A Modified Spectral Gradient Projection Method for Solving Non-Linear Monotone Equations With Convex Constraints and Its Application

LI ZHENG, LEI YANG*, (Member, IEEE), AND YONG LIANG

Faculty of Information Technology, Macau University of Science and Technology, Taipa, Macau

Corresponding author: Lei Yang (leiyang@must.edu.mo)

This work was supported in part by the Science and Technology Development Fund, Macau, under Grant 101/2016/A3 and Grant 0070/2019/A2, and in part by the National Natural Science Foundation of China under Grant 11701598.

ABSTRACT In this paper, we propose a derivative free algorithm for solving non-linear monotone equations with convex constraints. The proposed algorithm combines the method of spectral gradient and the projection method. We also modify the backtracking line search technique. The global convergence of the proposed method is guaranteed, under the mild conditions. Further, the numerical experiments show that the large-scale non-linear equations with convex constraints can be effectively solved with our method. The $L_1$-norm regularized problems in signal reconstruction are studied by using our method.

INDEX TERMS Non-linear equations, non-smooth equation, derivative free method, spectral gradient method, projection method, signal reconstruction.

I. INTRODUCTION

In this paper, we focus on finding the solution of the following non-linear systematic equations

$$F(x) = 0, \quad x \in \Omega;$$

(1)

where $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a given functions and $\Omega$ is a non-empty convex set.

If $\Omega = \mathbb{R}^n$, Eq.(1) is an unconstrained problem, which could be solved by many methods, such as Newton method, Quasi-Newton method, Conjugate Gradient method, Fixed Point method and their variants. The details could refer to [1]–[5].

If $\Omega \neq \mathbb{R}^n$ and $\Omega$ is a non-empty closed set, the problem is a constrained problem. Great efforts have been made to find the solutions, including the projection Newton method [6], Trust Region method [7], [8], Lagrangian method [9]. These methods perform well both in numerical experiments and practical applications, based on the information of the first and the second derivative of $F$. However, many problems in application are not differentiable and the above mentioned methods are not available directly.

In recent years, many efforts have been put in finding the solution of non-smooth or semi-smooth system of (1), in which $F$ is non-differentiable, such as [10]–[13]. By approximating the secant condition of Quasi-Newton methods, Barzilai and Borwein proposed the BB step size in [14], which has been widely used in the unconstrained problem. There are many other modified BB step size methods, such as [15]–[18]. In [15], Li and Wan proposed an adaptive and composite BB step size by integrating the advantages of such existing step sizes and in [16], Huang et al. extended the nonmonotone line search method and developed a class of spectral gradient algorithms by combining with the spectral step-size. The Spectral Gradient (SG) method is one of extensions of the BB method and is not a descent method such that its convergent analysis is a non-trivial task. To overcome this problem, Huang and Wan proposed a spectral residual method in [19] for non-smooth and non-linear equations, and Svaiter and Solodov [6] proposed a hybrid projection method for solving the unconstrained monotone problems. By constructing an appropriate separation hyperplane in the Newton direction, under non regular assumptions, Svaiter and Solodov’s method is sufficiently descent and globally convergent, and it is suitable for solving smooth equations as it depends on the gradient information of the equations. Consequently, Zhang and Zhou [20] combined the SG method and projection method to propose a new framework for solving non-linear unconstrained monotone equations with derivative free feature. Illustrated by the work in [21],
Li and Guo [22] proposed another derivative free method to solve the non-monotone and non-linear problem with no constraints. Due to the non-expansion of the projection operator, Wang et al. [23] proposed a method to solve the convex constrained problem of (1) by projecting the predictor-corrector point in the constraint sets. And then, Yu et al. [24] extended the works of [20] and [23] to solve monotone equations with convex constraints. By modifying the search direction to be sufficiently descent, their method performs more effectively in numerical experiments. Subsequently, Awwal et al. [25] proposed a modified Hestenes-Stiefel (HS) spectral conjugate gradient (CG) method to solve monotone non-linear equations with convex constraints efficiently. Xiao and Zhu [26] combined the CG-Descent method [27] and projection method to solve the image deconvolution problems. More recently, many related works have been done to solve the constraint or unconstrained problem of (1) and could be referred to [28]–[35] for details.

In addition, it is noted that the emerging neural networks [36]–[42] provide different approaches for solving this kind of problems, which deserves the deep exploration.

In this paper, inspired by the works of [24] and [25], we further investigate the projection method with a modified SG parameter and propose a derivative free algorithm for solving non-linear monotone equations with convex constraints. The aim of this work is to improve the efficiency of the projection method and expand the scope of its applications. The proposed derivative free method is suitable to solve large-scale non-linear equations as no sub-problem needs to be solved, which is different from the classical projection method. In our proposed algorithm, the line search technique is also modified to obtain the descent property. The global convergence is guaranteed with a new line search technique under the mild conditions. Furthermore, we present some preliminary numerical experiments to illustrate the efficiency of the proposed method. In addition, we apply our proposed method to solve the $L_1$-norm regularized problems in signal reconstruction.

This paper is organized as follows. In Section 2, we introduce related preliminary knowledge and propose our method. In Section 3, the global convergence result of the proposed method is established. In Section 4, numerical experiments are implemented by comparing our proposed method with the other two methods, and then application of the method in signal reconstruction problem is constructed. Conclusion is given in Section 5. Throughout this paper, $\| \cdot \|$ denotes the Euclidean norm of a vector and $\langle \cdot, \cdot \rangle$ denotes the vector product. Moreover, for convenience, $F_k$ is denoted as an abbreviation of $F(x_k)$.

### II. ALGORITHM

If $\Omega$ is a closed convex subset of $\mathbb{R}^n$, the projection operator $P(\cdot)$ is a map, which is for any $x \in \mathbb{R}^n$,

$$P_\Omega(x) = \arg \min_x \| x - y \|, \quad y \in \Omega.$$  

We firstly recall some basic properties of the projection operator in [43] and [44].

**Proposition 1**: Let $\Omega$ be a closed convex subset of $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$ and $y \in \Omega$, the following inequality holds

$$\langle P_\Omega(x) - x, y - P_\Omega(x) \rangle \geq 0.$$  

**Proposition 2**: Let $\Omega \subset \mathbb{R}^n$ be a closed convex subset. Then it holds that

$$\| x - P_\Omega(y) \|^2 \leq \| x - y \|^2 - \| y - P_\Omega(y) \|^2$$  

for any $x, y \in \mathbb{R}^n$.

From (2) and (3), it is easy to find

$$\| P_\Omega(x) - P_\Omega(y) \| \leq \| x - y \|,$$  

which indicates that the projection operator is a non-expansive operator. In the spectral gradient method, the iterative scheme for solving (1) has a general form as

$$x_{k+1} = x_k + t_k d_k,$$

where $t_k$ is the step-size and $d_k$ is the search direction obtained by some formulation term.

In [24], the search direction $d_k$ is given by

$$d_k = -\tilde{\theta}_k F(x_k), \quad k = 1, \ldots, n,$$  

where $\tilde{\theta}_k$ is the direction scalar parameter, given by

$$\tilde{\theta}_k = \frac{s_k^T \tilde{s}_k}{s_k^T F_k s_k},$$  

where $\tilde{s}_k = x_k - x_{k-1}$, $\tilde{y}_k = F(x_k) - F(x_{k-1}) + r_s \tilde{s}_k$.

In [25], a bounded parameter of the search direction is given by

$$\theta'_k = 1 - \frac{(F_k ^T d_{k-1})^2}{\| F_k \|^2 \| d_{k-1} \|^2},$$  

Thus the value of $\theta'_k$ depends on the angle between $F_k, d_{k-1}$, and we have $0 \leq \theta'_k \leq 1$.

Motivated by [24] and [25], we propose a new method in which the search direction is given as,

$$d_k = \begin{cases} -F_k, & k = 0, \\ -\beta_k F_k, & k > 0, \end{cases}$$  

where

$$\beta_k = \gamma_k \cdot \theta_k,$$  

$$\theta_k = 1 - \mu \cdot \frac{(d_{k-1}^T F(x_k))^2}{\| d_{k-1} \|^2 \| F(x_k) \|^2},$$  

$$\gamma_k = \frac{(F(x_{k-1}) - F(x_k) + v s_{k-1})^T (s_{k-1})}{\| F(x_{k-1}) - F(x_k) + v s_{k-1} \|^2},$$

where $s_{k-1} = z_{k-1} - x_{k-1}$, $\tilde{F}(x_k) = \max(\| F_k - \| F_k \|)$. $v$ and $\mu$ are constants. Here, $z_k$ is a point depending on $x_k$ and $d_k$. Notice that if we set $\mu = 0$ and replace $z_{k-1}$ by $x_k$, the parameter $\theta_k$ will be $\theta_k$ in [24].

Due to the modified direction we denote our method as Algorithm MSGP with the following steps.
Algorithm MSGP

Step 1 (Initialization): Choose μ, ρ ∈ (0, 1), v, λ, σ > 0, stop tolerance ε ≥ 0 and the initial point x₀ ∈ Rⁿ. Set k = 0.

Step 2 (Stop criterion): If ∥F(x)∥ ≤ ε, then stop the iteration. Otherwise, go to the next step.

Step 3 (Search direction): If k = 0, compute d_k = −F(x_k). Otherwise, compute d_k by (7)-(9).

Step 4 (Line search): Compute the point z_k = x_k + α_k d_k, where α_k = λρ^{m_k} with m_k being the smallest non-negative integer such that

\[ F(z_k)^T d_k ≥ σα_k ∥d_k∥∥F(z_k)∥. \]  

Step 5 (Projection): Compute the predictor-corrector point by \( \bar{z}_k = x_k - ζ_k F(z_k) \), where

\[ ζ_k = \frac{F(z_k)^T (x_k - z_k)}{∥F(z_k)∥^2}. \]

And then compute next iterative point by the following projection operator,

\[ x_{k+1} = P_Ω(\bar{z}_k). \]

Step 6 (Update x_k): Set k := k + 1, go to step 2.

III. CONVERGENCE ANALYSIS

In this section, the global convergence of the proposed method is analyzed. Assume S is a solution set of (1) and \( F_k \neq 0 \) otherwise the solution has been obtained. First of all, some assumptions on the objective function \( F \) will be given.

Assumption 3: The function \( F : R^n → R^n \) is Lipschitz continuous, i.e., there exists a constant \( L > 0 \) such that

\[ ∥F(x) − F(y)∥ ≤ L∥x − y∥, \forall x, y ∈ R^n. \]  

Assumption 4: \( F \) satisfies monotonicity, i.e., for \( ∀x, y ∈ R^n \), it holds that

\[ (F(x) − F(y), (x − y)) ≥ 0. \]  

Lemma 5: Suppose the Assumption 3 holds. The sequence \{d_k\} is generated by Algorithm MSGP. We denote \( Γ_1 = \frac{1 − μ}{v + λ} \) and \( Γ_2 = \frac{1}{v} \). Then the following inequalities can be obtained as

\[ Γ_1 ∥F(x_k)∥ ≤ ∥d_k∥ ≤ Γ_2∥F(x_k)∥. \]  

The proof of Lemma 5 is in Section VI.

Lemma 6: Suppose the sequence \{d_k\} is generated by (6). In each iteration, the search direction \( d_k \) always is the descent direction, i.e., there exists a positive constant \( M = \min\{1, Γ_1\} \) such that

\[ d_k^T F(x_k) ≤ −M∥F(x_k)∥^2. \]  

Proof: If \( k = 0 \), by the definition of \( d_k \), we can easily find that

\[ d_k^T F(x_k) = −∥F(x_k)∥^2 ≤ −M∥F(x_k)∥^2. \]

If \( k ≠ 0 \), we have

\[ d_k^T F(x_k) = −β_k∥F(x_k)∥^2 ≤ −Γ_1∥F(x_k)∥^2 ≤ −M∥F(x_k)∥^2, \]  

where the first part of the inequality follows from (25). The proof is complete.

From Lemma 6, we have

\[ d_k^T F(x_k) < 0, \]  

which implies that the point \( z_k \), obtained in Step 4 of Algorithm MSGP, exists in each step.

Lemma 7: Suppose the Assumptions 3 and 4 hold for any \( x \in S \). Then the sequences \{x_k − x\}, \{F(x_k)\} and \{z_k\}, generated in Algorithm MSGP, are bounded. In addition,

\[ \lim_{k→+∞} α_k ∥d_k∥ = 0. \]  

Lemma 7 is proved in Section VI.

Theorem 8: Suppose the Assumptions 3 and 4 hold, and \{x_k\} is a sequence of points generated by Algorithm MSGP. The sequence \{F(x_k)\} converges globally, i.e.,

\[ \lim_{k→+∞} \inf ∥F(x_k)∥ = 0. \]  

The proof of Theorem 8 is given in Section VI.

IV. NUMERICAL EXPERIMENTS AND APPLICATION

In this section, in our numerical experiments, we consider 8 problems to implement our proposed method and make comparison with MHSP [25] method and PDY [45] method. And then we apply the proposed method in the signal reconstruction problem as application.

A. NUMERICAL EXPERIMENTS

In this subsection, to apply our proposed method in the numerical experiments, we consider the following 8 examples which are the samples in the related works of several references. Meanwhile, we make a numerical comparison of our method with MHSP [25] method and PDY [45] method. Both of the two methods have good performance in numerical experiments, and specially PDY method is well known for its stable performance on many problems.

In the numerical results, the number of iteration is denoted by \( NT \); the number of evaluation of \( F(x) \) in iterations is denoted by \( NF \); the CPU running time is denoted as \( Time \) and recorded in second; the norm of the residual value of \( F(x) \) is denoted as \( Norm \). In order to make the relative fair comparisons in our numerical experiments, we set 8 different initial points as follows: \( x_1 = (\frac{2−1}{n}, \frac{n−2}{n}, \ldots, 0)^T \), \( x_2 = (0, \frac{1}{n}, \ldots, \frac{n−1}{n})^T \), \( x_3 = (2, 2, \ldots, 2)^T \), \( x_4 = (8, 8, \ldots, 8)^T \), \( x_5 = (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})^T \), \( x_6 = (\frac{1}{4}, \frac{1}{4}, \ldots, \frac{1}{4})^T \), \( x_7 = (1, 1, \ldots, 1)^T \), \( x_8 = (1, \frac{1}{2}, \ldots, \frac{1}{2})^T \). The iteration stops if \( ∥F(x_k)∥ < 10^{-6} \) or \( NF ≥ 5000 \).

The parameters of MHSP method and PDY method are chosen from [25] and [45] respectively. In MSGP method, we set the parameter \( σ = 0.001, λ = 1, ρ = 0.8, μ = 0.85 \) and \( ν = 0.001 \). The symbol ‘-’ is denoted the method fails when \( NF ≥ 5000 \) on the relevant problems. The numerical results by applying three methods are shown in Tables 1 and 2.
**Problem 9:** A strictly convex function taken from [23], i.e.,

\[ F_i(x) = e^{\Omega x_i} - 1, \quad i = 1, 2, \ldots, n, \]

where \( \Omega = R^+ \).

**Problem 10:** A trigonometric function taken from [46], i.e.,

\[ F_i(x) = 2x_i - \sin x_i, \quad i = 1, 2, \ldots, n, \]

where \( \Omega = (-2, +\infty) \).

**Problem 11:** A strictly convex function modified from [23], i.e.,

\[ F_i(x) = \frac{ie^{\Omega x_i}}{n} - 1, \quad i = 1, 2, \ldots, n, \]

where \( \Omega = R^+ \).

**Problem 12:** A tridiagonal function taken from [47], i.e.,

\[ F_1(x) = x_1 - e^{\cos \frac{x_1 + x_3}{n} + \frac{1}{n}}, \]
\[ F_i(x) = x_i - e^{\cos \frac{x_i - x_{i-1}}{n} + \frac{1}{n} + \frac{1}{n}}, \quad i = 2, 3, \ldots, n - 1, \]
\[ F_n(x) = x_n - e^{\cos \frac{x_n - x_{n-2}}{n} + \frac{1}{n} + \frac{1}{n}}, \]

where \( \Omega = R^+ \).

**Problem 13:** A modified logarithmic function taken from [21], i.e.,

\[ F_i(x) = \log(x_i + 1) - \frac{x_i}{n}, \quad i = 1, 2, \ldots, n, \]

where \( \Omega = (-1, +\infty) \).

**Problem 14:** A function taken from [45], i.e.,

\[ F_1(x) = 2x_1 + \frac{0.5(x_1 + \frac{1}{n^2} + 1)^3}{(n + 1)^2} - x_2, \]
\[ F_i(x) = 2x_i + \frac{0.5(x_i + \frac{i}{n^2} + 1)^3}{(n + 1)^2} - x_{i-1} + x_{i+1}, \]
\[ F_n(x) = 2x_n + \frac{0.5(x_n + \frac{n}{n^2} + 1)^3}{(n + 1)^2} - x_{n-1}, \]

where \( i = 2, 3, \ldots, n - 1, \)

**Problem 15:** A non-smooth function taken from [48], i.e.,

\[ F_i(x) = \min\{\min(|x_i|, x_i^2), \max(|x_i|, x_i^3)\}, \quad i = 1, 2, \ldots, n, \]

where \( \Omega = R^+ \).

**Problem 16:** An exponential function taken from [49], i.e.,

\[ F_1(x) = e^{\Omega x_1} - 1, \]
\[ F_i(x) = e^{\Omega x_i} + x_{i-1} - 1, \quad i = 2, 3, \ldots, n, \]

where \( \Omega = R^+ \).

From a practical point of view, both stability and effectiveness are important for a good appropriate algorithm. In our numerical experiments, although good results have been achieved by using MHSP method on some specific problems, such as Problems 10, 12, and 13, PDY method is more stable than MHSP method. In comparison, it can be seen from Tables 1 and 2 that our proposed MSGP method not only guarantees the good performance that MHSP method shows on some problems, but also maintains the stability compared with the PDY method. Furthermore, MSGP method can guarantee the effectiveness and stability of the algorithm with very few iterations and running time, and one obvious fact is that problems on which the method fails are rare. In order

| Problem | N | MSGP | MSGP | PDY |
|---------|---|------|------|-----|
| Problem 1 | | | | |
| Problem 2 | | | | |
| Problem 3 | | | | |
| Problem 4 | | | | |
| Problem 5 | | | | |
| Problem 6 | | | | |
| Problem 7 | | | | |
| Problem 8 | | | | |
| Problem 9 | | | | |
| Problem 10 | | | | |
| Problem 11 | | | | |
| Problem 12 | | | | |
| Problem 13 | | | | |
| Problem 14 | | | | |
| Problem 15 | | | | |
| Problem 16 | | | | |

TABLE 1. Numerical performance with respect to the dimension of the problem N = 5000.
TABLE 2. Numerical performance with respect to the dimension of the problem $N = 50000$.

| Problem | $N$ | MSGP | MHSP | PDY | Time/Sec |
|---------|-----|------|------|-----|----------|
| Problem 1 | 10  | 15  | 0.64 | 0.75 | 0.95 |
|          | 20  | 15  | 0.64 | 0.75 | 0.95 |
|          | 40  | 15  | 0.64 | 0.75 | 0.95 |
|          | 80  | 15  | 0.64 | 0.75 | 0.95 |
|          | 160 | 15  | 0.64 | 0.75 | 0.95 |
|          | 320 | 15  | 0.64 | 0.75 | 0.95 |

FIGURE 1. (a), (b) and (c) show $\log_2$ scaled performance profiles of methods MSGP, MHSP and PDY with respect to $NT$, $NF$ and CPU Time respectively. Overall, our proposed method MSGP is more applicable in many problems.

B. APPLICATION IN SIGNAL RECONSTRUCTION

Sparse representations of signals has attracted much attention from researchers in the fields of signal processing, image processing, computer vision and machine learning, specially sparse signal reconstruction problems. The standard problem of sparse representation is closed to solve a $l_2 - l_1$ problem as following:

$$\min \frac{1}{2} ||Ax-b||_2^2 + \tau ||x||_1.$$  

(18)

where $A$ is a $k \times n$ matrix, $b$ is a vector and $\tau$ is a non-negative parameter.

Many different algorithms have been proposed to solve (18). In [51], these algorithms are divided into four categories.
including greedy strategy approximation, homotopy algorithm, proximate algorithm and constrained optimization. One of the earlier greedy algorithm is Matching Pursuits (MP) [52] algorithm. It is a simple method involving the computation of the inner product between the signal and dictionary column or basis by dividing signals on these dictionaries or basis. The homotopy algorithm [53], which requires at least as many iterations as the number of nonzero arguments of $x$, is a competitive method when the signal is sparse enough. There are various approximate algorithms, such as Iterative Shrinkage/Thresholding (IST) [54] method, the sparse reconstruction by Separable Approximation [55] (SpaRSA) method and so on. The main idea of these approximate algorithms is to approximate the problem by solving the sub-problems iteratively, and they are usually used to solve the non-smooth convex problems.

Our proposed MSGP method can be viewed as a modified Gradient Projection Sparse Representation (GPSR) method, which is a standard approach for the constrained optimization problem. Comparing with the original GPSR method, instead of searching in the direction of negative value of the gradient, our iterations are obtained by the following two steps. Firstly, a point is obtained by searching in the direction of the negative value of the function. The second step is to project the obtained point onto the separation hyper-plane. By considering the robustness of the signal reconstruction problem, the signal is with noise or error. And the strategy of our method is to reconstruct new signal with length $n$ on the $k$ observations.

As in [56], we split vector $x$ into two parts as

$$x = u - v, \quad u \geq 0, \quad v \geq 0,$$

where $u$ and $v$ are the positive and the negative parts of $x$, respectively. Then, from (18), we have

$$\min_{u,v} \frac{1}{2} ||A(u - v) - b||^2 + \tau E_n u + \tau E_n v,$$

with the constraints $u \geq 0, v \geq 0$, where $E_n = [1, 1, \ldots, 1]$. By the properties in [26] and [57], the above problem is equivalent to the following form

$$F(z) = \min_z \{z, Hz + C\} = 0,$$

where

$$z = \begin{bmatrix} u \\ v \end{bmatrix}, \quad H = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix}, \quad C = \tau E_{2n} + \begin{bmatrix} -A^T b \\ A^T b \end{bmatrix}.$$ 

By Lemma 3 in [58] and Lemma 2 in [59], we have that (20) is Lipschitz continuous and monotone. Thus, our method is applied to solve (20) in this paper. Furthermore, Wan et al. [60] proposed a new reformulation of (18), and they took the constraint $u^T v = 0$ into account in (19).

In our numerical experiments, $A$ is a random Gaussian matrix with $k = 1024, n = 16384$. We denote the origin signal as $\tilde{x}$, which is a random signal with non-zero elements $\hat{n} = 256$. We reconstruct the signal $\tilde{x}$ from $b$ by $k = 1024$ observations and set $f(x) = \frac{1}{2}||Ax - b||^2 + \tau ||x||_1$, where $b = A\tilde{x} + \tilde{b}$ and $\tilde{b}$ is white Gaussian noise with parameter $0.05, \tau = 0.0197762$.

We compare the efficiency of our proposed method against the PCG method [29], and IST method [61], which is a well known and widely applied method in Lasso problem (18). To indicate the efficiency of our method, we consider three elements: the MSE, which is equivalent to $\frac{||x - \tilde{x}||^2}{n}$; the
CPU time, which is counted in second; the totally number of iteration abbreviated by "Iter". We set \( \nu = 0.05, \mu = 0.8, \lambda = 10, \rho = 0.6 \) and \( \sigma = 0.001 \) in MSGP method. The parameter of PCG method follows [29]. The initial guess of \( x_0 \) in all methods is set to be 0. The iterations in the numerical experiments terminate when the relative change of the result satisfies \( \frac{|f(x_k) - f(x_{k-1})|}{|f(x_{k-1})|} \leq 10^{-4} \).

From the perspective of sparse reconstruction, our method and PCG method can be viewed as special cases of GPSR method. A large number of experiments show that the GPSR method benefits from a so-called warm start [62]. Thus, illustrated by it, we employ the regularization parameter \( \tau \) as a descending sequence as follows: \( \{800\tau, 800^2\tau, 800^3\tau, 800^4\tau, \tau\} \) in the numerical experiments by applying these methods to solve the problem (18).

Fig. 2 demonstrates the original signal, noised signal, the reconstructed signal by PCG method, MSGP method and IST method, respectively from the top to the bottom. Fig. 3 shows results of the iterations vs MSE in (a) and results of CPU time vs MSE in (b) respectively. From Figs. 2 and 3, we can clearly find out that, for comparing with the other two methods on the condition of the same order of magnitude of MSE, our method can reconstruct the original signal in less iterations, which means the lower cost of CPU running time.

V. CONCLUSION

In this paper, a new derivative free algorithm MSGP for solving non-linear equations with convex constraints is proposed by combining the method of spectral gradient and the projection method. In our method, we don’t need to solve sub-problem, which is different from the classical projection method. More specifically, the line search technique is modified in our method. The global convergence of the MSGP method is guaranteed analytically with certain conditions. The numerical experiments show that the large-scale non-linear equations with convex constraints can be effectively solved with our method. Furthermore, our proposed MSGP method is applicable for solving the \( L_1 \)-norm regularized problems in signal reconstruction. As one future work, we will explore the application of our method in the research area of the neural network. It is noted that the projection neural networks proposed in [36], [37] can also be used to solve the problem in this paper. We will report the comparison between our method and those projection neural networks in the forthcoming paper.

VI. PROOFS OF THE MAIN RESULTS

A. A PROOF OF LEMMA 5

Proof: From Assumption 3 and the definition of \( d_k \), we have

\[
(F(z_k) - F(x_k)) + v s_k T s_k \leq (\nu + L)|s_k|^2. \tag{21}
\]

In addition, \( F \) is monotonic, i.e., (12) holds. Then we have

\[
(F(z_k) - F(x_k)) + v s_k T s_k \geq \nu |s_k|^2. \tag{22}
\]

From (9), (21) and (22), \( \gamma_k \) satisfies

\[
\frac{1}{\nu + L} \leq \gamma_k \leq \frac{1}{\nu}. \tag{23}
\]

And then we discuss the value of \( \theta_k \) in two cases. If \( |F(x_{k-1})| \geq |F(x_k)| \), from the definition of \( \tilde{F} \) and the Cauchy-Schwarz inequality, we have

\[
0 \leq \frac{(d_{k-1}^T F(x_k))^2}{\|d_{k-1}\|^2 \|\tilde{F}(x_{k-1})\|^2} = \frac{(d_{k-1}^T F(x_k))^2}{\|d_{k-1}\|^2 \|F(x_{k-1})\|^2} \leq 1.
\]

If \( |F(x_{k-1})| \leq |F(x_k)| \), by the Cauchy-Schwarz inequality, we have

\[
0 \leq \frac{(d_{k-1}^T F(x_k))^2}{\|d_{k-1}\|^2 \|F(x_{k-1})\|^2} = \frac{(d_{k-1}^T F(x_k))^2}{\|d_{k-1}\|^2 \|F(x_k)\|^2} \leq 1.
\]

Both of the above two inequalities imply that

\[
1 - \mu \leq \theta_k \leq 1. \tag{24}
\]

Here, we denote \( \Gamma_1 = \frac{1 - \mu}{\nu + L}, \Gamma_2 = \frac{1}{\nu} \). From (7), (23) and (24), consequently we obtain

\[
\Gamma_1 \leq \beta_k \leq \Gamma_2. \tag{25}
\]

Combining the definition of \( d_k \), we have (13) holding. \( \square \)

B. A PROOF OF LEMMA 7

Proof: We first deduce the boundedness of \( \{x_k - \bar{x}\} \) and \( \{x_k\} \).

\[
|x_{k+1} - \bar{x}|^2 = \|P_{\Omega}[x_k - \xi_k F(z_k)] - P_{\Omega}(\bar{x})\|^2
\]

\[
\leq \|x_k - \bar{x} - \xi_k F(z_k)\|^2
\]

\[
= \|x_k - \bar{x}\|^2 - 2\xi_k F(z_k)^T (x_k - \bar{x}) + \xi_k^2 \|F(z_k)\|^2
\]

\[
\leq \|x_k - \bar{x}\|^2 - 2\xi_k F(z_k)^T (x_k - \bar{x}) + \xi_k^2 \|F(z_k)\|^2
\]

\[
= \|x_k - \bar{x}\|^2 - \frac{(F(z_k)^T (x_k - \bar{x}))^2}{\|F(z_k)\|^2}, \tag{26}
\]

where the first and second inequalities follow (4) and Assumption 4 respectively, and the last equality follows the definition of \( \xi_k \).

The inequality (26) yields that \( \{x_k - \bar{x}\} \) is a non-increasing and bounded sequence, and thus \( \{x_k - \bar{x}\} \) is convergent. Since

\[
|x_k - \bar{x}| \leq \|x_0 - \bar{x}\|, \tag{27}
\]

the sequence \( \{x_k\} \) is bounded too. Then, by the continuity of \( F \), we obtain the boundedness of sequence \( \{F(x_k)\} \), i.e., there is a constant \( M' \) such that

\[
|F(x_k)| \leq M'.
\]

From (13), we have

\[
|d_k| \leq \Gamma_2 M'. \tag{28}
\]
By (26) and (28), we obtain
\[
(\Gamma_2 M')^2 ||x_0 - \bar{x}||^2 \geq \sum_{k=0}^{+\infty} \frac{||d_k||^2 (F'(z_k)^T (x_k - z_k))^2}{||F'(z_k)||^2} \geq \sigma^2 \sum_{k=0}^{+\infty} ||d_k||^4,
\]
where the last inequality is obtained from (10). Hence, the sequence \(\{d_k||d_k]\) is bounded and
\[
\lim_{k \to +\infty} \alpha_k||d_k|| = 0. \tag{29}
\]

**C. A PROOF OF THEOREM 8**

**Proof:** Assume the limit (17) dose not hold, which implies that there is a positive constant \(\epsilon\), such that the inequality \(||F_k|| > \epsilon\) holds for all \(k \in N\). Then by (13), we have \(||d_k|| > \Gamma_1 \epsilon\) as well. By the boundedness of \(\{\alpha_k d_k\}\) and \(\{x_k\}\), we have the sequence \(x_k + \rho^{-1} \alpha_k d_k\) is bounded as well. Thus according to the continuity of \(F\), there is a constant \(M\), such that
\[
||F(x_k + \rho^{-1} \alpha_k d_k)|| \leq M. \tag{30}
\]

From the process of Step 4 in Algorithm MSGP and the (30), the following inequality holds that
\[
-F(x_k + \rho^{-1} \alpha_k d_k)^T d_k < \sigma \alpha_k \rho^{-1} ||d_k|| ||F(x_k + \rho^{-1} \alpha_k d_k)|| \leq \sigma \alpha_k \rho^{-1} ||d_k|| M. \tag{31}
\]

By the definition of \(d_k\), we have
\[
\Gamma_1 ||F_k||^2 = -d_k^T F_k = (d_k^T F(x_k + \rho^{-1} \alpha_k d_k) - d_k^T F(x_k + \rho^{-1} \alpha_k d_k)) \leq ||d_k|| ||F(x_k + \rho^{-1} \alpha_k d_k)|| + ||\sigma \alpha_k \rho^{-1} ||d_k|| ||M|| \leq \sigma \alpha_k \rho^{-1} ||d_k||^2 + \sigma \alpha_k \rho^{-1} \Gamma_1 ||F_k|| M. \tag{32}
\]

where the inequalities are obtained from (13), (31) and Assumption 3 respectively. By (32) and the assumption \(||F_k|| > \epsilon\), we have
\[
\alpha_k \geq \frac{\rho \Gamma_1}{L \Gamma_2^2 + \sigma M \Gamma_1^2 ||F_k||} \geq \frac{\rho \Gamma_1}{L \Gamma_2^2 + \sigma M \epsilon \Gamma_1^2}.
\]

The above inequality implies that
\[
\alpha_k ||d_k|| \geq \frac{\rho \Gamma_1}{L \Gamma_2^2 + \sigma M \epsilon ||F_k||},
\]
which contradicts with (29). The proof is completed.

**REFERENCES**

[1] M. Á. Fernández and M. Moubachir, “A Newton method using exact Jacobians for solving fluid–structure coupling,” Comput. Struct., vol. 83, nos. 2–3, pp. 127–142, Jan. 2005.

[2] E. G. Birgin, N. Krejčí, and J. M. Martínez, “Globally convergent inexact quasi-Newton methods for solving nonlinear systems,” Numer. Algorithms, vol. 32, nos. 2–4, pp. 249–260, Apr. 2003.

[3] K. Scheinberg and X. Tang, “Practical inexact proximal quasi-Newton method with global complexity analysis,” Math. Program., vol. 160, nos. 1–2, pp. 495–529, Nov. 2016.

[4] R. Fletcher, “Function minimization by conjugate gradients,” Comput. J., vol. 7, no. 2, pp. 149–154, Feb. 1964.

[5] Y.-H. Dai, L.-Z. Liao, and D. Li, “On restart procedures for the conjugate gradient method,” Numer. Algorithms, vol. 35, nos. 2–4, pp. 249–260, Apr. 2004.

[6] M. V. Solodov and B. F. Svaiter, “A globally convergent inexact Newton method for systems of monotone equations,” in Reformulation: Non-smooth, Piecewise Smooth, Semismooth and Smoothing Methods. Boston, MA, USA: Springer, 1998, pp. 355–369.

[7] J. J. Moré and D. C. Sorensen, “Computing a trust region step,” SIAM J. Sci. Comput., vol. 4, no. 3, pp. 553–572, Sep. 1983.

[8] T. F. Coleman and Y. Li, “An interior trust region approach for nonlinear minimization subject to bounds,” SIAM J. Optim., vol. 6, no. 2, pp. 418–445, May 1996.

[9] L. Q. Qi and J. Sun, “A nonsmooth version of Newton’s method,” Math. Program., vol. 58, nos. 1–3, pp. 353–367, 1993.

[10] J. K. Liu, J. L. Xu, and L. Q. Zhang, “Partially symmetrical derivative-free Liu-Storey projection method for convex constrained equations,” Int. J. Comput. Math., vol. 96, no. 9, pp. 1787–1798, Sep. 2019.

[11] A. B. Abubakar, P. Kumam, H. Mohammad, A. M. Awwal, and K. Sitthithakerngkiet, “A modified Fletcher–Reeves conjugate gradient method for monotone nonlinear equations with some applications,” Mathematics, vol. 7, no. 8, p. 745, 2019.

[12] P. Gao, C. He, and Y. Liu, “An adaptive family of projection methods for constrained monotone nonlinear equations with applications,” Appl. Math. Comput., vol. 359, pp. 1–16, Oct. 2019.

[13] A. H. Ibrahim, P. Kumam, A. B. Abubakar, W. Jirakitpuwapat, and J. Abubakar, “A hybrid conjugate gradient algorithm for constrained monotone equations with application in compressive sensing,” Helv. Phys. Acta, vol. 63, no. 3, Mar. 2020, Art. no. e03466.

[14] J. Barzilai and J. M. Borwein, “Two-point step size gradient methods,” IMA J. Numer. Anal., vol. 8, no. 1, pp. 141–148, 1988.

[15] T. Li and Z. Wan, “New adaptive barzilai-borwein step size and its application in solving large-scale optimization problems,” ANZIAM J., vol. 61, p. 76, Mar. 2019.

[16] S. Huang, Z. Wan, and J. Zhang, “An extended nonmonotone line search technique for large-scale unconstrained optimization,” J. Comput. Appl. Math., vol. 330, pp. 586–604, Mar. 2018.

[17] Y.-H. Dai, Y. Huang, and X.-W. Liu, “A family of spectral gradient methods for optimization,” J. Optim. Theory Appl., vol. 74, pp. 43–65, Sep. 2019.

[18] E. G. Birgin, J. M. Martínez, and M. Raydan, “Spectral projected gradient methods: Review and perspectives,” J. Stat. Softw., vol. 160, nos. 2–3, pp. 127–142, Jan. 2005.

[19] S. Huang and Z. Wan, “A new nonmonotone spectral residual method for nonsmooth nonlinear equations,” J. Comput. Appl. Math., vol. 269, pp. 59–69, Oct. 2015.

[20] T. Li and Z. Wan, “Line searches for large-scale nonlinear systems of equations,” Comput. Optim. Appl., vol. 74, no. 1, pp. 1–21, 2014.

[21] S. Huang and Z. Wan, “A new nonmonotone spectral residual method for nonsmooth nonlinear equations,” J. Comput. Optim. Math., vol. 313, pp. 82–101, Mar. 2017.

[22] L. Zhang and W. Zhou, “Spectral gradient projection method for solving large-scale unconstrained optimization,” J. Comput. Appl. Math., vol. 196, no. 2, pp. 478–484, Nov. 2006.

[23] W. La Cruz, J. M. Martínez, and M. Raydan, “Spectral residual method without gradient information for solving large-scale nonlinear systems of equations,” Math. Program., vol. 75, no. 255, pp. 1429–1449, Apr. 2006.

[24] X. Li and X. Guo, “Spectral residual methods with two new non-monotone line searches for large-scale nonlinear systems of equations,” Appl. Math. Comput., vol. 269, pp. 59–69, Oct. 2015.

[25] C. Wang, Y. Wang, and C. Xu, “A projection method for a system of nonlinear monotone equations with convex constraints,” Math. Methods Oper. Res., vol. 66, no. 1, pp. 33–46, Jul. 2007.
[24] Z. Yu, J. Lin, J. Sun, Y. Xiao, L. Liu, and Z. Li, “Spectral gradient projection method for monotone nonlinear equations with convex constraints,” *Appl. Numer. Math.*, vol. 59, no. 10, pp. 2416–2423, Oct. 2009.

[25] A. M. Awwal, P. Kumam, and A. B. Abubakar, “A modified conjugate gradient method for monotone nonlinear equations with convex constraints,” *Appl. Numer. Math.*, vol. 145, pp. 507–520, Nov. 2019.

[26] Y. Xiao and H. Zhu, “A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing,” *J. Math. Anal. Appl.*, vol. 405, no. 1, pp. 310–319, Sep. 2013.

[27] W. W. Hager and H. Zhang, “A new conjugate gradient method with guaranteed descent and an efficient line search,” *SIAM J. Optim.*, vol. 16, no. 1, pp. 170–192, Jan. 2005.

[28] S.-Y. Liu, Y.-Y. Huang, and H.-W. Jiao, “Sufficient descent conjugate gradient methods for solving convex constrained nonlinear monotone equations,” in *Abstract and Applied Analysis*. London, U.K.: Hindawi, 2014.

[29] J. K. Liu and S. J. Li, “A projection method for convex constrained monotone nonlinear equations with applications,” *Comput. Math. Appl.*, vol. 70, no. 10, pp. 2442–2453, Nov. 2015.

[30] J. Guo and Z. Wan, “A modified spectral PRP conjugate gradient projection method for solving large-scale monotone equations and its application in compressed sensing,” *Math. Problems Eng.*, vol. 2019, pp. 1–17, Apr. 2019.

[31] J. K. Liu and S. J. Li, “Spectral DY-type projection method for nonlinear monotone systems of equations,” *J. Comput. Math.*, vol. 33, no. 4, pp. 341–355, Jun. 2015.

[32] H. Mohammad and A. Bala Abubakar, “A descent derivative-free algorithm for nonlinear monotone equations with convex constraints,” *RAIRO: Optim. Res.*, vol. 54, no. 2, pp. 489–505, Mar. 2020.

[33] J. Liu and Y. Duan, “Two spectral gradient projection methods for constrained equations and their linear convergence rate,” *J. Inequalities Appl.*, vol. 2015, no. 1, p. 8, Dec. 2015.

[34] H. Mohammad and A. B. Abubakar, “A positive spectral gradient-like method for large-scale nonlinear monotone equations,” *Bull. Comput. Appl. Math.*, vol. 5, no. 1, 2017.

[35] G. H. Yu, S. Z. Niu, and J. H. Ma, “Multivariate spectral gradient projection method for nonlinear monotone equations with convex constraints,” *Ind. Manage. Optim.*, vol. 9, no. 1, pp. 117–129, 2013.

[36] Y. Xia, H. Leung, and J. Wang, “A projection neural network and its application to constrained optimization problems,” *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 4, pp. 447–458, Apr. 2002.

[37] Y. Xia and G. Feng, “A new neural network for solving nonlinear projection equations,” *Neural Netw.*, vol. 20, no. 5, pp. 577–589, Jul. 2007.

[38] S. Zhang, Y. Xia, and J. Wang, “A complex-valued projection neural network for constrained optimization of real functions in complex variables,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3227–3238, Dec. 2015.

[39] Y. Xia, G. Feng, and J. Wang, “A novel recurrent neural network for solving nonlinear optimization problems with inequality constraints,” *IEEE Trans. Neural Netw.*, vol. 19, no. 8, pp. 1340–1353, Aug. 2008.

[40] Y. Xia, “An extended projection neural network for constrained optimization,” *Neural Comput.*, vol. 16, no. 4, pp. 863–883, Apr. 2004.

[41] S. Zhang and Y. Xia, “Solving nonlinear optimization problems of real functions in complex variables by complex-valued iterative methods,” *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 277–287, Jan. 2018.

[42] Y. Xia, J. Wang, and W. Guo, “Two projection neural networks with reduced model complexity for nonlinear programming,” *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 16, 2019, doi: 10.1109/TNNLS.2019.2927639.

[43] E. H. Zarantonello, “Projections on convex sets in Hilbert space and spectral theory: Part I. Projections on convex sets: Part II. Spectral theory,” in *Contributions to Nonlinear Functional Analysis*. Amsterdam, The Netherlands: Elsevier, 1971, pp. 237–424.

[44] N. Xu, “Convergence properties of projection and contraction methods for variational inequality problems,” *Appl. Math. Optim.*, vol. 43, no. 2, pp. 147–168, Jan. 2001.

[45] J. Liu and Y. Feng, “A derivative-free iterative method for nonlinear monotone equations with convex constraints,” *Numer. Algorithms*, vol. 82, no. 1, pp. 245–262, Sep. 2019.

[46] W. J. Zhou and D. H. Li, “Limited memory BFGS method for nonlinear monotone equations,” *J. Comput. Math.*, vol. 25, no. 1, pp. 89–96, 2007.

[47] Y. Bing and G. Lin, “An efficient implementation of Merrill’s method for sparse or partially separable systems of nonlinear equations,” *SIAM J. Optim.*, vol. 1, no. 2, pp. 206–221, May 1991.

[48] W. L. Cruz and M. Raydan, “Nonmonotone spectral methods for large-scale nonlinear systems,” *Optim. Methods Softw.*, vol. 18, no. 5, pp. 583–599, Oct. 2003.

[49] E. D. Dolan and J. J. Moré, “Benchmarking optimization software with performance profiles,” *Math. Program.*, vol. 91, no. 2, pp. 201–213, Jan. 2002.

[50] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, “A survey of sparse representation: Algorithms and applications,” *IEEE Access*, vol. 3, pp. 490–530, 2015.

[51] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[52] D. M. Malioutov, M. Cetin, and A. S. Willsky, “Homotopy continuation for sparse signal representation,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, May 2005, pp. v733–v736.

[53] I. D. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[54] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.

[55] J.-J. Fuchs, “Multipath time-delay detection and estimation,” *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 237–243, 1st Quart., 1999.

[56] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[57] J.-S. Pang, “Inexact Newton methods for the nonlinear complementarity problem,” *Math. Program.*, vol. 36, no. 1, pp. 54–71, Oct. 1986.

[58] Y. Xiao, Q. Wang, and Q. Hu, “Non-smooth equations based method for norm problems with applications to compressed sensing,” *Nonlinear Anal., Theory, Methods Appl.*, vol. 74, no. 11, pp. 3570–3577, Jul. 2011.

[59] Z. Wan, J. Guo, J. Liu, and W. Liu, “A modified spectral conjugate gradient projection method for signal recovery,” *Signal, Image Video Process.*, vol. 12, no. 8, pp. 1455–1462, Nov. 2018.

[60] M. A. T. Figueiredo and R. D. Nowak, “A bound optimization approach to wavelet-based image deconvolution,” in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Sep. 2005, pp. II-782.

[61] E. T. Hale, W. Yin, and Y. Zhang, “A fixed-point continuation method for t1-regularized minimization with applications to compressed sensing,” *Rice Univ.*, Houston, TX, USA, CAAM Tech. Rep. TR07-07, 2007, vol. 43, p. 44.
LEI YANG (Member, IEEE) received the Ph.D. degree in mathematics from the Hong Kong University of Science and Technology, Hong Kong, in 2008. From 2008 to 2012, she worked as a Postdoctoral Researcher and a Lecturer with the Department of Mathematics, Hong Kong Baptist University. She is currently an Assistant Professor with the Faculty of Information and Technology, Macau University of Science and Technology. Her research interests include computational mathematics and related applications.

YONG LIANG received the B.Sc. and M.Sc. degrees in applied mathematics from Xi’an Jiaotong University, Xi’an, China, in 1996 and 1999, respectively, and the Ph.D. degree in computer science from the Chinese University of Hong Kong, Hong Kong, in 2003. From 2004 to 2006, he worked as a Postdoctoral Researcher with the Department of Computer Science and Engineering, Chinese University of Hong Kong and later joined Shantou University, China. He joined the Faculty of Information Technology, Macau University of Science and Technology, in 2007, where he is currently a Professor. He has authored and coauthored over 50 articles. His research interests include computational intelligent, including machine learning, big data analysis, evolutionary computation, and bioinformatics.