

RESEARCH ARTICLE

# A Modified BFGS Formula Using a Trust Region Model for Nonsmooth Convex Minimizations

Zengru Cui<sup>1</sup>, Gonglin Yuan<sup>1,2\*</sup>, Zhou Sheng<sup>1</sup>, Wenjie Liu<sup>2,3</sup>, Xiaoliang Wang<sup>1</sup>, Xiabin Duan<sup>1</sup>

**1** Guangxi Colleges and Universities Key Laboratory of Mathematics and Its Applications, College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi 530004, China, **2** School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China, **3** Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing 210044, China

\* [glyuan@gxu.edu.cn](mailto:glyuan@gxu.edu.cn)



OPEN ACCESS

**Citation:** Cui Z, Yuan G, Sheng Z, Liu W, Wang X, Duan X (2015) A Modified BFGS Formula Using a Trust Region Model for Nonsmooth Convex Minimizations. PLoS ONE 10(10): e0140606. doi:10.1371/journal.pone.0140606

**Editor:** Lixiang Li, Beijing University of Posts and Telecommunications, CHINA

**Received:** April 8, 2015

**Accepted:** September 27, 2015

**Published:** October 26, 2015

**Copyright:** © 2015 Cui et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data are available and they are listed in the paper.

**Funding:** This work is supported by the Program for Excellent Talents in Guangxi Higher Education Institutions (Grant No. 201261), Guangxi NSF (Grant No. 2012GXNSFAA053002), China NSF (Grant No. 11261006 and 11161003), the Guangxi Science Fund for Distinguished Young Scholars (No. 2015GXNSFGA139001), NSFC No. 61232016, NSFC No. U1405254, and PAPD issue of Jiangsu advantages discipline.

## Abstract

This paper proposes a modified BFGS formula using a trust region model for solving nonsmooth convex minimizations by using the Moreau-Yosida regularization (smoothing) approach and a new secant equation with a BFGS update formula. Our algorithm uses the function value information and gradient value information to compute the Hessian. The Hessian matrix is updated by the BFGS formula rather than using second-order information of the function, thus decreasing the workload and time involved in the computation. Under suitable conditions, the algorithm converges globally to an optimal solution. Numerical results show that this algorithm can successfully solve nonsmooth unconstrained convex problems.

## Introduction

Consider the following convex problem:

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1}$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is a possibly nonsmooth convex function. In general, this problem has been well studied for several decades when  $f$  is continuously differentiable, and a number of different methods have been developed for its solution [Eq \(1\)](#) (for example, numerical optimization method [\[1–3\]](#) etc, heuristic algorithm [\[4–6\]](#) etc). However, when  $f$  is a nondifferentiable function, the difficulty of solving this problem increases. Recently, such problems have arisen in many medical, image restoration and optimal control applications (see [\[7–13\]](#) etc). Some authors have previously studied nonsmooth convex problems (see [\[14–18\]](#) etc).

**Competing Interests:** The authors have declared that no competing interests exist.

Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  be the so-called Moreau-Yosida regularization of  $f$ , which is defined by

$$F(x) := \min_{z \in \mathbb{R}^n} \left\{ f(z) + \frac{1}{2\lambda} \|z - x\|^2 \right\}, \tag{2}$$

where  $\lambda$  is a positive parameter and  $\|\cdot\|$  denotes the Euclidean norm. The problem Eq (1) is equivalent to the following problem

$$\min_{x \in \mathbb{R}^n} F(x). \tag{3}$$

It is well known that the problems Eqs (1) and (3) of the solution sets are the same. As we know, one of the most effective methods for problems Eq (3) is the trust region method.

The trust region method plays an important role in the area of nonlinear optimization, and it has been proven to be a very efficient method. Levenberg [19] and Marquardt [20] first applied this method to nonlinear least-squares problems, and Powell [21] established a convergence result for this method for unconstrained problems. Fletcher [22] first proposed a trust region method for composite nondifferentiable optimization problems. Over the past decades, many authors have studied the trust region algorithm to minimize nonsmooth objective function problems. For example, Sampaio, Yuan and Sun [23] used the trust region algorithm for nonsmooth optimization problems; Sun, Sampaio and Yuan [24] proposed a quasi-Newton trust region algorithm for nonsmooth least-squares problems; Zhang [25] used a new trust region algorithm for nonsmooth convex minimization; and Yuan, Wei and Wang [26] proposed a gradient trust region algorithm with a limited memory BFGS update for nonsmooth convex minimization problems. For other references on trust region methods, see [27–35], among others. In particular, for the problem we address in this study, as we can compute the exact Hessian, the trust region method could be very efficient. However, it is difficult to compute the Hessian at every iteration, which increases the computational workload and time.

The purpose of this paper is to present an efficient trust region algorithm to solve Eq (3). With the use of the Moreau-Yosida regularization (smoothing) and the new quasi-Newton equation, the given method has the following good properties: (i) the Hessian makes use of not only the gradient value but also the function value and (ii) the subproblem of the proposed method, which possesses the form of an unconstrained trust region subproblem, can be solved using existing methods.

The remainder of this paper is organized as follows. In the next section, we briefly review some basic results in convex analysis and nonsmooth analysis and state a new quasi-Newton secant equation. In section 3, we present a new algorithm for solving problem Eq (3). In section 4, we prove the global convergence of the proposed method. In section 5, we report numerical results and present some comparisons for the existing methods to solve problem Eq (1). We conclude our paper in Section 6.

Throughout this paper, unless otherwise specified,  $\|\cdot\|$  denotes the Euclidean norm of vectors or matrices.

## Initial results

In this section, we first state some basic results in convex analysis and nonsmooth analysis. Let

$$\theta(z, x) = f(z) + \frac{1}{2\lambda} \|z - x\|^2,$$

and denote  $p(x) := \operatorname{argmin}_{z \in \mathbb{R}^n} \theta(z, x)$ . Then,  $p(x)$  is well defined and unique, as  $\theta(z, x)$  is

strongly convex. By Eq (2),  $F$  can be rewritten as

$$F(x) = f(p(x)) + \frac{1}{2\lambda} \| p(x) - x \|^2.$$

In the following, we denote  $g(x) = \nabla F(x)$ . Some important properties of  $F$  are given as follows:

1.  $F$  is finite-valued, convex and everywhere differentiable with

$$g(x) = \nabla F(x) = \frac{x - p(x)}{\lambda}. \tag{4}$$

2. The gradient mapping  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is globally Lipschitz continuous with modulus  $\lambda$ , i.e.,

$$\| g(x) - g(y) \| \leq \frac{1}{\lambda} \| x - y \|, \quad \forall x, y \in \mathbb{R}^n. \tag{5}$$

3.  $x$  solves Eq (1) if and only if  $\nabla F(x) = 0$ , namely,  $p(x) = x$ .

It is obvious that  $F(x)$  and  $g(x)$  can be obtained through the optimal solution of  $\operatorname{argmin}_{z \in \mathbb{R}^n} \theta(z, x)$ . However, the minimizer of  $\theta(z, x)$ ,  $p(x)$  is difficult or even impossible to solve for exactly. Thus, we cannot compute the exact value of  $p(x)$  to define  $F(x)$  and  $g(x)$ . Fortunately, for each  $x \in \mathbb{R}^n$  and any  $\epsilon > 0$ , there exists a vector  $p^\alpha(x, \epsilon) \in \mathbb{R}^n$  such that

$$f(p^\alpha(x, \epsilon)) + \frac{1}{2\lambda} \| p^\alpha(x, \epsilon) - x \|^2 \leq F(x) + \epsilon. \tag{6}$$

Thus, we can use  $p^\alpha(x, \epsilon)$  to define respective approximations of  $F(x)$  and  $g(x)$  as follows, when  $\epsilon$  is small,

$$F^\alpha(x, \epsilon) := f(p^\alpha(x, \epsilon)) + \frac{1}{2\lambda} \| p^\alpha(x, \epsilon) - x \|^2 \tag{7}$$

and

$$g^\alpha(x, \epsilon) := \frac{x - p^\alpha(x, \epsilon)}{\lambda}, \tag{8}$$

The papers [36, 37] describe some algorithms to calculate  $p^\alpha(x, \epsilon)$ . The following remarkable feature of  $F^\alpha(x, \epsilon)$  and  $g^\alpha(x, \epsilon)$  is obtained from [38].

**Proposition 2.1** Let  $p^\alpha(x, \epsilon)$  be a vector satisfying Eq (6), and  $F^\alpha(x, \epsilon)$  and  $g^\alpha(x, \epsilon)$  are defined by Eqs (7) and (8), respectively. Then, we obtain

$$F(x) \leq F^\alpha(x, \epsilon) \leq F(x) + \epsilon, \tag{9}$$

$$\| p^\alpha(x, \epsilon) - p(x) \| \leq \sqrt{2\lambda\epsilon}, \tag{10}$$

and

$$\| g^\alpha(x, \epsilon) - g(x) \| \leq \sqrt{\frac{2\epsilon}{\lambda}}. \tag{11}$$

The relations Eqs (9), (10) and (11) imply that  $F^\alpha(x, \epsilon)$  and  $g^\alpha(x, \epsilon)$  may be made arbitrarily close to  $F(x)$  and  $g(x)$ , respectively, by choosing the parameter  $\epsilon$  to be small enough.

Second, recall that when  $f$  is smooth, the quasi-Newton secant method is used to solve problem Eq (1). The iterate  $x_k$  satisfies  $\nabla f_k + B_k(x_{k+1} - x_k) = 0$ , where  $\nabla f_k = \nabla f(x_k)$ ,  $B_k$  is an approximation Hessian of  $f$  at  $x_k$ , and the sequence of matrix  $\{B_k\}$  satisfies the secant equation as

follows.

$$B_{k+1}s_k = y_k, \tag{12}$$

where  $y_k = \nabla f_{k+1} - \nabla f_k$  and  $s_k = x_{k+1} - x_k$ . However, the function values are not exploited in Eq (12), which the method solves by only using the gradient information. Motivated by the above observations, we hope to develop a method that uses both the gradient information and function information. This problem has been studied by several authors. In particular, Wei, Li and Qi [39] proposed an important modified secant equation by using not only the gradient values but also the function values, and the modified secant is defined as

$$B_{k+1}s_k = v_k, \tag{13}$$

where  $v_k = y_k + \beta_k s_k$ ,  $f_k = f(x_k)$ ,  $\nabla f_k = \nabla f(x_k)$ , and  $\beta_k = \frac{(\nabla f_{k+1} + \nabla f_k)^T s_k + 2(f_k - f_{k+1})}{\|s_k\|^2}$ . When  $f$  is twice continuously differentiable and  $B_{k+1}$  is updated by the BFGS formula [40–43], where  $B_k = I$  is a unit matrix if  $k = 0$ , this secant Eq (13) possesses the following remarkable property:

$$f_k = f_{k+1} + \nabla f_{k+1}^T s_k + \frac{1}{2} s_k^T B_{k+1} s_k$$

This property holds for all  $k$ . Based on the result of Theorem 2.1 [39], Eq (13) has an advantage over Eq (12) in this approximate relation.

### The new model

In this section, we present a modified BFGS formula using trust region model for solving Eq (1), which is motivated by the Moreau-Yosida regularization (smoothing), general trust region method and the new secant Eq (13). First, we describe the trust region method. In each iteration, a trial step  $d_k$  is generated by solving an adaptive trust region subproblem, in which the values of the gradient of  $F(x)$  at  $x_k$  and Eq (13) are used:

$$\begin{aligned} \min \quad & q_k(d) = g^z(x_k, \epsilon_k)^T d + \frac{1}{2} d^T B_k d, \\ \text{s.t.} \quad & \|d\| \leq \Delta_k, \end{aligned} \tag{14}$$

where the scalar  $\epsilon_k > 0$  and  $\Delta_k$  describe the trust region radius.

Let  $d_k$  be the optimal solution of Eq (14). The actual reduction is defined by

$$\text{Are } d_k := F^z(x_k, \epsilon_k) - F^z(x_k + d_k, \epsilon_{k+1}), \tag{15}$$

and we define the predict reduction as

$$\text{Pre } d_k := -g^z(x_k, \epsilon_k)^T d_k - \frac{1}{2} d_k^T B_k d_k. \tag{16}$$

Then, we define  $r_k$  to be the ratio between  $\text{Are } d_k$  and  $\text{Pre } d_k$

$$r_k := \frac{\text{Are } d_k}{\text{Pre } d_k}. \tag{17}$$

Based on the new secant Eq (13) and with  $B_{k+1}$  being updated by the BFGS formula, we propose a modified BFGS formula. The  $B_{k+1}$  is defined by

$$B_{k+1} := \begin{cases} B_k, & \text{if } s_k^T v_k \leq 0, \\ B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{v_k v_k^T}{v_k^T s_k}, & \text{if } s_k^T v_k > 0, \end{cases} \tag{18}$$

where  $s_k = x_{k+1} - x_k, y_k = g^\alpha(x_{k+1}, \epsilon_{k+1}) - g^\alpha(x_k, \epsilon_k), v_k = y_k + \beta_k s_k$  and

$$\beta_k = \frac{(g^\alpha(x_{k+1}, \epsilon_{k+1}) + g^\alpha(x_k, \epsilon_k))^T s_k + 2(F^\alpha(x_k, \epsilon_k) - F^\alpha(x_{k+1}, \epsilon_{k+1}))}{\|s_k\|^2},$$

if  $k = 0$ , then  $B_k = I$ , and  $I$  is a unit matrix.

We now list the steps of the modified trust region algorithm as follows.

**Algorithm 1.**

**Step 0.** Choose  $x_0 \in \mathbb{R}^n, 0 < \sigma_1 < \sigma_2 < 1, 0 < \eta_1 < 1 < \eta_2, \lambda > 0, 0 \leq \epsilon \ll 1, \Delta_{max} \geq \Delta_0 > 0$  is called the maximum value of trust region radius,  $B_0 = I$ , and  $I$  is the unit matrix. Let  $k := 0$ .

**Step 1.** Choose a scalar  $\epsilon_{k+1}$  satisfying  $0 < \epsilon_{k+1} < \epsilon_k$ , and calculate  $p^\alpha(x_k, \epsilon_k), g^\alpha(x_k, \epsilon_k) = \frac{x_k - p^\alpha(x_k, \epsilon_k)}{\lambda}$ . If  $x_k$  satisfies the termination criterion  $\|g^\alpha(x_k, \epsilon_k)\| \leq \epsilon$ , then stop. Otherwise, go to Step 2.

**Step 2.**  $d_k$  solves the trust region subproblem Eq (14).

**Step 3.** Compute  $Are d_k, Pre d_k, r_k$  using Eqs (15), (16) and (17).

**Step 4.** Regulate the trust region radius. Let

$$\Delta_{k+1} := \begin{cases} \eta_1 \Delta_k, & \text{if } r_k < \sigma_1, \\ \Delta_k, & \text{if } \sigma_1 \leq r_k \leq \sigma_2, \\ \min\{\eta_2 \Delta_k, \Delta_{max}\}, & \text{if } r_k > \sigma_2. \end{cases}$$

**Step 5.** If the condition  $r_k \geq \sigma_1$  holds, then let  $x_{k+1} = x_k + d_k$ , update  $B_{k+1}$  by Eq (18), and let  $k := k + 1$ ; go back to Step 1. Otherwise, let  $x_{k+1} := x_k$  and  $k := k + 1$ ; return to Step 2.

Similar to Dennis and Moré [44] or Yuan and Sun [45], we have the following result.

**Lemma 1** *If and only if the condition  $s_k^T v_k > 0$  holds,  $B_{k+1}$  will inherit the positive property of  $B_k$ .*

*Proof* “ $\Rightarrow$ ” If  $B_{k+1}$  is symmetric and positive definite, then

$$\begin{aligned} s_k^T B_{k+1} s_k &= s_k^T \left[ B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{v_k v_k^T}{v_k^T s_k} \right] s_k \\ &= s_k^T B_k s_k - \frac{s_k^T B_k s_k s_k^T B_k s_k}{s_k^T B_k s_k} + \frac{s_k^T v_k v_k^T s_k}{v_k^T s_k} \\ &= s_k^T v_k \\ &> 0. \end{aligned}$$

“ $\Leftarrow$ ” For the proof of the converse, suppose that  $s_k^T v_k > 0$  and  $B_k$  is symmetric and positive definite for all  $k \geq 0$ . We shall prove that  $x^T B_{k+1} x > 0$  holds for arbitrary  $x \neq 0$  and  $x \in \mathbb{R}^n$  by induction. It is easy to see that  $B_0 = I$  is symmetric and positive definite. Thus, we have

$$\begin{aligned} x^T B_{k+1} x &= x^T B_k x - \frac{x^T B_k s_k s_k^T B_k x}{s_k^T B_k s_k} + \frac{x^T v_k v_k^T x}{v_k^T s_k} \\ &= x^T B_k x - \frac{(x^T B_k s_k)^2}{s_k^T B_k s_k} + \frac{(x^T v_k)^2}{v_k^T s_k}. \end{aligned} \tag{19}$$

Because  $B_k$  is symmetric and positive definite for all  $k \geq 0$ , there exists a symmetric and positive definite matrix  $B_k^{\frac{1}{2}}$  such that  $B_k = B_k^{\frac{1}{2}} B_k^{\frac{1}{2}}$ . Thus, by using the Cauchy-Schwartz inequality, we

obtain

$$\begin{aligned}
 (x^T B_k s_k)^2 &= \left[ x^T B_k^{\frac{1}{2}} B_k^{\frac{1}{2}} s_k \right]^2 = \left[ \left( B_k^{\frac{1}{2}} x \right)^T \left( B_k^{\frac{1}{2}} s_k \right) \right]^2 \\
 &\leq \left\| B_k^{\frac{1}{2}} x \right\|^2 \left\| B_k^{\frac{1}{2}} s_k \right\|^2 \\
 &= \left( B_k^{\frac{1}{2}} x \right)^T \left( B_k^{\frac{1}{2}} x \right) \left( B_k^{\frac{1}{2}} s_k \right)^T \left( B_k^{\frac{1}{2}} s_k \right) \\
 &= (x^T B_k x) (s_k^T B_k s_k).
 \end{aligned} \tag{20}$$

It is not difficult to prove that the above inequality holds true if and only if there exists a real number  $\gamma_k \neq 0$  such that  $B_k^{\frac{1}{2}} x = \gamma_k B_k^{\frac{1}{2}} s_k$ , namely,  $x = \gamma_k s_k$ .

Hence, if Eq (20) strictly holds (and note that  $s_k v_k^T > 0$ ), then from Eq (19), we have

$$\begin{aligned}
 x^T B_{k+1} x &> x^T B_k x - \frac{(x^T B_k s_k)^2}{s_k^T B_k s_k} + \frac{(x^T v_k)^2}{v_k^T s_k} \\
 &= \frac{(x^T v_k)^2}{v_k^T s_k} > 0.
 \end{aligned}$$

Otherwise,  $(x^T B_k s_k)^2 = (x^T B_k x) (s_k^T B_k s_k)$ ; then, there exists  $\gamma_k$  such that  $x = \gamma_k s_k$ . Thus,

$$\begin{aligned}
 x^T B_{k+1} x &= \frac{[(\gamma_k s_k)^T v_k]^2}{v_k^T s_k} \\
 &= \gamma_k^2 s_k^T v_k > 0.
 \end{aligned}$$

Therefore, for each  $0 \neq x \in \mathbb{R}^n$ , we have  $x^T B_{k+1} x > 0$ . This completes the proof.

Lemma 1 states that if  $s_k^T v_k > 0$ , then the matrix sequence  $\{B_k\}$  is symmetric and positive definite, which is updated by the BFGS formula of Eq (18).

### Convergence analysis

In this section, the global convergence of Algorithm 1 is established under the assumption that the following conditions are required.

**Assumption A.**

1. Let the level set  $\Omega$

$$\Omega = \{x \in \mathbb{R}^n | F^z(x, \epsilon) \leq F^z(x_0, \epsilon), \quad \forall x_0 \in \mathbb{R}^n\}.$$

2.  $F$  is bounded from below.
3. The matrix sequence  $\{B_k\}$  is bounded on  $\Omega$ , which means that there exists a positive constant  $M$  such that

$$\| B_k \| \leq M \quad \forall k.$$

4. The sequence  $\{\epsilon_k\}$  converges to zero.

Now, we present the following lemma.

**Lemma 2** If  $d_k$  is the solution of Eq (14), then

$$\text{Pre } d_k = q_k(0) - q_k(d_k) \geq \frac{1}{2} \| g^z(x_k, \epsilon_k) \| \min \left\{ \Delta_k, \frac{\| g^z(x_k, \epsilon_k) \|}{\| B_k \|} \right\}. \tag{21}$$

*Proof* Similar to the proof of Lemma 7(6.2) in Ma [46]. Note that the matrix sequence  $\{B_k\}$  is symmetric and positive definite; then, we present  $d_k^c$  to be a Cauchy point at iteration point  $x_k$ , which is defined by

$$d_k^c = -\mu_k \frac{\Delta_k}{\|g^z(x_k, \epsilon_k)\|} g^z(x_k, \epsilon_k),$$

where  $\mu_k = \min \left\{ \frac{\|g^z(x_k, \epsilon_k)\|^3}{\Delta_k g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k)}, 1 \right\}$ . It is easy to verify that the Cauchy point is a feasible point, i.e.,  $\|d_k^c\| \leq \Delta_k$ .

If  $\frac{\|g^z(x_k, \epsilon_k)\|^3}{\Delta_k g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k)} > 1$ , then

$$\|g^z(x_k, \epsilon_k)\|^3 > \Delta_k g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k),$$

and

$$d_k^c = -\frac{\Delta_k}{\|g^z(x_k, \epsilon_k)\|} g^z(x_k, \epsilon_k).$$

Thus, we obtain

$$\begin{aligned} \text{Pre } d_k^c &= -q_k \left( -\frac{\Delta_k}{\|g^z(x_k, \epsilon_k)\|} g^z(x_k, \epsilon_k) \right) \\ &= -g^z(x_k, \epsilon_k)^T \left( -\frac{\Delta_k}{\|g^z(x_k, \epsilon_k)\|} g^z(x_k, \epsilon_k) \right) \\ &\quad - \frac{1}{2} \left( -\frac{\Delta_k}{\|g^z(x_k, \epsilon_k)\|} g^z(x_k, \epsilon_k) \right)^T B_k \left( -\frac{\Delta_k}{\|g^z(x_k, \epsilon_k)\|} g^z(x_k, \epsilon_k) \right) \\ &= \frac{\Delta_k}{\|g^z(x_k, \epsilon_k)\|} \|g^z(x_k, \epsilon_k)\|^2 - \frac{1}{2} \frac{\Delta_k^2}{\|g^z(x_k, \epsilon_k)\|^2} g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k) \\ &\geq \frac{1}{2} \Delta_k \|g^z(x_k, \epsilon_k)\| \\ &\geq \frac{1}{2} \|g^z(x_k, \epsilon_k)\| \min \left\{ \Delta_k, \frac{\|g^z(x_k, \epsilon_k)\|}{\|B_k\|} \right\}. \end{aligned}$$

Otherwise, we have  $d_k^c = -\frac{\|g^z(x_k, \epsilon_k)\|^2}{g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k)} g^z(x_k, \epsilon_k)$ . Thus, we obtain

$$\begin{aligned} \text{Pre } d_k^c &= -g^z(x_k, \epsilon_k)^T \left( -\frac{\|g^z(x_k, \epsilon_k)\|^2}{g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k)} g^z(x_k, \epsilon_k) \right) \\ &\quad - \frac{1}{2} \left( -\frac{\|g^z(x_k, \epsilon_k)\|^2}{g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k)} g^z(x_k, \epsilon_k) \right)^T B_k \left( -\frac{\|g^z(x_k, \epsilon_k)\|^2}{g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k)} g^z(x_k, \epsilon_k) \right) \\ &= \frac{1}{2} \frac{\|g^z(x_k, \epsilon_k)\|^4}{g^z(x_k, \epsilon_k)^T B_k g^z(x_k, \epsilon_k)} \\ &\geq \frac{1}{2} \frac{\|g^z(x_k, \epsilon_k)\|^2}{\|B_k\|} \\ &\geq \frac{1}{2} \|g^z(x_k, \epsilon_k)\| \min \left\{ \Delta_k, \frac{\|g^z(x_k, \epsilon_k)\|}{\|B_k\|} \right\}. \end{aligned}$$

Let  $d_k$  be the solution of Eq (14). Because  $q_k(d_k^c) \geq q_k(d_k)$ , we have

$$Pre d_k = q_k(0) - q_k(d_k) \geq \frac{1}{2} \|g^z(x_k, \epsilon_k)\| \min \left\{ \Delta_k, \frac{\|g^z(x_k, \epsilon_k)\|}{\|B_k\|} \right\}.$$

This completes the proof.

**Lemma 3** Let Assumption A hold true and the sequence  $\{x_k\}$  be generated by Algorithm 1. If  $d_k$  is the solution of Eq (14), then

$$|Are d_k - Pre d_k| = o(\|d_k\|^2). \tag{22}$$

*Proof* Let  $d_k$  be the solution of Eq (14). By using Taylor expansion,  $F^\alpha(x_k + d_k, \epsilon_{k+1})$  can be expressed by

$$F^z(x_k + d_k, \epsilon_{k+1}) = F^z(x_k, \epsilon_k) + g^z(x_k, \epsilon_k)^T d_k + \frac{1}{2} d_k^T B_k d_k + o(\|d_k\|^2), \tag{23}$$

Note that with the definitions of  $Are d_k$  and  $Pre d_k$  and by using Eq (23), we have

$$\begin{aligned} |Are d_k - Pre d_k| &= |F^z(x_k, \epsilon_k) - F^z(x_k + d_k, \epsilon_{k+1}) + q_k(d_k)| \\ &= o(\|d_k\|^2). \end{aligned}$$

The proof is complete.

**Lemma 4** Let Assumption A hold. Then, Algorithm 1 does not circle in the inner cycle infinitely.

*Proof* Suppose, by contradiction to the conclusion of the lemma, that Algorithm 1 cycles between Steps 2 and 5 infinitely at iteration point  $x_k$ , i.e.,  $r_k < \sigma_1$  and that there exists a scalar  $\rho > 0$  such that  $\|g^\alpha(x_k, \epsilon_k)\| \geq \rho$ . Thus, noting that  $0 < \eta_1 < 1$ , we have

$$\|d_k\| \leq \Delta_k = \eta_1^k \Delta_0 \rightarrow 0, \text{ for } k \rightarrow \infty.$$

By using the result Eq (22) of Lemma 3 and the definition of  $r_k$ , we obtain

$$\begin{aligned} |r_k - 1| &= \frac{|Are d_k - Pre d_k|}{|Pre d_k|} \\ &\leq \frac{2o(\|d_k\|^2)}{\|g^z(x_k, \epsilon_k)\| \min \left\{ \Delta_k, \frac{\|g^z(x_k, \epsilon_k)\|}{\|B_k\|} \right\}} \rightarrow 0, \text{ for } k \rightarrow \infty. \end{aligned}$$

which means that we must have  $r_k \geq \sigma_1$ ; this contradicts the assumption that  $r_k < \sigma_1$ , and the proof is complete.

Based on the above lemmas, we can now demonstrate the global convergence of Algorithm 1 under suitable conditions.

**Theorem 1 (Global Convergence).** Suppose that Assumption A holds and that the sequence  $\{x_k\}$  is generated by Algorithm 1. Let  $d_k$  be the solution of Eq (14). Then,  $\liminf_{k \rightarrow \infty} \|g_k\| = 0$  holds, and any accumulation point of  $x_k$  is an optimal solution of Eq (1).

*Proof* We first prove that

$$\liminf_{k \rightarrow \infty} \|g^z(x_k, \epsilon_k)\| = 0. \tag{24}$$

Suppose that  $g^z(x_k, \epsilon_k) \neq 0$ . Without loss of generality, by the definition of  $r_k$ , we have

$$|r_k - 1| = \left| \frac{F^z(x_k + d_k, \epsilon_{k+1}) - F^z(x_k, \epsilon_k) - q_k(d_k)}{q_k(d_k)} \right|. \tag{25}$$



Using Taylor expansion, we obtain

$$F^z(x_k + d_k, \epsilon_{k+1}) = F^z(x_k, \epsilon_k) + g^z(x_k, \epsilon_k)^T d_k + \int_0^1 d_k^T [g^z(x_k + td_k, \epsilon_{k+1}) - g^z(x_k, \epsilon_k)] dt.$$

When  $\Delta_k > 0$  and small enough, we have

$$\begin{aligned} & |F^z(x_k + d_k, \epsilon_{k+1}) - F^z(x_k, \epsilon_k) - q_k(d_k)| \\ &= \left| \frac{1}{2} d_k^T B_k d_k - \int_0^1 d_k^T [g^z(x_k + td_k, \epsilon_{k+1}) - g^z(x_k, \epsilon_k)] dt \right| \\ &\leq \frac{1}{2} M \|d_k\|^2 + o(\|d_k\|). \end{aligned} \tag{26}$$

Suppose that there exists  $\omega_0 > 0$  such that  $\|g^z(x_k, \epsilon_k)\| \geq \omega_0$ . By contradiction, using Eqs (25) and (26) and Lemma 2, we have

$$\begin{aligned} |r_k - 1| &\leq \frac{\frac{1}{2} M \|d_k\|^2 + o(\|d_k\|)}{\frac{1}{2} \|g^z(x_k, \epsilon_k)\| \min\left\{\Delta_k, \frac{\|g^z(x_k, \epsilon_k)\|}{\|B_k\|}\right\}} \\ &\leq \frac{M \Delta_k^2 + o(\Delta_k)}{\omega_0 \min\left\{\Delta_k, \frac{\omega_0}{M}\right\}} \\ &= O(\Delta_k). \end{aligned} \tag{27}$$

which means that there exists sufficiently small  $\hat{\Delta} > 0$  such that  $\Delta_k \leq \hat{\Delta}$  for each  $k$ , and we have  $|r_k - 1| < 1 - \sigma_2$ , i.e.,  $r_k > \sigma_2$ . Then, according to the Algorithm 1, we have  $\Delta_{k+1} \geq \Delta_k$ .

Thus, there exists a positive integer  $k_0$  and a constant  $\rho_0$  for arbitrary  $k \geq k_0$  and satisfying  $\Delta_k \leq \hat{\Delta}$ , for which we have

$$\Delta_k \neq \rho_0 \hat{\Delta}. \tag{28}$$

On the other hand, because  $F$  is bounded from below, and supposing that there exists an infinite number  $k$  such that  $r_k > \sigma_1$ , by the definition of  $r_k$  and Lemma 2, for each  $k \geq k_0$ ,

$$\begin{aligned} & F^z(x_k, \epsilon_k) - F^z(x_k + d_k, \epsilon_{k+1}) \\ &> \sigma_1 [q_k(0) - q_k(d_k)] \\ &\geq \frac{\sigma}{2} \omega_0 \min\left\{\Delta_k, \frac{\omega_0}{M}\right\}. \end{aligned}$$

which means that  $\Delta_k \rightarrow 0$  for  $k \rightarrow \infty$ ; this is a contradiction to Eq (28).

Moreover, suppose that for sufficiently large  $k$ , we have  $r_k < \sigma_1$ . Then,  $\Delta_k = \eta_1^k \Delta_0$ , and we can see that  $\Delta_k \rightarrow 0$  for  $k \rightarrow \infty$ ; this is also a contradiction to Eq (28). The contradiction shows that Eq (24) holds.

We now show that  $\liminf_{k \rightarrow \infty} \|g_k\| = 0$  holds. By using Eq (11), we have

$$\|g^z(x_k, \epsilon_k) - g(x_k)\| \leq \sqrt{\frac{2\epsilon_k}{\lambda}}.$$

Together with Assumption A(iv), this implies that

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \tag{29}$$

Finally, we make a final assertion. Let  $x^*$  be an accumulation point of  $\{x_k\}$ . Then, without loss of generality, there exists a subsequence  $\{x_k\}_K$  satisfying

$$\lim_{k \rightarrow \infty, k \in K} x_k = x^*. \tag{30}$$

From the properties of  $F$ , we have

$$g(x_k) = \frac{x_k - p(x_k)}{\lambda}.$$

Thus, by using Eqs (29) and (30), we have  $x^* = p(x^*)$ . Therefore,  $x^*$  is an optimal solution of Eq (1). The proof is complete.

Similar to Theorem 3.7 in [25], we can show that the rate of convergence of Algorithm 1 is Q-superlinear. We omit this proof here (the proof of the Q-superlinear convergence can be found in [25]).

**Theorem 2 (Q-superlinear Convergence)** [25] *Suppose that Assumption A(ii) holds, that the sequence  $\{x_k\}$  is generated by Algorithm 1, which has a limit point  $x^*$ , and that  $g$  is BD-regular and semismooth at  $x^*$ . Furthermore, suppose that  $\epsilon_k = o(\|g(x_k)\|^2)$ . Then,*

1.  $x^*$  is the unique solution of Eq (1);
2. the entire sequence  $\{x_k\}$  converges to  $x^*$  Q-superlinearly, i.e.,

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

## Results

In this section, we test our modified BFGS formula using a trust region model for solving non-smooth problems. The type of nonsmooth problems addressed in Table 1 can be found in [47–53]. The problem dimensions and optimum function values are listed in Table 1, where “No.” is the number of the test problem, “Dim” is the dimension of the test problem, “Problem” is the name of the test problem, “ $x_0$ ” is the initial point, and “ $f_{ops}(x)$ ” is the optimization function evaluation. Here, the modified algorithm was implemented using MATLAB 7.0.4, and all numerical experiments were run on a PC with CPU Intel CORE(TM) 2 Duo T6600 2.20 GHZ, with 2.00 GB of RAM and with the Windows 7 operating system.

**Table 1. Problem descriptions for test problems.**

No.	Dim	Problem	$x_0$	$f_{ops}(x)$
1	2	Rosenbrock [47]	(-1.2, 1.0)	0
2	2	Crescent [47]	(-1.5, 2.0)	0
3	2	CB2 [48]	(1.0, -0.1)	1.9522245
4	2	CB3 [48]	(2.0, 2.0)	2.0
5	2	DEM [49]	(1.0, 1.0)	-3.0
6	2	QL [50]	(-1.0, 5.0)	7.20
7	2	LQ [50]	(-0.5, -0.5)	-1.4142136
8	2	Mifflin 2 [51]	(-1.0, -1.0)	-1.0
9	5	Shor [52]	(0.0, 0.0, 0.0, 0.0, 1.0)	22.600162
10	50	MXHILB [53]	ones(50, 1)	0
11	50	LIHILB [53]	ones(50, 1)	0

doi:10.1371/journal.pone.0140606.t001

**Table 2. Test results.**

No.	PBL NI/NF/ $f(x)$	LGTR NI/NF/ $f(x)$	BT NI/NF/ $f(x)$	Algorithm 1 NI/NF/ $f(x)$
1	42/45/ $3.81 \times 10^{-5}$	—	79/88/ $1.30 \times 10^{-10}$	26/66/ $4.247136 \times 10^{-6}$
2	18/20/ $6.79 \times 10^{-5}$	10/10/ $3.156719 \times 10^{-5}$	24/27/ $9.44 \times 10^{-5}$	13/13/ $2.521899 \times 10^{-5}$
3	32/34/1.9522245	10/11/1.952225	13/16/1.952225	4/6/1.952262
4	14/16/2.0	2/3/2.000217	13/21/2.0	3/4/2.000040
5	17/19/-3.0	3/3/-2.999700	9/13/-3.0	4/24/-2.999922
6	13/15/7.2000015	19/119/7.200001	12/17/7.200009	9/9/7.200043
7	11/12/-1.4142136	1/1/-1.207068	10/11/-1.414214	2/2/-1.414214
8	66/68/-0.99999941	3/3/-0.9283527	6/13/-1.0	4/4/-0.9978547
9	27/29/22.600162	42/443/22.62826	29/30/22.600160	8/9/22.600470
10	19/20/ $4.24 \times 10^{-7}$	12/12/9.793119 $\times 10^{-3}$	—	23/108/5.228012 $\times 10^{-3}$
11	19/20/ $9.90 \times 10^{-8}$	20/63/9.661137 $\times 10^{-3}$	—	7/7/2.632534 $\times 10^{-3}$
Total	278/298	164/1111	353/412	103/252

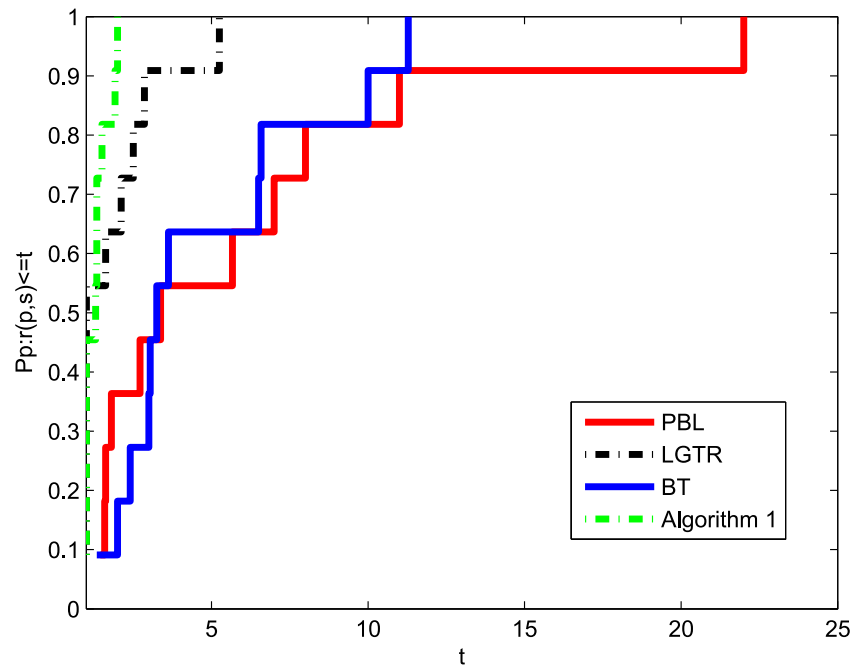
doi:10.1371/journal.pone.0140606.t002

To test the performance of the given algorithm for the problems listed in Table 1, we compared our method with the trust region concept (BT) of paper [15], the proximal bundle method (PBL) of paper [17] and the gradient trust region algorithm with limited memory BFGS update (LGTR) described in [26]. The parameters were chosen as follows:  $\sigma_1 = 0.45$ ,  $\sigma_2 = 0.75$ ,  $\eta_1 = 0.5$ ,  $\eta_2 = 4$ ,  $\lambda = 1$ ,  $\Delta_0 = 0.5 < \Delta_{max} = 100$  and  $\epsilon_k = \frac{1}{(2+k)^2}$  (where  $k$  is the iterate number). We stopped the algorithm when the condition  $\|g^\alpha(x, \epsilon)\| \leq 10^{-6}$  was satisfied. Based on the idea of [26], we use the function *fminsearch* in MATLAB for solving  $\min \theta(z, x)$ . Then, we obtained the solution  $p(x)$ ; moreover, we obtained  $g^\alpha(x, \epsilon)$ , which is computed using Eq (8). Meanwhile, we also listed the results of PBL, LGTR, BT and our modified algorithm in Table 2. The numerical results of PBL and BT can be found in [17], and the numerical results of LGTR can be found in [26]. The following notations are used in Table 2: “NI” is the number of iterations; “NF” is the number of the function evaluations; “ $f(x)$ ” is the function value at final iteration; “—” indicates that the algorithm fails to solve the problem; and “Total” denotes the sum of the NI/NF.

The numerical results show that the performance of our algorithm is superior to those of the methods in Table 2. It can be seen clearly that the sum of our algorithm relative to NI and NF is less than the other three algorithms. The paper [54] provides a new tool for analyzing the efficiency of these four algorithms. Figs 1 and 2 show the performances of these four methods relative to NI and NF of Table 2, respectively. These two figures prove that Algorithm 1 provides a good performance for all the problems tested compared to PBL, LGTR and BT. In sum, the preliminary numerical results indicate that the modified method is efficient for solving nonsmooth convex minimizations.

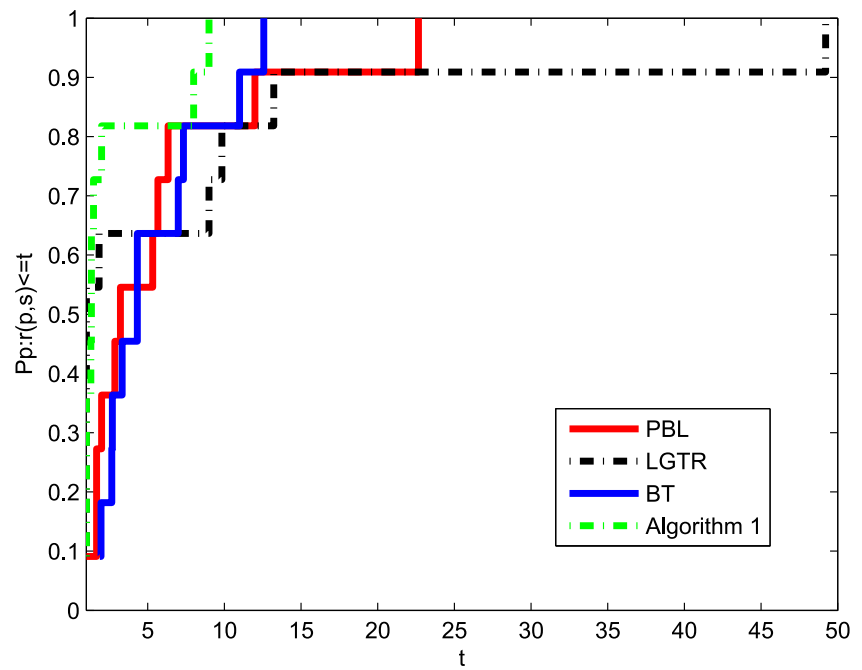
### Conclusion

The trust region method is one of the most efficient optimization methods. In this paper, by using the Moreau-Yosida regularization (smoothing) and a new secant equation with the BFGS formula, we present a modified BFGS formula using a trust region model for solving nonsmooth convex minimizations. Our algorithm does not compute the Hessian of the objective function at every iteration, which decrease the computational workload and time, and it uses the function information and the gradient information. Under suitable conditions, global convergence is established, and we show that the rate of convergence of our algorithm is Q-



**Fig 1. Performance profiles of these methods (NI).**

doi:10.1371/journal.pone.0140606.g001



**Fig 2. Performance profiles of these methods (NF).**

doi:10.1371/journal.pone.0140606.g002

superlinear. Numerical results show that this algorithm is efficient. We believe that this algorithm can be used in future applications to solve non smooth convex minimizations.

## Acknowledgments

This work is supported by China NSF (Grant No. 11261006 and 11161003), the Guangxi Science Fund for Distinguished Young Scholars (No. 2015GXNSFGA139001), NSFC No. 61232016, NSFC No. U1405254, and PAPD issue of Jiangsu advantages discipline. The authors wish to thank the editor and the referees for their useful suggestions and comments which greatly improve this paper.

## Author Contributions

Conceived and designed the experiments: ZC GY ZS. Performed the experiments: ZC GY ZS. Analyzed the data: ZC GY ZS WL XW XD. Contributed reagents/materials/analysis tools: ZC GY ZS WL XW XD. Wrote the paper: ZC GY ZS.

## References

1. Steihaug T. The conjugate gradient method and trust regions in large scale optimization, *SIAM Journal on Numerical Analysis*, 20, 626–637 (1983) doi: [10.1137/0720042](https://doi.org/10.1137/0720042)
2. Dai Y, Yuan Y. A nonlinear conjugate gradient with a strong global convergence properties, *SIAM Journal on Optimization*, 10, 177–182 (2000) doi: [10.1137/S1052623497318992](https://doi.org/10.1137/S1052623497318992)
3. Wei Z, Li G and Qi L. New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems, *Applied Mathematics and Computation*, 179, 407–430 (2006) doi: [10.1016/j.amc.2005.11.150](https://doi.org/10.1016/j.amc.2005.11.150)
4. Li L, Peng H, Kurths J, Yang Y and Schellnhuber H.J. Chaos-order transition in foraging behavior of ants, *PNAS*, 111, 8392–8397 (2014) doi: [10.1073/pnas.1407083111](https://doi.org/10.1073/pnas.1407083111) PMID: [24912159](https://pubmed.ncbi.nlm.nih.gov/24912159/)
5. Peng H, Li L, Yang Y and Liu F. Parameter estimation of dynamical systems via a chaotic ant swarm, *Physical Review E*, 81, 016207, (2010) doi: [10.1103/PhysRevE.81.016207](https://doi.org/10.1103/PhysRevE.81.016207)
6. Wan M, Li L, Xiao J, Wang C and Yang Y. Data clustering using bacterial foraging optimization, *Journal of Intelligent Information Systems*, 38, 321–341 (2012) doi: [10.1007/s10844-011-0158-3](https://doi.org/10.1007/s10844-011-0158-3)
7. Chan C, Katsaggelos A.K and Sahakian A.V. Image sequence filtering in quantum noise with applications to low-dose fluoroscopy, *IEEE Transactions on Medical Imaging*, 12, 610–621 (1993) doi: [10.1109/42.241890](https://doi.org/10.1109/42.241890) PMID: [18218455](https://pubmed.ncbi.nlm.nih.gov/18218455/)
8. Banham M.R, Katsaggelos A.K. Digital image restoration, *IEEE Signal Processing Magazine*, 14, 24–41 (1997) doi: [10.1109/79.581363](https://doi.org/10.1109/79.581363)
9. Gu B, Sheng V. Feasibility and finite convergence analysis for accurate on-line v-support vector learning, *IEEE Transactions on Neural Networks and Learning Systems*, 24, 1304–1315 (2013) doi: [10.1109/TNNLS.2013.2250300](https://doi.org/10.1109/TNNLS.2013.2250300)
10. Li J, Li X, Yang B and Sun X. Segmentation-based image copy-move forgery detection scheme, *IEEE Transactions on Information Forensics and Security*, 10, 507–518 (2015) doi: [10.1109/TIFS.2014.2381872](https://doi.org/10.1109/TIFS.2014.2381872)
11. Wen X, Shao L, Fang W, and Xue Y. Efficient feature selection and classification for vehicle detection, *IEEE Transactions on Circuits and Systems for Video Technology*, (2015) doi: [10.1109/TCSVT.2014.2358031](https://doi.org/10.1109/TCSVT.2014.2358031)
12. Zhang H, Wu J, Nguyen T and Sun M. Synthetic aperture radar image segmentation by modified student's t-mixture model, *IEEE Transaction on Geoscience and Remote Sensing*, 52, 4391–4403 (2014) doi: [10.1109/TGRS.2013.2281854](https://doi.org/10.1109/TGRS.2013.2281854)
13. Fu Z. Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing, *IEICE Transactions on Communications*, E98-B, 190–200 (2015) doi: [10.1587/transcom.E98.B.190](https://doi.org/10.1587/transcom.E98.B.190)
14. Yuan G, Wei Z and Li G, A modified Polak-Ribière-Polyak conjugate gradient algorithm for nonsmooth convex programs, *Journal of Computational and Applied Mathematics*, 255, 86–96 (2014) doi: [10.1016/j.cam.2013.04.032](https://doi.org/10.1016/j.cam.2013.04.032)

15. Schramm H, Zowe J. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results, *SIAM Journal on Optimization*, 2, 121–152 (1992) doi: [10.1137/0802008](https://doi.org/10.1137/0802008)
16. Haarala M, Miettinen K and Mäkelä M.M. New limited memory bundle method for large-scale non-smooth optimization, *Optimization Methods and Software*, 19, 673–692 (2004) doi: [10.1080/10556780410001689225](https://doi.org/10.1080/10556780410001689225)
17. Lukšan L, Vlček J. A bundle-Newton method for nonsmooth unconstrained minimization, *Mathematical Programming*, 83, 373–391 (1998) doi: [10.1016/S0025-5610\(97\)00108-1](https://doi.org/10.1016/S0025-5610(97)00108-1)
18. Wei Z, Qi L and Birge J.R. A new methods for nonsmooth convex optimization, *Journal of Inequalities and Applications*, 2, 157–179 (1998)
19. Levenberg K. A method for the solution of certain nonlinear problem in least squares, *Quarterly Journal of Mechanics and Applied Mathematics*, 2, 164–166 (1944)
20. Martinet B. Régularisation d'inéquations variationnelles par approximations successives, *Rev. Fr. Inform. Rech. Oper.*, 4, 154–159 (1970)
21. Powell M.J.D. Convergence properties of a class of minimization algorithms. In: Mangasarian Q.L., Meyer R.R., Robinson S.M. (eds.) *Nonlinear Programming*, vol. 2. Academic Press, New York (1975)
22. Fletcher R. A model algorithm for composite nondifferentiable optimization problems, *Math. Program. Stud.*, 17, 67–76 (1982) doi: [10.1007/BFb0120959](https://doi.org/10.1007/BFb0120959)
23. Sampaio R.J.B, Yuan J and Sun W. Trust region algorithm for nonsmooth optimization, *Applied Mathematics and Computation*, 85, 109–116 (1997) doi: [10.1016/S0096-3003\(96\)00112-9](https://doi.org/10.1016/S0096-3003(96)00112-9)
24. Sun W, Sampaio R.J.B and Yuan J. Quasi-Newton trust region algorithm for non-smooth least squares problems, *Applied Mathematics and Computation*, 105, 183–194 (1999) doi: [10.1016/S0096-3003\(98\)10103-0](https://doi.org/10.1016/S0096-3003(98)10103-0)
25. Zhang L. A new trust region algorithm for nonsmooth convex minimization, *Applied Mathematics and Computation* 193, 135–142 (2007) doi: [10.1016/j.amc.2007.03.059](https://doi.org/10.1016/j.amc.2007.03.059)
26. Yuan G, Wei Z and Wang Z. Gradient trust region algorithm with limited memory BFGS update for non-smooth convex minimization, *Computational Optimization and Application*, 54, 45–64 (2013) doi: [10.1007/s10589-012-9485-8](https://doi.org/10.1007/s10589-012-9485-8)
27. Yuan G, Lu X and Wei Z. BFGS trust-region method for symmetric nonlinear equations, *Journal of Computational and Applied Mathematics*, 230, 44–58 (2009) doi: [10.1016/j.cam.2008.10.062](https://doi.org/10.1016/j.cam.2008.10.062)
28. Qi L, Sun J. A trust region algorithm for minimization of locally Lipschitzian functions, *Mathematical Programming*, 66, 25–43 (1994) doi: [10.1007/BF01581136](https://doi.org/10.1007/BF01581136)
29. Bellavia S, Maccini M and Morini B. An affine scaling trust-region approach to bound-constrained nonlinear systems, *Applied Numerical Mathematics*, 44, 257–280 (2003) doi: [10.1016/S0168-9274\(02\)00170-8](https://doi.org/10.1016/S0168-9274(02)00170-8)
30. Akbari Z, Yousefpour R and Reza Peyghami M. A new nonsmooth trust region algorithm for locally Lipschitz unconstrained optimization problems, *Journal of Optimization Theory and Applications*, 164, 733–754 (2015) doi: [10.1007/s10957-014-0534-6](https://doi.org/10.1007/s10957-014-0534-6)
31. Bannert T. A trust region algorithm for nonsmooth optimization, *Mathematical Programming*, 67, 247–264 (1994) doi: [10.1007/BF01582223](https://doi.org/10.1007/BF01582223)
32. Amini K, Ahookhosh M. A hybrid of adjustable trust-region and nonmonotone algorithms for unconstrained optimization, *Applied Mathematical Modelling*, 38, 2601–2612 (2014) doi: [10.1016/j.apm.2013.10.062](https://doi.org/10.1016/j.apm.2013.10.062)
33. Zhou Q, Hang D. Nonmonotone adaptive trust region method with line search based on new diagonal updating, *Applied Numerical Mathematics*, 91, 75–88 (2015) doi: [10.1016/j.apnum.2014.12.009](https://doi.org/10.1016/j.apnum.2014.12.009)
34. Yuan G, Wei Z and Lu X. A BFGS trust-region method for nonlinear equations, *Computing*, 92, 317–333 (2011) doi: [10.1007/s00607-011-0146-z](https://doi.org/10.1007/s00607-011-0146-z)
35. Lu S, Wei Z and Li L. A trust region algorithm with adaptive cubic regularization methods for nonsmooth convex minimization, *Computational Optimization and Application*, 51, 551–573 (2012) doi: [10.1007/s10589-010-9363-1](https://doi.org/10.1007/s10589-010-9363-1)
36. Correa R, Lemaréchal C. Convergence of some algorithms for convex minimization, *Mathematical Programming*, 62, 261–273 (1993) doi: [10.1007/BF01585170](https://doi.org/10.1007/BF01585170)
37. Dennis J.E. Jr, Li S.B and Tapia R.A. A unified approach to global convergence of trust region methods for nonsmooth optimization, *Mathematical Programming*, 68, 319–346 (1995) doi: [10.1016/0025-5610\(94\)00054-W](https://doi.org/10.1016/0025-5610(94)00054-W)
38. Fukushima M, Qi L. A global and superlinearly convergent algorithm for nonsmooth convex minimization, *SIAM Journal on Optimization*, 6, 1106–1120 (1996) doi: [10.1137/S1052623494278839](https://doi.org/10.1137/S1052623494278839)

39. Wei Z, Li G and Qi L. New quasi-Newton methods for unconstrained optimization problems, *Applied Mathematics and Computation*, 175, 1156–1188 (2006) doi: [10.1016/j.amc.2005.08.027](https://doi.org/10.1016/j.amc.2005.08.027)
40. Broyden C.G. The convergence of a class of double rank minimization algorithms: the new algorithm, *Journal of the Institute of Mathematics and its Applications*, 6, 222–131 (1970) doi: [10.1093/imamat/6.3.222](https://doi.org/10.1093/imamat/6.3.222)
41. Fletcher R. A new approach to variable metric algorithms, *Computer Journal*, 13, 317–322 (1970) doi: [10.1093/comjnl/13.3.317](https://doi.org/10.1093/comjnl/13.3.317)
42. Goldfarb D. A family of variable metric methods derived by variational means, *Mathematics of Computation*, 24, 23–26 (1970) doi: [10.1090/S0025-5718-1970-0258249-6](https://doi.org/10.1090/S0025-5718-1970-0258249-6)
43. Shanno D.F. Conditioning of quasi-Newton methods for function minimization, *Mathematics of Computation*, 24, 647–650 (1970)
44. Dennis J.E, Moré J.J. A characterization of superlinear convergence and its application to quasi-Newton methods, *Mathematics of Computation*, 28, 549–560 (1974) doi: [10.1090/S0025-5718-1974-0343581-1](https://doi.org/10.1090/S0025-5718-1974-0343581-1)
45. Yuan Y, Sun W. *Optimization theory and methods*, Science Press, Beijing (1997)
46. Ma C. *Optimization method and the Matlab programming*, Science Press, Beijing (2010)
47. Mäkelä M.M, Neittaanmäki P. *Nonsmooth Optimization*. World Scientific, London (1992)
48. Charalambous J, Conn A.R. An efficient method to solve the minimax problem directly, *SIAM Journal on Numerical Analysis*, 15, 162–187 (1978) doi: [10.1137/0715011](https://doi.org/10.1137/0715011)
49. Demyanov V.F, Malozemov V.N. *Introduction to Minimax*. Wiley, New York (1974)
50. Womersley J. *Numerical methods for structured problems in nonsmooth optimization*. Ph.D. thesis. Mathematics Department, University of Dundee, Dundee, Scotland (1981)
51. Gupta N. *A higher than first order algorithm for nonsmooth constrained optimization*. P.h. thesis, Department of Philosophy, Washington State University, Pullman, WA (1985)
52. Short N.Z. *Minimization methods for nondifferentiable function*. Springer, Berlin (1985)
53. Kiwiel K.C. An ellipsoid trust region bundle method for nonsmooth convex minimization, *SIAM Journal on Control and Optimization*, 27, 737–757 (1989) doi: [10.1137/0327039](https://doi.org/10.1137/0327039)
54. Dolan E.D, Moré J.J. Benchmarking optimization software with performance profiles, *Mathematical Programming*, 91, 201–213 (2002) doi: [10.1007/s101070100263](https://doi.org/10.1007/s101070100263)