A NEW ALGORITHM AND ITS GLOBAL CONVERGENCE FOR CONVEX COMPOSITE OPTIMIZATION PROBLEMS

Meng Zhang, Wenling Zhao* and Ximin Guo
School of Mathematics and Statistics, Shandong University of Technology. Zibo, China, 255000
E-mail: zm1924330152@163.com, *zwlsdj@163.com, 15163325081@163.com

ABSTRACT
In this paper, we present a new algorithm to solve the convex composite optimization problems. The main idea of this algorithm is to find the optimal solution by combination of Newton method and projection technique. And we prove the algorithm own good global convergence.

Keywords: *convex optimization problem; Newton algorithm; projection algorithm; global convergence*.

1. Introduction

Convex composite optimization problems have widely applied in many fields[1]. Many optimization problems, such as location problems, transportation problems and economic problems, their objective function are mostly convex composite functions, such as maximin optimization problem, constrained optimization problems where the objective function is a composite of convex function and linear operator can all be regarded as a special case of convex composite optimization problems[2]. So the convex composite model provides a unified framework for many optimization problems.

How to solve the convex composite optimization problems, Burke and Ferris[2] proposed the Gauss-Newton method, and made important progress in convergence studies, the method didn't require that the solution set of the problem be a single point set, it even allow the solution set to be unbounded. Li, Chong et al[3] in the case where constraint set is a set of weak sharp minimal, they extend the results of Burke et al, and established the local quadratic convergence of Gauss-Newton method. Solodov et al[4] certified that the convex composite optimization problems have good local convergence of Gauss-Newton method. Patrinos et al[5] proposed two quasi-Newton methods for solving convex composite optimization problem. Huang et al[6] proposed a composite splitting algorithm of convex composite optimization problem. Aybat[7] proposed divergent approximation method of convex composite optimization problem, etc. Although there are many methods to solve convex composite optimization problem. But up to now, we haven't seen an algorithm combining Newton Method and projection method to solve the convex composite optimization problem.

The idea of Newton's method [8, 9, 10] is to approximate non-linear equations with linear equations, it has at least second-order local convergence and fast convergence speed. Especially when the current iteration point is close enough to the exact solution, the convergence speed is faster. Projection method [11, 12] is one of the basic methods to solve the optimization problem. The method is simple and feasible, the algorithm is stable, the computation is small, and the convergence is strong. So, projection method has become an important tool for solving approximate solutions of nonlinear programming and variational inequalities. Its main idea is to use the concept of projection to establish the equivalence between related problems and fixed point problems.

In this paper, we present a new algorithm based on the idea of Newton method and projection method, and we have proved that the algorithm has good convergence. We consider the convex composite optimization

$$min f(x) x \in C, (1.1)$$

where $f(x) = h(F(x))$ is convex function, C is a nonempty closed convex subset of R^n. By reference[13], we know when $h: R^m \to R^n$ is a monotone non-decreasing continuous differentiable convex functions, $F: R^n \to R^m$ is continuous differentiable convex function.

If $x \in C$ is a stability point of (1.1), that is

$$\langle \nabla f(x), y - x \rangle \geq 0, \forall y \in C.$$

where $\langle \cdot,\cdot \rangle$ represents the inner product, $\nabla f(x)$ is the gradient of $f(x)$.

Suppose $x \in R^n$, the projection of x on a closed set C is

$$P_C(x) = argmin_{y \in C} \parallel y - x \parallel.$$

The distance between x and C is given by

$$dist(x, C) = \inf_{y \in C} \parallel y - x \parallel.$$

If C is closed set, we have

$$dist(x, C) = \parallel P_C(x) - x \parallel.$$

This paper has three sections. In section 2, we introduce a new algorithm, and in section 3, we prove the global convergence of the algorithm.

2. Algorithm

In this section, we give a new algorithm to solve the convex composite optimization.

For the iteration point $x^k \in C$, we select a positive semidefinite matrix G_k and a regularization parameter $\mu_k > 0$, let

$$\phi_k(z) = \nabla h(F(x^k))\nabla F(x^k) + (G_k + \mu_k I)(z - x^k).$$

Considering the linearized subproblem: find $\hat{z}^k \in C$, such that $\forall z \in C$ we have

$$\langle \phi_k(\hat{z}^k), z - \hat{z}^k \rangle \geq 0. (2.1)$$

As $G_k + \mu_k I$ is positive definite, so (2.1) has a unique solution $\hat{z}^k$, at the same time it satisfy the following formula[13],

$$\hat{z}^k - P_C(\hat{z}^k - \phi_k(\hat{z}^k)) = 0.$$

So, for $\forall \varepsilon_k > 0$, (2.1) always has a inexact solution z^k such that

$$||e^k|| \leq \varepsilon_k.$$

where

$$e^k = z^k - P_C(z^k - \phi_k(z^k)).$$

Algorithm

Step 0 $\forall x^0 \in C$, choose parameters $\eta, \beta, \alpha \in (0,1)$, let $k := 0$.

Step 1 Select a positive semidefinite matrix G_k, regularization parameter $\mu_k > 0$, $\rho_k \in (0,1)$, Compute a inexact solution z^k of the subproblem (2.1), such that

$$||e^k|| \leq \rho_k \mu_k \parallel |z^k - x^k||. (2.2)$$

$$\langle e^k, \phi_k(z^k) + z^k - x^k \rangle \leq \rho_k \mu_k ||z^k - x^k||^2. (2.3)$$

If $z^k = x^k$, stop

Step 2 Take

$$y_c^k = P_C(z^k - \phi_k(z^k))$$

$$v^k = \nabla h(F(y_c^k))\nabla F(y_c^k) - \phi_k(z^k) + e^k$$

If

$$\langle v^k, x^k - y_c^k \rangle \geq \eta \mu_k ||y_c^k - x^k||^2. (2.4)$$

let

$$y^k = y_c^k$$

$$v^k = \nabla h(F(y^k))\nabla F(y^k) - \phi_k(z^k) + e^k$$

go to step 4.

Step 3 Take $y^k = x^k + \lambda_k(z^k - x^k)$, $\lambda_k = \beta^{m_k}$, m_k is the smallest nonnegative integer such that

$$\parallel |\nabla h(F(y^k))\nabla F(y^k) - \nabla h(F(x^k))\nabla F(x^k)|| \leq \alpha(1 - \rho_k)\mu_k ||z^k - x^k||.$$

Take $v^k = \nabla h(F(y^k))\nabla F(y^k)$.

Step 4 Compute:

$$\hat{x}^k = x^k - \frac{\langle v^k, x^k - y^k \rangle}{||v^k||^2} v^k.$$

$$x^{k+1} = P_C(\hat{x}^k).$$

Let $k := k + 1$, go to Step1.

Next, we prove the global convergence of the algorithm.

3. Global convergence

We give the following assumption.

(A) Existence constant $M > 0, \delta \in (0,1)$ such as

$$||G_k|| \leq M. ||r(x^k)|| \leq \mu_k \leq M, \rho_k \leq \min(1, \tfrac{1}{M})\delta.$$

Let

$$K_N = \{k | y^k = P_C(z^k - \phi_k(z^k))\},$$
$$K_A = \{k | y^k = x^k + \lambda_k(z^k - x^k)\}.$$

Lemma 3.1 The following inequality holds

$$\langle \nabla h(F(x^k))\nabla F(x^k), x^k - z^k \rangle \geq (1-\rho_k)\mu_k ||z^k - x^k||^2. (3.1)$$

Proof From the properties of projection, we have

$$\langle y_c^k - z^k + \phi_k(z^k), x^k - y_c^k \rangle \geq 0.$$

$$\begin{aligned}
& \langle \nabla h(F(x^k))\nabla F(x^k), x^k - z^k \rangle \\
\geq\ & \langle \nabla h(F(x))\nabla F(x) - \phi_k(z^k) + e^k, x^k - z^k + e^k \rangle \\
& -\langle \nabla h(F(x))\nabla F(x), e^k \rangle \\
\geq\ & \langle -(G_k + \mu_k I)(z^k - x^k), x^k - z^k + e^k \rangle \\
& +\langle e^k, (G_k + \mu_k I)(z^k - x^k) \rangle \\
& +\langle e^k, x^k - z^k - \nabla h(F(x^k))\nabla F(x^k) \rangle \\
\geq\ & \langle (G_k + \mu_k I)(z^k - x^k), z^k - x^k \rangle - (||\nabla h(F(x^k)\nabla F(x^k)|| \\
& +(||G_k||) + \mu_k + 1)\rho_k\mu_k \,\|\, |z^k - x^k||^2 \\
\geq\ & \langle (G_k + (1-\rho_k)\mu_k I(z^k - x^k), z^k - x^k \rangle \\
\geq\ & (1-\rho_k)\mu_k \,\|\, |z^k - x^k||^2.
\end{aligned}$$

Lemma 3.2 If $\underset{x\in C}{\operatorname{argmin}} f(x) \neq \emptyset$, for $\forall x^* \in \underset{x\in C}{\operatorname{argmin}} f(x)$, we have

$$\frac{\langle v^k, x^k - y^k \rangle^2}{||v^k||^2} \leq ||x^k - x^*||^2 - ||x^{k+1} - x^*||^2 \forall k. (3.3)$$

Proof For $\forall \bar{x} \in C$, we have

$$\begin{aligned}
& ||\hat{x}^k - \bar{x}||^2 = ||x^k - \bar{x}||^2 + \frac{\langle v^k, x^k - y^k \rangle^2}{||v^k||^2} \\
& -2\frac{\langle v^k, x^k - y^k \rangle^2}{||v^k||^2}\langle v^k, x^k - \bar{x} \rangle \\
=\ & \frac{||\hat{x}^k - \bar{x}||^2}{||v^k||}(\langle v^k - x^k - y^k - 2\bar{x} - 2y^k \rangle) \\
=\ & \|\, |x^k - \bar{x}||^2 - \frac{\langle v^k, x^k - y^k \rangle^2}{||v^k||^2} \\
& +2\frac{\langle v^k, x^k - y^k \rangle^2}{||v^k||^2}\langle v^k, \bar{x} - y^k \rangle.
\end{aligned}$$

We let $\bar{x} = x^*$ in formula We have

$$Q_k = \frac{\langle v^k, x^k - y^k \rangle^2}{||v^k||^2}\langle v^k, x^* - y^k \rangle. (3.4)$$

For $x^* \in \arg min_{x\in C} f(x)$, using $y^k \in C$, we obtain

$$\langle \nabla h(F(x^*)\nabla F(x^*), y^k - x^* \rangle \geq 0.$$

From the hypothesis of h and F, we have $\nabla h(F(x))\nabla F(x)$ is monotonic.

$$\langle \nabla h(F(y^k))\nabla F(y^k), y^k - x^* \rangle \geq 0. (3.5)$$

Next we have two situations are discussed in (3.4).

(1) If $k \in K_N$, from (2.1), we have

$$\langle v^k, x^k - y^k \rangle \geq 0. (3.6)$$

From the definition of y^k, and the properties of projection, we have

$$\langle \phi_k(z^k) - e^k, x^* - y^k \rangle \geq 0. (3.7)$$

By the definition of v^k and(3.7), we have $Q_k \leq 0$.

(2) If $k \in K_A$, from (3.5), we obtain $Q_k \leq 0$,

So, $\forall k$, $Q_k \leq 0$, we have

$$||x^{k+1} - x^*||^2 \leq ||\hat{x}^k - x^*||^2 \leq ||x^k - x^*||^2 - \frac{\langle v^k, x^k - y^k \rangle^2}{||v^k||^2}.$$

So the conclusion is hold.

Lemma 3.3 Suppose that $argmin_{x\in C} f(x) \neq \emptyset$, and assume that (A) holds, there is a constant $m > 0$, for $\forall k$, we have

$$\mu_k \geq m. (3.8)$$

Then

$$\lim_{k\to\infty} ||z^k - x^k|| = 0, \lim_{k\to\infty} ||y^k - x^k|| = 0.$$

Proof According to Lemma 3.2, we have sequence $||\hat{x}^k - x^*||$ is monotonicity£¬so $\{x^k\}$ is bounded, and

$$\lim_{k\to\infty} \frac{\langle v^k, x^k - y^k\rangle}{||v^k||} = 0. (3.9)$$

From (3.1),(3.9), and assumption (A) holds, we have

$$||\nabla h(F(x^k))\nabla F(x^k)|| \geq m(1-\delta)||z^k - x^k||. (3.10)$$

from $\nabla h(F(x^k))\nabla F(x^k)$ is continuous and $\{x^k\}$ is bounded, we have $\{z^k\}$ is bounded by (3.10), using (3.9) and assumption (A), we obtain $\{y^k\}, \{v^k\}$ is bounded, by(3.9), we obtain

$$\lim_{k\to\infty} \langle v^k, x^k - y^k\rangle = 0. (3.11)$$

(1) If $k \subseteq K_N$, according to (3.1) and (3.8), we get

$$\langle v^k, x^k - y^k\rangle \geq m\eta||y^k - x^k||^2. (3.12)$$

Take the limit on both sides for (3.12)

$$\lim_{k\in K_N, k\to\infty} ||x^k - y^k|| = 0. (3.13)$$

By (2.2) we have

$$||z^k - x^k|| \leq \frac{1}{1-\delta}||y^k - x^k||.$$

We obtain

$$\lim_{k\in K_N, k\to\infty} ||z^k - x^k|| = 0.$$

(2) If $k \in K_A$, from the definition of v^k, (3.1), and Step3, we have

$$\langle v^k, x^k - y^k\rangle \geq \lambda_k(1-\alpha)(1-\rho_k)\mu_k||z^k - x^k||^2. (3.14)$$

By (3.11),(3.14), we have

$$\lim_{k\in K_A, k\to\infty} \lambda_k||z^k - x^k||^2 = 0. (3.15)$$

Suppose that

$$\lim_{k\in K_A, k\to\infty} ||z^k - x^k|| > 0. (3.16)$$

Because $\{x^k\}$,$\{z^k\}$,$\{\rho_k\}$,$\{\mu_k\}$are bounded, suppose

$$\lim_{k\in K_A, k\to\infty} z^k = \bar{z}, \lim_{k\in K_A, k\to\infty} x^k = \bar{x},$$

$$\lim_{k\in K_A, k\to\infty} \rho_k = \bar{\rho}, \lim_{k\in K_A, k\to\infty} \mu_k = \bar{\mu}.$$

From (3.15), we have $\bar{z} \neq \bar{x}$, according to (3.14),(3.15), we obtain

$$\lim_{k\in K_A, k\to\infty} \lambda_k = 0. (3.17)$$

From Step3, we know, when $k \in K_A, k \to \infty$, we have

$$\langle \nabla f(x^k + \beta^{-1}\lambda_k(z^k - x^k)), x^k - z^k\rangle < \alpha(1-\rho_k)\mu_k||z^k - x^k||^2.$$

From the above formula and (3.18) we get

$$\langle \nabla h(F(\bar{x}))\nabla F(\bar{x}), \bar{x} - \bar{z}\rangle \leq \alpha(1-\bar{\rho})\bar{\mu}||\bar{z} - \bar{x}||^2. (3.18)$$

For (3.1), passing onto the limit as $k \to \infty$, $k \subseteq K_A$, let

$$\langle \nabla h(F(\bar{x}))\nabla F(\bar{x}), \bar{x} - \bar{z}\rangle \geq (1-\bar{\rho})\bar{\mu}||\bar{z} - \bar{x}||^2. (3.19)$$

Because $\alpha \in (0,1)$, so (3.17) and (3.18) are contradicted. So we get $\lambda_k \neq 0$, that is

$$\lim_{k\in K_A, k\to\infty} ||z^k - x^k|| = 0.$$

And from Step3, we have

$$\lim_{k\in K_A, k\to\infty} ||y^k - x^k|| = 0.$$

So the conclusions are hold.

Next we proves the convergence of the algorithm.

Let the natural residual is

$$r(x) = x - P_C(x - \nabla h(F(x))\nabla F(x)).$$

Theorem3.1 If (A) holds, we have

(1) If $argmin_{x\in C} f(x) \neq \emptyset$, $\{x^k\}$ is bounded, and

$$\lim_{k\to\infty} x^k = x^* \in \operatorname{argmin}_{x\in C} f(x). (3.20)$$

(2) If $\{x^k\}$ is bounded, we have $\operatorname{argmin}_{x\in C} f(x) \neq \emptyset$ and (3.20) holds.

Proof (1) If $argmin_{x\in C} f(x) \neq \emptyset$, we have $||\hat{x}^k - x^*||$ is monotonic, by Lemma 3.2, and $\{x^k\}$ is bounded, to prove the (3.20) holds, we need to prove $\lim_{k\to\infty}||r(x^k)||=0$.

$$\begin{aligned} ||x^k-z^k|| &= ||x^k-P_C(z^k-\phi_k(z^k))||-||e^k|| \\ &\geq ||x^k-P_C(x^k-\nabla h(F(x^k))\nabla F(x^k)||-||e^k|| \\ &\quad -||P_C(x^k-\nabla h(F(x^k))\nabla F(x^k)-P_C(z^k-\phi_k(z^k))|| \\ &\geq ||r(x)||-||x^k-z^k-\nabla h(F(x^k))\nabla F(x^k)+\phi_k(z^k)||-||e^k|| \\ &\geq ||r(x)||-(1+\rho_k\mu_k)||x^k-z^k||-||(G_k+\mu_k I)(x^k-z^k)|| \\ &\geq ||r(x)||+(1+G_k+2\mu_k)||x^k-z^k||. \end{aligned}$$

So we get

$$||x^k-z^k|| \geq ||r(x^k)||.$$

By Lemma 3.3 we have

$$\lim_{k\to\infty}||r(x^k)||=0.$$

So the conclusion (1) is hold.

(2) Suppose that $\{x^k\}$ is bounded, we only need to prove

$$\lim_{k\to\infty}||r(x^k)||>0. \quad (3.21)$$

We have $argmin_{x\in C}f(x)\neq\emptyset$. Suppose that (3.21) holds, according to (A) we know there exists $m>0$ such as (3.8). From Lemma 3.1 we get $\{x^k\},\{z^k\},\{y^k\}$ is bounded. We select a constant M_0 , for $\forall k$, such that

$$\max\{||x^k||,||z^k||,||y^k||\}\leq M_0.$$

We arbitrarily take an $\bar{x}\in C$, Éè

$$Y=\{x\in R^n|||x||\leq M_0+||\bar{x}||\}\cap C.$$

So $Y\subseteq C$ is nonempty and compact set, $\underset{x\in C}{\operatorname{argmin}} f(x)\neq\emptyset$. From conclusion (1) we have (3.20) holds.

In this paper, we introduce a new algorithm of convex composite optimization problems, and have proved that the algorithm has global convergence.

Acknowledgement

This research was supported by National Natural Science Foundations of China (11771255) and Shandong Province Natural Science Foundation (ZR2016AM07).

References

[1]. Xiao X , Li Y , Wen Z , et al. A Regularized Semi-Smooth Newton Method With Projection Steps for Composite Convex Programs[J]. Journal of Scientific Computing, 2016:1-26.
[2]. Burke, J.V. Ferris, M.C. (1995): A Gauss-Newton method for convex composite optimization. Math.Program. 71, 179¨C194
[3]. Li,Chong, and X.Wang. On convergence of the Gauss-Newton method for convex composite optimization. Mathematical Programming 91.2(2002):349-356.
[4]. Solodov M V , Svaiter B F . A Truly Globally Convergent Newton-Type Method for the Monotone Nonlinear Complementarity Problem[M]. Society for Industrial and Applied Mathematics, 1999.
[5]. Patrinos P , Stella L , Bemporad A . Forward-backward truncated Newton methods for convex composite optimization[J]. Mathematics, 2014, 27(3):1.
[6]. Huang J , Zhang S , Li H , et al. Composite splitting algorithms for convex optimization[J]. Computer Vision Image Understanding, 2011, 115(12):1610-1622.
[7]. Aybat N S , Iyengar G , Wang Z . A Distributed Proximal Method for Composite Convex Optimization[J]. Eprint Arxiv, 2014.
[8]. P. Marcotte and J.-P. Dussault, A note on a globally convergent Newton method for solving monotone variational inequalities, Oper. Res. Lett., 6 (1987), pp. 35¨C42.
[9]. Qi L , Sun J . A nonsmooth version of Newton's method[J]. Mathematical Programming, 1993, 58(1-3):353-367.
[10]. K. Taji, M. Fukushima, and T. Ibaraki, A globally convergent Newton method for solving strongly monotone variational inequalities, Math. Programming, 58 (1993), pp. 369¨C383.
[11]. M. V. Solodov and B. F. Svaiter, A new projection method for variational inequality problems, SIAM J. Control Optim., 37 (1999), pp. 765¨C776.
[12]. M. Solodov and B. Svaiter, A hybrid projection¨Cproximal point algorithm, J. Convex Anal.,6 (1999), pp. 59¨C70.
[13]. (Japan)Masao Fukushima. Translated by Guihua Lin. Nonlinear optimization basis[M]. BeiJing£ºScience Press, 2011.05.