Recent advances in trust region algorithms

Ya-xiang Yuan

Volume 151 · Number 1 · June 2015

Mathematical Programming SERIES B

A Publication of the Mathematical Optimization Society

International Symposium on Mathematical
Programming, Pittsburgh, July 2015

Edited by Gérard Cornuéjols · Andrew J. Schaefer

Springer

Springer

Springer

CrossMark

Recent advances in trust region algorithms

Ya-xiang Yuan[1]

Abstract Trust region methods are a class of numerical methods for optimization. Unlike line search type methods where a line search is carried out in each iteration, trust region methods compute a trial step by solving a trust region subproblem where a model function is minimized within a trust region. Due to the trust region constraint, nonconvex models can be used in trust region subproblems, and trust region algorithms can be applied to nonconvex and ill-conditioned problems. Normally it is easier to establish the global convergence of a trust region algorithm than that of its line search counterpart. In the paper, we review recent results on trust region methods for unconstrained optimization, constrained optimization, nonlinear equations and nonlinear least squares, nonsmooth optimization and optimization without derivatives. Results on trust region subproblems and regularization methods are also discussed.

Keywords Trust region algorithms · Nonlinear optimization · Subproblem · Complexity · Convergence

Mathematics Subject Classification 65K05 · 90C30

1 Introduction

Trust region algorithms are a class of numerical methods for optimization, which have been extensively studied for many decades. Historical development of trust region methods can be traced back to Levenberg [72], where a modified Gauss–Newton method is given for solving nonlinear least square problems. Levenberg's method

✉ Ya-xiang Yuan
yyx@lsec.cc.ac.cn

[1] State Key Laboratory of Scientific/Engineering Computing, Institute of Computational
Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems
Science, Chinese Academy of Sciences, Zhong Guan Cun Donglu 55, Beijing 100190, China

was re-discovered independently by Morrison [80] and Marquardt [76]. The method was popularized due to the influential paper of Marquardt [76], thus it is called as Levenberg–Marquardt method (see [48]). Though Morrison's paper was not widely known because it was published in a conference proceeding instead of a journal, many leading experts now favor the name "Levenberg–Morrison–Marquadt method" as an acknowledgment of the important contribution of Morrison [80]. Levenberg–Marquardt method modifies Gauss–Newton method by adding a damping term to restrict the length of the Gauss–Newton step, in order to avoid a too long step when the Jacobian matrix of the nonlinear functions is nearly singular at the current point. Thus, a more imaginable name for trust region method is restricted step method which was used by Fletcher [48]. Pioneer researches on trust region methods were given by Powell [88–91], Fletcher [47], Hebden [67], Madsen [74], Osborne [86], Moré [78], Toint [118–122], Dennis and Mei [37], Sorensen [113,114], and Steihaug [115]. Moré [79] gave an excellent survey of early works on trust region methods, which popularized trust region methods and standardized the term "trust region". An invited plenary talk on trust region methods was given by Yuan [148] at the Fourth International Congress on Industrial and Applied Mathematics, which was held in Edinburgh, July 1999. The huge comprehensive monograph by Conn et al. [24] is the first published book on trust region methods, which gives a very good summary of works being done until the time of publication and is regarded as an encyclopedia of the subject.

Unlike line search algorithms, where a line search is carried out along a search direction in each iteration, trust region algorithms obtain the new iterate point by searching in a trust region which is normally a neighbourhood of the current iterate point. To be more exact, at the kth iteration, a trust region algorithm for the general optimization problem

$$\min_{x \in X} f(x), \tag{1.1}$$

where $f(x)$ is the objective function to be minimized and $X \subset \Re^n$ is the feasible set, obtains the trial step s_k by solving the following trust region subproblem

$$\min_{d \in X_k} \quad m_k(d) \tag{1.2}$$

$$\text{s.t.} \quad \|d\|_{W_k} \le \Delta_k, \tag{1.3}$$

where $m_k(d)$ is a model function that approximates the objective function $f(x_k + d)$ near the current iteration point x_k, X_k is an approximation to the shifted feasible set $X - x_k$, $\|.\|_{W_k}$ is a norm in $\Re^n$ and $\Delta_k > 0$ is the trust region radius.

The framework of a trust region method for optimization problem (1.1) is as follows

---

Framework of a trust region method for optimization

Step 1   Initialization, given x_1, construct $m_1(d)$, $\|.\|_{W_1}$ and $\Delta_1 > 0$, set $k := 1$.

Step 2   If convergence, then stop;
         Solve (1.2)–(1.3) giving the trial step s_k.

Step 3   Decide whether the trial step should be accepted (whether to set $x_{k+1} = x_k + s_k$ or
         $x_{k+1} = x_k$).

Step 4   Construct $m_{k+1}(d)$, choose Δ_{k+1} and $\|.\|_{W_{k+1}}$;
         Set $k := k + 1$; go to Step 2.

Thus, the essential parts of a trust region algorithm are the construction of the trust region subproblem and how to judge the trial step. In order to define the trust region subproblem, we need to build up an approximate model $m_k(d)$, and to select the trust region $\{d \in \Re^n \mid \|d\|_{W_k} \leq \Delta_k\}$. Once the trial step s_k is obtained by solving the trust region subproblem exactly or inexactly, we can use the information at the trial point $x_k + s_k$ to judge the quality of the trial step s_k. Based on certain criteria, if the trial point $x_k + s_k$ is (sufficiently) better than the current iterate x_k, we accept the trial step. At the end of each iteration, we need to choose the trust region for the next iteration and construct the new model function $m_{k+1}(d)$.

Due to the trust region constraint (1.3), the feasible region of a trust region subproblem is always a bounded set, which enables us to use nonconvex approximation model functions $m_k(d)$. This is one of the advantages of trust region methods over line search methods. The trust region constraint can be viewed as an implicit regularization and preconditioning can be used by selecting proper norm $\|.\|_{W_k}$. These make trust region algorithms reliable and robust, as they can be applied to nonconvex and ill-conditioned problems.

An advantage of trust region methods is that it is easier to establish convergence results for trust region algorithms than for line search algorithms. For example, in order to prove the global convergence of line search type quasi-Newton methods, one needs to use sophisticated techniques to estimate the growing of the quasi-Newton matrices (see Powell [92]), while the global convergence of trust region type quasi-Newton methods can be easily established as long as the trace of the quasi-Newton matrices does not increase faster than linearly (see Powell [90]). In [128], it is explained also why the assumptions for line searches are stronger. Another advantage of trust region methods over line search methods is that trust region methods using second order information can normally converges to second-order stationary points while line search type methods could converges to saddle points (for example, see [24] and [117]).

In this paper, we review recent results on trust region methods. In Sect. 2, we consider trust region methods for unconstrained optimization, particularly on results about subspace techniques, non-standard updating techniques for trust region radius, and non-standard shaped trust regions that are not defined by Euclidean balls. In Sect. 3, we discuss trust region methods for constrained optimization, where emphases are given on subspace techniques, and recent results on augmented Lagrangian function methods. Trust region techniques for nonlinear equations and nonlinear least squares are presented in Sect. 4. In Sect. 5, trust region algorithms for nonsmooth optimization are discussed. Trust region methods for optimization without derivatives are reviewed in Sect. 6 and some regularization methods are discussed in Sect. 7. A brief discussion is given at the end of the paper.

2 Unconstrained optimization

In this section, we consider unconstrained optimization, that is the case where the feasible region X is the whole space $\Re^n$. This case is fundamental and relatively

simple, because we can simply let $X_k = X = \Re^n$. Thus, the trust region subproblem in the kth iteration has the following form:

$$\min_{d \in \Re^n} \quad m_k(d) \tag{2.1}$$

$$\text{s.t.} \quad \|d\|_{W_k} \leq \Delta_k. \tag{2.2}$$

Let s_k be a trial step which is an exact or inexact solution of (2.1)–(2.2). We compute the predicted reduction $Pred_k = m_k(0) - m_k(s_k)$, which is the reduction in the model function, and the actual reduction of the objective function $Ared_k = f(x_k) - f(x_k + s_k)$. In general, if x_k is not a stationary point of $f(x)$, the predicted reduction $Pred_k > 0$. For example, a quadratic model function with the Euclidean norm trust region constraint leads to the standard trust region subproblem (TRS):

$$\min_{d \in \Re^n} \quad m_k(d) = f(x_k) + g_k^T d + \frac{1}{2} d^T B_k d \tag{2.3}$$

$$\text{s.t.} \quad \|d\|_2 \leq \Delta_k, \tag{2.4}$$

where $g_k = g(x_k) = \nabla f(x_k)$ and $B_k \in \Re^{n \times n}$ is a symmetric matrix approximating the Hessian of the objective function. Subproblem (2.3)–(2.4) is widely used in many trust region algorithms as it is easy to solve and it has nice theoretical properties including that the Hessian of the Lagrangian function of (2.3)–(2.4) is positive semi-definite [53,113], and that it has at most one non-global local minimizer [77]. For convex quadratic model functions $m_k(d)$, it is shown that the truncated conjugate gradient method can obtain a very good inexact solution of (2.3)–(2.4), namely the reduction of the objective function obtained by the truncated conjugate gradient method will be at least half of the maximum reduction in the trust region [149]. Let s_k be the exact solution of (2.3)–(2.4), we have that

$$Pred_k \geq \frac{1}{2} \|g_k\|_2 \min \left\{ \Delta_k, \ \frac{\|g_k\|_2}{\|B_k\|_2} \right\}, \tag{2.5}$$

which was first proved by Powell [91]. However, it is not necessary for s_k to be the exact solution of (2.3)–(2.4) to satisfy (2.5). For example, a minimizer of the quadratic model $m_k(d)$ in (2.3) along the steepest descent direction within the trust region (2.4) also satisfies (2.5). Inequalities of the type (2.5) indicate that certain sufficient reduction can be achieved in the approximate model provided that x_k is not a stationary point of the objective function. The quadratic model (2.3) is not the only possible model for $m_k(d)$. For example, the following conic model

$$m_k(d) = \frac{g_k^T d}{1 - h_k^T d} + \frac{1}{2} \frac{d^T B_k d}{(1 - h_k^T d)^2} \tag{2.6}$$

is used in a trust region algorithm given by Di and Sun [38], where $h_k \in \Re^n$. No matter what kind of model function $m_k(d)$ is chosen, we normally require that the trial step

s_k obtained satisfies some "sufficient descent" condition similar to (2.5), in order to establish the global convergence of the algorithm.

The ratio between the actual reduction and the predicted reduction

$$r_k = \frac{Ared_k}{Pred_k} \tag{2.7}$$

plays a very important role in trust region algorithms, since it gives a measure for judging how good the trial step s_k is. Intuitively, a larger r_k corresponds to a better trial step because it achieves a lower objective function value for the same predicted reduction $Pred_k$. Thus, a general strategy is to accept the trial step as long as r_k is larger than a preset non-negative constant.

The value of the trust region radius Δ_k decides the size of the trust region in which we replace the original problem by the approximate model. A larger trust region radius shows that we are more confident about our model while a smaller trust region indicates that we are more conservative. Thus, the ratio (2.7) is also used to adjust the new trust region radius.

For unconstrained optimization problem, a model trust region method is given as follows.

A Model Trust Region Method for Unconstrained Optimization

Step 1 Given $x_1 \in \Re^n$, $\Delta_1 > 0$, $\epsilon \geq 0$, $B_1 \in \Re^{n \times n}$ symmetric;
$0 < \tau_3 < \tau_4 < 1 < \tau_1$, $0 \leq \tau_0 \leq \tau_2 < 1$, $\tau_2 > 0$, $k := 1$.

Step 2 If $\|g_k\|_2 \leq \epsilon$ then stop; Solve (2.1)–(2.2) giving s_k.

Step 3 Compute $r_k = Ared_k/Pred_k$;

$$x_{k+1} = \begin{cases} x_k & \text{if } r_k \leq \tau_0 , \\ x_k + s_k & \text{otherwise ;} \end{cases} \tag{2.8}$$

Choose Δ_{k+1} that satisfies

$$\Delta_{k+1} \in \begin{cases} [\tau_3 \|s_k\|_{W_k}, \ \tau_4 \Delta_k] & \text{if } r_k < \tau_2, \\ [\Delta_k, \ \tau_1 \Delta_k] & \text{otherwise;} \end{cases} \tag{2.9}$$

Step 4 Choose a new model function $m_{k+1}(d)$ and select a norm $\|.\|_{W_{k+1}}$;
$k := k + 1$; go to Step 2.

Different choices for the parameters τ_i $(i = 0, 1, 2, 3, 4)$ are discussed in Yuan [148], Chapter 17.1 of Conn et al. [24], and Gould et al. [55]. Global convergence of the trust region method can be easily established as long as the trial step s_k satisfies a sufficient descent condition similar to (2.5). To prove the local superlinear convergence of a trust region method, the general approach is to show that the trust region constraint $\|d\|_{W_k} \leq \Delta_k$ is inactive for all sufficiently large k. In order to ensure $\|s_k\|_{W_k} < \Delta_k$, the standard technique is to show that $s_k \to 0$ and Δ_k is bounded away from zero. This explains why we set $\Delta_{k+1} \geq \Delta_k$ when $r_k \geq \tau_2$ in the model algorithm.

However, Δ_k bounded away from zero is only a sufficient condition for the local superlinear convergence, but not a necessary one. Therefore, it is worthwhile to study

techniques that allow $\Delta_k \to 0$. The reason for preferring $\Delta_k \to 0$ is that for ill-conditioned and even degenerate problems, it is more reliable and robust to have $\Delta_k \to 0$. Moreover, in trust region methods for derivative-free optimization, Δ_k also goes to zero, because of the so-called criticality step (for example, see [30]).

One approach is to let the new trust region radius Δ_{k+1} depend on s_k. Hei [68] suggests to update the trust region radius by

$$\Delta_{k+1} = R_\eta(r_k)\|s_k\|_2, \tag{2.10}$$

where $R_\eta(t)$ is a monotonically increasing function satisfying

$$0 < \beta \leq R_\eta(t) \leq 1 - \gamma_1 < 1, \quad \forall t \in (-\infty, \eta),$$
$$1 < 1 + \gamma_2 \leq R_\eta(t) \leq M < +\infty, \quad \forall t \in [\eta, +\infty),$$

with $0 < \beta < 1, 0 < \gamma_1 < 1 - \beta, \gamma_2 > 0, M > 1 + \gamma_2, 0 < \eta < 1$.

Suppose the algorithm converges to a local minimizer x^* of $f(x)$ at which the second order sufficient condition is satisfied. The algorithm is locally superlinearly convergent if we can prove that s_k is close to Newton's step $-(\nabla^2 f(x_k))^{-1} g_k$, namely,

$$\|s_k + (\nabla^2 f(x_k))^{-1} g_k\|_2 = o(\|s_k\|_2), \tag{2.11}$$

and that the trust region constraint is inactive ($\|s_k\|_2 < \Delta_k$). Assuming that the minimum eigenvalue of $\nabla^2 f(x^*)$ is ξ, (2.11) implies that $\|s_k\|_2 \leq \frac{2}{\xi}\|g_k\|_2$ for all sufficiently large k. Thus, it is reasonable to let Δ_k depend on $\|g_k\|_2$. Fan and Yuan [45] suggested to use

$$\Delta_k := \mu_k \|g_k\|_2 \tag{2.12}$$

where

$$\mu_{k+1} = \begin{cases} c_1 \mu_k & \text{if } r_k < \eta_2, \\ c_2 \mu_k & \text{if } r_k \geq \eta_2 \text{ and } \|s_k\|_2 > \frac{1}{2}\Delta_k, \\ \mu_k & \text{otherwise}, \end{cases}$$

with $0 < \eta_2 < 1, 0 < c_1 < 1 < c_2$ and $\mu_1 > 0$ being given (2.12) can be generalized to

$$\Delta_k = \mu_k \|g_k\|_2^\alpha,$$

where $\alpha \in (0, 1]$ is a parameter.

Recently, motivated by applications in adaptive techniques which exploit the information made available during the optimization process in order to vary the accuracy of the objective function computation, Bastin et al. [5] proposed an interesting way of updating the trust region radius. In their retrospective trust region algorithm, they use

the model function at the current iteration to compute the ratio of the actual reduction and the predicted reduction in the previous trial step, and use this ratio to update the trust region radius. Namely, if $x_{k+1} = x_k + s_k$, a new model function m_{k+1} is constructed. Then, the new ratio

$$\tilde{\rho}_k = \frac{f(x_k) - f(x_k + s_k)}{m_{k+1}(-s_k) - m_{k+1}(0)} \tag{2.13}$$

is computed, and the trust region radius is updated by the following formula:

$$\Delta_{k+1} = \begin{cases} \max[\gamma_2 \|s_k\|, \Delta_k] & \text{if } \tilde{\rho}_k \geq \tilde{\eta}_2, \\ \Delta_k & \text{if } \tilde{\rho}_k \in [\tilde{\eta}_1, \tilde{\eta}_2], \\ \gamma_1 \|s_k\| & \text{if } \tilde{\rho}_k \in [0, \tilde{\eta}_1), \\ \min[\gamma_1 \|s_k\|, \max[\gamma_0, \tilde{\theta}_k]\Delta_k] & \text{if } \tilde{\rho}_k < 0, \end{cases} \tag{2.14}$$

where

$$\tilde{\theta}_k = \frac{-(1 - \tilde{\eta}_2)\langle \nabla_x f(x_{k+1}), s_k \rangle}{(1 - \tilde{\eta}_2)[f(x_{k+1}) - \langle \nabla_x f(x_{k+1}), s_k \rangle] + \tilde{\eta}_2 m_{k+1}(-s_k) - f(x_k)},$$

and the constants satisfy $0 < \gamma_0 < \gamma_1 \leq \gamma_2 > 1$ and $0 < \tilde{\eta}_1 \leq \tilde{\eta}_2 < 1$. $\tilde{\theta}_k \Delta_k$ is used to ensure $\tilde{\rho}_{k+1} \geq \tilde{\eta}_2$. Please notice that $\tilde{\theta}_k$ uses the gradient at x_{k+1}.

For large-scale optimization problems, namely when n is very large, trust region subproblems are also large-scale. Thus, it is valuable to investigate subspace properties of trust region subproblems. Consider the standard trust region quasi-Newton method where the model functions $m_k(d)$ are quadratic functions with the Hessian matrices being updated by quasi-Newton formulae. Wang and Yuan [130] gave the subspace properties of the trial steps s_k:

Lemma 1 *Suppose $B_1 = \sigma I$, $\sigma > 0$. The matrix updating formula is any one chosen from PSB and Broyden family (where the updates may be singular), and B_k is the kth update matrix. Let s_k be a solution of subproblem* (2.3)–(2.4)*, $g_k = \nabla f(x_k)$, $\mathcal{G}_k = Span\{g_1, g_2, \ldots, g_k\}$ and $x_{k+1} = x_k + s_k$ ($x_1 \in \Re^n$ is any given initial point). Then for all $k \geq 1$, $s_k \in \mathcal{G}_k$. Moreover for any $z \in \mathcal{G}_k$ and any $u \in \mathcal{G}_k^\perp$, we have*

$$B_k z \in \mathcal{G}_k, \qquad B_k u = \sigma u. \tag{2.15}$$

This result extends a similar result for line search type quasi-Newton methods given by Siegel [112]. Based on this result, it can be shown that the full space trust region quasi-Newton method is equivalent to the subspace counterpart provided that the approximate Hessian is updated by quasi-Newton formulae and the initial Hessian approximation is appropriately chosen. In the subspace algorithm, the trial step can be obtained by solving the trust region subproblem in the subspace spanned by all the gradient vectors computed so far. To be more exact, the trial step can be defined by minimizing the quasi-Newton quadratic model in the subspace spanned by all the computed gradients. The advantage of subspace trust region algorithms is that the subproblems are defined in lower dimension, thus they are more suitable for large scale

problems. Subproblems of general subspace trust region algorithms for unconstrained optimization have the following form

$$\min_{d \in \Re^n} \quad m_k(d) \tag{2.16}$$

$$\text{s.t.} \quad \|d\|_{W_k} \le \Delta_k, \quad d \in \mathcal{S}_k, \tag{2.17}$$

where $\mathcal{S}_k \subset \Re^n$ is a low dimensional subspace. Possible choices for $\mathcal{S}_k$ are

$$\mathcal{S}_k = Span\{g_1, g_2, ..., g_k\}, \tag{2.18}$$

$$\mathcal{S}_k = Span\{g_k, s_{k-1}, ..., s_{k-r}\}, \tag{2.19}$$

$$\mathcal{S}_k = Span\{g_k, y_{k-1}, ..., y_{k-r}\}, \tag{2.20}$$

and

$$\mathcal{S}_k = Span\left\{e_{k(1)}, e_{k(2)}, ..., e_{k(r)}\right\}, \tag{2.21}$$

where $r \ll n$ is a non-negative integer, $y_i = g_{i+1} - g_i$, and e_i is the ith coordinate unit vector in $\Re^n$. For more detailed discussions on subspace selection, please see Yuan [151].

Regarding the choices for the norm $\|.\|_{W_k}$ in the trust region subproblem, it is common to use the Euclidean norm $\|.\|_2$, since the Euclidean norm trust region subproblem (2.3)–(2.4) is widely studied and easy to solve. However, in some special cases, we can use different norm $\|.\|_{W_k}$ other than the Euclidean norm $\|.\|_2$. For example, for the quadratic model function (2.3) with B_k being updated by limited memory quasi-Newton updates (for example, see [73]), the matrix B_k has the form

$$B_k = \sigma_k I + P_k D_k P_k^T, \tag{2.22}$$

where $\sigma_k > 0$, $P_k \in \Re^{n \times r}$ satisfying $P_k^T P_k = I_r$ and $D_k \in \Re^{r \times r}$ is diagonal and positive definite. Normally, r is much smaller than n. Burdakov et al. [9] suggested to use the norm

$$\|d\|_{W_k} = \max\left\{\|P_k^T d\|_\infty, \ \|(P_k)_\perp^T d\|_2\right\}, \tag{2.23}$$

where $(P_k)_\perp^T$ is the projection onto the subspace orthogonal to $Range(P_k)$. In this case, the solution s_k for the trust region subproblem (2.1)–(2.2) can be expressed (by orthogonal decomposition) as $P_k s_1 + (P_k)_\perp s_2$. Both s_1 and s_2 have closed-form solutions [151], due to the fact that s_1 is the solution of the box constrained QP (in $\Re^r$)

$$\min_{s \in \Re^n} \quad \frac{1}{2} s^T (\sigma_k I + D_k) s + s^T \left(P_k^T g_k\right) \tag{2.24}$$

$$\text{s.t.} \quad \|s\|_\infty \le \Delta_k \tag{2.25}$$

while s_2 solves the 2-norm constrained special QP

$$\min_{s \in \Re^n} \frac{1}{2}\sigma_k s^T s + s^T (P_k)_{\perp}^T g_k \tag{2.26}$$

$$\text{s.t.} \quad \|s\|_2 \leq \Delta_k . \tag{2.27}$$

Encouraging numerical results for trust region algorithms using the cylinder-shape trust region (2.23) were reported in Gong [54] and Zikrin [158].

A class of recursive trust region methods for multiscale nonlinear optimization, often arising in systems governed by partial differential equations, is given by by Gratton et al. [63]. The standard trust region method for unconstrained optimization is a special case of this class. A remarkable result obtained by Gratton, Sartenaer and Toint [63] is that the unconstrained trust region algorithm needs at most $O(\epsilon^{-2})$ iterations to achieve $||g(x_k)||$ below ϵ.

3 Constrained optimization

Normally, trust region methods for constrained optimization problem (1.1) are more complicated than line search methods due to the difficulty that the approximate feasible region X_k may not have points in the trust region of the current iteration, namely $x_k + d \notin X_k$ as long as $\|d\|_{W_k} \leq \Delta_k$. Consider the case that the feasible set X is defined by equality and inequality constraints

$$X = \{x \mid c_i(x) = 0, \ i = 1, ..., \ m_e; c_i(x) \geq 0, \ i = m_e + 1, ..., m\} . \tag{3.1}$$

At the kth iteration point x_k, line search type SQP methods normally compute the search direction d_k by solving the QP subproblem:

$$\min \ m_k(d) \tag{3.2}$$

$$\text{s.t.} \ d \in X_k \tag{3.3}$$

where $m_k(d)$ is a quadratic function approximating the Lagrangian function and

$$X_k = \{d \mid c_i(x_k) + d^T \nabla c_i(x_k) = 0, i = 1, ..., m_e;$$
$$c_i(x_k) + d^T \nabla c_i(x_k) \geq 0, i = m_e + 1, ..., m\} \tag{3.4}$$

is the feasible set of the linearized constraints. To overcome the possible difficulty that $X_k \cap \{d \mid \|d\|_{W_k} \leq \Delta_k\} = \emptyset$, instead of using (1.2)–(1.3), we can use the following trust region subproblem

$$\min_{d \in \Re^n} \ m_k(d) \tag{3.5}$$

$$\text{s.t.} \quad \|d\|_{W_k} \leq \Delta_k \tag{3.6}$$

$$d \in \theta_k X_k, \tag{3.7}$$

for some $\theta_k \in (0, 1]$. It is easy to see that $\theta_k X_k \cap \{d \mid \|d\|_{W_k} \leq \Delta_k\} \neq \emptyset$ for sufficiently small $\theta_k \in (0, 1]$ provided that $\Delta_k > 0$ and X_k is nonempty. In the special case of equality constrained optimization problems, subproblem (3.5)–(3.7) is equivalent to null-space approach, detailed discussions can be found in Bryd et al. [12], Vardi [126] and Omojukun [85]. Normally, the model function $m_k(d)$ in (3.5) is a quadratic function. Similar to unconstrained case, it is also possible to use non-quadratic models such as conic models [117].

Intuitively, we should choose $\theta_k \in (0, 1]$ in (3.7) as large as possible since a larger θ_k corresponding to the case that $\theta_k X_k$ is a better approximation to the shifted feasible set $X - x_k$. Thus, one way to achieve this is to replace θ_k by a variable $\theta \in (0, 1]$ and penalize the term $(1 - \theta)$ in the object. Namely we can use the following subproblem

$$\min_{d \in \Re^n} \quad m_k(d) + \sigma(1 - \theta)^2 \tag{3.8}$$

$$\text{s.t.} \quad \|d\|_{W_k} \leq \Delta_k, \tag{3.9}$$

$$d \in \theta X_k, \quad \theta \in (0, 1]. \tag{3.10}$$

The shifted linearized feasible region θX_k, can be viewed as a null space of the linearized constraints. All the trust region algorithms that obtain the trial step by computing a range-space step (also called vertical step or normal step) and a null-space step (also called horizontal step or tangential step) can be regarded as trying to satisfy (3.7). Along this direction, a trust-funnel algorithm is given by Gould and Toint [58] where neither a penalty function nor a filter [52] is used. The algorithm converges to first-order stationary points by using different trust regions to cope with the nonlinearities of the objective function and the constraints and allowing inexact SQP steps that do not lie exactly in the null space of the constraints. Recently, a new algorithm is given by Curtis et al. [33] where global convergence is achieved by combining a trust region methodology with a funnel mechanism similar to [58]. The new algorithm has the additional capability that it solves problems with both equality and inequality constraints, and has prominent features such as that the subproblems that define each search direction may be solved approximately, that inexact sequential quadratic optimization steps may be utilized when advantageous, and that it may be implemented matrix-free so that derivative matrices need not be formed or factorized so long as matrix-vector products with them can be performed.

The condition $d \in \theta X_k$ is one particular way of approximating the linearized constraints. For trust region SQP algorithms that use null-space approach, the normal steps try to reduce the constraint violations whilst the tangential steps try to reduce the objective function or the Lagrangian function. In order to ensure convergence, conditions have to be imposed on normal steps and tangential steps. A detailed study was made on these condition in [69], where convergence results can be extended to the case that allow the use of inexact problem information originating from inexact first-order derivative information.

Another type of trust region subproblems has the following form

$$\min_{d \in \Re^n} \quad \phi_k(d) \tag{3.11}$$

$$\text{s.t.} \quad \|d\|_{W_k} \leq \Delta_k, \tag{3.12}$$

where $\phi_k(d)$ is a model function approximating some penalty function. For example, Fletcher [50] suggests the following SL_1QP type trust region subproblem

$$\min_{d \in \Re^n} \quad g_k^T d + \frac{1}{2} d^T B_k d + \sigma_k \left\| \left(c_k + A_k^T d \right)^{(-)} \right\|_1 \qquad (3.13)$$

$$\text{s.t.} \quad \|d\|_\infty \leq \Delta_k, \qquad (3.14)$$

where $(c_k + A_k^T d)^{(-)}$ is the vector of violations of the linearized constraints defined by

$$(c_k + (A_k)^T d)_i^{(-)} = \begin{cases} c_i(x_k) + d^T \nabla c_i(x_k) & i = 1, ..., m_e, \\ \min\{0, \ c_i(x_k) + d^T \nabla c_i(x_k)\} & i = m_e + 1, ..., m, \end{cases} \qquad (3.15)$$

where A_k is the Jacobian of the constraints computed at x_k, The above trust region subproblem for constrained optimization can be viewed as the trust region subproblem for the L_1 exact penalty of the constrained problem. A similar subproblem based on the L_∞ exact penalty function is given by Yuan [146]. The L_∞ exact penalty function, comparing with the L_1 exact penalty, can reduce the maximum violation of constraints more quickly, whilst possibly increasing the violations overall. Trust region subproblems based on exact penalty functions are closely related to subproblems of trust region algorithms for nonlinear equations [40, 147] and nonlinear least squares that will be discussed in the next section. Trust region algorithms that use (3.13)–(3.14) are also similar to trust region algorithms for composite nonsmooth optimization [50, 51, 140–142].

For equality constrained optimization problems, Niu and Yuan [82] suggested a quadratic model function $Q_k(d)$ which approximates the augmented Lagrangian function at the kth iteration, and used the subproblem

$$\min_{d \in \Re^n} \quad Q_k(d) \qquad (3.16)$$

$$\text{s.t.} \quad \|d\|_2 \leq \Delta_k. \qquad (3.17)$$

One nice property of this approach is that (3.16)–(3.17) is exactly in the form of the standard trust region subproblem for unconstrained optimization (2.3)–(2.4). The model function in (3.16) is given by

$$Q_k(d) = (g_k - A_k \lambda_k)^T d + \frac{1}{2} d^T B_k d + \sigma_k \|c_k + A_k^T d\|_2^2, \qquad (3.18)$$

where λ_k is the current approximate Lagrange multiplier and $\sigma_k > 0$ is the penalty parameter. In the algorithm of Niu and Yuan [82], the trial step is computed by solving (3.16)–(3.17). Then, the augmented Lagrange function is also used as a merit function to decide whether the trial step should be accepted. Their method extends the traditional trust region approach by combining a filter technique into the rules for accepting trial steps so that a trial step could still be accepted even when it is rejected by the traditional rule based on the merit function reduction. An estimate of the Lagrange multiplier is

updated at each iteration, and the penalty parameter is updated to force sufficient reduction in the norm of the constraint violations. An active set technique is used to handle the inequality constraints. Numerical results for a set of constrained problems from the CUTEr collection [56] indicate that this approach is very efficient. However, no convergence results are given in Niu and Yuan [82] about their algorithm. Another possible way of handling inequality constraints is to use

$$Q_k(d) = (g_k - A_k \lambda_k)^T d + \frac{1}{2} d^T B_k d + \sigma_k \|(c_k + A_k^T d)^{(-)}\|_2^2. \tag{3.19}$$

The idea of approximating augmented Lagrangian by a quadratic function is also used in the trust region algorithm by Curtis et al. [34], in which global convergence is analyzed. However, when constraint violations converge to zero and the penalty parameter varies infinitely many times, first-order criticality could not be guaranteed, unless a much stronger assumption is enforced on the update of penalty parameters.

A modification to Niu and Yuan's algorithm is given by Wang and Yuan [132] for equality constrained optimization, where a condition is introduced to decide whether the Lagrange multiplier should be updated. It also suggests a new strategy for adjusting the penalty parameters. Global convergence of the modified method is established under mild conditions by Wang and Yuan [132], in which the behavior of penalty parameters is analyzed. Recently, an extended trust region algorithm is given by Wang and Zhang [133] for nonlinear optimization with both equality constraints and simple bounds. At each iteration, an affine scaling trust region subproblem is constructed, where affine scaling technique is applied to handle bounds constraints. By adaptive update strategies of Lagrange multipliers and penalty parameters, global convergence can be achieved under mild conditions.

The shifted feasible set X_k defined by (3.4) is the set containing all the points that satisfy the linearized constraints. We can equivalently write it as

$$X_k = \left\{ d \mid (c_k + A_k^T d)^{(-)} = 0 \right\}. \tag{3.20}$$

Thus, in order to handle the case when X_k has no intersection with the trust region, instead of requiring zero residual of the linearized constraints, we can set an upper bound for the sum of squares of violations of the linearized constraints. Namely, we can let

$$X_k = \left\{ d \mid \|(c_k + A_k^T d)^{(-)}\|_2^2 \le \xi_k \right\}, \tag{3.21}$$

where $\xi_k \ge 0$ is a parameter. If

$$\xi_k \ge \min_{\|d\|_{W_k} \le \Delta_k} \|(c_k + A_k^T d)^{(-)}\|_2^2,$$

X_k defined by (3.21) has nonempty intersection with the trust region. Thus, it is reasonable to consider the following trust region subproblem

$$\min_{d \in \Re^n} \quad m_k(d) \tag{3.22}$$

$$\text{s.t.} \quad \|(c_k + A_k^T d)^{(-)}\|_2^2 \le \xi_k \tag{3.23}$$

$$\|d\|_{W_k} \le \Delta_k . \tag{3.24}$$

The above subproblem with quadratic model $m_k(d)$ and $\|.\|_{W_k} = \|.\|_2$, called Celis–Dennis–Tapia (CDT) subroblem, is given by Celis et al. [18] and used by Powell and Yuan [107]. Properties of the CDT subproblem are discussed by many authors [2,19,87,116,144], and algorithms for solving the CDT subproblem are given by [145,156]. Recently, Bienstock [6] shows that the CDT problem is polynomial-time solvable.

Subspace properties of the CDT problem are studied by Grapiglia, Yuan and Yuan [59], which establishes the following result.

Lemma 2 *Let S_k be an r ($1 \le r \le n$) dimensional subspace in $\Re^n$, and $Z_k \in \Re^{n \times r}$ be an orthonormal basis matrix of S_k, namely*

$$S_k = span\{Z_k\}, \quad Z_k^T Z_k = I_r.$$

Suppose that

$$\{\nabla c_1(x_k), ..., \nabla c_m(x_k), g_k\} \subset S_k$$

and $B_k \in \Re^{n \times n}$ is a symmetric matrix satisfying

$$B_k u = \sigma u, \quad \forall u \in S_k^{\perp},$$

where $\sigma > 0$. Then, the CDT subproblem

$$\min_{d \in \Re^n} \quad g_k^T d + \frac{1}{2} d^T B_k d \tag{3.25}$$

$$\text{s.t.} \quad \|c_k + A_k^T d\|_2 \le \xi_k \tag{3.26}$$

$$\|d\|_2 \le \Delta_k \tag{3.27}$$

is equivalent to the following problem:

$$\min_{\bar{d} \in \Re^r} \quad \bar{g}_k^T \bar{d} + \frac{1}{2} \bar{d}^T \bar{B}_k \bar{d} \tag{3.28}$$

$$\text{s.t.} \quad \|c_k + \bar{A}_k^T \bar{d}\|_2 \le \xi_k \tag{3.29}$$

$$\|\bar{d}\|_2 \le \Delta_k, \tag{3.30}$$

where $\bar{g}_k = Z_k^T g_k$, $\bar{B}_k = Z_k^T B_k Z_k$, and $\bar{A}_k = Z_k^T A_k$.

Due to the above lemma, a subspace version of the Powell-Yuan trust region algorithm is presented in Grapiglia, Yuan and Yuan [59]. Comparing with the full space implementation, the amount of computation of the subspace version algorithm can be

reduced when the number of constraints is much smaller than the number of variables, especially in the early iterations. The subspace properties of the CDT subproblem described in the above lemma can be used in the same way to construct a subspace version of the CDT trust region algorithm for equality constrained optimization proposed by Celis et al. [18], as well of any algorithm based on the CDT subproblem.

Trust region interior-point methods are important class of numerical methods for general constrained optimization. Such methods are based on solving the KKT systems of barrier functions with trust region techniques. For more detailed discussions on trust region interior-point methods, please see Chapter 13 of [24]. The famous software package KNITRO [11] is an excellent example of the success of trust region interior-point methods. Other recent works on trust region interior-point methods include [129], [139] and [71].

When the constraints are simple bounds, we have the general bound constrained optimization problem

$$\min_{x \in \Re^n} \quad f(x) \tag{3.31}$$

$$\text{s.t.} \quad l \leq x \leq u, \tag{3.32}$$

where $l < u$ are two vectors in $\Re^n$. If x_k is an interior point, namely

$$l < x_k < u, \tag{3.33}$$

Coleman and Li [22,23] suggested the following affine scaling trust region subproblem

$$\min_{d \in \Re^n} \quad m_k(d) \tag{3.34}$$

$$\text{s.t.} \quad \|D_k^{-1} d\|_2 \leq \Delta_k \tag{3.35}$$

where D_k is a diagonal positive definite matrix. In Coleman and Li [22,23], the affine scaling matrix D_k is defined by

$$(D_k)_{ii} = \sqrt{|v^{(i)}(x_k)|}, \quad i = 1, \ldots, n, \tag{3.36}$$

where $v^{(i)}(x)$ is defined by

$$v^{(i)}(x) = \begin{cases} x^{(i)} - u^{(i)} & \text{if } (g(x))_i < 0 \quad \text{and} \quad u^{(i)} < \infty, \\ x^{(i)} - l^{(i)} & \text{if } (g(x))_i \geq 0 \quad \text{and} \quad l^{(i)} > -\infty, \\ -1 & \text{if } (g(x))_i < 0 \quad \text{and} \quad u^{(i)} = \infty, \\ 1 & \text{if } (g(x))_i \geq 0 \quad \text{and} \quad l^{(i)} = -\infty. \end{cases} \tag{3.37}$$

The derivation of the scaling matrix (3.36) is based on the Newton's method for the diagonal system $D(x)^2 g(x) = 0$.

Wang and Yuan [131] suggested a new scaling matrix:

$$(D_k)_{ii} = \begin{cases} t_k \cdot \sqrt{\dfrac{a_k^{(i)}}{(g_k)_i}}, & i \in \mathcal{S}_k^1, \\[3mm] t_k \cdot \sqrt{\dfrac{b_k^{(i)}}{|(g_k)_i|}}, & i \in \mathcal{S}_k^2, \\[3mm] 1, & i \notin \mathcal{S}_k \triangleq \mathcal{S}_k^1 \cup \mathcal{S}_k^2. \end{cases} \tag{3.38}$$

where

$$t_k = \frac{\sqrt{\displaystyle\sum_{i \in \mathcal{S}_k^1} a_k^{(i)}(g_k)_i + \sum_{i \in \mathcal{S}_k^2} b_k^{(i)}|(g_k)_i|}}{\Delta_k}, \tag{3.39}$$

with the vectors a_k and b_k being defined componentwise by

$$a_k^{(i)} = x_k^{(i)} - l^{(i)}, \quad b_k^{(i)} = u^{(i)} - x_k^{(i)}, \quad i = 1, \ldots, n,$$

and the two index sets are defined by

$$\mathcal{S}_k^1 = \left\{ i \mid a_k^{(i)} \le \Delta_k, \ (g_k)_i \ge \epsilon a_k^{(i)} \right\}, \quad \mathcal{S}_k^2 = \left\{ i \mid b_k^{(i)} \le \Delta_k, \ -(g_k)_i \ge \epsilon b_k^{(i)} \right\}, \tag{3.40}$$

with $\epsilon \ll 1$ being a small positive constant. The scaling matrix (3.38) depends on the distances of the current iterate to the boundaries, the gradient of the objective function and the trust region radius. The motivation for choosing the scaling matrix by (3.38) is that the trust region step s_k would head to the solution if the objective function is linear, which is proved in [131]. To be more exact, if the object function $f(x) = c^T x$ with $c_i \ne 0$ for all $i = 1 : n$, the minimizer s_k of problem (3.34)–(3.35) with $m_k(d) = c^T d$ and the scaling matrix D_k given by (3.38) implies that $x_k + s_k = x^*$ when x_k is sufficiently close to x^* which is the solution of (3.31)–(3.32).

4 Nonlinear equations and nonlinear least squares

In this section we consider trust region algorithms for nonlinear equations

$$F(x) = (f_1(x), f_2(x), \ldots, f_m(x))^T = 0, \tag{4.1}$$

where $F(x)$ is a nonlinear mapping from $\Re^n$ to $\Re^m$, and a closely related problem, the nonlinear least squares problem,

$$\min_{x \in \Re^n} \| F(x) \|_2^2. \tag{4.2}$$

Normally, a trust region algorithm for nonlinear Eq. (4.1) tries to minimize some penalty function $h(F(x))$. Thus, trust region algorithms for composite nonsmooth

optimization [50,51,140–142] can be applied. The penalty function $h(\cdot)$ satisfies that $h(0) = 0$ and

$$h(F) > 0, \qquad \forall F \neq 0, \ F \in \Re^m . \tag{4.3}$$

Thus, a general trust region subproblem for nonlinear equations is:

$$\min_{d \in \Re^n} \ h(F_k + J_k d) + \frac{1}{2} d^T B_k d \tag{4.4}$$

$$\text{s.t.} \quad \|d\|_{W_k} \leq \Delta_k, \tag{4.5}$$

where $F_k = F(x_k)$ and $J_k = [\nabla f_1(x_k), \nabla f_2(x_k), ..., \nabla f_m(x_k)]^T \in \Re^{m \times n}$ is the Jacobian matrix of $F(x)$ computed at x_k. Pioneer works on trust region algorithms for nonlinear equations and nonlinear least squares are given by Powell [90], Madsen [74], Moré [78], Duff et al. [40]. In these early papers, B_k is set to be the zero matrix for all k, and $h(\cdot)$ is chosen as a simple penalty function such as $\|.\|_2$, $\|.\|_1$ or $\|.\|_\infty$. El Hallabi and Tapia [41] considered the case where $h(F) = \|F\|_\alpha$ is a general norm of F. The advantage of using a zero B_k is that the subproblem is easy to solve as it can be converted into a linear programming problem when $\|d\|_{W_k} = \|d\|_\infty$ or $\|d\|_{W_k} = \|d\|_1$. However, the price that we have to pay for $B_k = 0$ is that normally the convergence rate is at most linear. Powell and Yuan [106] studied conditions for superlinear convergence when $h(F) = \|F\|_1$ and $h(F) = \|F\|_\infty$.

Under certain non-degeneracy conditions, fast convergence results can be proved even if we let $B_k \equiv 0$ for all k. Consider the special case $h(F) = \|F\|_2^2$ and $\|.\|_{W_k} = \|.\|_2$, the trust region subproblem (for nonlinear equations and nonlinear least squares) becomes

$$\min_{d \in \Re^n} \ \|F_k + J_k d\|_2^2 \tag{4.6}$$

$$\text{s.t.} \quad \|d\|_2 \leq \Delta_k . \tag{4.7}$$

Instead of using the standard technique [such as (2.9)] for updating trust region radii, Fan [42] suggests to use

$$\Delta_k = \mu_k \|F_k\|_2^\delta, \tag{4.8}$$

where $\delta \in (\frac{1}{2}, 1)$. Let s_k be the solution of (4.6)–(4.7) with Δ_k being defined by (4.8), and let

$$\rho_k = \frac{Ared_k}{Pred_k} = \frac{\|F_k\|_2^2 - \|F(x_k + s_k)\|_2^2}{\|F_k\|_2^2 - \|F_k + J_k s_k\|_2^2} .$$

Fan [42] updates the parameter μ_{k+1} by

$$\mu_{k+1} = \begin{cases} c_5\mu_k & \text{if } \rho_k < c_2, \\ \mu_k & \text{if } \rho_k \in [c_2, c_4], \\ \min\{c_6\mu_k, M\} & \text{otherwise,} \end{cases}$$

where $0 < c_2 < c_4 < 1$, $0 < c_5 < 1 < c_6$, $M \gg c_6\mu_1$ and $\mu_1 > 0$ are given. Fan [42] proves that the algorithm has the local 2δ order convergence rate in the sense that

$$\text{dist}(x_{k+1}, X^*) \le \text{dist}(x_k, X^*)^{2\delta}, \tag{4.9}$$

where X^* is the solution set of $F(x) = 0$ if the local error bound condition holds. A similar technique is given by Zhang and Wang [154] in which the trust region radius has the form

$$\Delta_k = \alpha^p \|F_k\|_2^\delta, \tag{4.10}$$

where $\alpha \in (0, 1)$ and p is the smallest positive integer that gives an acceptable trial step s_k. (4.10), as well as (4.8), implies that the trust region radius converges to zero when $k \to \infty$. Fan and Pan [44] proposed to use

$$\Delta_k = \alpha^p M \|F_k\|_2,$$

where $0 < \alpha < 1$, $M \gg 0$, and obtained local quadratic convergence rate:

$$\|x_{k+1} - x^*\| \le O(\|x_k - x^*\|^2).$$

Recently, Fan and Lu [43] presented a new trust region algorithm which improves the algorithm given in Fan [42]. At each iteration, the new algorithm needs to compute the function value at an extra point and to solve an additional trust region subproblem. The algorithm is described as follows. At the kth iteration, the algorithm obtains s_k by solving the trust region subproblem:

$$\min_{s \in \mathbb{R}^n} \|F_k + J_k s\|_2^2$$
$$s.t. \ \|s\| \le \Delta_k := \mu_k \|F_k\|_2^\delta$$

where $\frac{1}{2} < \delta < 1$. Denote $y_k = x_k + s_k$. Then, another subproblem:

$$\min_{s \in \mathbb{R}^n} \|F(y_k) + J_k s\|_2^2,$$
$$s.t. \ \|s\| \le \bar{\Delta}_k := \mu_k \|F(y_k)\|_2^\delta$$

is solved to obtain $\bar{s}_k$. Define the new predicted reduction by

$$Pred_k = \|F_k\|_2^2 - \|F_k + J_k s_k\|_2^2 + \|F(y_k)\|_2^2 - \|F(y_k) + J_k \bar{s}_k\|_2^2,$$

and $\rho_k = \dfrac{Ared_k}{Pred_k} = \dfrac{\|F_k\|_2^2 - \|F(x_k + s_k + \bar{s}_k)\|_2^2}{Pred_k}$. The algorithm then updates x_{k+1} and μ_{k+1} by

$$x_{k+1} = \begin{cases} x_k + s_k & \text{if } \rho_k \geq p_0, \\ x_k & \text{otherwise,} \end{cases}$$

and

$$\mu_{k+1} = \begin{cases} 4\mu_k & \text{if } \rho_k < \eta_1, \\ \mu_k & \text{if } \rho_k \in [\eta_1, \eta_2], \\ \max\left\{\dfrac{\mu_k}{4}, m\right\} & \text{if } \rho_k > \eta_2, \end{cases}$$

where $m > 0$ is a small constant. Under the local error bound condition, it can be shown that the modified trust region algorithm has the following property:

$$\text{dist}(x_{k+1}, X^*) \leq \text{dist}(x_k, X^*)^{2\delta^2 + \delta}. \tag{4.11}$$

This result improves (4.9) because $2\delta^2 + \delta > 2\delta$ and the modified trust region algorithm is superlinearly convergent (in the sense of the rate of the distance $\text{dist}(x_k, X^*)$ converging to zero) for all $\delta \in (\frac{1}{2}, 1)$. Moreover, the convergence rate of $\text{dist}(x_k, X^*)$ is nearly cubic when δ is very close to 1.

Convergence results of several methods, including those allow the trust region radius to converge to zero, can be covered by a unified convergence theory given by Toint [124] under the framework of nonlinear stepsize control.

A derivative-free algorithm for nonlinear least squares was give by Zhang, Conn and Scheinberg [153]. Its local convergence properties are studied in [152].

5 Nonsmooth optimization

Non-smooth optimization problems are special problems where either the objective function or some constraints are nonsmooth. Trust region methods for nonsmooth optimization were first given by Fletcher [49–51], where the following composite nonsmooth optimization problem

$$\min_{x \in \Re^n} h(F(x)) \tag{5.1}$$

is studied, where $F(x) = (f_1(x), ..., f_m(x))^T$ is a continuously differentiable mapping from $\Re^n$ to $\Re^m$ and $h(\cdot)$ is a nonsmooth convex function defined on $\Re^m$. Other pioneering researches on trust region algorithms for nonsmooth optimization include Powell [93], Burke [10] and Yuan [140–143]. A trust region subproblem for composite nonsmooth optimization problem (5.1) is

$$\min_{d \in \Re^n} \ h(F(x_k) + J_k d) + \frac{1}{2} d^T B_k d \tag{5.2}$$

$$\text{s.t.} \ \|d\|_{W_k} \le \Delta_k, \tag{5.3}$$

where $J_k = [\nabla f_1(x_k), \dots \nabla f_m(x_k)]^T$ is the Jacobian matrix of $F(x)$ at x_k. The composite nonsmooth optimization (5.1) is studied by Sampaio, Yuan and Sun [110] where it is assumed that $F(x)$ is a nonsmooth locally Lipschitzian function while $h(\cdot)$ is a continuously differentiable function.

For the general nonsmooth optimization problem

$$\min_{x \in \Re^n} f(x) \tag{5.4}$$

where $f(x) : \Re^n \to \Re$ is a nonsmooth function, we can build at each iteration a model $m_k(d) = m(x_k, p_k)(d)$ which is an approximation of $f(x_k + d)$ for small d, and $p_k \in \Re^l$ is an l-dimensional parameter vector which may change from iteration to iteration [36]. Consequently, the trust region subproblem at the kth iteration is

$$\min \ m(x_k, p_k)(d) \tag{5.5}$$

$$\text{s.t.} \ \|d\|_{W_k} \le \Delta_k. \tag{5.6}$$

The Clarke generalized directional derivative [21] of f at x in the direction of $d \in \Re^n$ is defined by

$$f^\circ(x; d) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + td) - f(y)}{t}. \tag{5.7}$$

Under some conditions which include

$$m(x, p)(0) = f(x), \tag{5.8}$$

and

$$m(x, p)^\circ(0; d) = f^\circ(x; d) \quad \forall d \in \Re^n \backslash \{0\}, \tag{5.9}$$

Dennis et al. [36] proved the convergence of a general trust region method for problem (5.4), where the trail step s_k is obtained by solving subproblem (5.5)–(5.6) approximately. The results in [36] are extended by Qi and Sun [108] where the trust region subproblem

$$\min \ f(x_k) + \phi(x_k, d) + \frac{1}{2} d^T B_k d \tag{5.10}$$

$$\text{s.t.} \ \|d\|_2 \le \Delta_k \tag{5.11}$$

is used. It is assumed in [108] that the function $\phi(x_k, d)$ satisfies

$$f(x_k + s_k) - f(x_k) \le \phi(x_k, s_k) + o(\|s_k\|_2) \tag{5.12}$$

if $s_k \to 0$, for any convergent subsequence $\{x_k\}$.

By converting nonsmooth convex minimization problems into differentiable convex optimization problems through Moreau-Yosida regularation, Zhang [155] gives a trust region algorithm for nonsmooth convex optimization. The Moreau-Yosida regularization of $f(x)$ is defined by

$$f_\lambda(x) = \min_{z \in \Re^n} \left\{ f(z) + \frac{1}{2\lambda} \|z - x\|_2^2 \right\}, \tag{5.13}$$

where $\lambda > 0$ is a parameter.

In recent years, regularized minimization problems with nonconvex, nonsmooth, perhaps non-Lipschitz penalty functions have attracted considerable attention, owing to their wide applications in image restoration, signal reconstruction, and variable selection. Chen et al. [20] studied the following nonsmooth unconstrained minimization problem

$$\min_{x \in \Re^n} f(x) := \theta(x) + \lambda \sum_{i=1}^m \varphi\left(|d_i^T x|\right), \tag{5.14}$$

where $\theta : \Re^n \to \Re_+$ is twice continuously differentiable $\lambda \in \Re_+$, $d_i \in \Re^n$, $i = 1, \ldots, m$, and $\varphi : \Re_+ \to \Re_+$ may not be convex, differentiable, and perhaps not even Lipschitz. Chen et al. [20] proposed a smoothing trust region Newton method for solving (5.14). At each iteration, the method obtains the trial step by solving a model that approximates the smoothing problem

$$\min_{x \in \Re^n} \tilde{f}(x, \mu) := \theta(x) + \lambda \sum_{i=1}^m \tilde{\varphi}(s(d_i^T x, \mu)), \tag{5.15}$$

where $\tilde{\varphi}$ is a smoothing function of φ, and $s(d_i^T x, \mu)$ is a smoothing term for $|d_i^T x|$. For example, one possible choice of $s(t, \mu)$ is $\mu\ln(2 + e^{t/\mu} + e^{-t/\mu})$. The smoothing parameter μ is updated in order to ensure convergence. Assume that the penalty function φ satisfies the following assumption.

Assumption 1 1. φ is differentiable in $(0, \infty)$ and φ' is locally Lipschitz continuous in $(0, \infty)$.
2. φ is continuous at 0 with $\varphi(0) = 0$, $\varphi'(0^+) > 0$ and $\varphi'(t) \geq 0$ for all $t > 0$.

Chen et al. [20] showed that the sequence generated by the smoothing trust region Newton method for (5.14) converges to a point satisfying the second order necessary optimality condition from any starting point.

Cartis et al. [13] studied the worst-case complexity of minimizing an unconstrained, nonconvex composite objective with a structured nonsmooth term by means of some first-order methods. [13] finds that it is unaffected by the nonsmoothness of the objective in the sense that a first-order trust region or quadratic regularization method (where the subproblem is to minimize the approximation function plus a quadratic regularization term $\sigma\|d\|_2^2$) applied to it takes at most $O(\epsilon^{-2})$ function-evaluations to reduce the size of a first-order criticality measure below ϵ.

6 Optimization without derivatives

In this section, we consider optimization without derivatives. Numerical methods for optimization problems without using derivatives are also called derivative-free methods. Derivative-free trust region methods date back to the algorithm of Winfield [137,138], which is one of the pioneering works in trust region methods. The study of derivative-free trust region methods is very active in recent years. Excellent reviews on trust region algorithms for optimization without derivatives are given in [95,97] and [29].

Comparing with derivative-based algorithms, derivative-free trust region algorithms have to build the model function $m_k(d)$ without using derivatives. Therefore, we can not apply Taylor expansions which are normally used in derivative-based algorithms. Instead, it is natural for derivative-free algorithms to construct model functions by interpolation.

The simplest nonlinear model is the quadratic model, which can be expressed by

$$m_k(d) = Q_k(x_k + d) = \hat{f}_k + \hat{g}_k^T d + \frac{1}{2} d^T B_k d, \tag{6.1}$$

where $\hat{f}_k \in \Re$, $\hat{g}_k \in \Re^n$, $B_k \in \Re^{n \times n}$ symmetric. Because B_k is symmetric, the number of independent elements is $\frac{1}{2} n(n + 1)$. Thus, the total number of independent parameters in (6.1) is $\frac{1}{2}(n + 1)(n + 2)$. Hence it is reasonable to use $\frac{1}{2}(n + 1)(n + 2)$ function values and the interpolation conditions

$$Q_k(x_j^{(k)}) = f(x_j^{(k)}) \quad j = 1, ..., \frac{1}{2}(n + 1)(n + 2), \tag{6.2}$$

to set the values $\hat{f}_k$, $\hat{g}_k$ and B_k, where $x_j^{(k)}(j = 1, ..., \frac{1}{2}(n + 1)(n + 2))$ are the interpolation points at the kth iteration. Methods using (6.2) are given by Winfield [137,138] and Powell [96].

However, even for a medium-size n (for example, several hundred), evaluating the objective function at $\frac{1}{2}(n + 1)(n + 2)$ points can be prohibitively expensive. Thus, modern model-based trust region algorithms often build quadratic interpolation models using less than $(n + 1)(n + 2)/2$ points. Assume that the interpolation set has only m points $x_j^{(k)}(j = 1, ..., m)$ with $m < \frac{1}{2}(n + 1)(n + 2)$ at the kth iteration. Consider the interpolation conditions

$$Q_k(x_j^{(k)}) = f\left(x_j^{(k)}\right) \quad j = 1, ..., m. \tag{6.3}$$

Normally we should choose $m \geq n + 1$, since even for a linear model, we need at least $(n + 1)$ function evaluations to define the model function by interpolation uniquely. Let $\mathcal{Q}$ be the linear space of all quadratic functions from $\Re^n$ to $\Re$. It is easy to see that the dimension of $\mathcal{Q}$ is $\frac{1}{2}(n + 1)(n + 2)$. Hence, the interpolation conditions (6.3) give an under-determined system for the parameters $\hat{f}_k$, $\hat{g}_k$ and B_k. Therefore, we need to impose more conditions on the model function in addition to interpolation conditions

(6.3) in order to obtain the parameters $\hat{f}_k$, $\hat{g}_k$ and B_k, for example,

$$\min \|\hat{g}_k\|_F^2 + \|B_k\|_F^2$$

by Conn and Toint [31],

$$\min \|B_k\|_F^2$$

by Conn et al. [25,26] and Conn, Scheigberg and Vicente (Chapter 5 of [29]), and

$$\min \|\text{vec}(B_k)\|_1,$$

where $\text{vec}(B_k)$ is a vector containing all the upper-triangular entries of B_k, by Bandeira, Scheinberg and Vicente [3].

Powell [97–99] suggests

$$\min \|B_k - B_{k-1}\|_F^2, \tag{6.4}$$

which is used in a series of software [100–102]. An extension of (6.4) is given by Powell [104] where the model function is obtained by

$$\min \|B_k - B_{k-1}\|_F^2 + \sigma_k \|\hat{g}_k - \nabla m_{k-1}(x_k - x_{k-1})\|^2, \tag{6.5}$$

subject to the interpolation conditions (6.3). Zhang [157] studies this extension using a Sobolev semi-norm which is defined by

$$|f|_{H^1(\Omega)} := \left[\int_\Omega \|\nabla f(x)\|_2^2 dx\right]^{1/2}. \tag{6.6}$$

It is shown by Zhang [157] that

$$\min_{Q(x)\in\mathcal{Q}} \|\nabla^2 Q\|_F^2 + \sigma\|\nabla Q(\bar{x})\|_2^2 \tag{6.7}$$

$$\text{s.t.} \quad Q(x) = f(x), x \in \mathcal{S}, \tag{6.8}$$

where $\mathcal{S}$ is the set of interpolation points, is equivalent to

$$\min_{Q\in\mathcal{Q}} |Q(x)|_{H^1(\mathcal{B}_r)} \tag{6.9}$$

$$\text{s.t.} \quad Q(x) = f(x), x \in \mathcal{S}, \tag{6.10}$$

where $\mathcal{B}_r = \{x|\|x-\bar{x}\| \le r\}$ with $r = \sqrt{(n+2)/\sigma}$. Based on this observation, Zhang [157] proposes a simple and efficient method to determine σ_k in (6.5).

In some applications, it can also happen that the interpolation conditions (6.3) are overdetermined ($m > (n+1)(n+2)/2$), especially when the function evaluation is not very expensive but contaminated by noise. In this case, instead of requiring the

exact interpolation conditions (6.3), one can define the quadratic model $m_k(d)$ to be a weighted least squares approximation of f by solving

$$\min \sum_{i=1}^{m} w_i^2 \left(m_k \left(x_j^{(k)} - x_k \right) - f(x_j^{(k)}) \right)^2, \tag{6.11}$$

which is used in the trust region method based on regression models [7,28], and is also discussed in Chapter 4 of [29].

In addition to quadratic models, radial basis functions (RBF) models are also very popular for constructing derivative-free trust region algorithms. A radial basis function model has the following form

$$m_k(d) = \sum_{j=1}^{m} \lambda_j \phi(\|d - x_j^{(k)}\|) + p(d), \tag{6.12}$$

where $\phi : R_+ \to R$ is a univariate function and $p \in \mathcal{P}_{d-1}^n$ is a polynomial in the space of $n-$variate polynomials of total degree no more than $d-1$. Algorithms using radial basis function models are discussed in [84,135,136].

A special issue for derivative-free trust region methods based on interpolation and regression is the geometry of the interpolation/regression set, which is critical to the quality of the model. This interesting problem is discussed in [27,28]. In order to keep the interpolation/regression set from becoming degenerate, many methods (for instances, DFO, UOBYQA, and NEWUOA) use explicit "geometry-improving" steps. Marazzi and Nocedal [75] proposed another strategy in which a "wedge constraint" is imposed to the trust region subproblem so that the geometry of the interpolation set will not be destroyed after the trial point is included into the set. Fasano, Morales and Nocedal [46] studied the numerical performance of a trust region method using interpolation models but without any "geometry phase", where surprisingly good results are obtained. However, it is shown in [111] that the "geometry phase" cannot be completely omitted, otherwise the convergence will be jeopardized. In [111] they also propose a convergent trust region method that does not have any explicit "geometry-improving" step or wedge constraint but exploits "self-correcting geometry" of the interpolation set.

Convergence theory is more difficult to be established for a derivative-free trust region algorithm than a derivative-based trust region algorithm because no derivative information is used in the construction of the algorithm. Convergence results for derivative-free trust region algorithms are given in [26,30,103], and [136].

Recently, an interesting paper [4] was published in which it proves the almost-sure convergence of trust region method assuming that the models are "probabilistically fully linear". Another recent paper [62] shows that the algorithm studied in [4] drives the gradient to zero at the rate of $\mathcal{O}(1/\sqrt{k})$ with "overwhelmingly high" probability, matching the rate of the deterministic trust region method using derivatives.

Derivative-free trust region algorithms for constrained problems were first discussed by Powell [94], where linearly constrained derivative-free optimization is studied. Powell [102] also studies another special case for bound constrained problems. An

active set derivative-free trust region method is given for solving nonlinear bound constrained optimization by [64]. Recently, Troeltzsch [125] gives a derivative-free trust region algorithm based on SQP for equality constrained optimization.

Sampaio and Toint [109] gave a derivative-free trust-funnel method for general equality constrained nonlinear optimization problems. The algorithm is an generalization of the trust-funnel method of Gould and Toint [58], and preserves the main features of the method in [58] without using derivatives.

Trust region framework provides a way to rigorously handle surrogate models [65]. Grapiglia, Yuan and Yuan [61] presented a derivative-free trust region algorithm for composite nonsmooth optimization (5.1). The proposed algorithm is an adaptation of the derivative-free trust region algorithm of Conn, Scheinberg and Vicente [30] for unconstrained smooth optimization, with the techniques for composite nonsmooth optimization proposed by Fletcher [50,51] and Yuan [141,142]. Under suitable conditions, global convergence of the algorithm is proved. Furthermore, by introducing a slightly modified update rule for the trust region radius and assuming that the interpolation models are fully linear at all the iterations, it is proved that the function-evaluation complexity of the algorithm is $O(\epsilon^{-2})$ for the algorithm to reduce the first order criticality measure below ϵ.

Hybrid of trust region and other methodology is also studied in the context of derivative-free optimization. For instances, [35] and [39] incorporate trust region techniques into direct search and stochastic search, and [65] studies a surrogate management framework using trust region steps for derivative-free optimization.

7 Regularization methods

One interesting result of the standard trust region subproblem is that the minimizer of (2.3)–(2.4) is also a global minimizer of the model function $m_k(d)$ with a suitable quadratic penalty term. Namely (2.3)–(2.4) is equivalent to

$$\min_{d \in \Re^n} g_k^T d + \frac{1}{2} d^T B_k d + \frac{1}{2} \lambda_k \|d\|_2^2, \tag{7.1}$$

for a proper penalty parameter $\lambda_k \geq 0$. Thus, trust region algorithms can be viewed as a regularization by using the quadratic penalty function. Consequently, extensions of trust region algorithms by other regularization techniques have been attracting attention from many researchers. Cartis et al. [14] proposed an adaptive regularisation algorithm using cubics (ARC) for unconstrained optimization. At each iteration of the ARC algorithm, the following cubic model

$$m_k(d) = f(x_k) + g_k^T d + \frac{1}{2} d^T B_k d + \frac{1}{3} \sigma_k \|d\|_2^3 \tag{7.2}$$

is used to obtain the trial step s_k. To be more exact, s_k satisfies

$$m_k(s_k) \leq m_k(s_k^C) \tag{7.3}$$

where $s_k^C = -\alpha_k^C g_k$ is the Cauchy point with

$$\alpha_k^C = \arg \min_{\alpha \in \Re_+} m_k(-\alpha g_k) \, .$$

The ARC algorithm uses an adaptive estimation of the local Lipschitz constant and approximations to the global model-minimizer which remain computationally-viable even for large-scale problems. It is shown in [14] that the excellent global and local convergence properties obtained by Nesterov and Polyak are retained, and sometimes extended to a wider class of problems, by the ARC approach. Worst-case function- and derivative-evaluation complexity of the ARC method is studied in [15]. Recently, Grapiglia, Yuan and Yuan [60] extended the results of [15] by relaxing the boundedness of B_k to the following weaker condition

$$\|B_k\| \le \delta_1 + \delta_2 k.$$

In [16], the ARC method is adapted to the constrained optimization problem where a nonlinear, possibly nonconvex, smooth objective function is minimized over a convex domain.

The generalization of the model function (7.2) gives the following general regularised version of trust region subproblem:

$$\min_{d \in \Re^n} m(d) = g^T d + \frac{1}{2} d^T B d + \frac{\sigma}{p} \|d\|_2^p, \tag{7.4}$$

where $\sigma > 0$, $p > 2$. (7.4) with $p = 3$ is the cubic reguarised subproblem used by Cartis et al. [14] in their ARC algorithm. This special case was first due to Griewank [66] and was considered later by many authors, see Nesterov and Polyak [81], Weiser et al. [134], and Cartis et al. [14]. Recently, Gould and Toint [57], and Hsia et al. [70] studied properties of subproblem (7.4) for the general case $p > 2$.

The necessary global optimality condition for (7.4) was established by Cartis et al. [14, Thm. 3.1] and by Nesterov and Polyak [81, Thm. 10 when p = 3], see also [57]. Very recently, Hsia et al. [70] proved that the presented necessary condition is also sufficient. In the same paper, the ℓ_2 norm of the global minimizers is shown to be unique.

Theorem 1 (Necessary and sufficient global optimality condition) *The point d^* is a global minimizer of (7.4) if and only if it is a critical point satisfying $\nabla m(d^*) = 0$ and $B + \sigma \|d^*\|_2^{p-2} I \succeq 0$. Moreover, the ℓ_2 norms of all the global minimizers are equal.*

Following Martínez's characterization [77] on necessary and sufficient conditions for the non-global local minimizer of (2.3)–(2.4), Hsia et al. [70] established the necessary and sufficient condition for the non-global local minimizer of (7.4) as follows.

Theorem 2 (Necessary and sufficient condition for non-global local minimizer) *$\underline{d}$ is a non-global local minimizer of (7.4) if and only if*

$$\underline{d} = -\left(\sigma \underline{t}^* I + B\right)^{-1} g, \tag{7.5}$$

where $\underline{t}^*$ is a root of the secular function

$$h(t) = \| (\sigma t I + B)^{-1} g \|^2 - t^{\frac{2}{p-2}}, \quad t \in \left(\max \left\{ -\frac{\alpha_2}{\sigma}, 0 \right\}, -\frac{\alpha_1}{\sigma} \right) \qquad (7.6)$$

such that $h'(\underline{t}^*) > 0$, where $\sigma_1 \leq \sigma_2$ are the two smallest eigenvalues of B.

Notice that, the secular function for (2.3)–(2.4) [cf. $h(t)$ in (7.6)] is defined by

$$\phi(\lambda) = \|(B_k + \lambda I)^{-1} g_k\|^2.$$

Martínez [77] proved that, if $\underline{d}$ is a non-global local minimizer of (2.3)–(2.4), then $\underline{d}$ satisfies $(B_k + \lambda^* I)\underline{d} = -g_k$ with $\lambda^* \in (-\alpha_2, -\alpha_1)$, $\lambda^* \geq 0$ and $\phi'(\lambda^*) \geq 0$. And, this necessary condition becomes sufficient when $\phi'(\lambda^*) > 0$. To the best of our knowledge, it is unknown whether the gap between the necessary and sufficient conditions can be closed as in Theorem 2.

As a corollary of Theorem 2, it is shown in [70] that

Theorem 3 (7.4) *with $p > 2$ has at most one non-global local minimizer.*

Interestingly, Theorem 3 is applied by Hsia et al. [70] to show that problem (7.4) with m linear inequality constraints:

$$\min_{d \in \mathfrak{R}^n} \quad g^T d + \frac{1}{2} d^T B d + \frac{\sigma}{p} \|d\|_2^p \qquad (7.7)$$

$$\text{s.t.} \quad l_i \leq a_i^T d \leq u_i, \quad i = 1, \ldots, m, \qquad (7.8)$$

can be solved in polynomial time when $p = 4$ and m is a fixed number. Notice that, when m is also an input, (7.7)–(7.8) is generally NP-hard even when $p = 4$, see a proof in [70].

In a recent paper, Cartis et al. [17] proved that a general nonlinear nonconvex optimization problem can be solved to the accuracy of ϵ by a second-order method using cubic regularization in at most $O(\epsilon^{-3/2})$ function evaluations, the same order bound as in the unconstrained case. This result is obtained by first showing that the same result holds for inequality constrained nonlinear least-squares.

8 Discussions

In this paper, trust region and regularization algorithms for unconstrained optimization, constrained optimization, nonlinear equations and nonlinear least squares, nonsmooth optimization, and optimization without derivatives are reviewed. Trust region algorithms have been flourishing in nonlinear optimization over the past three decades, they can also been applied to other types of optimization problems.

The application of trust region methods to Nash equilibrium problems is studied in Yuan [150], which proposes a Jacobi-type trust region method for their solutions. The method includes different trust regions for each player, and the trial step for each player is computed and accepted (or rejected) based on each individual utility function.

In order to ensure global convergence, an overall merit function is used to update the trust regions.

A trust region algorithm for unconstrained multi-objective optimization is given in Villacorta, Oliveira and Soubeyran [127]. At each iteration, instead of treating each object separably, the algorithm constructs a single nonsmooth trust region subproblem. It is shown that the algorithm generates a sequence that converges to Pareto critical points.

Recently, a Riemannian trust region method for low-rank matrix completion is given by Boumal and Absil [8], which studies the problem of recovering large scale low-rank matrices when most of the entries are unknown. It adapts an approach that exploits the geometry of the low-rank constraint to recast the problem as an unconstrained optimization problem on the Grassmann manifold. Consequently, the derived unconstrained optimization problem is solved by applying first- and second-order Riemannian trust region methods, which are studied in Absil, Baker and Gallivan [1].

Trust region is a powerful technique for ensuring global convergence and is particularly suitable for handling nonconvex and ill-conditioned problems. Trust region can be used with other techniques to construct efficient numerical algorithms, for example its combination with line search technique [83], and with a non-monotone technique [123]. Recently, in optimization, particularly for sparse problems, alternating direction minimization (ADM) techniques are attracting much attention from researchers. Yet, most ADM methods are globalized through line search techniques. It is worthwhile to explore the possibility of using trust region techniques in ADM type algorithms.

References

1. Absil, P.-A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds. Found. Comput. Math. **7**, 303–330 (2007)
2. Ai, W.B., Zhang, S.Z.: Strong duality for the CDT subproblem: a necessary and sufficient condition. SIAM J. Optim. **19**, 1735–1756 (2009)
3. Bandeira, A.S., Scheinberg, K., Vicente, L.N.: Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization. Math. Program. **134**, 223–257 (2012)
4. Bandeira, A.S., Scheinberg, K., Vicente, L.N.: Convergence of trust region methods based on probabilistic models. SIAM J. Optim. **24**, 1238–1264 (2014)
5. Bastin, F., Malmedy, F., Mouffe, M., Toint, PhL, Tomanos, D.: A retrospective trust-region method for unconstrained optimization. Math. Program. **123**, 395–418 (2010)
6. Bienstock, D.: A Note on Polynomial Solvability of the CDT Problem. Technical Report, Columbia University, USA, arXiv:1406.6429v2 (2013)
7. Billups, S.C., Larson, J., Graf, P.: Derivative-free optimization of expensive functions with computational error using weighted regression. SIAM J. Optim. **23**, 27–53 (2013)
8. Boumal, N., Absil, P.-A.: RTRMC: A Riemannian Trust-Region Method for Low-Rank Matrix Completion. Technical Report, ICTEAM Institute Universit e catholique de Louvain B-1348 Louvain-la-Neuve (2014) (http://www.optimization-online.org/DB_HTML/2012/09/3619.html)
9. Burdakov, O., Gong, L.J., Yuan, Y., Zikrin, S.: On Efficiently Combining Limited Memory and Trust-Region Techniques. Techinical Report, Linkoping University, Sweden (2014)

10. Burke, J.: Descent methods for composite nondifferentiable optimization problem. Math. Program. **33**, 260–279 (1985)
11. Byrd, R.H., Nocedal, J., Waltz, R.A.: An Integrated Package for Nonlinear Optimization. In: Di Pillo, G., Roma, M. (eds.) Large-scale nonlinear optimization, pp. 35–59. Springer, Berlin (2006)
12. Byrd, R., Schnabel, R.B., Shultz, G.A.: A trust region algorithm for nonlinear constrained optimization. SIAM J. Numer. Anal. **24**, 1152–1170 (1987)
13. Cartis, C., Gould, N.I.M., Toint, Ph.L.: On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. SIAM J. Optim. **21**, 1721–1739 (2011)
14. Cartis, C., Gould, N.I.M., Toint, Ph.L.: Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Math. Program. **127**, 245–295 (2011)
15. Cartis, C., Gould, N.I.M., Toint, Ph.L.: Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. Math. Program. **130**, 295–319 (2011)
16. Cartis, C., Gould, N.I.M., Toint, Ph.L.: An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity. IMA J. Numer. Anal. **32**(4), 1662–1695 (2012)
17. Cartis, C., Gould, N.I.M., Toint, Ph.L.: On the Evaluation Complexity of Constrained Nonlinear Least-Squares and General Constrained Nonlinear Optimization Using Second-Order Methods. Report naXys-01-2013, Dept of Mathematics, FUNDP, Namur (B) (2013)
18. Celis, M.R., Dennis, J.E., Tapia, R.A.: A trust region algorithm for nonlinear equality constrained optimization. In: Boggs, P.T., Byrd, R.H., Schnabel, R.B. (eds.) Numerical Optimization, pp. 71–82. SIAM, Philadelphia (1985)
19. Chen, X.D., Yuan, Y.: On local solutions of the CDT subproblem. SIAM J. Optim. **10**, 359–383 (1999)
20. Chen, X.J., Niu, L.F., Yuan, Y.: Optimality conditions and smoothing trust region Newton method for non-Lipschitz optimization. SIAM J. Optim. **23**(3), 1528–1552 (2013)
21. Clarke, F.H.: Optimization and Nonsmooth Analysis. SIAM, Philadelphia (1980)
22. Coleman, T.F., Li, Y.: On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. Math. Program. **67**, 189–224 (1994)
23. Coleman, T.F., Li, Y.: An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Opt. **6**, 418–445 (1996)
24. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. MPS-SIAM Series on Optimization. SIAM, Philedalphia (2000)
25. Conn, A.R., Scheinberg, K., Toint, Ph.L.: Recent progress in unconstrained nonlinear optimization without derivatives. Math. Program. **79**, 397–414 (1997)
26. Conn, A.R., Scheinberg, K., Toint, Ph.L.: On the convergence of derivative-free methods for unconstrained optimization. In: Buhmann, M.D., Iserles, A. (eds.) Approximation Theory and Optimization Tributes to M.J.D. Powell, pp. 83–108. Cambridge University Press, Cambridge (1997)
27. Conn, A.R., Scheinberg, K., Vicente, L.N.: Geometry of interpolation sets in derivative free optimization. Math. Program. **111**, 141–172 (2008)
28. Conn, A.R., Scheinberg, K., Vicente, L.N.: Geometry of sample sets in derivative-free optimization: polynomial regression and underdetermined interpolation. IMA J. Numer. Anal. **28**, 721–748 (2008)
29. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. MPS-SIAM Series on Optimization. SIAM, Philadelphia (2009)
30. 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. **20**, 387–415 (2009)
31. Conn, A.R., Toint, Ph.L.: An algorithm using quadratic interpolation for unconstrained derivative free optimization. In: Di Pillo, G., Gianessi, F. (eds.) Nonlinear Optimization and Applications, pp. 27–47. Plenum, New York (1996)
32. Conn, A.R., Vicente, L.N.: Bilevel derivative-free optimization its application to robust optimization. Optim. Methods Softw. **27**(3), 559–575 (2012)
33. Curtis, F., Gould, N.I.M., Robinson, D., Toint, Ph.L.: An Interior-Point Trust-Funnel Algorithm for Nonlinear Optimization. Report naXys-02-2014, Dept. of Mathematics, UNamur, Namur (B), (2014)
34. Curtis, F., Jiang, H., Robinson, D.P.: An adaptive augmented Lagrangian method for large-scale constrained optimization. Math. Program. Ser. A (2014). doi:10.1007/s10107-014-0784-y

35. Custodio, A.L., Rocha, H., Vicente, L.N.: Incorporating minimum Frobenius norm models in direct search. Comput. Optim. Appl. **46**, 265–278 (2010)
36. Dennis, J.E., Li, S.-B., Tapia, R.A.: A unified approach to global convergence of trust region methods for nonsmooth optimization. Math. Program. **68**, 319–346 (1995)
37. Dennis, J.E., Mei, H.H.: Two new unconstrained optimization algorithms which use function and gradient values. J. Optim. Theory Appl. **28**, 453–482 (1979)
38. Di, S., Sun, W.Y.: Trust region method for conic model to solve unconstrained optimization problems. Optim. Methods Softw. **6**, 237–263 (1996)
39. Diouanne, Y., Gratton, S., Vicente, L.N.: Globally Convergent Evolutionary Strategies for Constrained Optimization. Technical Report TR-PA-14-50, CERFACS, Toulouse, France (2014)
40. Duff, I.S., Nocedal, J., Reid, J.K.: The use of linear programming for the solution of sparse sets of nonlinear equations. SIAM J. Sci. Stat. Comput. **8**, 99–108 (1987)
41. El Hallabi, M., Tapia, R.A.: A Global Convergence Theory for Arbitrary Norm Trust Region Methods for Nonlinear Equations. Technical Report 87–25, Dept. Math. Sciences, Rice University, USA (1987)
42. Fan, J.Y.: Convergence rate of the trust region method for nonlinear equations under local error bound condition. Comput. Optim. Appl. **34**, 215–227 (2006)
43. Fan, J.Y., Lu, N.: On the modified trust region algorithm. Optim. Methods Softw. (2014). doi:10.1080/10556788.2014.932943
44. Fan, J.Y., Pan, J.Y.: An improved trust region algorithm for nonlinear equations. Comput. Optim. Appl. **48**, 59–70 (2011)
45. Fan, J.Y., Yuan, Y.: A new trust region algorithm with trust region radius converging to zero. in: D. Li ed. Proceedings of the 5th International Conference on Optimization: Techniques and Applications (December 2001, Hongkong), pp. 786–794 (2001)
46. Fasano, G., Morales, J.L., Nocedal, J.: On the geometry phase in model-based algorithms for derivative-free optimization. Optim. Methods Softw. **24**, 145–154 (2009)
47. Fletcher, R.: An Efficient, Global Convergent Algorithm for Unconstrained and Linearly Constrained Optimization Problems. Technical Report TP 431, AERE Harwell Laboratory, Oxfordshire, England (1970)
48. Fletcher, R.: Practical Methods of Optimization, Volume 1: Unconstrained Optimization. Wiley, Chichester (1980)
49. Fletcher, R.: Practical Methods of Optimization, Volume 2: Constrained Optimization. Wiley, Chichester (1981)
50. Fletcher, R.: A model algorithm for composite NDO problem. Math. Prog. Study **17**, 67–76 (1982)
51. Fletcher, R.: Second order correction for nondifferentiable optimization. In: Watson, G.A. (ed.) Numerical Analysis, pp. 85–115. Springer, Berlin (1982)
52. Fletcher, R., Leyffer, S.: Nonlinear programming without a penalty function. Math. Program. **91**, 239C269 (2002)
53. Gay, D.M.: Computing optimal local constrained steps. SIAM J. Sci. Comput. **2**, 186–197 (1981)
54. Gong, L.J.: Trust Region Algorithms for Unconstrained Optimization. Ph.D. thesis, Academy of Mathematics and Systems Science, Chinese Academy of Sciences (2011)
55. Gould, N.I.M., Orban, D., Sartenar, A., Toint, PhL: Sensitive of trust-region algorithms on their parameters. 4OR, Q. J. Italian Fr. Belgian OR Soc **3**, 227–241 (2005)
56. Gould, N.I.M., Orban, D., Toint, PhL: CUTEr and SifDec: a constrained and unconstrained testing environment, revisited. ACM Trans. Math. Softw. (TOMS) **29**(4), 373–394 (2003)
57. Gould, N.I.M., Robinson, D.P., Thorne, H.S.: On solving trust-region and other regularised subproblems in optimization. Math. Program. Comput. **2**, 21–57 (2010)
58. Gould, N.I.M., Toint, PhL: Nonlinear programming without a penalty function or a filter. Math. Program. **122**, 155–196 (2010)
59. Grapiglia, G.N., Yuan, J.Y., Yuan, Y.: A subspace version of the Powell–Yuan trust region algorithm for equality constrained optimization. J. Oper. Res. Soc. China **1**(4), 425–451 (2013)
60. Grapiglia, G.N., Yuan, J.Y., Yuan, Y.: On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization. Math. Program. (2014). doi:10.1007/s10107-014-0794-9
61. Grapiglia, G.N., Yuan, J.Y., Yuan, Y.: A derivative-free trust-region algorithm for composite nonsmooth optimization. Comput. Appl. Math. (2014). doi:10.1007/s40314-014-0201-4
62. Gratton, S., Royer, C.W., Vicente, L.N., Zhang, Z.: Direct Search Based on Probabilistic Descent. Technical Report 14–11, Dept. Math., University of Coimbra, Portugal (2014)

63. Gratton, S., Sartenaer, A., Toint, Ph.L.: Recursive trust-region methods for multiscale nonlinear optimization. SIAM J. Optim. **19**, 414–444 (2008)

64. Gratton, S., Toint, PhL, Troeltzsch, A.: An active-set trust region method for derivative-free nonlinear bound-constrained optimization. Optim. Methods Softw. **26**, 873–894 (2011)

65. Gratton, S., Vicente, L.N.: A surrogate management framework using rigorous trust-region steps. Optim. Methods Softw. **29**, 10–23 (2014)

66. Griewank, A.: The Modification of Newtons Method for Unconstrained Optimization by Bounding Cubic Terms. Technical Report NA/12, DAMTP, University of Cambridge, England (1981)

67. Hebden, M.D.: An Algorithm for Minimization Using Exact Second Order Derivatives. Technical Report, T.P. 515, AERE Harwell Laboratory, Harwell, Oxfordshire, England (1973)

68. Hei, L.: A self-adaptive trust region algorithm. J. Comput. Math. **21**, 229–236 (2003)

69. Heinkenschloss, M., Vicente, L.N.: Analysis of inexact trust-region SQP algorithms. SIAM J. Optim. **12**, 283–302 (2001)

70. Hsia, Y., Sheu, R.L., Yuan, Y.: On the p-Regularized Trust Region Subproblem. arXiv:1409.4665 (2014)

71. Kanzow, C., Klug, A.: An interior-point affine-scaling trust-region method for semismooth equations with box constraints. Comput. Optim. Appl. **37**, 329C353 (2007)

72. Levenberg, K.: A method for solution of certain problems in least squares. Q. J. Appl. Math. **2**, 164–168 (1944)

73. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program. **45**, 503–528 (1989)

74. Madsen, K.: An algorithm for the minimax solution of overdetermined systems of nonlinear equations. J. Inst. Math. Its Appl. **16**, 321–328 (1975)

75. Marazzi, M., Nocedal, J.: Wedge trust region methods for derivative free optimization. Math. Program. **91**, 289–305 (2002)

76. Marquardt, D.: An algorithm for least-squares estimation on nonlinear parameters. SIAM J. Appl. Math. **11**, 431–441 (1963)

77. Martinez, J.M.: Local minimizers of quadratic functions on Euclidean balls and spheres. SIAM J. Optim. **4**, 159–176 (1994)

78. Moré, J.: The Levenberg–Marquardt algorithm: implementation and theory. In: Watson, G.A. (ed.) Numerical Analysis, pp. 105–116. Springer, Berlin (1978)

79. Moré, J.J.: Recent developments in algorithms and software for trust region methods. In: Bachem, A., Grötschel, M., Korte, B. (eds.) Mathematical Programming: The State of the Art, pp. 258–287. Springer, Berlin (1983)

80. Morrison, D.D.: Methods for nonlinear least squares problems and convergence proofs. In: Lorell, J., Yagi, F. (eds.) Proceedings of the Seminar on Tracking Programs and Orbit Determination, pp. 1–9. Jet Propulsion Laboratory, Pasadena (1960)

81. Nesterov, Yu., Polyak, B.T.: Cubic regularization of Newtons method and its global performance. Math. Program. **108**, 177–205 (2006)

82. Niu, L.F., Yuan, Y.: A new trust region algorithm for nonlinear constrained optimization. J. Comput. Math. **28**, 72–86 (2010)

83. Nocedal, J., Yuan, Y.: Combining trust region and line search techniques. In: Yuan, Y. (ed.) Advances in Nonlinear Programming, pp. 153–175. Kluwer, Netherlands (1998)

84. Oeuvray, R., Bierlaire, M.: BOOSTER: a derivative-free algorithm based on radial basis functions. Inter. J. Model. Simul. **29**, 349–371 (2009)

85. Omojokun, E.O.: Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints. Ph.D. Thesis, University of Colorado at Boulder, USA (1989)

86. Osborne, M.R.: Nonlinear least squares—the Levenberg–Marquardt algorithm revisited. J. Aust. Math. Soc. B **19**, 343–357 (1976)

87. Peng, J., Yuan, Y.: Optimality conditions for the minimization of a quadratic with two quadratic constraints. SIAM J. Optim. **7**, 579–594 (1997)

88. Powell, M.J.D.: A hybrid method for nonlinear equations. In: Rabinowitz, P. (ed.) Numerical Methods for Nonlinear Algebraic Equations, pp. 87–114. Gordon and Breach, London (1970)

89. Powell, M.J.D.: A Fortran subroutine for solving systems of nonlinear algebraic equations. In: Robinowitz, P. (ed.) Numerical Methods for Nonlinear Algebraic Equations, pp. 115–161. Gordon and Breach, London (1970)

90. Powell, M.J.D.: A new algorithm for unconstrained optimization. In: Rosen, J.B., Mangasarian, O.L., Ritter, K. (eds.) Nonlinear Programming, pp. 31–65. Academic Press, London (1970)
91. Powell, M.J.D.: Convergence properties of a class of minimization algorithms. In: Mangasarian, O.L., Meyer, R.R., Robinson, S.M. (eds.) Nonlinear Programming 2, pp. 1–27. Academic Press, New York (1975)
92. Powell, M.J.D.: Some global convergence properties of a variable metric algorithm for minimization without exact line searches. In: Cottle, R.W., Lemke, C.E. (eds.) Nonlinear Programming, SIAM-AMS Proceedings, vol. IX, pp. 53–72. SIAM, Philadelphia (1976)
93. Powell, M.J.D.: General algorithms for discrete nonlinear approximation calculations. In: Schumaker, L.L. (ed.) Approximation Theory IV, pp. 187–218. Academy Press, New York (1984)
94. 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 Numercial Analysis. Mathematics and Its Applications, vol. 275, pp. 51–67. Springer, Netherland (1994)
95. Powell, M.J.D.: Direct search algorithms for optimization calculations. Acta Numer. **7**, 287–336 (1998)
96. Powell, M.J.D.: UOBYQA: unconstrained optimization by quadratic approximation. Math. Program. **92**, 555–582 (2002)
97. Powell, M.J.D.: On trust region methods for unconstrained minimization without derivatives. Math. Program. **97**, 605–623 (2003)
98. Powell, M.J.D.: Least Frobenius norm updating of quadratic models that satisfy interpolation conditions. Math. Program. **100**, 183–215 (2004)
99. Powell, M.J.D.: On the use of quadratic models in unconstrained minimization without derivatives. Optim. Methods Softw. **19**, 399–411 (2004)
100. Powell, M.J.D.: The NEWOUA software for unconstrained optimization without derivatives. In: Di Pillo, G., Roma, M. (eds.) Large Scale Nonlinear Optimization. Nonconvex Optimization and Its Applications, vol. 83, pp. 255–297. Springer, Berlin (2006)
101. Powell, M.J.D.: Developments of NEWUOA for minimization without derivatives. IMA J. Numer. Anal. **28**, 649–664 (2008)
102. Powell, M.J.D.: The BOBYQA Algorithym for Bound Constrained Optimization Without Derivatives. Technical Report DAMTP 2009/NA06, CMS, University of Cambridge, UK (2009)
103. Powell, M.J.D.: On the convergence of trust region algorithm for unconstrained minimization without derivatives. Comput. Optim. Appl. **53**, 527–555 (2012)
104. Powell, M.J.D.: Beyond symetric Broyden for updating quadratic models in minimization without derivatives. Math. Program. **138**, 475–500 (2013)
105. Powell, M.J.D.: On fast trust region methods for quadratic models with linear constraints. DAMTP 2014/NA02, CMS, University of Cambridge (2014)
106. Powell, M.J.D., Yuan, Y.: Conditions for superlinear convergence in l_1 and l_∞ solutions of overdetermined non-linear equations. IMA J. Numer. Anal. **4**, 241–251 (1984)
107. Powell, M.J.D., Yuan, Y.: A trust region algorithm for equality constrained optimization. Math. Program. **49**, 189–211 (1991)
108. Qi, L., Sun, J.: A trust region algorithm for minimization of locally Lipschitzian functions. Math. Program. **66**, 25–43 (1994)
109. Sampaio, Ph.R., Toint, Ph.L.: A Derivative-Free Trust-Funnel Method for Equality-Constrained Nonlinear Optimization. Report naXys-08-2014, Dept of Mathematics, UNamur, Namur (B), (2014)
110. Sampaio, R.J.B., Yuan, J.Y., Sun, W.Y.: Trust region algorithm for nonsmooth optimization. Appl. Math. Comput. **85**, 109–116 (1997)
111. Scheinberg, K., Toint, Ph.L.: Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. SIAM J. Optim. **20**, 3512–3532 (2010)
112. Siegel, D.: Implementing and Modifying Broyden Class Updates for Large Scale Optimization. Report DAMPT1992/NA12, University of Cambridge, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England (1992)
113. Sorensen, D.C.: Newton's method with a model trust region modification. SIAM J. Numer. Anal. **19**, 409–426 (1982)
114. Sorensen, D.C.: Trust region methods for unconstrained optimization. In: Powell, M.J.D. (ed.) Nonlinear Optimization 1981, pp. 29–39. Academic Press, London (1982)
115. Steihaug, T.: The conjugate gradient method and trust regions in large scale optimization. SIAM J. Numer. Anal. **20**, 626–637 (1983)

116. Sturm, J.F., Zhang, S.Z.: On cones of nonnegative quadratic functions. Math. Oper. Res. **28**, 246–267 (2003)
117. Sun, W.Y., Yuan, Y.: A conic trust-region method for nonlinearly constrained optimization. Ann. Oper. Res. **103**, 175–191 (2001)
118. Toint, Ph.L.: Some numerical result using a sparse matrix updating formula in unconstrainded optimization. Math. Comput. **32**, 839–851 (1978)
119. Toint, Ph.L.: On the superlinear convergence of an algorithm for solving a sparse minimization problem. SIAM J. Numer. Anal. **16**, 1036–1045 (1979)
120. Toint, Ph.L.: Sparsity exploiting quasi-Newton methods for unconstrained optimization. In: Dixon, L.C.W., Spedicato, E., Szego, G.P. (eds.) Nonlinear Optimization: Theory and Algorithms, pp. 65–90. Birkhauser, Belgium (1980)
121. Toint, Ph.L.: Convergence Properties of a Class of Minimization Algorithms That Use a Possibly Unbounded Sequences of Quadratic Approximation. Technical Report 81/1, Dept. Math., University of Namur, Belgium (1981)
122. Toint, Ph.L.: Towards an efficient sparsity exploiting Newton method for minimization. In: Duff, I.S. (ed.) Sparse Matrices and Their Uses, pp. 57–88. Academic Press, Lodon (1978)
123. Toint, Ph.L.: A non-monotone trust region algorithm for nonlinear optimization subject to convex constraints. Math. Program. **77**, 69–94 (1997)
124. Toint, Ph.L.: Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization. Optim. Methods Sofw. **28**, 82–95 (2013)
125. Troeltzsch, A.: A sequential quadratic programming algorithm for equality-constrained optimization without derivatives. Optim. Lett. (2014). doi:10.1007/s11590-014-0830-y
126. Vardi, A.: A trust region algorithm for equality constrained minimization: convergence properties and implementation. SIAM J. Numer. Anal. **22**, 575–591 (1985)
127. Villacorta, K.D.V., Oliveira, P.R., Soubeyran, A.: A trust-region method for unconstrained multi-objective problems with applications in satisficing processes. J. Opt. Theory Appl. **160**, 865–889 (2014)
128. Vicente, L.N.: A comparison between line searches and trust regions for nonlinear optimization. Investigacao Operacional **16**, 173–179 (1996)
129. Waltz, R.A., Morales, J.L., Nocedal, J., Orban, D.: An interior algorithm for nonlinear optimization that combines line search and trust region steps. Math. Program. **107**, 391C408 (2006)
130. Wang, Z.H., Yuan, Y.: A subspace implementation of quasi-Newton trust region methods for unconstrained optimization. Numer. Math. **104**, 241–269 (2006)
131. Wang, X., Yuan, Y.: A trust region method based on a new affine scaling technique for simple bounded optimization. Optim. Methods Soft. **28**, 871–888 (2013)
132. Wang, X., Yuan, Y.: An augmented Lagrangian trust region method for equality constrained optimization. Optim. Methods Softw. (2014). doi:10.1080/10556788.2014.940947
133. Wang X., Zhang H.: An augmented Lagrangian affine scaling method for nonlinear programming. https://www.math.lsu.edu/hozhang/papers/ALAS, (2014)
134. Weiser, M., Deuflhard, P., Erdmann, B.: Affine conjugate adaptive Newton methods for nonlinear elastomechanics. Optim. Methods Softw. **22**(3), 413–431 (2007)
135. Wild, S.M., Regis, R.G., Shoemaker, C.A.S.: ORBIT: Optimization by radial basis function interpolation in trust-regions. SIAM J. Sci. Comput. **30**, 3197–3219 (2008)
136. Wild, S.M., Shoemaker, C.A.S.: Global convergence of radial basis function trust-region algorithms for derivative-free optimization. SIAM Rev. **55**, 349–371 (2013)
137. Winfield, D.H.: Function and Functional Minimization by Interpolation in Data Tables. Ph.D. Thesis, Harvard Unviersity, Cambridge, MA, USA (1969)
138. Winfield, D.H.: Function minimization by interpolation in date table. IMA J. Numer. Anal. **12**, 339–347 (1973)
139. Yamashita, H., Yabe, H., Tanabe, T.: A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. Math. Program. **102**, 111–151 (2005)
140. Yuan, Y.: An example of only linearly convergence of trust region algorithms for nonsmooth optimization. IMA J. Numer. Anal. **4**, 327–335 (1984)
141. Yuan, Y.: Conditions for convergence of trust region algorithms for nonsmooth optimization. Math. Program. **31**(2), 220–228 (1985)

142. Yuan, Y.: On the superlinear convergence of a trust region algorithm for nonsmooth optimization. Math. Program. **31**(3), 269–285 (1985)
143. Yuan, Y.: Some Theories and Algorithms in Nonlinear Programming. Ph.D. Thesis, University of Cambridge, UK (1985)
144. Yuan, Y.: On a subproblem of trust region algorithms for constrained optimization. Math. Program. **47**, 53–63 (1990)
145. Yuan, Y.: A dual algorithm for minimizing a quadratic function with two quadratic constraints. J. Comput. Math. **9**, 348–359 (1991)
146. Yuan, Y.: On the convergence of a new trust region algorithm. Numer. Math. **70**, 515–539 (1995)
147. Yuan, Y.: Trust region algorithms for nonlinear equations. Information **1**, 7–20 (1998)
148. Yuan, Y.: A review of trust region algorithms for optimization. In: Ball, J.M., Hunt, J.C.R. (eds.) ICIAM 99, Proceedingss of the Fourth International Cogress on Industrial and Applied Mathematics, pp. 271–282. Oxford University Press, Oxford (2000)
149. Yuan, Y.: On the truncated conjugate gradient method. Math. Program. **87**, 561–571 (2000)
150. Yuan, Y.: A trust region algorithm for Nash equilibrium problems. Pac. J. Optim. **7**, 125–138 (2011)
151. Yuan, Y.: A review on subspace methods for nonlinear optimization. In: Proceedings of the International Congress of Mathematics 2014, Seoul, Korea, pp. 807–827 (2014)
152. Zhang, H.C., Conn, A.R.: On the local convergence of a derivative-free algorithm for least-squares minimization. Comput. Optim. Appl. **51**(2), 481–507 (2012)
153. Zhang, H.C., Conn, A.R., Scheinberg, K.: A derivative-free algorithm for least squares minimization. SIAM J. Optim. **20**(6), 3555–3576 (2012)
154. Zhang, J.: A new trust region method for nonlinear equations. Math. Methods Oper. Res. **58**, 283–298 (2003)
155. Zhang, L.P.: A new trust region algorithm for nonsmooth convex minimization. Appl. Math. Comput. **193**, 135–142 (2007)
156. Zhang, Y.: Computing a Celis–Dennis–Tapia trust region step for equality constrained optimization. Math. Program. **55**, 109–124 (1992)
157. Zhang, Z.: Sobolev seminorm of quadratic functions with applications to derivative-free optimization. Math. Program. **146**, 77–96 (2014)
158. Zikrin, S.: Large-Scale Optimization Mehods with Applications to Design of Filter Networks. Linkoping Studies in Science and Technology, Dissertation No. 151, Likoping University, Sweden (2014)