problems with separable structure, where derivatives of the objective function are not available and cannot be directly approximated. At each iteration, we construct a quadratic interpolation model of the objective function around the current iterate. The new iterates are generated by minimizing the augmented Lagrangian function of this model over the trust region. The filter technique is used to ensure the feasibility and optimality of the iterative sequence. Global convergence of the proposed algorithm is proved under some suitable assumptions.

Keywords constrained optimization, derivative-free optimization, multivariate interpolation, separable optimization, global convergence

MSC(2010) 65K05, 90C30

Citation: Xue D, Sun W Y. On convergence analysis of a derivative-free trust region algorithm for constrained optimization with separable structure. Sci China Math, 2014, 57: 1287–1302, doi: 10.1007/s11425-013-4677-y

1 Introduction

In this paper, we consider the separable optimization problem

$$\begin{aligned} \min \quad & f_1(x) + f_2(y) \\ \text{s.t.} \quad & c(x,y) = 0, \end{aligned} \tag{1.1}$$

where $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^s \to \mathbb{R}$ are continuously differentiable functions, but none of their first-order or second-order derivatives is explicitly available,

$$c(x,y) = 0$$

is a general constraint and $c(x,y)$ is continuously differentiable.

The problem (1.1) has numerous applications in compressive sensing, signal/image processing and statistics, etc. When the derivatives of f_1 and f_2 are available, several numerical methods have been studied extensively, such as alternating projection method [31] and alternating direction method [13, 14, 16].

*Corresponding author

However, when the derivatives of f_1 and f_2 are unavailable, this problem belongs to derivative-free nonlinear separable optimization. It seems that now there are not available suitable methods for solving it. The aim of this paper is to propose a derivative-free trust region method for solving (1.1).

Existing derivative-free optimization (DFO) methods can be divided into four different classes. The first class of DFO algorithms is the direct search or pattern search methods which are based on the exploration of the variable space by using sample points from a class of geometric patterns and using either the Nelder-Mead simplex algorithm [21] or parallel direct search algorithm [1, 32]. Some of these methods do not assume the smoothness of the objective function and therefore they can be useful for non-smooth optimization problems. However, a large number of function evaluations are required. The second class of methods are line search methods which consist of a sequence of $n + 1$ one-dimensional searches introduced by Powell [22], Lucidi and Sciandrone [18], and Grippo et al. [15]. The third class of methods is to resort to finite difference approximation of the derivatives (see [10]). In general, it can be much too expensive to evaluate the Hessian or gradient of the objective functions. One can use quasi-Newton Hessian approximation instead. However, this class of methods often cause disastrous loss of accuracy in pathological cases when the finite differences are used to approximate derivatives. So, usually, it is not robust, especially in the presence of noise. The last class of the methods is based on modeling the objective function by multivariate interpolation. These methods have been pioneered by Winfield [33] and Powell [23]. The derivative-free trust-region methods building a model by polynomial interpolation have been developed by [6, 8, 24, 25, 30]. The global convergence properties are established (see [7–9]). Some recent researches [20, 35] indicate that the fourth class method, i.e., the derivative-free trust region model-based method, is frequently superior compared with the first three classes of methods, even for noisy and piecewise-smooth problems. Nevertheless, the first three approaches are still widely used in the engineering community. There are still several disadvantages and challenging issues in the fourth class of methods. The algorithm discussed in this paper belongs to the fourth class of methods.

Some authors developed methods by using the problem structure in order to treat the special optimization problems. Colson and Toint [3] exploit band structure in unconstrained derivative-free optimization problems. Colson and Toint [4] handle partially separable unconstrained optimization problems by using derivative-free trust-region strategy. In this paper, we consider the separable optimization problems with general constraints by using the derivative-free technique. Our work is an extension of the derivative-free optimization methods from unconstrained separable optimization to constrained separable optimization.

Fletcher and Leyffer [12] proposed a filter method for nonlinear constrained optimization. The idea is referred to the concept of filter and is motivated by taking aim of avoiding use of penalty function and penalty parameters whose choice and adjustment may generate numerical difficulty. Now, the filter-type methods are used widely in unconstrained and constrained optimization, for example, filter-trust region methods for unconstrained optimization problems [19, 28], filter-SLP, filter-SQP methods for nonlinear programming [2, 11], filter-SL method, filter-SSP method for nonlinear semidefinite programming [17, 34]. So, in this paper, we would like to exploit the filter strategy in trust region framework to ensure the feasibility and optimality, in which the step is accepted if it either reduces the objective function or the constrained violation.

In this paper, we propose a globally convergent derivative-free trust region algorithm for nonlinear constrained optimization with separable structure, where the trust-region models are constructed by interpolation. At each iteration, we construct the quadratic interpolation models and improve the interpolation sets. The new iterative steps are generated by minimizing the augmented Lagrangian function of the interpolation model over the trust region. We update the trust region radii following the idea of structured trust region in [5]. The filter strategy is used to decide whether to accept new iterates. Global convergence analysis is given under some suitable conditions.

The paper is organized as follows. In Section 2, we give an introduction to constructing trust-region framework, multivariate interpolation and the filter technique in our problem. In Section 3, we state the new algorithm using the derivative-free filter technique for solving the separable optimization problems with nonlinear constraints. Global convergence of the algorithm is established in Section 4. Finally, some conclusions are given in Section 5.

2 The trust-region framework and polynomial interpolation

2.1 The trust-region framework

Let C be the feasible region, i.e., $C = \{(x, y) \mid x \in \mathbb{R}^n, y \in \mathbb{R}^s, c(x, y) = 0\}$.

The objective function is defined as $f(x, y) = f_1(x) + f_2(y)$ at each iteration k of a trust-region method, and the model $m_{i,k}$ are defined with each function f_i $(i = 1, 2)$ in the trust region $B_{i,k}$, where $B_{1,k} = \{x \in \mathbb{R}^n \mid \|x - x_k\| \leqslant \triangle_{1,k}\}$, $B_{2,k} = \{y \in \mathbb{R}^s \mid \|y - y_k\| \leqslant \triangle_{2,k}\}$ and $\triangle_{i,k} > 0$ $(i = 1, 2)$ are the trust region radii. In the paper, the norm is chosen to be Euclidean norm.

The quadratic models are considered as follows,

$$m_{1,k}(x_k + s_1) = f_1(x_k) + \langle g_{1,k}, s_1\rangle + \frac{1}{2}\langle s_1, H_{1,k}s_1\rangle,$$
$$m_{2,k}(y_k + s_2) = f_2(y_k) + \langle g_{2,k}, s_2\rangle + \frac{1}{2}\langle s_2, H_{2,k}s_2\rangle,$$

where $H_{i,k}$ is set to be $\nabla^2 f_i$, the Hessian of f_i, or symmetric approximation to $\nabla^2 f_i$. So, the overall model is $m_k(x, y) = m_{1,k}(x)+m_{2,k}(y)$ for all (x, y) in the overall trust region defined by $B_k = B_{1,k}\times B_{2,k}$. In unconstrained optimization, the classical procedure of trust-region methods yield trial steps $s_{1,k}$ and $s_{2,k}$ by minimizing the model over the trust region. The trial points $x_k + s_{1,k}$ and $y_k + s_{2,k}$ are accepted as the new iterates if the ratio

$$\rho_k = \frac{f(x_k, y_k) - f(x_k + s_{1,k}, y_k + s_{2,k})}{m_k(x_k, y_k) - m_k(x_k + s_{1,k}, y_k + s_{2,k})} \tag{2.1}$$

is sufficiently positive. In this case, we say that the iteration is successful, the models are updated and the trust-region radii are possibly increased. Otherwise, the trial points are rejected and the radii are decreased.

2.2 Polynomial interpolation

In our work, the models will be computed by interpolation, such that the interpolation conditions

$$f_1(z_{1,j}) = m_{1,k}(z_{1,j}), \quad f_2(z_{2,j}) = m_{2,k}(z_{2,j}), \quad \text{for all } z_{i,j} \in Z_{i,k}, \quad i = 1, 2. \tag{2.2}$$

The sets $Z_{i,k}$ $(i = 1, 2)$ are the interpolation sets. If we want to obtain the accurate model, some additional conditions will be needed on the interpolation set. In this section, we cover the basic facts on polynomial interpolation and the subscript k is dropped in the following description for clarity.

Note that the derivatives of the objective function are not available to obtain Taylor models. In this case, we require that the models satisfy Taylor-like error bounds on the function value and the gradient. Now, we give the requirements on the models that we use in our algorithm. Without loss of generality, we take the construction of $m_{1,k}$ for example, and the construction of $m_{2,k}$ is similar.

Given x_0, from the motivation of our algorithm, we hope that all iterates belong to the level set $\mathcal{L}(x_0) = \{x \in \mathbb{R}^n \mid f_1(x) \leqslant f_1(x_0)\}$. However, when we consider models based on sampling, it might be possible (especially at the early iterations) that the function f_1 is evaluated outside $\mathcal{L}(x_0)$. Thus in this paper, we assume that f_1 is restricted to regions of the form

$$\mathcal{L}_{\text{enl}}(x_0, x) = \{z \in \mathbb{R}^n \mid \|x - z\| \leqslant \triangle_{\max} \text{ for some } x \text{ with } f_1(x) \leqslant f_1(x_0)\}, \tag{2.3}$$

where $\triangle_{\max}$ is a given positive constant.

Assumption 2.1. *The objective function f_1 and its gradient and Hessian are uniformly bounded on $\mathcal{L}_{\text{enl}}(x_0, x)$ for all x in a neighborhood of x_0.*

We first state the definition of fully linear model.

Definition 2.1 (Fully linear model, see [9]). Let a function $f_1 : \mathbb{R}^n \to \mathbb{R}$ for which the gradient ∇f_1 is Lipschitz continuous on $\mathcal{L}_{\text{enl}}(x_0, x)$ be given. Let positive constants κ_{ef} and κ_{eg} be given and fixed.

Suppose that a model m_1 is continuously differentiable on $\mathbb{R}^n$. For any given $\triangle_1 \in (0, \triangle_{\max}]$ and for any given $x_c \in \mathcal{L}(x_0)$, the model m_1 is said to be fully linear on $B_1(x_c; \triangle_1)$ with respect to κ_{ef} and κ_{eg} if the following conditions hold:

- the error between the gradient of the model and the gradient of the function satisfies

$$\|\nabla f_1(x) - \nabla m_1(x)\| \leqslant \kappa_{eg}\triangle_1, \quad \forall\, x \in B_1(x_c; \triangle_1). \tag{2.4}$$

- the error between the model and the function satisfies

$$|f_1(x) - m_1(x)| \leqslant \kappa_{ef}\triangle_1^2, \quad \forall\, x \in B_1(x_c; \triangle_1). \tag{2.5}$$

Now, we discuss some assumptions on the models required in our algorithm.

Assumption 2.2. *If a model m_1 is fully linear on $B_1(x_c; \widetilde{\triangle}_1)$ with respect to some (large enough) constants κ_{ef} and κ_{eg} for some $\widetilde{\triangle}_1 \leqslant \triangle_{\max}$, then it is also fully linear on $B_1(x_c; \triangle_1)$ for any $\triangle_1 \in [\widetilde{\triangle}_1, \triangle_{\max}]$, with the same constants.*

If the function evaluations are exact, fully linear models can be defined by interpolation [6]. In our derivative-free method, the quadratic model is chosen to interpolate the value of a function f_1 at a set $Z_1 = \{z_{1,0}, z_{1,1}, \ldots, z_{1,p}\}$. For any basis $\Phi_1 = \{\phi_{1,j}(\cdot)\}_{j=0}^{p}$ of the linear space of n-dimensional quadratic functions, the polynomial $m_1(x)$ can be written uniquely as $m_1(x) = \sum_{j=0}^{p} \alpha_j \phi_{1,j}(x)$, where α_j $(j = 0, \ldots, p)$ are real coefficients. The model $m_1(x)$ must satisfy the following interpolation conditions:

$$f_1(z_{1,j}) = m_1(z_{1,j}), \quad \text{for all } z_{1,j} \in Z_1. \tag{2.6}$$

Then the coefficients can be determined by solving the linear system

$$\sum_{j=0}^{p} \alpha_j \phi_{1,j}(z_1) = f_1(z_1), \quad \text{for all } z_1 \in Z_1.$$

For the above system to have a unique solution, the matrix $M(Z_1)$ must be nonsingular, where

$$M(Z_1) = \begin{pmatrix} \phi_{1,0}(z_{1,0}) & \cdots & \phi_{1,p}(z_{1,0}) \\ \vdots & \ddots & \vdots \\ \phi_{1,0}(z_{1,p}) & \cdots & \phi_{1,p}(z_{1,p}) \end{pmatrix}.$$

Definition 2.2 (Poisedness [9])**.** The set $Z_1 = \{z_{1,0}, z_{1,1}, \ldots, z_{1,p}\}$ is poised for polynomial interpolation if the corresponding matrix $M(Z_1)$ is nonsingular for some basis Φ_1.

Let $p_1 = p+1 = |Z_1|$ be a positive integer defining the number of points in the interpolation set, then p_1 should be $\frac{1}{2}(n+1)(n+2)$ to ensure the model can be entirely determined by (2.6). However, the above conditions are not sufficient to guarantee the existence of an interpolation, and some geometry on Z_1 (for example well-poised) must be added to ensure existence and uniqueness of the quadratic interpolation.

The interpolation technique used in this paper is based on Newton fundamental polynomial (NFP) (see [26, 27]). Conn et al. [7] just used Newton polynomial interpolation to construct the derivative-free method for unconstrained optimization. Since Newton polynomial interpolation is efficient in theory and practice, we also employ this approach in our derivative-free trust-region algorithm for constrained optimization with separable structure.

For a given interpolation set Z_1, the set of interpolation points is partitioned into three subsets $Z_1^{[0]}$, $Z_1^{[1]}$ and $Z_1^{[2]}$, which correspond to constant terms, linear terms and quadratic terms of a quadratic polynomial, respectively. Hence, $Z_1^{[0]}$ has a single element, $Z_1^{[1]}$ has n elements and $Z_1^{[2]}$ has $n(n+1)/2$ elements. The basis of NFP $\{N_{1,j}(\cdot)\}$ is also partitioned into three blocks $\{N_{1,j}^0(\cdot)\}$, $\{N_{1,j}^1(\cdot)\}$ and $\{N_{1,j}^2(\cdot)\}$, with the appropriate number of elements in each block. Moreover, the unique element of $\{N_{1,j}^0(\cdot)\}$ is a polynomial of degree zero, each of the n elements of $\{N_{1,j}^1(\cdot)\}$ is a polynomial of degree one, and each of the $n(n+1)/2$

elements of $\{N_{1,j}^2(\cdot)\}$ is a polynomial of degree two. Each point $z_{1,j}^{[l]} \in Z_1^{[l]}$ ($l = 0, 1, 2$) corresponds to a single Newton fundamental polynomial of degree l satisfying the conditions

$$N_{1,t}^{[l]}(z_{1,j}^{[m]}) = \delta_{tj}\delta_{lm}, \quad \text{for all } \ z_{1,j}^{[m]} \in Z_1^{[m]} \ \text{ with } \ m \leqslant l.$$

Obtaining the NFP basis can be done if and only if the set Z_1 is poised. We can use an analogue of Gramm-Schmidt orthogonalization method starting with any basis, for example, the basis of monomials

$$\{1, x_1, x_2, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_{n-1}x_n, x_n^2\}$$

and applying the following pivoting procedure.

Procedure 2.1 (Construct Newton fundamental polynomials). For any given Z_1, initialize $N_{1,j}^{[l]}$ ($j = 1, \ldots, |Z_1^{[l]}|, l = 0, 1, 2$). Set $\overline{Z}_1 = \emptyset$.

For $l = 0, 1, 2$, for $j = 1, \ldots, |Z_1^{[l]}|$,

choose some $z_{1,j}^{[l]} \in Z_1 \setminus \overline{Z}_1$ such that $|N_{1,j}^{[l]}(z_{1,j}^{[l]})| \neq 0$,

if no such $z_{1,j}^{[l]}$ exists in $Z_1 \setminus \overline{Z}_1$, reset $Z_1 = \overline{Z}_1$ and stop (the basis of Newton polynomials is incomplete),

$$\overline{Z}_1 \leftarrow \overline{Z}_1 \cup \{z_{1,j}^{[l]}\},$$

normalize the current polynomial by

$$N_{1,j}^{[l]}(x) \leftarrow N_{1,j}^{[l]}(x)/N_{1,j}^{[l]}(z_{1,j}^{[l]}),$$

update all Newton polynomials in block l and above by

$$\begin{aligned} N_{1,t}^{[l]}(x) &\leftarrow N_{1,t}^{[l]}(x) - N_{1,t}^{[l]}(z_{1,j}^{[l]})N_{1,j}^{[l]}(x), \quad t \neq j, \quad t = 1, \ldots, |Z_1^{[l]}|, \\ N_{1,t}^{[k]}(x) &\leftarrow N_{1,t}^{[k]}(x) - N_{1,t}^{[k]}(z_{1,j}^{[l]})N_{1,j}^{[l]}(x), \quad t = 1, \ldots, |Z_1^{[l]}|, \quad k = l+1, \ldots, 2, \end{aligned}$$

end

End (the basis of Newton polynomials is complete).

In this case, the set Z_1 is poised if and only if pivots $|N_{1,j}^{[l]}(z_{1,j}^{[l]})| \neq 0$. However, for numerical purpose, it is important that $|N_{1,j}^{[l]}(z_{1,j}^{[l]})|$ are sufficiently large, which is equivalent to well-poisedness. Because small pivot values result in very large coefficients of the Newton fundamental polynomials and lead to numerical instability, we are only interested in well-poised sets. More detailed descriptions of the Newton interpolation can be consulted in [7, 26, 27].

Having computed Newton fundamental polynomials, following [6], the interpolation polynomial $m_1(x)$ is given by

$$m_1(x) = \sum_{l=0}^{2} \sum_{j=1}^{|Z_1^{[l]}|} \lambda_{1,l}(z_{1,j}^{[l]})N_{1,j}^{[l]}(x),$$

where the coefficients $\lambda_{1,l}(z_{1,j}^{[l]})$ are generalized finite differences applied on f_1, defined as

$$\lambda_{1,0}(x) = f_1(x), \quad \lambda_{1,l+1}(x) = \lambda_{1,l}(x) - \sum_{j=1}^{|Z_1^{[l]}|} \lambda_{1,l}(z_{1,j}^{[l]})N_{1,j}^{[l]}(x), \quad l = 0, 1.$$

2.3 Improve the geometry of interpolation sets

Consider the problem (1.1). At iteration k, let $f_{1,k} = f_1(x_k)$, $f_{2,k} = f_2(y_k)$, $m_{1,k} = m_{1,k}(x_k)$, $m_{2,k} = m_{2,k}(y_k)$, then $f_k = f_{1,k} + f_{2,k}, m_k = m_{1,k} + m_{2,k}$, $g_k = \binom{g_{1,k}}{g_{2,k}} = \binom{\nabla m_{1,k}}{\nabla m_{2,k}}, \nabla f_k = \binom{\nabla f_{1,k}}{\nabla f_{2,k}}$. Let

$$\theta(x, y) = \frac{1}{2}\|c(x, y)\|^2 \tag{2.7}$$

be the constraint violation, $\theta_k \triangleq \theta(x_k, y_k)$. Let

$$q_k(x,y) = m_k(x,y) + p_k^{\mathrm{T}} c(x,y) \tag{2.8}$$

be the Lagrangian function of the model, where p_k is a tentative multiplier. Then we define the augmented Lagrangian function

$$Q_k(x,y) = q_k(x,y) + u\theta(x,y), \tag{2.9}$$

where u is a positive penalty parameter.

The new step $s_k = (s_{1,k}^{\mathrm{T}}, s_{2,k}^{\mathrm{T}})^{\mathrm{T}}$ is obtained by solving the following problem,

$$\begin{aligned} \min \quad & Q_k(x_k + s_1, y_k + s_2) \\ \text{s.t.} \quad & \|s_1\| \leqslant \triangle_{1,k}, \quad \|s_2\| \leqslant \triangle_{2,k}. \end{aligned} \tag{2.10}$$

Let $w_k \triangleq (x_k, y_k)$. If we define the smallest nonnegative integer $j = j_c$ such that the point

$$w_k(j) \triangleq w_k - \kappa_{bck}^j \frac{\triangle_k}{\|\nabla Q_k\|} \nabla Q_k$$

satisfies the condition $Q_k(w_k(j)) \leqslant Q_k(w_k) + \kappa_{ubs}\langle \nabla Q_k, w_k(j) - w_k\rangle$, then we define the approximate Cauchy point as $w_k^{AC} \triangleq w_k(j_c)$, where $\kappa_{bck} \in (0,1)$ and $\kappa_{ubs} \in (0, \frac{1}{2})$ are given constants, and $\triangle_k = \sqrt{\triangle_{1,k}^2 + \triangle_{2,k}^2}$.

A fundamental result that derives trust-region methods to first-order criticality is described below (see [5, Theorem 6.3.3] and [29, Theorem 6.1.4]).

Theorem 2.3. *Suppose that Q_k is twice differentiable, then the approximate Cauchy point w_k^{AC} is well-defined in the sense that j_c is finite. Moreover, there exists a constant $\kappa_{dcp} \in (0,1)$ such that*

$$Q_k(w_k) - Q_k(w_k^{AC}) \geqslant \kappa_{dcp} \pi_k \min\left\{\frac{\pi_k}{\beta_k}, \triangle_k\right\},$$

where

$$\pi_k = \|\nabla Q_k\|, \quad \beta_k = 1 + \max_{(x,y)\in B_k} \|\nabla^2 Q_k\|. \tag{2.11}$$

From Theorem 2.3, we can see that the new step will provide at least a fraction of the decrease obtained at the approximate Cauchy point. Therefore, there exists a constant $\kappa_{qd} \in (0,1]$ such that

$$Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k}) \geqslant \kappa_{qd} \pi_k \min\left\{\frac{\pi_k}{\beta_k}, \triangle_k\right\}. \tag{2.12}$$

Now, we consider the way to use the new points to improve the geometry of interpolation sets. Without loss of generality, we still take the interpolation set Z_1.

Let the new point $x_k^+ = x_k + s_{1,k}$, and the associated function value $f_1(x_k^+)$. We consider finding the best way to make x_k^+ play a role in the next iterations when building the quadratic interpolation model. If $|Z_1| < p_1$ and if the inclusion of x_k^+ in Z_1 does not destroy poisedness, we may simply add x_k^+ to Z_1, which allows to progressively complete the set of interpolation points. Otherwise, if $|Z_1| = p_1$, we try to find a point in Z_1, say z_-, which will be replaced by x_k^+. This replacement is performed in a way that makes the pivots as large as possible in order to obtain a well-poised interpolation problem. Similar to the technique used in [4], we deal with the case $|Z_1| = p_1$ as follows:

- If x_k^+ can be advantageously added to Z_1, we proceed by the following way.

We define $S = \{k \mid x_{k+1} = x_k + s_{1,k},\ y_{k+1} = y_k + s_{2,k}\}$ the set of successful iterations and let p_1 denote the size of a complete interpolation set for function f_1. We first initialize the radius $\triangle_{1,g}$ as follows:

$$\triangle_{1,g} = \begin{cases} \min(\|s_{1,k}\|, \triangle_{1,k}), & \text{if the iteration is successful,} \\ \min(\|s_{1,k}\|, \triangle_{1,k})/\gamma_{1,d}, & \text{if the iteration is unsuccessful} \\ & \text{and the interpolation set is well-poised,} \\ \min(\|s_{1,k}\|, \triangle_{1,r})/\gamma_{1,l}, & \text{otherwise,} \end{cases}$$

where we can set $\gamma_{1,d} = 1.75$, $\gamma_{1,l} = \max\{10, p_1\}\gamma_{1,d}$ and $\triangle_{1,r}$ the trust-region radius of the most recent successful iteration.

We then look for the point z_- whose distance to the basis is the largest.

If $\|z_- - x_{k+1}\|$ is not too small (e.g., larger than $1.5\triangle_{1,k}$) and the value of the fundamental polynomial associated to z_- evaluated at x_k^+ is larger than $2(\frac{\triangle_{1,g}}{\|z_- - x_{k+1}\|})^2$, then we replace z_- by x_k^+.

Otherwise, we choose z_- to be the point associated to the fundamental polynomial whose absolute value is maximal at x_k^+, and we replace z_- by x_k^+.

• The geometry of Z_1 deteriorates as x_k^+ is accepted.

Compute new point z_+ such that $|N_{1,j}^{[l]}(z_+)|$ is larger, for example, $z_+ = \arg\max_{x \in B_{1,k}} |N_{1,j}^{[l]}(x)|$, which replaces the point $z_- = z_{1,j}^{[l]}$.

2.4 The filter as a criterion to accept trial points

As an alternative to penalty functions, we adopt the filter technique which is introduced by Fletcher and Leyffer [12]. Filter methods treat the optimization problem as a biobjective and attempt to minimize the objective function and the constraint violation function, respectively. Some definitions about filter are needed here.

Definition 2.4. (x_i, y_i) is said to dominate (x_j, y_j) if and only if both $\theta_i \leqslant \theta_j$ and $f_i \leqslant f_j$.

Thus, if iterate (x_i, y_i) dominates iterate (x_j, y_j), the latter is of no real interests to us since (x_i, y_i) is at least as good as (x_j, y_j) with respect to both measures. All we need to do now is to remember iterates that are not dominated by any other iterates in current filter.

Definition 2.5. A filter is a list $\mathcal{F}$ of pairs of the form (θ_i, f_i) such that

$$\text{either} \quad \theta_i < \theta_j \quad \text{or} \quad f_i < f_j \quad \text{for} \quad i \neq j. \tag{2.13}$$

In fact, this definition of the filter is not adequate for proving convergence, as it allows points to accumulate in the neighborhood of a filter entry that has $\theta_i > 0$. Thus, we set a small envelope around the current filter, in which points are not accepted. Chin and Fletcher [2] proposed the following rule: (x, y) is acceptable for the filter if and only if

$$\text{either} \quad \theta(x, y) < (1-\gamma)\theta_j \quad \text{or} \quad f(x, y) < f_j - \gamma\theta_j \tag{2.14}$$

for all $(\theta_j, f_j) \in \mathcal{F}$, where γ is close to zero.

Having computed the step $(s_{1,k}, s_{2,k})$ from our current iterate (x_k, y_k), we need to decide whether the trial point $(x_k + s_{1,k}, y_k + s_{2,k})$ is better than (x_k, y_k). If we define

$$\mathcal{D}(\mathcal{F}) = \{(\theta, f) \mid \theta \geqslant (1-\gamma)\theta_j \text{ and } f \geqslant f_j - \gamma\theta_j, \text{ for some } (\theta_j, f_j) \in \mathcal{F}\}, \tag{2.15}$$

the part of the (θ, f)-space, this amounts to say that $(x_k + s_{1,k}, y_k + s_{2,k})$ could be accepted if (θ^t, f^t) $= (\theta(x_k + s_{1,k}, y_k + s_{2,k}), f(x_k + s_{1,k}, y_k + s_{2,k})) \notin \mathcal{D}(\mathcal{F}_k)$, where $\mathcal{F}_k$ denotes the filter at iteration k. The procedure to update the filter is simple: If (θ^t, f^t) does not belong to $\mathcal{D}(\mathcal{F}_k)$, then

$$\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k \cup (\theta^t, f^t) \setminus \{(\theta_j, f_j) \mid \theta_j \geqslant (1-\gamma)\theta^t, f_j \geqslant f^t - \gamma\theta^t\}, \tag{2.16}$$

while if $(\theta^t, f^t) \in \mathcal{D}(\mathcal{F}_k)$, then

$$\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k. \tag{2.17}$$

3 A derivative-free trust region algorithm

We now describe a derivative-free trust region algorithm for separable constrained optimization problem (1.1).

The following definitions will appear in Algorithm 3.1. Let

$$l_k(x,y) = f_1(x) + f_2(y) + p_k^{\mathrm{T}} c(x,y)$$

be the Lagrangian function of (1.1), and define the augmented Lagrangian function

$$L_k(x,y) = l_k(x,y) + u\theta(x,y). \tag{3.1}$$

Let

$$\begin{aligned}
&L_{1,k}(x,y) = f_1(x) + \frac{1}{2}p_k^{\mathrm{T}} c(x,y) + \frac{u}{2}\theta(x,y), \quad L_{2,k}(x,y) = f_2(y) + \frac{1}{2}p_k^{\mathrm{T}} c(x,y) + \frac{u}{2}\theta(x,y),\\
&Q_{1,k}(x,y) = m_{1,k} + \frac{1}{2}p_k^{\mathrm{T}} c(x,y) + \frac{u}{2}\theta(x,y), \quad Q_{2,k}(x,y) = m_{2,k} + \frac{1}{2}p_k^{\mathrm{T}} c(x,y) + \frac{u}{2}\theta(x,y),\\
&\delta L_k = L_k(x_k,y_k) - L_k(x_k+s_{1,k}, y_k+s_{2,k}),\\
&\delta L_{i,k} = L_{i,k}(x_k,y_k) - L_{i,k}(x_k+s_{1,k}, y_k+s_{2,k}), \quad i=1,2,\\
&\delta Q_k = Q_k(x_k,y_k) - Q_k(x_k+s_{1,k}, y_k+s_{2,k}),\\
&\delta Q_{i,k} = Q_{i,k}(x_k,y_k) - Q_{i,k}(x_k+s_{1,k}, y_k+s_{2,k}), \quad i=1,2.
\end{aligned}$$

Algorithm 3.1 (A derivative-free trust region algorithm for separable optimization problem)**.**

Step 0 (Initialization)**.** Choose initial point $(x_0,y_0) \in C$, $\triangle_{\max} > 0$, the initial trust region radius $\triangle_{i,0}^{icb} \in (0, \triangle_{\max}]$ is given for $i=1,2$. We initialize the parameters for geometry improvements $\gamma_{i,d}$, $\gamma_{i,l}$, and $\triangle_{i,r} = \triangle_{i,0}^{icb}$, $i=1,2$. Assume that for each $i=1,2$, there exists a set Z_i^{icb} of interpolation points.

The constants η_1, η_2, η_3, γ_1, γ_2, γ_3, μ_1, μ_2, ϵ_c, β, ν, α_1, α_2 and $\mathcal{F}$ are given and satisfy the conditions $0 < \eta_1 \leqslant \eta_2 < \eta_3 < 1$, $0 < \gamma_1 \leqslant \gamma_2 < 1 \leqslant \gamma_3$, $0 < \mu_1 < \mu_2 < 1$, $\eta_2 - \eta_1 \geqslant \mu_1 + \mu_2$, $\epsilon_c > 0$, $\nu > \beta > 0$, $\alpha_1 \in (0,1)$, $\alpha_2 \in (0,1)$ and $\mathcal{F} = \emptyset$. Set $k=0$.

Step 1 (Compute models)**.** For $i \in \{1,2\}$, compute quadratic interpolation polynomials $m_{i,k}^{icb}$.

Step 2 (Critical step)**.** If $\pi_k^{icb} \geqslant \epsilon_c$, then $m_{i,k} = m_{i,k}^{icb}$, $\triangle_{i,k} = \triangle_{i,k}^{icb}$, $i=1,2$.

If $\pi_k^{icb} \leqslant \epsilon_c$, then we proceed as follows. If at least one of the following conditions holds,

(1) the interpolation set Z_1^{icb} or Z_2^{icb} is not well-poised in $B(x_k, \triangle_{1,k}^{icb})$ or $B(y_k, \triangle_{2,k}^{icb})$,

(2) $\triangle_k^{icb} > \nu\pi_k^{icb}$,

then apply Algorithm 3.2 to construct interpolation sets $\widehat{Z}_1$ and $\widehat{Z}_2$, which are well-poised in $B(x_k, \widehat{\triangle}_{1,k})$ and $B(y_k, \widehat{\triangle}_{2,k})$ for some $\widehat{\triangle}_{i,k}$, $i=1,2$. Then compute the corresponding models $\widehat{m}_{1,k}$ and $\widehat{m}_{2,k}$. Set

$$m_{i,k} = \widehat{m}_{i,k} \quad \text{and} \quad \triangle_{i,k} = \min\{\max\{\widehat{\triangle}_{i,k}, \beta\widehat{\pi}_k\}, \triangle_{i,k}^{icb}\}, \quad i=1,2.$$

Otherwise set $m_{i,k} = m_{i,k}^{icb}$ and $\triangle_{i,k} = \triangle_{i,k}^{icb}$, $i=1,2$. Stop.

Step 3 (Determination of the step)**.** Compute steps $s_{1,k}$ and $s_{2,k}$ by solving the problem (2.10).

Step 4 (Test to accept the trial point)**.** If $(x_k+s_{1,k}, y_k+s_{2,k})$ is not acceptable by the filter, set $x_{k+1} = x_k$, $y_{k+1} = y_k$, $p_{k+1} = p_k$, $x^+ = x_k + s_{1,k}$, $y^+ = y_k + s_{2,k}$. Choose $\triangle_{i,k+1}^{icb} \in [\gamma_1\triangle_{i,k}, \gamma_2\triangle_{i,k}]$, $i=1,2$, and go to Step 7.

If

$$\delta Q_k < \kappa_\theta \theta_k^\psi, \tag{3.2}$$

then add (θ_k, f_k) to the filter $\mathcal{F}$ and update the filter, where κ_θ and ψ are positive constants. Set $x_{k+1} = x_k + s_{1,k}$, $y_{k+1} = y_k + s_{2,k}$ and $p_{k+1} = p_k + \lambda c(x_{k+1}, y_{k+1})$, $x^+ = x_k$, $y^+ = y_k$. Choose $\triangle_{i,k+1}^{icb} \in [\triangle_{i,k}, \gamma_3\triangle_{i,k}]$, $i=1,2$, and go to Step 7.

Step 5 (Test predicted and actual reduction)**.** Define

$$\rho_k = \frac{\delta L_k}{\delta Q_k}. \tag{3.3}$$

If $\rho_k \geqslant \eta_1$, then set $x_{k+1} = x_k + s_{1,k}$, $y_{k+1} = y_k + s_{2,k}$, $p_{k+1} = p_k + \lambda c(x_{k+1}, y_{k+1})$, $x^+ = x_k$, $y^+ = y_k$. Otherwise set $x_{k+1} = x_k$, $y_{k+1} = y_k$, $p_{k+1} = p_k$, $x^+ = x_k + s_{1,k}$, $y^+ = y_k + s_{2,k}$.

Step 6 (Update the trust region radii). We consider the following two cases:

Case I.

$$|\delta Q_{i,k}| > \frac{\mu_1}{2}\delta Q_k, \quad i \in \{1,2\}. \tag{3.4}$$

1. If

$$\delta L_{i,k} \geqslant \delta Q_{i,k} - \frac{1-\eta_3}{2}\delta Q_k \tag{3.5}$$

and $\rho_k \geqslant \eta_1$, then choose

$$\triangle_{i,k+1}^{icb} \in [\triangle_{i,k}, \gamma_3\triangle_{i,k}].$$

2. If (3.5) holds and $\rho_k < \eta_1$, then choose

$$\triangle_{i,k+1}^{icb} = \triangle_{i,k}.$$

If (3.5) fails but

$$\delta L_{i,k} \geqslant \delta Q_{i,k} - \frac{1-\eta_2}{2}\delta Q_k \tag{3.6}$$

holds, then choose

$$\triangle_{i,k+1}^{icb} \in [\gamma_2\triangle_{i,k}, \triangle_{i,k}].$$

3. If (3.6) fails, then choose

$$\triangle_{i,k+1}^{icb} \in [\gamma_1\triangle_{i,k}, \gamma_2\triangle_{i,k}].$$

Case II.

$$|\delta Q_{i,k}| \leqslant \frac{\mu_1}{2}\delta Q_k, \quad i \in \{1,2\}. \tag{3.7}$$

1. If

$$|\delta L_{i,k}| \leqslant \frac{\mu_2}{2}\delta Q_k, \tag{3.8}$$

and $\rho_k \geqslant \eta_1$, then choose

$$\triangle_{i,k+1}^{icb} \in [\triangle_{i,k}, \gamma_3\triangle_{i,k}].$$

2. If (3.8) holds and $\rho_k < \eta_1$, then choose

$$\triangle_{i,k+1}^{icb} = \triangle_{i,k}.$$

3. If (3.8) fails, then choose

$$\triangle_{i,k+1}^{icb} \in [\gamma_2\triangle_{i,k}, \triangle_{i,k}].$$

Step 7 (Improve the geometry). Update $\triangle_{i,g}, i = 1,2$. If x^+ and y^+ can make the geometry of interpolation set better respectively as described in Section 2, then modify $Z_i^{icb}, i = 1,2$ by replacing the existing point and update the corresponding interpolation model. Otherwise, compute a new point that can improve the geometry of Z_i^{icb}, modify $Z_i^{icb}, i = 1,2$ and update the corresponding interpolation model. If one of new points computed in this step has a better objective function value than (x_{k+1}, y_{k+1}), then replace it and update the corresponding interpolation model. Set $k := k+1$ and go to Step 1.

Algorithm 3.2 (Criticality step). **Initialization** Set $j = 0$ and $Z_i^0 = Z_i^{icb}$, $i = 1,2$.

Repeat Increase j by one, improve the geometry until Z_1^{j-1} and Z_2^{j-1} are well-poised in $B_1(x_k, \alpha_1^{j-1}\triangle_{1,k}^{icb})$ and $B_2(y_k, \alpha_2^{j-1}\triangle_{2,k}^{icb})$. Denote the new set by Z_i^j and the corresponding interpolation model by $m_{i,k}^j$ $(i = 1,2)$. Set $\widehat{\triangle}_{i,k} = \alpha_i^{j-1}\triangle_{i,k}^{icb}$ and $\widehat{m}_{i,k} = m_{i,k}^j$, $i = 1,2$.

Until $\widehat{\triangle}_k = \sqrt{\widehat{\triangle}_{1,k}^2 + \widehat{\triangle}_{2,k}^2} \leqslant \nu\pi_k^j$.

Let us discuss some particular remarks of the algorithm.

Remark 3.1. Note that if $\pi_k^{icb} \leqslant \epsilon_c$ in the criticality step of Algorithms 3.1 and 3.2, the model $m_{i,k}$ $(i = 1, 2)$ are fully linear on $B_1(x_k; \widehat{\triangle}_{1,k})$ and $B_2(y_k; \widehat{\triangle}_{2,k})$ separately with $\widehat{\triangle}_{i,k} \leqslant \triangle_{i,k}$ $(i = 1, 2)$. Then, by Assumption 2.2, $m_{i,k}$ $(i = 1, 2)$ are also fully linear on $B_1(x_k; \triangle_{1,k})$ and $B_2(y_k; \triangle_{2,k})$.

Remark 3.2. The role of condition (3.2) can be interpreted as follows. If it holds, one may think that the constraint violation is significant and one should aim to improve this situation by inserting the current point into the filter. Otherwise, the predicted reduction of augmented Lagrangian function of the model is more significant than the current constraint violation. In this case, we perform test (3.3).

Remark 3.3. To minimize a separable objective function subject to equality constraints successfully, structured trust region technique [5] is used in our algorithm and each of functions is modeled separately, which overcomes the restriction of using the basic trust region technique.

Before starting our global convergence analysis, we state some properties that result from the mechanism of structured trust region.

Lemma 3.4. *Let $M_k = \{i \mid |\delta Q_{i,k}| > \frac{\mu_1}{2}\delta Q_k\}$. Then, at each iteration k of the algorithm,*

1. *M_k contains at least one element. Furthermore,*

$$\left(1 - \frac{1}{2}\mu_1\right)\delta Q_k \leqslant \sum_{i\in M_k} \delta Q_{i,k} \leqslant \left(1 + \frac{1}{2}\mu_1\right)\delta Q_k. \tag{3.9}$$

2. *$\gamma_1\triangle_{i,k} \leqslant \triangle_{i,k+1} \leqslant \gamma_3\triangle_{i,k}$ for all $i \in \{1, 2\}$.*

Lemma 3.5. *At iteration k of the algorithm, if* (3.6) *holds for Case* I *or* (3.8) *holds for Case* II*, then iteration k is successful.*

The proofs of the above two lemmas are similar to the proofs of [5, Lemmas 10.2.1 and 10.2.2].

4 Convergence analysis

In this section, we establish the global convergence of the derivative-free trust-region algorithm. For our analysis, let $\Omega = \{k \mid (\theta_k, f_k) \text{ is added to the filter}\}$. In order to obtain the global convergence result, we start by the following assumptions.

Assumption 4.1. The sequence $\{(x_k, y_k)\}$ generated by Algorithm 3.1 remains in a bounded domain.

Assumption 4.2. There exits a constant $\kappa_{uqh} > 1$ such that

$$\beta_k \leqslant \kappa_{uqh}, \tag{4.1}$$

where β_k is defined in (2.11).

Assumption 4.3. The constraint function $c(x, y)$ and its gradient ∇c, Hessian $\nabla^2 c$ and $(\nabla c)^{\mathrm{T}} c$ are uniformly bounded on B_k.

Assumption 4.4. The multiplier vectors p_k $(k = 1, 2, \ldots)$ are uniformly bounded.

An important consequence of the assumptions is that Assumptions 2.1, 4.1 and 4.3 together directly ensure that, for all k,

$$f^{\min} \leqslant f(x_k, y_k) \leqslant f^{\max} \quad \text{and} \quad 0 \leqslant \theta(x_k, y_k) \leqslant \theta^{\max} \tag{4.2}$$

for some constants $f^{\min} \leqslant f^{\max}$ and $\theta^{\max} \geqslant 0$. Thus, the part of the (θ, f)-space in which the (θ, f)-pairs associated with the filter iterates lie is restricted to the rectangle $[0, \theta^{\max}] \times [f^{\min}, f^{\max}]$.

Assumptions 4.3 and 4.4 mean that for all k, there exist positive constants κ_c, κ_{cg}, κ_{ch}, κ_{cv} and κ_p, such that $\|c\| \leqslant \kappa_c$, $\|\nabla c\| \leqslant \kappa_{cg}$, $\|\nabla^2 c\| \leqslant \kappa_{ch}$, $\|(\nabla c)^{\mathrm{T}} c\| \leqslant \kappa_{cv}$, and $\|p_k\| \leqslant \kappa_p$.

We will need the following lemma, which can be obtained similarly from the well-known results (see [8, Lemma 5.1]).

Lemma 4.1. *If $\nabla L_k \neq 0$, Step 2 of Algorithm 3.1 will terminate in a finite number of improvement steps by applying Algorithm 3.2.*

We continue our analysis by showing that the iterations are successful when the trust-region radius is sufficiently small.

Lemma 4.2. *Under Assumption 2.1, if $Z_{i,k}$ $(i = 1, 2)$ are well-poised and*

$$\triangle_k \leqslant \min\left[\frac{1}{\kappa_{uqh}}, \frac{\kappa_{qd}(1-\eta_1)}{2\kappa_{ef}}\right]\pi_k, \tag{4.3}$$

then $\rho_k \geqslant \eta_1$.

Proof. By (4.3), we find that

$$\triangle_k \leqslant \frac{\pi_k}{\kappa_{uqh}}.$$

Then the Cauchy decrease condition (2.12) gives that

$$Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k}) \geqslant \kappa_{qd}\pi_k \min\left\{\frac{\pi_k}{\kappa_{uqh}}, \triangle_k\right\} = \kappa_{qd}\pi_k\triangle_k. \tag{4.4}$$

Since $Z_{i,k}$ $(i = 1, 2)$ are well-poised, from the bound (2.5), we have

$$\begin{aligned}
|\rho_k - 1| &\leqslant \frac{|L_k(x_k + s_{1,k}, y_k + s_{2,k}) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})|}{|Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})|} \\
&\quad + \frac{|L_k(x_k, y_k) - Q_k(x_k, y_k)|}{|Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})|} \\
&= \frac{|f(x_k + s_{1,k}, y_k + s_{2,k}) - m_k(x_k + s_{1,k}, y_k + s_{2,k})|}{|Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})|} \\
&\quad + \frac{|f(x_k, y_k) - m_k(x_k, y_k)|}{|Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})|} \\
&\leqslant \frac{2\kappa_{ef}(\triangle_{1,k}^2 + \triangle_{2,k}^2)}{\kappa_{qd}\pi_k\triangle_k} \\
&= \frac{2\kappa_{ef}\triangle_k}{\kappa_{qd}\pi_k}.
\end{aligned}$$

Obviously, the conclusion $\rho_k \geqslant \eta_1$ is obtained if Assumption (4.3) holds. □

The following lemma states that if the criticality measure is bounded away from zero, then so is the trust region radius.

Lemma 4.3. *Suppose that there exists a constant $\delta_1 > 0$ such that $\pi_k \geqslant \delta_1$ for all k. Then, there exists a constant $\delta_2 > 0$ such that, for all k,*

$$\triangle_k \geqslant \delta_2.$$

Proof. We know from Step 1 of Algorithm 3.1 that

$$\triangle_{i,k} \geqslant \min\{\beta\pi_k, \triangle_{i,k}^{icb}\}, \quad i = 1, 2.$$

Thus,

$$\triangle_{i,k} \geqslant \min\{\beta\delta_1, \triangle_{i,k}^{icb}\}, \quad i = 1, 2. \tag{4.5}$$

By Lemma 4.2 and the assumption that $\pi_k \geqslant \delta_1$ for all k, we know whenever

$$\triangle_k \leqslant \overline{\delta}_2 = \min\left[\frac{\delta_1}{\kappa_{uqh}}, \frac{\kappa_{qd}\delta_1(1-\eta_1)}{2\kappa_{ef}}\right],$$

either the k-th iteration is successful or the geometry of interpolation sets is improved. Hence, in this case, $\triangle_{i,k} \leqslant \overline{\delta}_2$. From Step 6, we have $\triangle_{i,k+1}^{icb} \geqslant \gamma_1 \triangle_{i,k}$ $(i = 1, 2)$, which, together with (4.5) and the rules of Step 6, gives $\triangle_{i,k} \geqslant \min\{\triangle_{i,0}^{icb}, \beta\delta_1, \gamma_1\overline{\delta}_2\}$ $(i = 1, 2)$. Therefore, there exists a constant $\delta_2 > 0$ such that $\triangle_k = \sqrt{\triangle_{1,k}^2 + \triangle_{2,k}^2} \geqslant \delta_2$. □

Now we concentrate on the case that there is no infinite subsequence of iterates being added to the filter. In this case, no further iterates are added to the filter for k sufficiently large. In what follows, we assume that $k_0 \geqslant 0$ is the last iteration for which (x_{k_0-1}, y_{k_0-1}) is accepted by the filter.

Lemma 4.4. *Suppose that Assumptions* 2.1, 4.1 *and* 4.3 *hold. Suppose further that* (3.2) *fails for all* $k \geqslant k_0$. *Then we have that*

$$\lim_{k\to\infty} \theta_k = 0.$$

Proof. Consider any successful iteration with $k \geqslant k_0$. Since the filter is not updated at iteration k, it follows from the algorithm that $\rho_k \geqslant \eta_1$ holds and thus

$$L_k(x_k, y_k) - L_k(x_k + s_{1,k}, y_k + s_{2,k}) \geqslant \eta_1[Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})] \geqslant \eta_1\kappa_\theta\theta_k^\psi \geqslant 0. \tag{4.6}$$

Thus, the augmented Lagrangian function of (1.1) does not increase for all successful iterations with $k \geqslant k_0$. Then, by Assumptions 2.1, 4.1 and 4.3, we must have that

$$\lim_{k\to\infty} L_k(x_k, y_k) - L_k(x_k + s_{1,k}, y_k + s_{2,k}) = 0. \tag{4.7}$$

The result (4.4) then immediately follows from (4.6). □

We now consider what happens when the number of successful iterations is finite.

Lemma 4.5. *Suppose that Assumptions* 2.1, 4.3 *and* 4.4 *hold and that* S *is finite. Then*

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

Proof. We consider iterations that come after the last successful iteration. We can have only a finite (uniformly bounded, say by N) number of geometry-improving iterations before the interpolation sets become well-poised. Hence, there is an infinite number of iterations that are either acceptable by the interpolation sets or unsuccessful, and in either case the trust region is reduced. Since there are no more successful iterations, then $\triangle_{i,k}$ $(i = 1, 2)$ are never increased for sufficiently large k. Moreover, $\triangle_{i,k}$ $(i = 1, 2)$ are decreased, which induces that $\{\triangle_{i,k}\}$ $(i = 1, 2)$ converges to zero.

For each j, let r be the index of the first iteration after the j-th iteration for which the interpolation sets $Z_{i,j}$ $(i = 1, 2)$ are well-poised. Then

$$\|x_j - x_r\| \leqslant N\triangle_{1,j} \to 0 \quad \text{and} \quad \|y_j - y_r\| \leqslant N\triangle_{2,j} \to 0 \tag{4.8}$$

as $j \to \infty$. Observe that

$$\|\nabla L_j(x_j, y_j)\| \leqslant \|\nabla L_j(x_j, y_j) - \nabla L_r(x_r, y_r)\| + \|\nabla L_r(x_r, y_r) - \nabla Q_r(x_r, y_r)\| + \|\nabla Q_r(x_r, y_r)\|.$$

By (2.9) and (3.1), we have that

$$\begin{aligned} \nabla L_r &= \nabla f + p_r^{\mathrm{T}}\nabla C + \lambda(\nabla C)^{\mathrm{T}}C, \\ \nabla Q_r &= \nabla m_r + p_r^{\mathrm{T}}\nabla C + \lambda(\nabla C)^{\mathrm{T}}C, \\ \nabla L_r - \nabla Q_r &= \nabla f - \nabla m_r. \end{aligned}$$

So we have

$$\|\nabla L_j(x_j, y_j)\| \leqslant \|\nabla L_j(x_j, y_j) - \nabla L_r(x_r, y_r)\| + \|\nabla f(x_r, y_r) - \nabla m_r(x_r, y_r)\| + \|\nabla Q_r(x_r, y_r)\|.$$

Now we show that all three terms on the right-hand side converge to zero. The first term converges to zero because of the Assumption 2.1 and (4.8). The second term converges to zero because of the bound (2.4). Finally, the third term can be shown to converge to zero by Lemma 4.2, since if $\|\nabla Q_r(x_r, y_r)\|$ was bounded away from zero for a subsequence, then for small enough $\triangle_r$, r would be a successful iteration, which would yield a contradiction. □

We continue our analysis by considering the case that the filter is updated in infinitely many times.

Lemma 4.6. *Suppose that Assumptions* 2.1 *and* 4.2 *hold and that* $|\Omega| = \infty$. *Then*

$$\lim_{k\to\infty, k\in\Omega} \theta_k = 0. \tag{4.9}$$

Proof. By contradiction. Suppose that there exists an infinite subsequence $K \subseteq \Omega$, such that $\theta_{k_i} \geqslant \epsilon$ for all $k_i \in K$ and for some $\epsilon > 0$. At each iteration k_i,

$$\mathcal{D}(\mathcal{F}_{k_i}) = \{(\theta, f) \mid \theta \geqslant (1-\gamma)\theta_j \text{ and } f \geqslant f_j - \gamma\theta_j, \text{ for some } (\theta_j, f_j) \in \mathcal{F}_{k_i}\}, \tag{4.10}$$

we can deduce that $(\theta, f) \in \mathcal{D}(\mathcal{F}_{k_i})$ can not be added to the filter and it is easy to see that

$$[(1-\gamma)\theta_j, \theta_j] \times [f_j - \gamma\theta_j, f_j] \subseteq \mathcal{D}(\mathcal{F}_{k_i}).$$

Now observe that the area of each of these squares is at least $\gamma^2\theta_j^2$. As a consequence,

$$\text{area}(\mathcal{D}(\mathcal{F}_{k_i})) \geqslant \gamma^2 \sum_{j=0, j\in\Omega}^{k_i-1} \theta_j^2 \geqslant k_i\gamma^2\epsilon^2.$$

However, (4.2) implies that, for any k_i, there is a positive constant κ_f independent of k_i such that area$(\mathcal{D}(\mathcal{F}_{k_i}))$ is bounded above by κ_f. Hence, we obtain that $k_i \leqslant \frac{\kappa_f}{\gamma^2\epsilon^2}$ and that k_i must also be finite. This contradicts the fact that the subsequence K is infinite. Hence, the assumption is impossible and the conclusion follows. □

We now prove an important property on $\triangle_k$, which gives a natural stopping criterion for the derivative-free trust region method.

Theorem 4.7. *Suppose that Assumptions* 2.1 *and* 4.1–4.4 *hold. Then*

$$\lim_{k\to\infty} \triangle_k = 0. \tag{4.11}$$

Proof. When S is finite, (4.11) is proved in the proof of Lemma 4.5. Let us consider the case when S is infinite. For any $k \in S$, we have

$$L_k(x_k, y_k) - L_k(x_k + s_{1,k}, y_k + s_{2,k}) \geqslant \eta_1[Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})].$$

By using the Cauchy decrease Condition (2.12), we have

$$Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k}) \geqslant \kappa_{qd}\pi_k \min\left\{\frac{\pi_k}{\beta_k}, \triangle_k\right\}.$$

Using Step 2 of Algorithm 3.1, we have $\pi_k \geqslant \min\{\epsilon_c, \nu^{-1}\triangle_k\}$. It then follows from (4.1) that

$$L_k(x_k, y_k) - L_k(x_k + s_{1,k}, y_k + s_{2,k}) \geqslant \eta_1\kappa_{qd} \min\{\epsilon_c, \nu^{-1}\triangle_k\} \min\left\{\frac{\min\{\epsilon_c, \nu^{-1}\triangle_k\}}{\kappa_{uqh}}, \triangle_k\right\}.$$

Since L_k does not increase and is bounded from below, the right-hand side of the above expression converges to zero. Hence $\lim_{k\in S}\triangle_k = 0$. Note that the $\triangle_k$ can increase only during a successful iteration. Let $k \notin S$ be the index of an iteration (the first one after the successful iteration). Then $\triangle_k \leqslant \gamma_3\triangle_{s_k}$, where s_k is the index of the last successful iteration before k. Since $\triangle_{s_k} \to 0$, then $\triangle_k \to 0$, for $k \notin S$. Hence, $\triangle_k \to 0$ for all k sufficiently large. □

By using Lemma 4.3 and Theorem 4.7, we immediately have the following lemma.

Lemma 4.8.

$$\liminf_{k\to\infty} \pi_k = 0. \tag{4.12}$$

The following lemma establishes a relationship between π_k and $\|\nabla L_k\|$.

Lemma 4.9. *For any subsequence* $\{k_i\}$, *if*

$$\liminf_{i\to\infty} \pi_{k_i} = 0, \tag{4.13}$$

then

$$\liminf_{i\to\infty} \|\nabla L_{k_i}\| = 0. \tag{4.14}$$

Proof. By (4.13), if i is sufficiently large, $\pi_{k_i} \leqslant \epsilon_c$. The criticality step of Algorithm 3.1 ensures that the interpolation sets $Z_{1,k_i} \subset B_1(x_{k_i}; \triangle_{1,k_i})$ (with $\triangle_{1,k_i} \leqslant \nu\pi_{k_i}$) and $Z_{2,k_i} \subset B_2(y_{k_i}; \triangle_{2,k_i})$ (with $\triangle_{2,k_i} \leqslant \nu\pi_{k_i}$) are well-poised for all large i. Then, using the bound (2.4) on the error between the gradients of the function and the model, we have

$$\|\nabla L_{k_i} - \nabla Q_{k_i}\| = \|\nabla f(x_{k_i}, y_{k_i}) - \nabla m_{k_i}(x_{k_i}, y_{k_i})\| \leqslant \kappa_{eg}(\triangle_{1,k_i} + \triangle_{2,k_i}) \leqslant 2\kappa_{eg}\nu\pi_{k_i}.$$

Thus we have

$$\|\nabla L_{k_i}\| \leqslant \|\nabla L_{k_i} - \nabla Q_{k_i}\| + \pi_{k_i} \leqslant (2\kappa_{eg}\nu + 1)\pi_{k_i},$$

and the result (4.14) immediately follows from $\pi_{k_i} \to 0$. □

Using the above lemma, we obtain the following global convergence result.

Theorem 4.10. *Suppose that Assumptions* 2.1–2.2 *and* 4.1–4.4 *hold. Then*

$$\liminf_{k\to\infty} \|\nabla L_k\| = 0.$$

This theorem follows directly from Lemmas 4.8 and 4.9.

Now we prove the following global convergence theorem.

Theorem 4.11. *Suppose that Assumptions* 2.1–2.2 *and* 4.1–4.4 *hold. Then*

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

Proof. When S is finite, the result follows from Lemma 4.5. So we now consider only the case when S is infinite. By contradiction we assume that there exists a subsequence $\{k_i\}$ such that

$$\|\nabla L_{k_i}\| \geqslant \delta_0 \tag{4.15}$$

for some $\delta_0 > 0$ and for all i. Then, from Lemma 4.9, there exists $\delta_1 > 0$ such that

$$\pi_{k_i} \geqslant \delta_1$$

for all i sufficiently large. Then, by Lemma 4.3, there exists a constant $\delta_2 > 0$, independent of k, such that

$$\triangle_k \geqslant \delta_2. \tag{4.16}$$

Then, since L_k does not increase and is bounded below, it follows from Algorithm 3.1, (2.12) and (4.1) that

$$\begin{aligned} +\infty &> \sum_{k=0}^{\infty}[L_k(x_k, y_k) - L_k(x_k + s_{1,k}, y_k + s_{2,k})] \\ &\geqslant \sum_{k\in S}[L_k(x_k, y_k) - L_k(x_k + s_{1,k}, y_k + s_{2,k})] \\ &\geqslant \sum_{k\in S}\eta_1[Q_k(x_k, y_k) - Q_k(x_k + s_{1,k}, y_k + s_{2,k})] \end{aligned}$$

$$\geqslant \sum_{k\in S} \eta_1 \kappa_{qd} \delta_1 \min\left\{\frac{\delta_1}{\kappa_{uqh}}, \triangle_k\right\}.$$

Hence,

$$\lim_{k\to\infty, k\in S} \|\triangle_k\| = 0.$$

Combining with Lemma 4.5, we have

$$\lim_{k\to\infty} \|\triangle_k\| = 0 \tag{4.17}$$

which contradicts (4.16). □

5 Conclusions

In this paper, we present a framework for a derivative-free trust region algorithms for minimizing constrained optimization with separable structure. At each iteration, we construct a quadratic interpolation model of the objective function based on Newton interpolation. The new iterates are generated by minimizing the augmented Lagrangian function of this model. We use the filter technique to ensure the feasibility and optimality of our algorithm. The poisedness of the interpolation set can be monitored by using Newton polynomials and maintained via appropriate exchange of new iterates and the points in the interpolation set. Global convergence of our algorithm is proved under suitable assumptions. In future works we will continue our investigation, for example, we will study the numerical behavior of the algorithm proposed in this paper, and consider further the case where the derivatives of both objective functions and constrained functions in separable optimization are unavailable.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 11071122 and 11171159), the Specialized Research Fund of Doctoral Program of Higher Education of China (Grant No. 20103207110002). The authors would like to thank referees for their beneficial comments and suggestions which improved the manuscript essentially.

References

1 Audet C, Dennis J E. A pattern search filter method for nonlinear programming without derivatives. SIAM J Optim, 2004, 13: 889–903
2 Chin C M, Fletcher R. On the global convergence of an SLP-filter algorithm that takes EQP steps. Math Program, 2003, 96: 161–177
3 Colson B, Toint Ph L. Exploiting band structure in unconstrained optimization without derivatives. Optim Eng, 2001, 2: 399–412
4 Colson B, Toint Ph L. Optimizing partially separable functions without derivatives. Optim Methods Softw, 2005, 20: 493–508
5 Conn A R, Gould N I M, Toint Ph L. Trust-Region Methods. Philadelphia: SIAM, 2000
6 Conn A R, Scheinberg K, Toint Ph L. Recent progress in unconstrained nonlinear optimization without derivatives. Math Program, 1997, 79: 397–414
7 Conn A R, Scheinberg K, Toint Ph L. On the convergence of derivative-free methods for unconstrained optimization. In: Iserles A, Buhmann M, eds. Approximation Theory and Optimization: Tributes to M.J.D. Pawell. Cambridge: Cambridge University Press, 1997, 83–108
8 Conn A R, Scheinberg K, Vicente L N. Global convergence of general derivative-free trust-region algorithms to first and second-order critical points. SIAM J Optim, 2009, 20: 387–415
9 Conn A R, Scheinberg K, Vicente L N. Introduction to Derivative-Free Optimization. Philadelphia: SIAM, 2009
10 Dennis J E, Schnabel R B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Englewood Cliffs, NJ: Prentice-Hall, 1983
11 Fletcher R, Gould N I M, Leyffer S, et al. Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. SIAM J Optim, 2002, 13: 635–659
12 Fletcher R, Leyffer S. Nonlinear programming without a penalty function. Math Program, 2002, 91: 239–269

13 Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Comput Math Appl, 1976, 2: 17–40
14 Glowinski R, Marrocco A. Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problems de Dirichlet non lineares. Revue Française d'Automatique, Informatique, et Recherche Opérationelle, 1975, 9: 41–76
15 Grippo L, Lampariello F, Lucidi S. Global convergence and stabilization of unconstrained minimization methods without derivatives. J Optim Theory Appl, 1988, 56: 385–406
16 He B S. Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities. Comput Optim Appl, 2009, 42: 195–212
17 Li C, Sun W. On filter-successive linearization methods for nonlinear semidefinite programming. Sci China Ser A, 2009, 52: 2341–2361
18 Lucidi S, Sciandrone M. Numerical results for unconstrained optimization without derivatives. In: Giannessi F, Di Pillo G, eds. Nonlinear Optimization and Applications. New York: Plenum Publishing, 1995, 261–269
19 Miao W, Sun W. A filter trust-region method for unconstrained optimization problems. Numer Math J Chin Univ, 2007, 29: 88–96
20 Moré J J, Wild S M. Benchmarking derivative-free optimization algorithms. SIAM J Optim, 2009, 20: 172–191
21 Nelder J A, Mead R. A simplex method for function minimization. Comput J, 1965, 7: 308–313
22 Powell M J D. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput J, 1964, 7: 155–162
23 Powell M J D. A direct search optimization method that models the objective and constraint functions by linear interpolation. In: Gomez S, Hennart J P, eds. Advances in Optimization and Numerical Analysis. Dordrecht: Kluwer Academic, 1994, 51–67
24 Powell M J D. Least Frobenius norm updating of quadratic models that satisfy interpolation conditions. Math Program Ser B, 2004, 100: 183–215
25 Scheinberg K, Toint P L. Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. SIAM J Optim, 2010, 20: 3512–3532
26 Stoer J, Bulirsch R. Introduction to Numerical Analysis. New York: Springer, 1980
27 Sun W, Du Q, Chen J. Computational Methods (in Chinese). Beijing: Science Press, 2007
28 Sun W, Xu D. A filter-trust-region method based on conic model for unconstrained optimization (in Chinese). Sci Sin Math, 2012, 42: 527–543
29 Sun W, Yuan Y. Optimization Theory and Methods: Nonlinear Programming. New York: Springer, 2006
30 Takaki J, Yamashita N. A derivative-free trust-region algorithm for unconstrained optimization with controlled error. Numer Algebra Contr Optim, 2011, 1: 117–145
31 Tseng P. Alternating projection-proximal methods for convex programming and variational inequalities. SIAM J Optim, 1997, 7: 951–965
32 Torczon V. On the convergence of pattern search algorithms. SIAM J Optim, 1997, 7: 1–25
33 Winfried D. Function minimization by interpolation in a data table. J Inst Math Appl, 1973, 12: 339–347
34 Xu Y, Sun W. A filter successive linear programming method for nonlinear semidefinite programming problems. Numer Algebra Contr Optim, 2012, 2: 193–206
35 Zhang H, Conn A R, Scheinberg K. A derivative-free algorithm for least-squares minimization. SIAM J Optim, 2010, 20: 3555–3576