Scientific
Research
Publishing

A Spectral Projected Gradient-Newton Two Phase Method for Constrained Nonlinear Equations

Yuezhe Zhang

College of Science, University of Shanghai for Science and Technology, Shanghai, China
Email: yuezhezhang@126.com

Abstract

In this paper, we proposed a spectral gradient-Newton two phase method for constrained semismooth equations. In the first stage, we use the spectral projected gradient to obtain the global convergence of the algorithm, and then use the final point in the first stage as a new initial point to turn to a projected semismooth asymptotically newton method for fast convergence.

Keywords

Constrained Semismooth Equations, Spectral Projected Gradient Method, Newton Method, Two-Phase

1. Introduction

In this paper, we consider the constrained nonlinear semismooth equations problem: finding a vector $x_* \in \Omega$ such that

$$H(x) = 0,$$
$$x \in \Omega := \left\{ x \in \mathbb{R}^n \mid l \le x \le u \right\}, \tag{1}$$

where

$$\Omega := \left\{ x \in \mathbb{R}^n \mid l \le x \le u \right\}, \ l_i \in \mathbb{R} \bigcup \{-\infty\}, \ u_i \in \mathbb{R} \bigcup \{-\infty\}, \ l_i < u_i, \ i = 1, \cdots, n$$

$H : \mathbb{R}^n \to \mathbb{R}^n$ is a semismooth mapping. The notation of semismoothness was introduced for the functionals by Mifflin [1] and extended to vector functions by Qi and Sun [2].

Systems of constrained semismooth equations arise in various application, for instance complementarity problems, the box constrained variational inequality problems, the KKT system of variational inequlity problems and so on. The solu-

tion of nonlinear equations can be transformed into solving the following constrained optimization problem:

$$\min f(x) = \frac{1}{2}\|H(x)\|^2 \tag{2}$$
$$\text{s.t.} \quad x \in \Omega$$

where $f: R^n \to R$ is continuously differentiable and its gradient denoted by $\nabla f(x)$. Many researchers have studied constrained optimization problems such as (2) and given many effective algorithms. For example, a new class of adaptive non-monotone spectral gradient method is given in reference [3], an active set projected trust region algorithm in [4]. The methods of optimization problems involve the first-order methods and the second-order methods. Classical first-order algorithms include gradient method, sub-gradient method, conjugate gradient method, etc. The main advantage of first-order method is its small storage, which is particularly suitable for large-scale problems. However, the disadvantage of first-order method is that its convergence speed is at most linear, and it can not meet the requirements of high precision. For the second-order method, it has the advantage of fast convergence speed. Under certain conditions, it can achieve superlinear convergence or even quadratic convergence. But its disadvantage is that it needs a good initial point, sometimes it even needs the initial point to approach the local optimal point.

Motivated by this, in this paper, we combine the advantages of the first-order method with those of the second-order method. We will consider the two-stage combination algorithm to solve the optimization problem. First, we use the first-order method to obtain the global convergence of the algorithm, and then use the final point obtained by the first-order method as the new initial point to turn to the second-order method to obtain the fast convergence speed. At the same time, we use projection technology to solve the constrained conditions.

2. Preliminaries

In this section, we present some definitions and theorems that are useful to our main result.

Suppose $H: R^n \to R^n$ is a locally Lipschitzian function, according to Rademacher theorem, H is differentiable almost everywhere. Denote the set of points at which H is differentiable by D_H. We write $H'(x_k)$ for the usual $n \times m$ Jacobian matrix of partial derivatives whenever x is a point at which the necessary partial derivatives exists. Let $\partial H(x)$ be the generalized Jacobian defined by Clarke in [5]. Then

$$\partial H(x) = C_0\left(\partial_B H(x)\right), \tag{3}$$

where the C_0 denotes the convex hull of a set, $\partial_B H(x) = \left\{ \lim_{\substack{x_j \to x \\ x_j \in D_H}} H'(x_j) \right\}$.

Definition 2.1 [2] Suppose $H: R^n \to R^n$ is a locally Lipschitzian function, we say that H is semismooth at x if

$$\lim_{\substack{V \in \partial H(x+th') \\ h' \to h, t \downarrow 0}} \{Vh'\} \tag{4}$$

exists for any $h \in R^n$.

Lemma 2.2 [2]: Suppose $H : R^n \to R^n$ is a locally Lipschitzian function, the following statements are equivalent:

1) H is semismooth at x;

2) For any $V \in \partial H(x+h), h \to 0$,

$$Vh - H'(x;h) = o(\|h\|), \tag{5}$$

$$H(x+h) - H(x) - Vh = o(\|h\|). \tag{6}$$

Lemma 2.3 [2]: Suppose $H : R^n \to R^n$ is a locally Lipschitzian function, H is semismooth at x if each component of H is semismooth at x.

Definition 2.4 [2]: Suppose $H : R^n \to R^n$ is a locally Lipschitzian function, If for any $V \in \partial H(x+h), h \to 0$

$$Vh - H'(x;h) = O(\|h\|^{1+p}), \tag{7}$$

where $0 < p \le 1$, then we call H is p-order semismooth at x.

Lemma 2.5 [2]: suppose $H : R^n \to R^n$ is a locally Lipschitzian function, we say H is strongly BD-regular at x if all $V \in \partial_B H(x)$ are nonsingular.

Lemma 2.6 [6]: Suppose that $H : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous and H is BD-regular at $x \in \mathbb{R}^n$. Then there exist a neighborhood $\mathbb{N}(x)$ of x and a constant K such that for any $y \in \mathbb{N}(x)$ and $V \in \partial_B H(y)$, V is nonsingular and $\|V^{-1}\| \le K$.

Lemma 2.7 [6]: Suppose that $H : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous and H is BD-regular at a solution x_* of $H(x) = 0$. If H is semismooth at x_*, then there exist a neighborhood $\mathbb{N}(x_*)$ of x_* and a constant $k > 0$ such that for any $x \in \mathbb{N}(x_*)$

$$\|H(x)\| \ge k\|x - x_*\|. \tag{8}$$

Lemma 2.8 [7]: The projection operator $\Pi_X(\cdot)$ satisfies.

1) For any $x \in X$, $[\Pi_X(z) - z]^T [\Pi_X(z) - x] \le 0$ for all $z \in \mathbb{R}^n$.

2) $\|\Pi_X(y) - \Pi_X(z)\| \le \|y - z\|$ for all $y, z \in \mathbb{R}^n$.

Lemma 2.9 [8]: Given $x \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$, the function ξ defined by

$$\xi(\lambda) = \|\Pi_X(x + \lambda d) - x\|/\lambda, \lambda \ge 0 \tag{9}$$

is nonincreasing.

Lemma 2.9 actually implies that if $x \in X$ is a stationary point of (2), then

$$\bar{d}_G(\lambda) = \Pi_X[x + \lambda d_G] - x = 0, \ \forall \lambda \ge 0 \tag{10}$$

3. Algorithm

In order to obtain the global convergence of the algorithm, in the first stage, we adopt the non-monotone spectral projection gradient method of the first-order

method. The one-dimensional search procedure of Algorithm 3.1 will be called SPG1 from now on and Algorithm 3.2 will be called SPG2 in the rest of the paper.

Given $z \in \mathbb{R}^n$, we define $P(z)$ as the orthogonal projection on Ω, denote $g(x) = \nabla f(x)$. $x_0 \in \Omega$, integer $M \geq 1$, a small parameter $\alpha_{\min} > 0$, a large parameter $\alpha_{\max} > \alpha_{\min}$, sufficient decrease parameter $\gamma \in (0,1)$, $0 < \sigma_1 < \sigma_2 < 1$, initially $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$, $x_0 \in \Omega$.

Algorithm 3.1 [9] (SPG1)

Step 1. If $\|P(x_k) - g(x_k) - x_k\| < \varepsilon_1$, stop, input x_k.

Step 2. (Backtracking)

Step 2.1 Set $\lambda = \alpha_k$.

Step 2.2 Set $x_+ = P(x_k - \lambda g(x_k))$.

Step 2.3 If

$$f(x_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \gamma \langle x_+ - x_k, g(x_k) \rangle, \tag{11}$$

Then define $\lambda_k = \lambda$, $x_{k+1} = x_+$, $s_k = x_{k+1} - x_k$, $y_k = g(x_{k+1}) - g(x_k)$, and go to step 3.

If (11) does not hold, define $\lambda_{new} \in [\sigma_1 \lambda, \sigma_2 \lambda]$. Set $\lambda = \lambda_{new}$, and go to step 2.2.

Step 3. compute $b_k = \langle s_k, y_k \rangle$, If $b_k \leq 0$, set $\alpha_{k+1} = \alpha_{\max}$; else compute $\alpha_k = \langle s_k, s_k \rangle$, $\alpha_{k+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, a_k/b_k\}\}$, and go to step 1.

Algorithm 3.2 [9] (SPG2)

Step 2. (Backtracking)

Step 2.1. Compute $d_k = P(x_k - \alpha_k g(x_k)) - x_k$, Set $\lambda = 1$.

Step 2.2. Set $x_+ = x_k + \lambda d_k$.

Step 2.3. If

$$f(x_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \gamma \lambda \langle d_k, g(x_k) \rangle, \tag{12}$$

Then define $\lambda_k = \lambda, x_{k+1} = x_+, s_k = x_{k+1} - x_k, y_k = g(x_{k+1}) - g(x_k)$, and go to step 3.

If (12) does not hold, define $\lambda_{new} \in [\sigma_1 \lambda, \sigma_2 \lambda]$. Set $\lambda = \lambda_{new}$, and go to step 2.2.

The output point of the first stage is used as the initial point of the next stage.

Algorithm 3.3 [10] (A Projected semismooth asymptotical newton method)

Step 0. Choose constants $\rho, \sigma, \eta \in (0,1), p_1 > 0, p_2 > 2$, Let $x_0 = x_N \in \Omega$, $k := 0$.

Step 1. Choose $V_k \in \partial_B H(x_k)$, compute $\nabla f(x_k) = V_k^T H(x_k)$.

Step 2. If x_k is a stationary point, stop. Otherwise let

$$d_G^k = -\gamma_k \nabla f(x_k), \tag{13}$$

where

$$\gamma_k = \min\left\{1, \eta f(x_k) / \|\nabla f(x_k)\|^2\right\}, \tag{14}$$

and go to step 3.

Step 3. If the linear system

$$H(x_k) + V_k d = 0 \tag{15}$$

has a solution d_N^k, and

$$-\nabla f(x_k)^{\mathrm{T}} d_N^k \geq p_1 \left\| d_N^k \right\|^{p_2}, \tag{16}$$

then use the direction d_N^k. Otherwise, set $d_N^k = d_G^k$.

Step 4. Let m_k be the smallest nonnegative integer m satisfying

$$f\left(x_k + \overline{d}^k(\rho^m)\right) \leq f(x_k) + \sigma \nabla f(x_k)^{\mathrm{T}} \overline{d}_G^k(\rho^m), \tag{17}$$

where for any $\lambda \in [0,1]$,

$$\overline{d}^k(\lambda) = t_k^*(\lambda) \overline{d}_G^k(\lambda) + \left[1 - t_k^*(\lambda)\right] \overline{d}_N^k(\lambda), \tag{18}$$

$$\overline{d}_G^k(\lambda) = \prod_X \left[x + \lambda d_G^k\right] - x_k, \quad \overline{d}_N^k(\lambda) = \prod_X \left[x + \lambda d_N^k\right] - x_k \tag{19}$$

and $t_k^*(\lambda)$ is an optimal solution to

$$\min_{t \in [0,1]} \frac{1}{2} \left\| H(x_k) + V_k \left[t \overline{d}_G^k(\lambda) + (1-t) \overline{d}_N^k(\lambda)\right] \right\|^2, \tag{20}$$

the optimal solution is

$$t^*(\lambda) = \max\left\{0, \min\left\{1, t(\lambda)\right\}\right\}, \tag{21}$$

let $\lambda_k = \rho^{m_k}$, $x_{k+1} = x_k + \overline{d}^k(\lambda_k)$.

Step 5. Let $k =: k+1$, and go to step 1.

4. Convergence Analysis

Theorem 4.1 [9]: Algorithm SPG1 is well defined, and any accumulation point of the sequence $\{x_k\}$ that is generates is a constrained stationary point.

Theorem 4.2 [9]: Algorithm SPG2 is well defined, and any accumulation point of, and any accumulation.

Theorem 4.3 [10]: Let $\{x_k\} \subset X$ be a sequence generated by Algorithm 3.3, then any accumulation point of $\{x_k\}$ is a station point of (2).

5. Application

Many practical problems can be solved by transforming them into constrained semi-smooth equations. For example, mixed complement problem (MCP):

$F : R^n \rightarrow R^n$ is a continuous differentiable function, finding a vectors $x \in X$ satisfies

$$F(x)^{\mathrm{T}}(y - x) \geq 0, \quad \forall y \in X, \tag{22}$$

The function $\psi_\alpha : \mathbb{R}^2 \rightarrow \mathbb{R}$ with $\alpha \in [0,1]$ is defined by

$$\psi_\alpha(a,b) := \left(\left[\phi_\alpha(a,b)_+\right]\right)^2 + \left(\left[-a\right]_+\right)^2, \tag{23}$$

where $[a]_+ := \max\{0, a\}$ for any $a \in \mathbb{R}$ and $\phi_\alpha : \mathbb{R}^2 \rightarrow \mathbb{R}$ is the penalized

Fischer-Burmeister function introduced by Chen *et al.* [11] and has the form:

$$\phi_\alpha(a,b) \coloneqq \alpha\phi_{FB}(a,b) + (1-\alpha)a_+b_+, \quad (24)$$

Here, $\phi_\alpha : \mathbb{R}^2 \to \mathbb{R}$ is an NCP function, which is given by

$$\phi_{FB}(a,b) \coloneqq (a+b) - \sqrt{a^2+b^2}, \quad (25)$$

The mixed complement problem can be transformed into a semi-smooth system of equations by the above functions.

Let $N = \{1, \cdots, n\}$

$$I_f \coloneqq \{i \,|\, l_i = -\infty, u_i = \infty, i \in N\}, \quad I_l \coloneqq \{i \,|\, l_i > -\infty, u_i = \infty, i \in N\},$$

$$I_u \coloneqq \{i \,|\, l_i = -\infty, u_i < \infty, i \in N\}, \quad I_{lu} \coloneqq N \setminus \{I_l \bigcup I_u \bigcup I_f\} \quad (26)$$

MCP can be reformulated as $H(x) = 0$ with

$$H_i(x) \coloneqq \begin{cases} |F_i(x)| & \text{if } i \in I_f \\ |\phi_\alpha(x_i - l_i, F_i(x))| & \text{if } i \in I_l \\ |\phi_\alpha(u_i - x_i, -F_i(x))| & \text{if } i \in I_u \\ \sqrt{\psi_\alpha(x_i - l_i, F_i(x)) + \psi_\alpha(x_i - l_i, F_i(x))} & \text{if } i \in I_{lu} \end{cases}, i = 1, \cdots, n \quad (27)$$

Then we can use the two phase method to solve this problem.

6. Conclusion

In this paper, we proposed a two-phase method for the constrained equations. We can also combine other first-order and second-order methods. In this paper, the iteration complexity analysis of the first-order method is a meaningful work, and we will do further research.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Mifflin, R. (2006) Semismooth and Semiconvex Functions in Constrained Optimization. *SIAM Journal on Control & Optimization*, **15**, 959-972. https://doi.org/10.1137/0315061

[2] Qi, L. and Sun, J. (1993) A Nonsmooth Version of Newton's Method. *Mathematical Programming*, **58**, 353-368. https://doi.org/10.1007/BF01581275

[3] Ji, L. and Yu, Z.S. (2009) New Class of Adaptive Nonmonotone Spectral Projected Gradient Method. *Journal of University of Shanghai for Science & Technology*.

[4] Qi, L.Q., Tong, X.J. and Li, D.H. (2004) Active-Set Projected Trust-Region Algorithm for Box-Constrained Nonsmooth Equations. *Journal of Optimization Theory and Applications*, **120**, 601-625. https://doi.org/10.1023/B:JOTA.0000025712.43243.eb

[5] Clarke, F.H. (1983) Optimization and Nonsmooth Analysis. John Wiley & Sons, New York.

[6] Qi, L. (1993) Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations. *Mathematics of Operations Research*, **18**, 227-244. https://doi.org/10.1287/moor.18.1.227

[7] Zarantonello, E.H. (1971) Projections on Convex Sets in Hilbert Space and Spectral Theory: Part I. Projections on Convex Sets: Part II. Spectral Theory. *Revista de la Unión Matemática Argentina*, **26**, 237-424.

[8] Powell, M.J.D. (1983) Variable Metric Methods for Constrained Optimization. *Mathematical Programming: The State of the Art*, Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-68874-4_12

[9] Birgin, E.G., Martínez, J.M. and Raydan, M. (2000) Nonmonotone Spectral Projected Gradient Methods on Convex Sets. *Society for Industrial and Applied Mathematic*, **10**, 1196-1211. https://doi.org/10.1137/S1052623497330963

[10] Sun, D., Womersley, R.S. and Qi, H. (2002) A Feasible Semismooth Asymptotically Newton Method for Mixed Complementarity Problems. *Mathematical Programming*, **94**, 167-187. https://doi.org/10.1007/s10107-002-0305-2

[11] Chen, B., Chen, X. and Kanzow, C. (2000) A Penalized Fischer-Burmeister NCP-Function. *Mathematical Programming*, **88**, 211-216. https://doi.org/10.1007/PL00011375