A new projection-based algorithm for solving a large-scale nonlinear system of monotone equations

Mushtak A. K. Shiker[*]

Department of Mathematics, University of Babylon
Babylon, Iraq
E-mail: <pure.mushtaq@uobabylon.edu.iq>

Keyvan Amini

Department of Mathematics, Faculty of Science, Razi University
Kermanshah, Iran
E-mail: <kamini@razi.ac.ir>

Abstract. Projection-based methods are an efficient and applicable family of derivative free methods for solving nonlinear monotone systems. This paper proposes a new projection method for solving a system of large-scale nonlinear monotone equations. The new algorithm, in each iteration, by using a modified conjugate gradient direction, constructs an appropriate hyperplane that strictly separates the current approximation from the solution set of the problem. Then the new approximation is determined by projecting the current point onto the separating hyperplane. The global convergence and the linear convergence rate of the proposed algorithm are proved under standard assumptions. Preliminary numerical experiments indicate that the proposed algorithm is promising.

1. Introduction

Consider the nonlinear system of equations:

$$F(x) = 0, \tag{1}$$

where $F : R^n \to R^n$ is a continuous and monotone function, i.e.:

$$\left(F(x) - F(y)\right)^T (x - y) \geq 0, \quad \forall x, y \in R^n. \tag{2}$$

---

[*] Corresponding author.

Nonlinear monotone systems of equations have many practical backgrounds in applied mathematics, as in: some monotone inequality problems [20], economic equilibrium problems and the first order necessary condition of the unconstrained optimization.

There are numerous methods for solving (1) where those are, generally, based on the following unconstrained optimization problem: $\min_{x \in R^n} f(x) = \frac{1}{2} \|F(x)\|^2$. Newton and quasi-Newton strategies [4, 5, 6, 7, 11, 15] are examples for such methods. They are attractive because they converge rapidly from a sufficient initial guess. But they are typically unattractive for large-scale nonlinear systems of equations because they need to solve a linear system using the Jacobian matrix or an approximation of it. In recent decades, as an efficient strategy, projection methods have received more attention so as to find solutions for monotone nonlinear systems. These methods usually use a search direction d_k which (for all $k \in N$) satisfies:

$$F_k^T d_k \leq -c\|F_k\|^2, \tag{3}$$

where c is a positive constant [1, 23]. It is observable that this condition is exactly the sufficient descant condition if $F(x)$ is a gradient vector for a real value function.

In 1998, Solodov and Svaiter [17] proposed the first projection algorithm for solving (1). They proved that their method is globally convergent to a solution of (1) without any additional regularity assumptions. Another attractive property of their method is that the sequence of the distances from the iterates to the solution set of the equation is decreasing. In 2005, Zhou and Toh [24] extended Solodov and Svaiter's method to solve a system of monotone equations with singular solutions, showing the superlinear convergence of their method. Because of the simplicity and the low storage, some authors decided to use the conjugate gradient methods in projection methods. In 2009, Cheng [8] proposed a derivative-free PRP method along with the projection method for solving monotone equations. In 2010, Yan et al. [19], by using modified HS directions, proposed two efficient derivative-free methods for solving nonlinear monotone equations. In 2015, Amini et al. [2] also proposed a double-projection-based algorithm for large-scale nonlinear system of monotone equations.

In this paper, inspired by some modified CG directions using the second order information, a new projection algorithm for solving problem (1) is proposed. The new algorithm has some advantages. First, the sequence of the distances from the iterates to the solution set of the equations is decreasing, and the approximates generated by the new algorithm are closer to the solution set of the equations in comparison to other projection methods. Second, the new algorithm is derivative-free so that it can be employed to solve nonsmooth monotone equations. Third, it does not impose any expenses computing to find the direction. Forth, the new method is globally convergent even when the solution of problem is not singleton. Under suitable conditions, we showed the global convergence and local linear convergence rate of the proposed method. The given preliminary numerical experiments clearly state that the algorithm is numerically robust and effective.

The rest of the paper is organized as follows: In section 2, after describing the idea of the projection-based methods, a new effective algorithm for solving problem (1) is presented. Section 3 presents the global convergence analysis of the proposed algorithm. Section 4 reports some numerical experiments to show the promising behavior of the approach especially when it

is used for large-scale nonlinear monotone equations. Finally, some concluding remarks are presented in section 5.

2. The new projection algorithm

As mentioned previously, there is a wide variety of approaches developed to solve a system of nonlinear equations. These approaches, generally, use an iterative process [16] for generating the next iterate by $x_{k+1} = x_k + \alpha_k d_k$, where α_k is a step length that is determined by some line search technique.

The projection methods firstly determine a direction d_k, then they use a line search technique to find a nonnegative step length α_k such that:

$$F(z_k)^T(x_k - z_k) > 0, \tag{4}$$

where

$$z_k = x_k + \alpha_k d_k. \tag{5}$$

After that, the hyperplane

$$H_k = \{x \in R^n \,|\, F(z_k)^T(x - z_k) = 0\}, \tag{6}$$

strictly separates the current approximation x_k from the solution set of the problem. Based on this fact, Solodov and Svaiter [17] advised that the next iterate, x_{k+1}, can be determined by projection z_k onto H_k, where

$$C_k = \{x \in R^n \,|\, F(z_k)^T(x - z_k) \leq 0\}. \tag{7}$$

Based on the fact that C_k contains the solutions of the problem, but $x_k \notin C_k$, by projecting x_k onto C_k, a better approximation for a solution of system (1) can be obtained. This proposes that the next approximation, x_{k+1}, can be determined as follows

$$x_{k+1} = P_c(x_k) = \mathrm{argmin}\{\|x - x_k\| \,|\, x \in C_k\} = x_k - \frac{F(z_k)^T(x_k - z_k)}{\|F(z_k)\|^2}F(z_k). \tag{8}$$

Theoretical properties as well as the numerical results demonstrate the efficiency and the robustness of the projection-based algorithms for monotone equations. It is clear that a proper line search technique to complete the projection algorithm is needed. Most of practical approaches exploit an inexact line search to identify a step length α_k guaranteeing the global convergence property in minimal cost. There exist some line search conditions for general nonlinear equations, for example [13, 19, 23], and also a particular backtracking line search proposed by Solodov and Svaiter in [17]. Recently, Amini and Kamandi [12] introduced a new line search processes some computational experiments on large-scale problems persuades us to take advantages of this line search, which requires $\alpha_k = \{\beta\rho^i : i = 0,1,2,\dots\}$ satisfies the following condition

$$-F(x_k + \alpha_k d_k)^T d_k \geq \frac{\sigma\,\alpha_k\,\|d_k\|^2}{1 + \|F(Z_k)\|}, \tag{9}$$

where σ is a positive constant, $\rho \in (0, 1)$ and β is an initial guess for α_k. It is easy to see that if d_k satisfies $F(x_k)^T d_k < 0$, then the inequality (9) holds for all sufficiently small $\alpha_k > 0$.

For any work that deals with large-scale nonlinear monotone equations, the low storage techniques will be very useful. So, because of their simplicity and low storage, many authors use the conjugate gradient methods combined with projection methods for solving large scale systems of monotone equations as well as optimization problems. Recently, the idea of conjugate gradient techniques has been widely developed for solving optimization problems and systems of nonlinear equations [3, 9, 14, 22, 24]. In 2009, Zhang and Zhou [22] proposed a spectral gradient projection method that is a combination of a modified spectral gradient method and Solodov and Svaiter's method. The attractive property of their method is that it can be applied for solving nonsmooth equations. Li and Li [14] in 2011, and Ahookhosh.et al [1] in 2013, also proposed some derivative-free projection approaches that use three-term CG directions appropriate to handle large-scale problems. The numerical results show that such methods are even appropriate to solve non-smooth monotone equations.

The theoretical and numerical results of the conjugate gradient methods motivating to be used to improve the projection method for solving systems of large-scale nonlinear monotone equations. Constructing and using conjugate gradient procedures in our algorithm gives us some advantages: these approaches are globally convergent without the differentiability assumptions; they decrease the computational cost of the algorithm and the number of iterations and function evaluations; and finally, they can improve the efficiency of the new algorithm. Based on the idea described by Yuan and Zhang [20], to generate a conjugate gradient direction, a new direction has been introducing as follows:

$$d_k = \begin{cases} -F_k + \beta_{k-1}^m \, d_{k-1} & \text{if } k \geq 1, \\ -F_k & \text{if } k = 0, \end{cases} \tag{10}$$

to use in a projection-based algorithm where

$$\beta_k^m = \text{Max}\left\{0, \; \frac{F_{k+1}^T y_k^m}{d_k^T y_k^m} - \frac{\mu \|y_k^m\|^2}{(d_k^T y_k^m)^2} F_{k+1}^T d_k\right\}, \tag{11}$$

with $\mu > \frac{1}{4}$ and

$$y_k^m = y_k + \frac{\max\{\rho_k, 0\}}{\|s_k\|^2} s_k, \tag{12}$$

where $y_k = F_{k+1} - F_k$ and $s_k = x_{k+1} - x_k$.

$$\rho_k = 2[\|F_k\| - \|F(x_k + \alpha_k d_k)\|] + (F(x_k + \alpha_k d_k) + F(x_k))^T s_k. \tag{13}$$

Now, we are ready to outline the new algorithm for nonlinear systems of monotone equations.

Algorithm 1: New Projection-Based Algorithm

Input: An initial point $x_0 \in R^n$, positive constants k_{max}, σ, ε, s and $\rho \in (0, 1)$.
Begin
    Set $k = 0$;
      $F_0 = F(x_0)$
      $d_0 = -F_0$.
      While $\|F_k\| > \varepsilon$.
            Step 1:
          Choose an initial step length β
          Set $\alpha_k = \beta$;
          Find the smallest index $i_k \in \{1,2,3,\ldots\}$ such that
$$-F(x_k + \alpha_k d_k)^T d_k \geq \frac{\sigma \alpha_k \|d_k\|^2}{1 + \|F(Z_k)\|},$$
          While $\alpha_k = \rho^{i_k} \alpha_k$;
            Set $z_k \leftarrow x_k + \alpha_k d_k$.
        End While
         Step 2: {New point calculation}
          If $\|F(z_k)\| \leq \varepsilon$ stop. Otherwise compute x_{k+1} by (8).
          Step 3: Compute the search direction d_k by (10).
        $F_{k+1} = F(x_{k+1})$;
          If $F_k^T d_k > -\varepsilon \|F_k\|^2$
            $d_k = -F_k$;
         End If
         $k \leftarrow k + 1$.
      End While
End

Remark 2.1 It is easy to see from Step 3 of Algorithm 1 that, for any k, $F_k^T d_k \leq -\varepsilon \|F_k\|^2$. Therefore, the direction defined by algorithm always satisfies the sufficient conditions.

3. Convergence properties

This section focuses on the global convergence of Algorithm 1. To reach this end, the following assumptions are needed.

Assumption 3.1
H1. The solution set of (1) is nonempty.
H2. $F(x)$ is a monotone and Lipschitz continuous function on R^n, i.e., there exists a positive constant $L > 0$ such that $\|F(x) - F(y)\| \leq L\|x - y\|, \forall x, y \in R^n$. Firstly, the projection operator is defined as a mapping from R^n to its nonempty closed convex set Ω [18] $P_\Omega[x] := \arg\min\{\|x - z\| \mid z \in \Omega\}, \forall x \in R^n$. In the sequel, we discuss some preliminary lemmas that show some important properties of orthogonal projection on a closed convex set and is necessary to analyze the convergence properties of Algorithm 1.

Lemma 3.1 [21] Let $\Omega \subseteq R^n$ be a nonempty closed convex set and $P_\Omega(x)$ be the projection of x onto Ω . For any $x, y \in R^n$, the following statements holds:

i) For any $i \in \Omega$, $\langle P_\Omega(x) - x, \ z - P_\Omega(x) \rangle \geq 0$.

ii) $\langle P_\Omega(x) - P_\Omega(y), \ x - y \rangle \geq 0$, and the inequality is strict when $P_\Omega(x) \neq P_\Omega(y)$.

iii) $\|P_\Omega(x) - P_\Omega(y)\| \leq \|x - y\|$.

Lemma 3.2 Let the assumptions H_1 and H_2 hold and the sequence $\{x_k\}$ is generated by Algorithm 1. For any x^* such that $F(x^*) = 0$, then

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2, \tag{14}$$

Also, the sequence $\{x_k\}$ is bounded. Furthermore, either the sequence $\{x_k\}$ is finite while the last iterate is a solution of (1) or the sequence is infinite and

$$\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0. \tag{15}$$

Moreover, $\{x_k\}$ converges to some solution of (1). Proof Theorem 2.1, in Solodov and Svaiter's [17], gives the result.

Lemma 3.3 Suppose that the assumption 3.1 holds and the sequences $\{x_k\}$ and $\{z_k\}$ are generated by algorithm 1, then

$$\alpha_k \geq \text{Min}\left\{\beta, \frac{\rho c\|F_k\|^2}{(L+\sigma)\|d_k\|^2}\right\}. \tag{16}$$

Proof By the line search rule (9), if $\alpha_k \neq \beta$ then $\widehat{\alpha}_k = \rho^{-1}\alpha_k$ does not satisfy (9). This means that $-F(x_k + \rho^{-1}\alpha_k d_k)^T d_k < \sigma\,\rho^{-1}\alpha_k\,\gamma_k\|d_k\|^2 \leq \sigma\,\widehat{\alpha}_k\|d_k\|^2$, where $\gamma_k = \frac{1}{1+\|F(z_k)\|}$ and $\widehat{z}_k = x_k + \widehat{\alpha}_k d_k$.

The Lipschitz continuity of F and (3) result in the following: $c\|F_k\|^2 \leq -F_k^T d_k = \left(F(\widehat{z}_k) - F(x_k)\right)^T d_k - F(\widehat{z}_k)^T d_k \leq \|F(\widehat{z}_k) - F(x_k)\|\|d_k\| + \sigma\,\widehat{\alpha}_k\|d_k\|^2 = \widehat{\alpha}_k(L+\sigma)\|d_k\|^2$. So, $\alpha_k \geq \frac{\rho c\|F_k\|^2}{(L+\sigma)\|d_k\|^2}$. This mean that (16) is correct, and the proof is complete.

As an obvious, it is resulted from Lemma 3.3 that the line search of algorithm 1 is well-defined.

Theorem 3.1 Suppose that the assumptions H_1 and H_2 hold and the sequence $\{x_k\}$ is generated by Algorithm 1, then

$$\lim_{k\to\infty} \|F_k\| = 0. \tag{17}$$

Proof The relation (8) and the line search (9) result that

$$\|x_{k+1} - x_k\|^2 \geq \frac{|F(z_k)^T(x_k - z_k)|}{\|F(z_k)\|} = \frac{-\alpha_k F(z_k)^T d_k}{\|F(z_k)\|} \geq \frac{\sigma\alpha_k^2\|d_k\|^2}{(1+\|F(z_k)\|)\|F(z_k)\|}. \tag{18}$$

The boundedness of the sequence $\{x_k\}$ along with the continuity of the mapping F result that $\|F(x_k)\|$ is bounded. So, there is a constant $M > 0$ such that $\|F(x_k)\| \leq M$. By the Lipschitz continuity of F, it can be concluded that

$$\|F(z_k)\| \leq \|F(z_k) - F(x_k)\| + \|F(x_k)\| \leq L(z_k - x_k) + M = L\alpha_k\|d_k\| + M. \quad (19)$$

The relation (18) together with (19) give $\|x_{k+1} - x_k\|^2 \geq \frac{\sigma\alpha_k^2\|d_k\|^2}{(1+L\alpha_k\|d_k\|+M)(L\alpha_k\|d_k\|+M)}$. So, $\lim\limits_{k\to\infty}\|x_{k+1} - x_k\|^2 \geq \lim\limits_{k\to\infty}\left(\frac{\sigma\alpha_k^2\|d_k\|^2}{(1+L\alpha_k\|d_k\|+M)(L\alpha_k\|d_k\|+M)}\right)$. It is easy to conclude that

$$\lim_{k\to\infty}\alpha_k\|d_k\| = 0. \quad (20)$$

Now, by using Cauchy Schwartz inequality along with (3), we get $c\|F_k\|^2 \leq -F_k^T d_k \leq \|F_k\|\|d_k\|$. So

$$c\|F_k\| \leq \|d_k\|. \quad (21)$$

On the other hand, multiplying (16) by $\|d_k\|^2$ results that

$$\alpha_k\|d_k\|^2 \geq \text{Min}\left\{\beta\|d_k\|^2, \frac{\rho c\|F_k\|^2}{L+\sigma}\right\}. \quad (22)$$

From (21) and (22), it can be resulted that

$$\sigma_0\|F_k\|^2 \leq \alpha_k\|d_k\|^2, \quad (23)$$

where $\sigma_0 = \text{Min}\left\{\beta c^2, \frac{\rho c}{L+\sigma}\right\}$. The relation (20) together with (23) conclude that $\lim\limits_{k\to\infty}\|F_k\| = 0$.

4. Numerical results

In this section, some numerical experiments are reported to compare the performance of the new method along with the following three algorithms:

DFPB1: This method is coming from Ahookhosh et al. [1], it uses a three-term PRP-based conjugate gradient direction.

MPRP: This method from Li and Li [14] uses a modified PRP- based direction.

MPRP*: This method takes the direction (2.8) in [14] with the line search (9) in this paper.

The experiments were run on a PC with CPU 2.20 GHz and 8 GB RAM. All of the codes were written in MATLAB R2012b programming environment. The running of the codes will check if the provided data for problems in all algorithms converges to the identical points. All of the algorithms terminate whenever $\|F_k\| \leq 10^{-4}$ or $\|F(z_k)\| \leq 10^{-4}$, or the total number of iterates exceeds 500000. In all of the algorithms, the parameters are specified as follows: $\mu = 0.26, \rho = 0.7, \sigma = 0.3, \epsilon = 10^{-4}$, and the initial adaptive step length is computed by $\beta = \frac{F_k^T d_k}{(F(x_k+td_k)-F_k)^T d_k/t}$, where $t = 10^{-8}$.

The performances of these methods are compared with respect to the number of iterations N_i, the number of function evaluations N_f and CPU time. In order to compare these

algorithms, some famous test problems in [1] and [14] are used where the dimensions are trapped between 5000-50000 for the following initial points:

$x_0 = (10,10,...,10)^T,$ $\quad x_1 = (-10,-10,...,-10)^T,$ $\quad x_2 = (1,1,...,1)^T,$

$x_3 = (-1,-1,...,-1)^T,$ $\quad x_4 = \left(1,\frac{1}{2},\frac{2}{3},,...,\frac{1}{n}\right)^T,$ $\quad x_5 = (0.1,0.1,...,0.1)^T,$

$x_6 = (\frac{1}{n},\frac{2}{n},...,1)^T,$ $\quad x_7 = (1-\frac{1}{n},1-\frac{2}{n},...,0)^T.$

The numerical results of running the algorithms are listed in Table 1 and 2. Table 1 contains the number of iterations and the evaluated functions while Table 2 contains the numerical results of CPU time.

To have a comprehensive comparison among the reported results, the performance profile introduced by Dolan and More [10] is used as a well-known tool for evaluating and comparing the algorithms and presented a wealth information, including efficiency and robustness. The proposed performance profiles of algorithms (New, DFPB1, MPRP and MPRP*) are exploited in the sense of the number of iterates, the number of function evaluations and CPU time in Figures 1-3 respectively. From these figures, it is easy to see that the new proposed method obtains the most wins on approximately 80%, 80% and 68% of problems respectively, and this clearly shows the superiority of the new method compared with the other methods for solving large-scale nonlinear monotone equations.
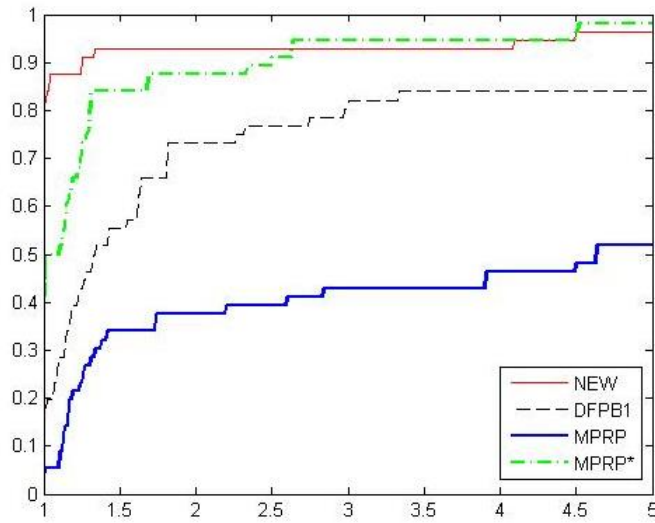

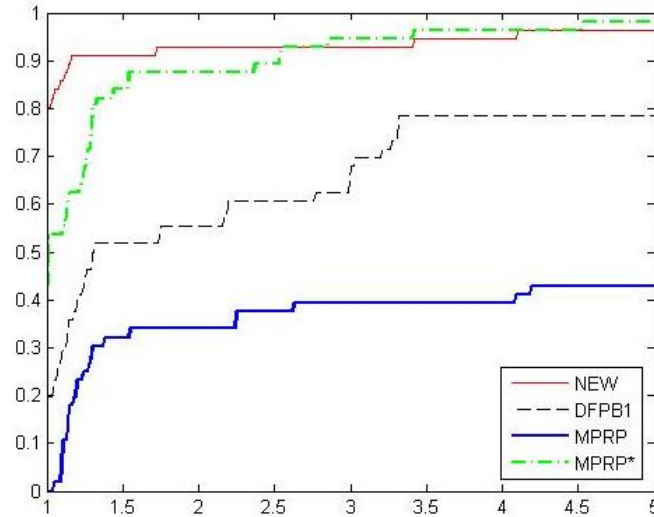Figure 1: *Performance profile of the iteration number*

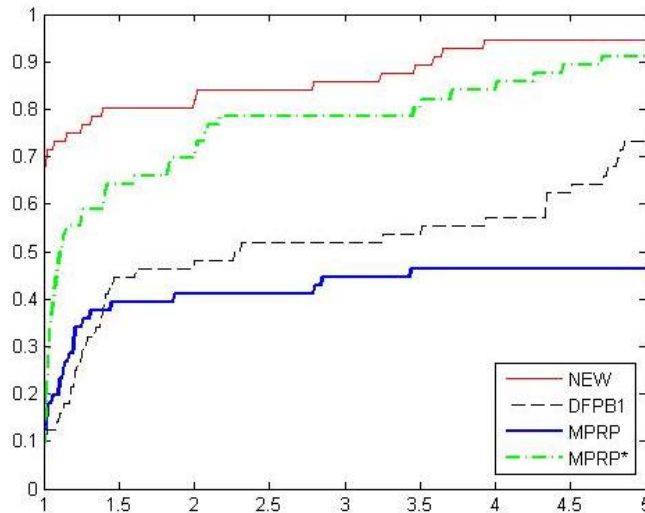Figure 2: *Performance profile of the function evaluations*

Figure 3: *Performance profile of CPU time*

From these results, it can be concluded that imposing the proposed conjugate gradient direction to projection algorithms can improve the promising behavior of algorithms for solving large-scale nonlinear monotone equations. Also, it can be concluded that Algorithm 1 is able to compute an accurate solution of problem (1) without employing any merit functions.

5. Conclusions

The presented work proposes a new projection method using a modified conjugate gradient direction for solving a system of large-scale nonlinear monotone equations. The projection-based algorithms are suitable for solving monotone equations because it belongs to the class of derivative-free function-value based methods, and it doesn't use any merit function and derivatives. Also, this method enables a simple globalization. The low memory requirement (i.e., the low computational cost) of derivative-free conjugate gradient strategies makes the

proposed algorithms prepare to face the large-scale problems. The global convergence of the proposed algorithm is proved under standard assumptions. Preliminary numerical experiments indicated that the proposed algorithm is very efficient, and it is competitive with some famous conjugate gradient methods.

References

[1] Ahookhosh, M., Amini, K. and Bahrami, S. (2013). Two derivative-free projection approaches for systems of large-scale nonlinear monotone equations. Numerical Algorithms, 64(1), 21–42.

[2] Amini, K., Kamandi, A. and Bahrami, S. (2015). A double-projection-based algorithm for large-scale nonlinear system of monotone equations. Numerical Algorithms, 68(2), 213–228.

[3] Amini, K., Shiker, M. A. K. and Kimiaei, M. (2016). A line search trust-region algorithm with nonmonotone adaptive radius for a system of nonlinear equations. Journal of operation research, 14(2), 133–152.

[4] Bellavia, S. and Morini, B. (2001). A globally convergent Newton-GMRES subspace method for systems of nonlinear equations. SIAM Journal of Scientific Computer, 23(3), 940–960.

[5] Birgin, E. G., Krejic, N. K. and Martínez, J. M. (2003). Globally convergent inexact quasi-Newton methods for solving nonlinear systems. Numerical Algorithms, 32(2-4), 249–260.

[6] Brown, P. N. and Saad, Y. (1994). Convergence theory of nonlinear Newton-Krylov algorithms. SIAM Journal on Optimization, 4(2), 297–330.

[7] Broyden, C. G. (1965). A class of methods for solving nonlinear simultaneous equations. Mathematical of Computation, 19(92), 577–593.

[8] Cheng, W. (2009). A PRP type method for systems of monotone equations. Mathematical and Computer Modeling, 50(1-2), 15–20.

[9] Dai, Y. (2003). Convergence properties of the BFGS algorithm. SIAM Journal on Optimization, 13(3), 693–701.

[10] Dolan, E. D. and More, J. J. (2002). Benchmarking optimization software with performance profiles. Mathematical Programming, 91(2), 201–213.

[11] Hager, W. W. and Zhang, H. (2005). A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM Journal on Optimization, 16(1), 170–192.

[12] Keyvan, A. and Ahmad, K. (2015). A new line search strategy for finding separating hyperplane in projection-based methods. Numerical Algorithms, 70(3), 559–570.

[13] La Cruz, W. and Raydan, M. (2003). Nonmonotone spectral methods for large-scale nonlinear systems. Optimization Methods and Software, 18(5), 583–599.

[14] Li, Q. and Li, D. H. (2011). A class of derivative-free methods for large-scale nonlinear monotone equations. IMA Journal of Numerical Analysis, 31(4), 1625–1635.

[15] Li, M. (2014). An Liu-Storey-Type Method for Solving Large-Scale Nonlinear Monotone Equations. Numerical Functional Analysis and Optimization, 35(3), 310–322.

[16] Ortega, J. M. and Rheinboldt, W. C. (1970). Iterative solution of nonlinear equations in several variables. New York: Academic Press.

[17] Solodov, M. and Svaiter, B. F. (1998). A globally convergent inexact Newton method for systems of monotone equations. In Fukushima, M. and Qi, L. (Eds.). Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Applied Optimization (pp. 355–369). Kluwer academic publishers.

[18] Wang, Y. J., Xiu, N. H. and Zhang, J. Z. (2003). Modified extra gradient method for variational inequalities and verification of solution existence. Journal of Optimization Theory and Applications, 119(1), 167–183.

[19] Yan, Q. R., Peng, X. Z. and Li, D. H. (2010). A globally convergent derivative-free method for solving large-scale nonlinear monotone equations. Journal of Computational and Applied Mathematics, 234(3), 649–657.

[20] Yuan, G. and Zhang, M. (2013). A modified Hestenes-Stiefel conjugate gradient algorithm for large- scale optimization. Numerical Functional Analysis and Optimization, 34(8), 914–937.

[21] Zarantonello, E. H. (1971). Projections on convex sets in Hilbert space and spectral theory. In Zarantonello, E. H. (Ed.). Contributions to Nonlinear Functional Analysis. New York: Academic Press.

[22] Zhang, L., Zhou, W. and Li, D. H. (2009). Some descent three-term conjugate gradient methods and their global convergence. Optimization Methods and Software, 22(4), 697–711.

[23] Zhang, L. and Zhou, W. J. (2006). Spectral gradient projection method for solving nonlinear monotone equations. Journal of Computational and Applied Mathematics, 196(2), 478–484.

[24] Zhou, G. and Toh, K. C. (2005). Superlinear convergence of a Newton-type algorithm for monotone equations. Journal of Optimization Theory and Applications, 125(1), 205–221.