Journal of Industrial and Management Optimization
Vol. 20, No. 1, January 2024, pp. 170-187
doi:10.3934/jimo.2023073

TWO DIAGONAL CONJUGATE GRADIENT LIKE METHODS FOR UNCONSTRAINED OPTIMIZATION

Hassan Mohammad*[1], Ibrahim Mohammed Sulaiman[2,3] and Mustafa Mamat[4]

[1]Numerical Optimization Research Group, Department of Mathematical Sciences, Faculty of Physical Sciences, Bayero University, Kano 700241, Nigeria

[2]School of Quantitative Science, Universiti Utara Malaysia (UUM), 06010, Sintok, Kedah, Malaysia

[3]Institute of Strategic Industrial Decision Modelling, UUM, 06010, Sintok, Kedah, Malaysia

[4]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terennganu, Malaysia

(Communicated by Xin Liu)

Abstract. In this paper, we studied a two member family of diagonal conjugate gradient like methods for unconstrained minimization problems. The proposed search directions of the methods are obtained by incorporating a diagonal Hessian approximation approach with the modifications of the Polak–Ribière–Polyak and Liu-Storey parameters. The resulting directions guarantees the sufficient decrease of the objective function free from any line search. The global convergence of the proposed methods using both the Wolfe and Armijo-type line search conditions without the convexity assumption on the function is provided. Furthermore, the R-linear convergence rate of the methods are analyzed under Wolfe line search condition. Numerical experiments demonstrated the performance of the method on general unconstrained minimization problems and image restoration problem.

1. **Introduction.** Two diagonal conjugate gradient like (`dcg`) approaches for the solution of unconstrained minimization problem are proposed. Nonlinear conjugate gradient (CG) methods are class of matrix-free approach for solving large-scale unconstrained minimization problem. Among the classical nonlinear CG parameters are the Hestenes-Stiefel (HS) [18], Fletcher-Reeves (FR) [15], Polak-Ribiére-Polyak (PRP) [25, 26], Liu-Storey (LS) [21], Dai-Yuan (DY) [10], and conjugate descent (CD) [14]. The HS, PRP, and LS methods are characterized by efficient numerical performance for unconstrained optimization problems, however, some of these methods fails to converge under certain line search procedure. On the other hand, the FR, DY, and CD algorithms possess good convergence properties, but, perform poorly on unconstrained optimization problems due to jamming phenomena [17].

2020 *Mathematics Subject Classification.* Primary: 49M37, 65K05, 90C26; Secondary: 90C52.
Key words and phrases. Unconstrained optimization, conjugate gradient methods, global convergence, numerical results.

*Corresponding author: Hassan Mohammad.

Recently, numerous studies have investigated different modifications of the classical methods. Yuan et al. [32] studied the performance of the PRP method using a modified weak Wolfe line search. Just recently, Yuan et al. [33] presented a modified PRP CG method for unconstrained optimization and its application in engineering Muskingum model and image restoration problems. Other modifications of these methods are presented in [1, 19, 22, 34, 36]. The study on the theoretical analysis of the CG methods is available in the following articles [3, 8–10, 16, 27]. Interested reader may also refer to the [6, 11, 17] for excellent reviews on this topic.

Dong et al. [13] presented a general form of the descent CG methods, and also classify the nonlinear CG methods into two categories. The first category consist of PRP, HS and LS CG methods, while the other category consist of FR, CD and DY CG methods. Lately, Mohammad and Santos [23] devised a diagonal Hessian approximation approach for solving the nonlinear least-squares (`NLS`) problems, a special kind of unconstrained minimization problem. Motivated by the development of the diagonal Hessian approach for general unconstrained minimization problem, this paper incorporate the first category of the conjugate gradient methods of Dong et al. [13] with the diagonal approximation approach presented in [23], to study the diagonal CG-like methods for solving unconstrained minimization problems. This approach may be considered as a preconditioned technique for nonlinear conjugate gradient method, where the preconditioner in this case is a safeguarded diagonal matrix. The contributions of this paper can be summarized as follows:

- Proposing two new diagonal conjugate gradient methods for solving large-scale unconstrained minimization problems.
- Combining the PRP and LS nonlinear conjugate gradient methods with a diagonal Hessian approximation approach to form the diagonal conjugate gradient methods.
- Showing that the proposed methods generate descent search directions irrespective of line search restrictions.
- Showing that if the exact line search is used, the proposed direction satisfy the conjugacy condition provided the diagonal matrix associated with the direction is a constant multiple of identity.
- Establishing the global convergence of the proposed methods using both the conditions of Wolfe and Armijo-type line search.
- Presenting the linear rate of convergence of the proposed methods under the Wolfe line search assuming convexity of the objective function.

The outstanding part of this paper is organized as follows: In Sec. 2 we discussed the derivation process of the proposed methods. In section 3, we analyze the global convergence of the proposed methods using both Wolfe and Armijo-type line searches. In section 4, we analyze the convergence rate of the methods using Wolfe line search. In section 5, we present some numerical results including the application of the proposed methods on image restoration problems. Conclusions are given in the last section. Throughout this paper, unless stated otherwise, the following notations $\|\cdot\|$ would denote the induced matrix norm and the vector Euclidean norm.

2. **Algorithm.**

Consider the nonlinear unconstrained optimization problem of the form

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

with f being a smooth mapping. Given an initial vector $x_0 \in \mathbb{R}^n$, the conjugate gradient (CG) methods produce a sequence of iterates $\{x_k\}$ and $\{d_k\}$ via

$$x_{k+1} := x_k + \alpha_k d_k, \tag{2}$$

$$d_k := \begin{cases} -g_k, & k = 0; \\ -g_k + \beta_k d_{k-1}, & k \geq 1, \end{cases} \tag{3}$$

where α_k is a steplength computed by a suitable line search procedure, $g_k = \nabla f(x_k)$ is the gradient of f at x_k, and β_k is a scalar usually referred as the CG parameter. Usually, the direction d_k given in (3) is required to satisfy the sufficient descent condition, that is there exists $c > 0$ such that

$$g_k^T d_k \leq -c\|g_k\|^2.$$

In addition, the direction is required to satisfy the conjugacy condition

$$d_k^T y_{k-1} = 0, \tag{4}$$

or the Dai-Liao [7] generalized conjugacy condition

$$d_k^T y_{k-1} = -r g_k^T s_{k-1},$$

where $y_{k-1} = g_k - g_{k-1}$, $s_{k-1} = x_k - x_{k-1}$ and $r \geq 0$ is a scalar.

In an attempt to develop nonlinear CG methods that possess the descent property without requiring any line search condition, Dong et al. [13] proposed a general form of CG-like parameter β_k with the generated search direction d_k satisfying

$$g_k^T d_k \leq -\left(1 - \frac{1}{4t}\right)\|g_k\|^2, \tag{5}$$

where t is a positive scalar greater than $\frac{1}{4}$. The generalized CG-like parameter by Dong et al. is given as

$$\bar{\beta}_k := \max\left\{0,\ \beta_k - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\beta_k^2\right\}, \tag{6}$$

where β_k is any choice of the CG parameter. We stress that by this choice of $\bar{\beta}_k$, the sufficient descent property is satisfied irrespective of the given parameter β_k. Indeed, if $\max\left\{0,\ \beta_k - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\beta_k^2\right\} \neq 0$, then

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \bar{\beta}_k d_{k-1} \\ &= -\|g_k\|^2 + \left(\beta_k - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\beta_k^2\right) g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{1}{4t}\|g_k\|^2 + \left(-\frac{1}{4t}\|g_k\|^2 + \beta_k g_k^T d_{k-1} - t\frac{(g_k^T d_{k-1})^2}{\|g_k\|^2}\beta_k^2\right) \end{aligned} \tag{7}$$

By applying the inequality $2xy \leq x^2 + y^2$ to the fourth term of (7) with $x = \frac{1}{2\sqrt{t}}\|g_k\|$ and $y = \sqrt{t}\beta_k \frac{g_k^T d_{k-1}}{\|g_k\|}$, we get (5).

In this work, we follow the idea of Dong et al. [13] which considers developing CG-like methods that possess the sufficient descent property and proposed alternative diagonal CG-like methods based on the generalized CG parameter (6), incorporating a diagonal matrix to the term $-g_k$ in (3). Our proposed direction is obtain without

necessary requiring the conjugacy condition (4). The proposed diagonal CG-like search direction for solving (1) is as follows:

$$d_k := \begin{cases} -Q_0 g_0, \text{ if } k = 0 \\ \\ -Q_k g_k + \bar{\beta}_k d_{k-1}, \text{ if } k \geq 1, \end{cases} \tag{8}$$

with $Q_0 = I_n$, Q_k is a positive definite diagonal matrix given by

$$Q_k := \text{diag}\Big(q_k^1,\ q_k^2,\ \ldots,\ q_k^n\Big), \tag{9}$$

$$q_k^i := \begin{cases} \frac{s_{k-1}^i}{y_{k-1}^i}, & \text{if } \frac{s_{k-1}^i}{y_{k-1}^i} > 0 \\ \\ 1, & \text{if } \frac{s_{k-1}^i}{y_{k-1}^i} \leq 0 \text{ or } y_{k-1}^i = 0. \end{cases} \tag{10}$$

here, I_n is a diagonal matrix of order n, and $i = 1, 2, \ldots n$.

Furthermore, to ensure the boundedness of the diagonal components of Q_k, we project its components into the given interval $[\ell,\ u]$, with $0 < \ell < 1 \leq u$. Therefore, the ith component of Q_k is

$$q_k^i := \begin{cases} \max\left\{\min\left\{\frac{s_{k-1}^i}{y_{k-1}^i},\ u\right\},\ \ell\right\}, & \text{if } y_{k-1} \neq 0 \\ \\ 1, & \text{if } y_{k-1}^i = 0. \end{cases}$$

For each i and k, the sequence $\{q_k^i\}$ of the diagonal matrix Q_k is uniformly bounded. Clearly, it follows that

$$\ell \leq q_k^i \leq u\ \forall i,\ \forall k. \tag{11}$$

The CG-like parameter $\bar{\beta}_k$, is given by any one of β_k^{DPRP} or β_k^{DLS} modified as follows:

$$\beta_k^{DPRP} := \max\left\{0, \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\left(\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}\right)^2\right\}. \tag{12}$$

$$\beta_k^{DLS} := \max\left\{0, \frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}} - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\left(\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}}\right)^2\right\}. \tag{13}$$

The lemma that follows show the direction d_k possesses the sufficient descent properties irrespective of the CG parameter β_k and without any line search condition.

Lemma 2.1. *Let the sequence $\{x_k\}$, $\{d_k\}$ be obtained via (2) and (8). Then, there exists $1 \leq u < 4t$ such that*

$$g_k^T d_k \leq -\left(\frac{1}{u} - \frac{1}{4t}\right)\|g_k\|^2,\ \forall k \geq 0. \tag{14}$$

Proof. For $k = 0$, it follows that

$$g_0^T d_0 = -g_0^T Q_0 g_0 = -\|g_0\|^2.$$

For $k \geq 1$, we consider two cases:

Case (i) If $\max\left\{0,\ \beta_k - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\beta_k^2\right\} = 0$ then

$$\begin{aligned} g_k^T d_k &= -g_k^T Q_k g_k \\ &= g_k^T \operatorname{diag}\left(-\frac{1}{q_k^1}, -\frac{1}{q_k^2}, ..., -\frac{1}{q_k^n}\right) g_k \\ &\leq g_k^T \operatorname{diag}\left(-\frac{1}{u}, -\frac{1}{u}, ..., -\frac{1}{u}\right) g_k \\ &= -\frac{1}{u}\|g_k\|^2. \end{aligned}$$

with the first inequality following from (11).

Case (ii) If $\max\left\{0,\ \beta_k - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\beta_k^2\right\} = \beta_k - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\beta_k^2$, then

$$\begin{aligned} g_k^T d_k &= g_k^T(-Q_k g_k + \bar{\beta}_k d_{k-1}) \\ &= g_k^T \operatorname{diag}\left(-\frac{1}{q_k^1}, -\frac{1}{q_k^2}, ..., -\frac{1}{q_k^n}\right) g_k + \left(\beta_k - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\beta_k^2\right) g_k^T d_{k-1} \\ &\leq -\frac{1}{u}\|g_k\|^2 + \beta_k g_k^T d_{k-1} - t\frac{(g_k^T d_{k-1})^2}{\|g_k\|^2}\beta_k^2 \\ &= \left(\frac{1}{u} - \frac{1}{4t}\right)\|g_k\|^2 - \frac{1}{4t}\|g_k\|^2 + \beta_k g_k^T d_{k-1} - t\frac{(g_k^T d_{k-1})^2}{\|g_k\|^2}\beta_k^2 \end{aligned} \tag{15}$$

By applying the inequality $2pq \leq p^2 + q^2$ to the third term of (15) with $p = \frac{1}{2\sqrt{t}}\|g_k\|$ and $q = \sqrt{t}\beta_k\frac{g_k^T d_{k-1}}{\|g_k\|}$, we have (14). □

Remark 2.2. We remark that the direction d_k described by (8) satisfy the descent condition (14) irrespective of the choice β_k parameter. Therefore, by choosing (12) and (13) as the β_k parameters, the condition (14) still holds.

The step-by-step procedure for the implementation of the diagonal CG-like methods is given below.

Remark 2.3.
The update rule for d_k at Step 2 of Algorithm 1, is given specifically to ensure global convergence of the methods.

Lemma 2.4. *Let the sequence $\{x_k\}$ and $\{d_k\}$ be defined by Algorithm 1. If there is $\varepsilon > 0$ such that $\|g(x_{k-1})\| \geq \varepsilon$ for $k \geq 1$, then there is $\varpi > 0$ such that*

$$\|d_k\| \leq \varpi\|g_k\| \quad \text{for all } k \in \mathbb{N} \cup \{0\}. \tag{16}$$

Proof. `Case 1:`
If $|g_k^T y_{k-1}|\|d_{k-1}\| \geq \eta\|g_k\|$, then

$$\begin{aligned} \|d_k\| &= \|Q_k g_k\| \\ &\leq \|Q_k\|\|g_k\| \leq \frac{\sqrt{n}}{\ell}\|g_k\|. \end{aligned}$$

`Case 2:` If $|g_k^T y_{k-1}|\|d_{k-1}\| < \eta\|g_k\|$, then it follows that

$$\begin{aligned} \|d_k\| &= \|-Q_k g_k + \bar{\beta}_k d_{k-1}\|\| \\ &\leq \|Q_k\|\|g_k\| + |\bar{\beta}_k|\|d_{k-1}\| \end{aligned} \tag{17}$$

Algorithm 1: Family of diagonal CG-like algorithms (`dcg`)

Input : Given $x_0 \in \mathbb{R}^n$, $0 < \ell \leq 1 \leq u,\ t > u/4$, and tol, $\eta > 0$.

Step 1 : **if** $\|g_k\| \leq tol$, **then**
stop.
end
Step 2 : **if** $k = 0$, **then**
set $d_k := -g_k$;
else
Calculate

$$d_k := \begin{cases} -Q_k g_k, & \text{if } |g_k^T y_{k-1}|\|d_{k-1}\| \geq \eta \|g_k\|, \\ -Q_k g_k + \bar{\beta}_k d_{k-1}, & \text{if } |g_k^T y_{k-1}|\|d_{k-1}\| < \eta \|g_k\|, \end{cases}$$

where Q_k is given by equations (9)-(10), $\bar{\beta}_k$ is computed using (6) with the choice of β_k from either (12) or (13), and $y_{k-1} = g_k - g_{k-1}$.
end
Step 3 : Compute the steplength α_k by a suitable line search.
Step 4 : Obtain the next iterate via $x_{k+1} = x_k + \alpha_k d_k$.
Step 5 : Set $k := k + 1$ and return to Step 1.

We first estimate $|\bar{\beta}_k|$. If $\bar{\beta}_k$ given by (12) is nonnegative, then we have

$$\begin{aligned} |\bar{\beta}_k| = |\beta_k^{DPRP}| &= \frac{|g_k^T y_{k-1}|}{\|g_{k-1}\|^2} + t\frac{|g_k^T d_{k-1}|}{\|g_k\|^2}\frac{(g_k^T y_{k-1})^2}{\|g_{k-1}\|^4} \\ &\leq \eta \frac{\|g_k\|}{\|d_{k-1}\|}\frac{1}{\|g_{k-1}\|^2} + t\frac{\|g_k\|\|d_{k-1}\|}{\|g_k\|^2}\eta^2 \frac{\|g_k\|^2}{\|d_{k-1}\|^2\|g_{k-1}\|^4} \\ &= \eta \frac{\|g_k\|}{\|d_{k-1}\|}\frac{1}{\|g_{k-1}\|^2} + t\frac{\|d_{k-1}\|}{\|g_k\|}\left(\frac{\eta\|g_k\|}{\|d_{k-1}\|\|g_{k-1}\|^2}\right)^2 \\ &\leq \left(\frac{\eta}{\varepsilon^2} + t\frac{\eta^2}{\varepsilon^4}\right)\frac{\|g_k\|}{\|d_{k-1}\|}. \end{aligned}$$

Also, if $\bar{\beta}_k$ is given by (13) is nonnegative, then

$$|\bar{\beta}_k| = |\beta_k^{DLS}| = \frac{|g_k^T y_{k-1}|}{|g_{k-1}^T d_{k-1}|} + t\frac{|g_k^T d_{k-1}|}{\|g_k\|^2}\left(\frac{(g_k^T y_{k-1})}{g_{k-1}^T d_{k-1}}\right)^2.$$

Observe that $1 \leq u$, $t > u/4$ and $0 < \frac{1}{u} - \frac{1}{4t} < 1$. We get from (14) that

$$\begin{aligned} |\bar{\beta}_k| &\leq \left(\frac{1}{u} - \frac{1}{4t}\right)^{-1}\frac{|g_k^T y_{k-1}|}{\|g_{k-1}\|^2} + t\left(\frac{1}{u} - \frac{1}{4t}\right)^{-2}\frac{|g_k^T d_{k-1}|}{\|g_k\|^2}\left(\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}\right)^2 \\ &\leq \left(\frac{1}{u} - \frac{1}{4t}\right)^{-2}\frac{|g_k^T y_{k-1}|}{\|g_{k-1}\|^2} + t\left(\frac{1}{u} - \frac{1}{4t}\right)^{-2}\frac{|g_k^T d_{k-1}|}{\|g_k\|^2}\left(\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}\right)^2 \\ &\leq \left(\frac{1}{u} - \frac{1}{4t}\right)^{-2}\left[\frac{|g_k^T y_{k-1}|}{\|g_{k-1}\|^2} + t\frac{|g_k^T d_{k-1}|}{\|g_k\|^2}\left(\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}\right)^2\right] \\ &= \left(\frac{1}{u} - \frac{1}{4t}\right)^{-2}|\beta_k^{DPRP}| \end{aligned}$$

$$\leq \left(\frac{1}{u}-\frac{1}{4t}\right)^{-2}\left(\frac{\eta}{\varepsilon^2}+t\frac{\eta^2}{\varepsilon^4}\right)\frac{\|g_k\|}{\|d_{k-1}\|}.$$

Let $N := \min\left\{\left(\frac{1}{u}-\frac{1}{4t}\right)^{-2}\left(\frac{\eta}{\varepsilon^2}+t\frac{\eta^2}{\varepsilon^4}\right),\ \left(\frac{\eta}{\varepsilon^2}+t\frac{\eta^2}{\varepsilon^4}\right)\right\}$, then we have for any choice of $\bar{\beta}_k$ from (12) and (13)

$$|\bar{\beta}_k| \leq N\frac{\|g_k\|}{\|d_{k-1}\|}, \quad \text{for all } k. \tag{18}$$

Substituting (18) into (17), we get

$$\begin{aligned}\|d_k\| &\leq \|Q_k\|\|g_k\| + |\bar{\beta}_k|\|d_{k-1}\| \\ &\leq \frac{\sqrt{n}}{\ell}\|g_k\| + N\|g_k\| \\ &= \left(\frac{\sqrt{n}}{\ell}+N\right)\|g_k\|.\end{aligned}$$

Therefore, if we let $\varpi := \frac{\sqrt{n}}{\ell} = \min\left\{\frac{\sqrt{n}}{\ell},\ \frac{\sqrt{n}}{\ell}+N\right\}$, equation (16) will immediately follows. □

Remark 2.5. From the proofs of Lemma 2.1 and Lemma 2.4, it is easy to see that if we replace $\bar{\beta}_k$ with

$$\beta_k^{DHS} := \max\left\{0, \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - t\frac{g_k^T d_{k-1}}{\|g_k\|^2}\left(\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}\right)^2\right\},$$

defined in [13], the results still hold. Thus, the subsequent convergence result will hold for the three modified CG parameters $\beta_k^{DPRP}, \beta_k^{DLS}$ and β_k^{DHS}. To obtain the convergence results for other CG parameters, such as FR, CD and DY, another update rule for the search direction is necessary (see [13] for details).

Remark 2.6. In general, the proposed diagonal conjugate gradient like direction defined in (8) does not satisfy the Dai-Liao conjugacy condition (4). Consequently, we cannot assure its quadratic termination property (see Theorem 5.1 of Nocedal and Wright [24]). However, if the exact line search is used (i.e. $g_k^T d_{k-1} = 0$ for all k) and $\bar{\beta}_k > 0$, then

$$\begin{aligned}d_k^T y_{k-1} &= -(Q_k g_k)^T y_{k-1} + \beta_k d_{k-1}^T y_{k-1} \\ &= -g_k^T Q_k y_{k-1} + \beta_k(d_{k-1}^T g_k - d_{k-1}^T g_{k-1}) \\ &= -g_k^T Q_k y_{k-1} - \beta_k d_{k-1}^T g_{k-1}.\end{aligned} \tag{19}$$

If $\beta_k = \beta_k^{PRP}$ in (19) then

$$d_k^T y_{k-1} = -g_k^T\left(Q_k - \frac{d_{k-1}^T g_{k-1}}{\|g_{k-1}\|}I\right)y_{k-1}.$$

If $\beta_k = \beta_k^{LS}$ or $\beta_k = \beta_k^{HS}$ in (19) then

$$d_k^T y_{k-1} = -g_k^T\left(Q_k - I\right)y_{k-1}.$$

Therefore, the proposed direction satisfy the conjugacy condition (4) if the diagonal matrix Q_k is a scalar multiple of identity.

The subsequent section is about the convergence analysis of the new methods under the conditions of both the standard Wolfe and the Armijo-type line search procedures. Throughout the article, we used the suppositions given below:

Supposition 2.7. The level set $\mathcal{U}(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded.

Supposition 2.8. In some vicinity of $\mathcal{U}$, the gradient g of a continuously differentiable function f is Lipschitz continuous. This implies, we can find $L > 0$ satisfying

$$\|g(x) - g(y)\| \leq L\|x - y\|, \text{ for all } x, y \text{ in the neighborhood of the level set } \mathcal{U}.$$

Based on suppositions 2.7 and 2.8, we can conclude that $\forall x \in \mathcal{U}$, $\exists\ m_1,\ m_2 > 0$ for which $\|x\| \leq m_1$ and $\|g(x)\| \leq m_2$. Moreover, by the continuity of f and the boundedness of $\mathcal{U}$, the sequence $\{f(x_k)\}$ is bounded below.

3. **Convergence results.** The Wolfe line search determine α_k by conditions defined below

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta_1 \alpha_k g_k^T d_k, \tag{20}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \delta_2 g_k^T d_k, \tag{21}$$

with $0 < \delta_1 \leq \delta_2 < 1$.

The result of the following lemma is called *Zoutendijk condition*, and its proof can be found in [29].

Lemma 3.1. *Let the suppositions 2.7 - 2.8 hold. If the Wolfe conditions* (20)-(21) *are satisfied then*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

Theorem 3.2. *Let $\{x_k\}$ be a sequence that follows from Algorithm 1 with the Wolfe line search conditions* (20)-(21). *Then*

$$\lim_{k \to \infty} \|g_k\| = 0. \tag{22}$$

Proof. It can be obtain from (14), (16) and Lemma 3.1 that

$$\sum_{k=0}^{\infty} \varpi^{-2} \|g_k\|^2 \leq \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \leq c^{-2} \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

This implies that

$$\sum_{k=0}^{\infty} \|g_k\|^2 < \infty.$$

Hence, equation (22) holds. □

Next, we analyze the convergence of the sequence of iterates generated by Algorithm 1 using Armijo-type line search condition. The Armijo-type line search determine $\alpha_k = \gamma\rho^i$, with i denoting least non-negative integer that satisfy

$$f(x_k + \alpha_k d_k) \leq (1 + \tau_k) f(x_k) - \sigma\alpha_k^2 \|d_k\|^2, \tag{23}$$

where $\sigma > 0,\ \gamma > 0\ \rho \in (0,\ 1)$ and τ_k is nonnegative sequence satisfying $\sum_{k=0}^{\infty} \tau_k < \infty$.

Theorem 3.3. *Suppose that suppositions 2.7 and 2.8 are satisfied. If $\{x_k\}$ is a sequence obtained via Algorithm 1 using the Armijo-type condition (23) then*

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

Proof. The proof is omitted because of close similarity with the proof of Theorem 3.2 of [35]. □

4. **Linear convergence rate.** In this section, we analyze the linear convergence rate of the Algorithm 1 for twice continuously differentiable and strongly convex function. That is there are some positive constants $m \leq M$ such that

$$m\|h\|^2 \leq h^T \nabla^2 f(x) h \leq M\|h\|^2, \quad \text{for all } x,\ h \in \mathbb{R}^n,$$

where $\nabla^2 f(x)$ denotes the Hessian of f at x. The analysis in section is restricted to the Wolfe line search condition, that is we investigate the linear convergence rate of the Algorithm 1 using the Wolfe line search condition only.

See Lemma 2.2.7 of [28] for the proof of the subsequent lemma.

Lemma 4.1. *Suppose f is twice continuously differentiable and strongly convex function. Then*

$$\frac{m}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{M}{2}\|x - x^*\|^2, \tag{24}$$

$$\|g(x)\| \geq m\|x - x^*\|, \tag{25}$$

where x^* is the minimizer of (1).

To prove the convergence rate of the Algorithm 1 we first require the following lemma, which guaranty that the stepsize α_k obtained using the Wolfe line search condition (21) is bounded away from zero.

Lemma 4.2. *Suppose f is twice continuously differentiable and the sequence $\{x_k\}$ be generated by Algorithm 1 with the Wolfe line search condition. Then*

$$\alpha_k \geq \frac{(1 - \delta_2)(\frac{1}{4} - \frac{1}{4t})}{L\varpi^2}. \tag{26}$$

Proof. Combining the Wolfe line search condition (21) together with (14) and supposition 2.8 we get

$$(1 - \delta_2)\left(\frac{1}{u} - \frac{1}{4t}\right)\|g_k\|^2 \leq (\delta_2 - 1) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq L\alpha_k \|d_k\|^2.$$

This implies

$$\begin{aligned}\alpha_k &\geq \frac{(1-\delta_2)\left(\frac{1}{4}-\frac{1}{4t}\right)\|g_k\|^2}{L\|d_k\|^2}\\ &\geq \frac{(1-\delta_2)(\frac{1}{4}-\frac{1}{4t})}{L\varpi^2} > 0,\end{aligned}$$

where the last partial inequality follows from (16). □

Theorem 4.3. *Let $\{x_k\}$ be a sequence generated by Algorithm 1 with the Wolfe line search. Suppose that f is twice continuously differentiable and strongly convex, x^* is the minimizer of (1). Then there exist constants $\mu > 0$ and $j \in (0,\ 1)$ such that*

$$\|x_k - x^*\| \leq \mu j^k. \tag{27}$$

Proof. From the Wolfe condition (21), inequalities (24), (25) and (26), we get

$$\begin{aligned}f(x_{k+1}) - f(x^*) &\leq f(x_k) - f(x^*) + \delta_1\alpha_k g_k^T d_k\\ &\leq f(x_k) - f(x^*) - \delta_1\alpha_k\left(\frac{1}{4}-\frac{1}{4t}\right)\|g_k\|^2\\ &\leq f(x_k) - f(x^*) - \delta_1\left(\frac{(1-\delta_2)(\frac{1}{4}-\frac{1}{4t})^2}{L\varpi^2}\right)m^2\|x_k - x^*\|^2\\ &\leq f(x_k) - f(x^*) - \left(\frac{2m^2\delta_1(1-\delta_2)(\frac{1}{4}-\frac{1}{4t})^2}{ML\varpi^2}\right)(f(x_k) - f(x^*))\\ &= \left(1 - \frac{2m^2\delta_1(1-\delta_2)(\frac{1}{4}-\frac{1}{4t})^2}{ML\varpi^2}\right)(f(x_k) - f(x^*)).\end{aligned}$$

This and inductive process gives

$$f(x_k) - f(x^*) \leq j^2(f(x_{k-1}) - f(x^*)) \leq \ldots \leq j^2 k(f(x_0) - f(x^*)), \tag{28}$$

where $j = \left(1 - \frac{2m^2\delta_1(1-\delta_2)\left(\frac{1}{4}-\frac{1}{4t}\right)^2}{ML\varpi^2}\right) \in (0,\ 1)$. Combining (24) with (28), we get

$$\|x_k - x^*\| \leq \sqrt{\frac{2}{m}\left(f(x_0) - f(x^*)\right)}j^k.$$

By letting $\mu = \sqrt{\frac{2}{m}\left(f(x_0) - f(x^*)\right)}$, we obtain (27). Hence, this means that the sequence $\{x_k\}$ be generated by Algorithm 1 converges R-linearly under Wolfe line search. □

5. **Numerical results.** In this part, we analyze the computational efficiency of the DPRP and DLS methods, by comparing their performance with some existing methods. The analyses was based on a set of forty nine (49) problems considered from [4] with dimensions ranging from 2 up to 10^5. The tabulated list of all 49 test problems used for this experiment is accessible through the following link https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:129f6056-7992-4aae-a126-bf0434479829. For the implementation of the classical PRP, LS, and the proposed DPRP and DLS methods, we used the Wolfe conditions with the parameters value as $\delta_1 = \delta_2 = 0.0001$ and Armijo conditions with parameters values as $\rho = 0.5$, $\gamma = 1$ and $\sigma = 0.0001$. We compare the proposed methods according to number of iterations (NOI), function evaluations (NFE), and CPUT metrics. The following existing methods are considered for the comparison purposes:

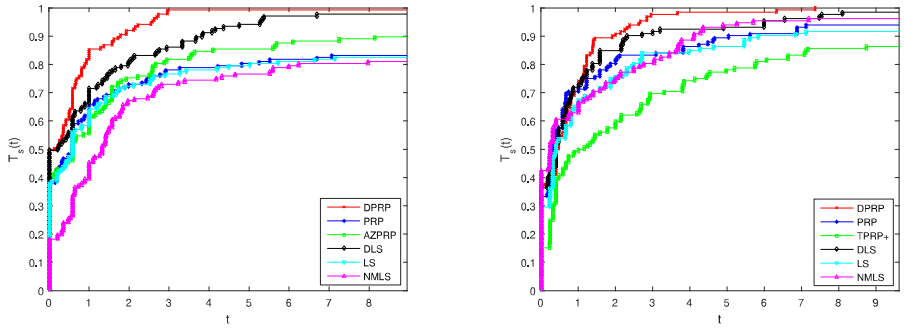
- The classical PRP method in [25, 26]
- The AZPRP method in [2].
- The TPRP+ method in [20].
- The classical LS method in [21].
- The NMLS method proposed in [1].

The recent methods, namely, AZPRP, TPRP+, and NMLS were implemented using the standard parameters values reported in the respective papers [1, 2, 20].

All coding were done on MATLAB R2019b software and ran on a core i5 Windows 10 PC with 8GB RAM. The iteration is terminated if $\|g_k\| \leq 10^{-6}$ and when the NOI exceed 2,000, we report a failure. The detailed numerical results is available at https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:df9a7715-020c-44f9-a623-35d211e13dbb.

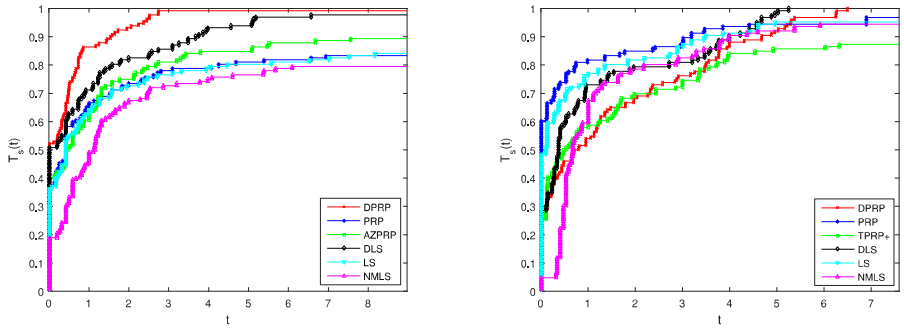
To present a visual illustration and comparison of the proposed methods with the classical as well as recent modifications, the study employed the performance profiles tool introduced by Dolan and Moré [12]. This tool states that the algorithm with the top curve is regarded as more efficient because it solved the most problems compare to other methods used in the comparison. Based on the outcome of our numerical experiment, we present the following results. Figs. 1 illustrate the experimental outcome based on number of iterations for AZPRP, TPRP+, PRP, DPRP, NMLS, LS, DLS methods under Wolfe line search (A) and Armijo-type line search (B) while Figs. 2 and 3 represent the performance based on number of function evaluation and CPU time respectively.

From Figs. 1(A), 2(A), and 3(A), we observed that the performance of DPRP and DLS obviously outperformed all the other methods under the Wolfe line search. Also, based on Fig. 1(B), it is evident that the new DPRP algorithm perform better based on NOI, while the DLS method was competing other methods. However, from Fig. 2(B) and 3(B), we can observe a close competition from the all the algorithms, particularly in terms of NFE and CPUT, yet, the proposed methods slightly outperforms the other methods. Since all the methods presented in Figs. 1(A), 2(A), and 3(A), are implemented using Wolfe procedure, we can draw a conclusion that the proposed DPRP and DLS methods perform better because they generated more efficient search directions compare to the other methods. Also, based on the Armijo-type line search as presented in Fig. 1(B), the DPRP and DLS curves are the top performer for a large number of problems despite the competition as shown in Figs. 2(B), and 3(B), thus, on average, we may conclude that the proposed methods generated more efficient search direction compare to PRP, TPRP+, LS, and NMLS methods respectively.



(A) (B)

FIGURE 1. Performance profile outputs based on number of iterations under Wolfe line search (A) and Armijo line search (B)



(A) (B)

FIGURE 2. Performance profile outputs based on number of function evaluation under Wolfe line search (A) and Armijo line search (B)

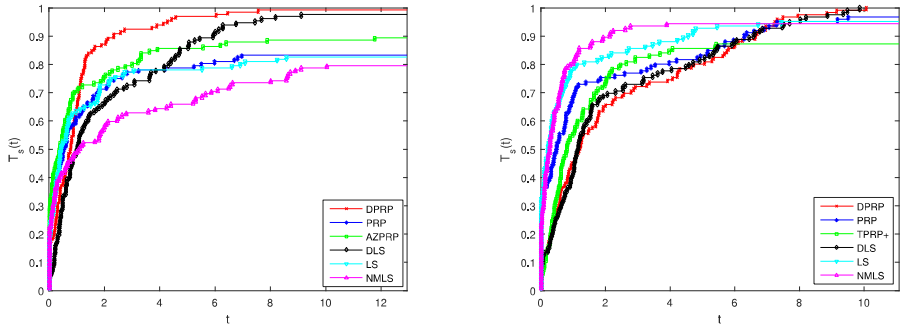
In the second part of the numerical experiments, the performance of the proposed DPRP and DLS methods was tested on image restoration problem. The problem of restoring images corrupted by impulse noise has recently gained a significant attention due to its important in real life applications [5,30,31]. There are two types of impulse noise, namely, salt-and-pepper and random-valued, here we consider noisy image corrupted by salt-and-pepper impulse noise. To achieve our goal, we consider the following image restoration problem similar to the problem used by Yu et al. in [30]:

$$\min \mathcal{G}(u),$$

where

$$\mathcal{G}(u) = \sum_{(i,j)\in K} \left\{ \sum_{(m,n)\in V_{i,j}/K} \phi_\alpha(u_{i,j} - \xi_{m,n}) + \frac{1}{2} \sum_{(m,n)\in V_{i,j}\cap K} \phi_\alpha(u_{i,j} - u_{m,n}) \right\},$$

where



(A) (B)

FIGURE 3. Performance profile outputs based on CPU time under Wolfe line search (A) and Armijo line search (B)

$$K = \{(i,j) \in W | \bar{\xi}_{ij} \neq \xi_{ij},\ \xi_{ij} = s_{\min} \text{ or } s_{\max}\},$$

is the index of the noise candidate, $W = \{1,2,\cdot,M\} \times \{1,2,\cdot,N\}$ and $V_{i\ j} = \{(i,\ j-1),(i,j+1),(i-1,j),(i+1,j)\}$, ξ is the observed noisy image corrupted by salt-and-pepper impulse noise, $\bar{\xi}$ is an adaptive median filter to the noisy image ξ. The maximum and minimum of a noisy pixel are represented by $s_{\max}$ and $s_{\min}$, respectively. Furthermore, ϕ_α is an edge-preserving functional taken as $\phi_\alpha(t) = \sqrt{t^2+\alpha}$ with $\alpha = 1$. We make performance comparison of DPRP and DLS with

TABLE 1. Image restoration outputs for EMPRP, DPRP and DLS based on CPUT, RelErr and PSNR metrics

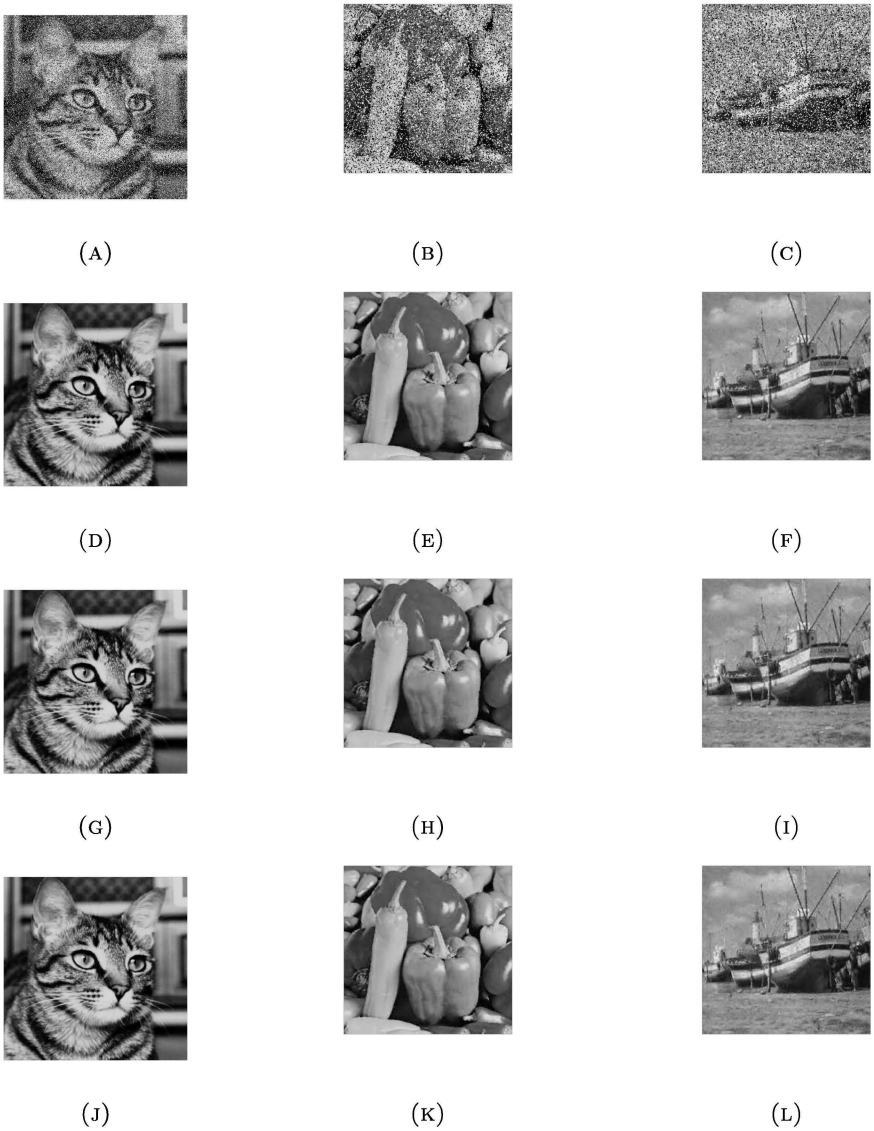
METHOD		EMPRP			DPRP			DLS		
IMAGE	**NOISE**	**CPUT**	**RelErr**	**PSNR**	**CPUT**	**RelErr**	**PSNR**	**CPUT**	**RelErr**	**PSNR**
CAT	30%	105.542	0.5781	28.475	105.493	0.5572	28.5543	105.5418	0.5743	28.425
	50%	104.651	0.7452	27.0691	104.761	0.8596	25.8077	104.7299	0.8661	25.7951
	70%	143.276	1.0233	25.3733	143.101	1.0159	25.3698	139.424	1.0269	125.3934
PEPPER	30%	50.0791	0.8131	34.6163	49.225	0.8031	34.2636	50.0232	0.801	34.2661
	50%	190.629	1.2806	30.9059	187.223	1.2086	30.1389	189.445	1.2327	30.9452
	70%	138.435	1.7082	28.1305	124.24	1.675	28.3416	128.345	1.699	28.1128
SHIP	30%	38.2418	0.8208	33.0966	33.6865	0.8036	32.5642	33.6957	0.8071	32.9238
	50%	63.9543	1.3115	29.508	58.934	1.3075	29.6171	59.1445	1.3012	29.6127
	70%	104.753	1.999	26.7453	98.666	1.9889	26.6844	98.732	1.9824	26.6772

related methods such as EMPRP in [5]. Three test images, namely, Cat, Pepper and Ship (both of them are 256×256 grey level images) were considered.

The quality of the images restored by the various methods are measured using the relative error (RelErr) and peak signal-to-noise ratio (PSNR) defined in [5]. The PSNR ratio between two pictures, measured in decibels, is computed by the PSNR block. This ratio is used to compare the quality between the original and compressed images. The quality of the restored image improves with a greater PSNR which is computed as follows:

$$\begin{aligned} PNSR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE), \end{aligned}$$

where MSE defines the mean square error and MAX_I denotes the image possible maximum pixel value.



(A) (B) (C)

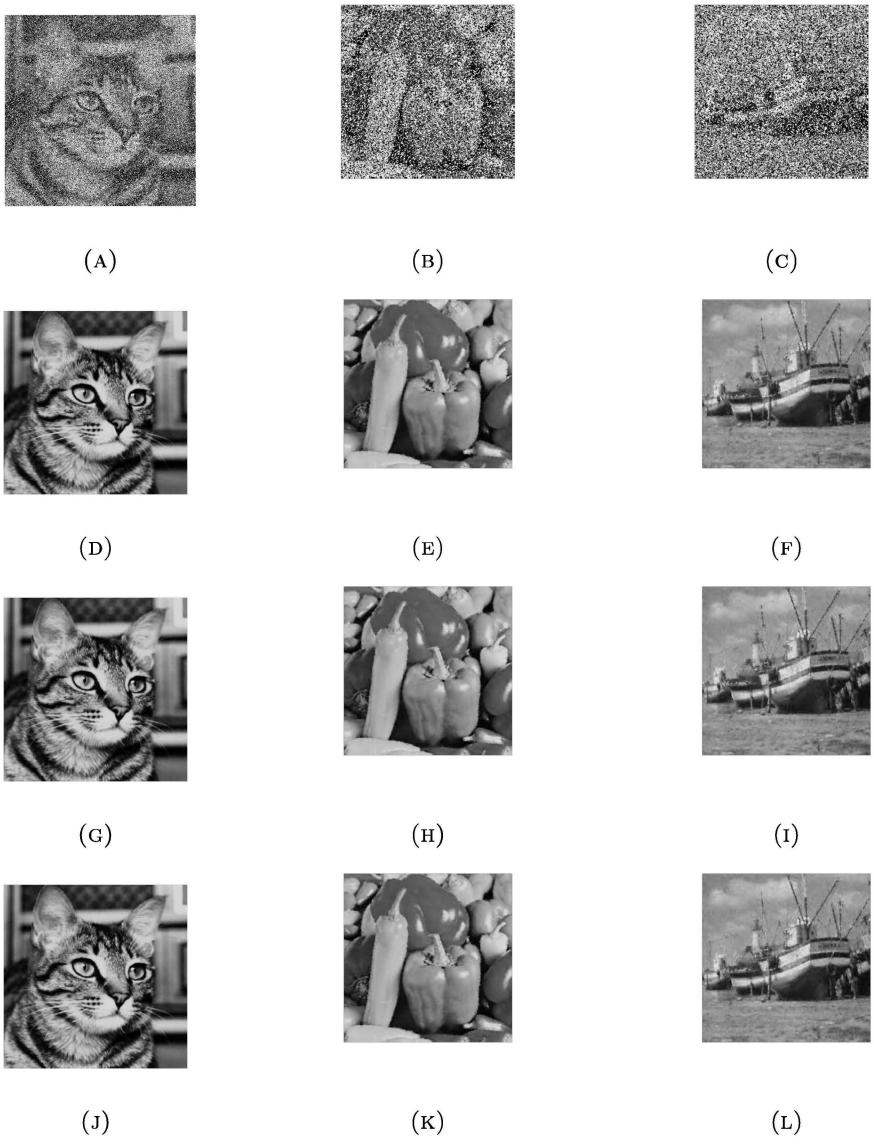
(D) (E) (F)

(G) (H) (I)

(J) (K) (L)

FIGURE 4. Images corrupted by 30 % salt-and-pepper noise: (A,B,C). The restored images via EMPRP: (D,E,F). The restored images via DPRP: (G,H,I), and the restored images via DLS: (J,K,L)

As observed from Table 1 and Figs. 4, 5 and 6, the proposed methods are capable of efficiently removing the noise from the corrupted images.



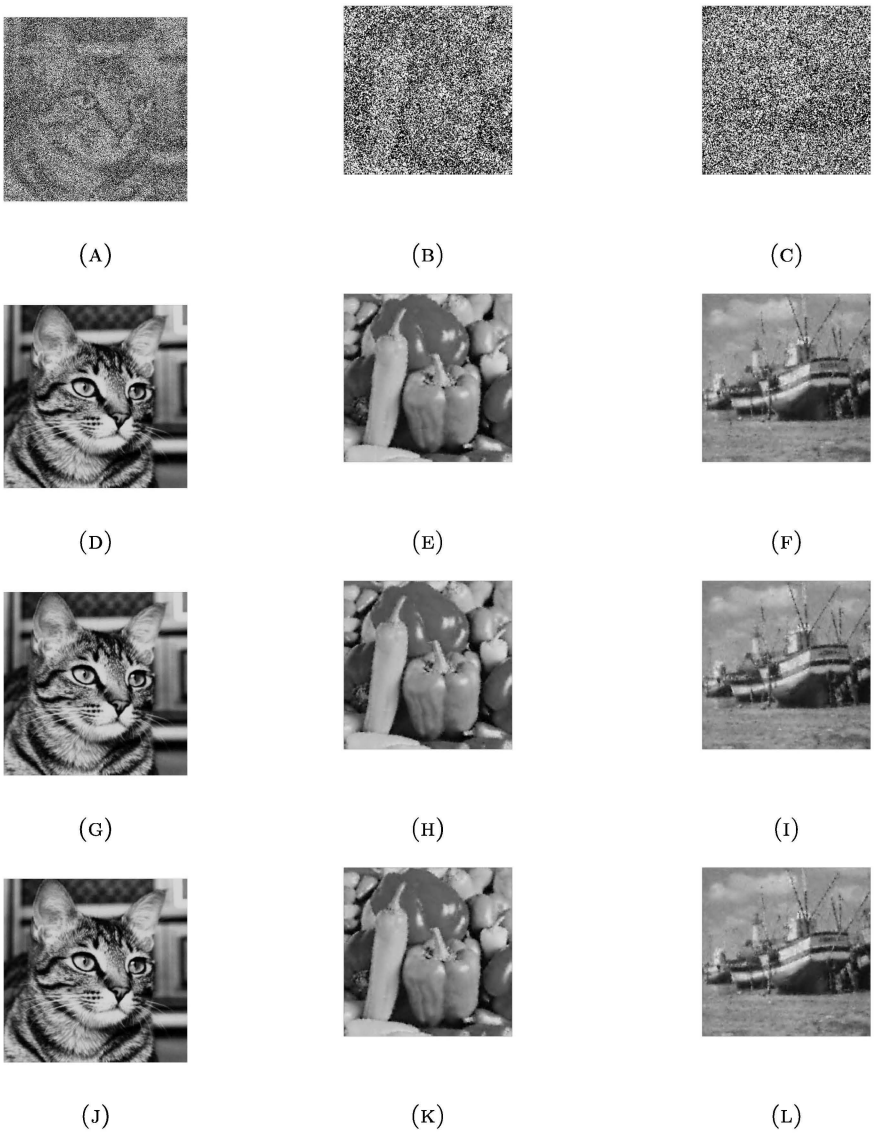
(A) (B) (C)

(D) (E) (F)

(G) (H) (I)

(J) (K) (L)

FIGURE 5. Images corrupted by 50 % salt-and-pepper noise: (A,B,C). The restored images via EMPRP: (D,E,F). The restored images via DPRP: (G,H,I), and the restored images via DLS: (J,K,L)



(A) (B) (C)

(D) (E) (F)

(G) (H) (I)

(J) (K) (L)

FIGURE 6. Images corrupted by 70 % salt-and-pepper noise: (A,B,C). The restored images via EMPRP: (D,E,F). The restored images via DPRP: (G,H,I), and the restored images via DLS: (J,K,L)

Acknowledgments. The authors are grateful to Professor Saman Babaie-Kafaki for providing the image restoration MATLAB implementation codes for the EMPRP method. We are also thankful to the anonymous reviewers for their constructive comments that improve the quality of this manuscript.

Data availability. The MATLAB codes for the implementation of the presented algorithms are available upon request. The datasets used to plot the graphs presented in Section 5 of this paper are openly available in the corresponding author's

repository https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:df9a7715-020c-44f9-a623-35d211e13dbb.

Declarations. Conflict of interest: The authors declare no competing interests.

REFERENCES

[1] A. B. Abubakar, M. Malik, P. Kumam, H. Mohammad, M. Sun, A. H. Ibrahim and A. I. Kiri, A Liu-Storey-type conjugate gradient method for unconstrained minimization problem with application in motion control, *Journal of King Saud University-Science*, **34** (2022), 101923.

[2] A. Ahmad, S. Zabidin, M. Mamat and M. Rivaie, An efficient modified Polak-Ribiére-Polyak conjugate gradient method with global convergence properties, *Optim. Methods Softw.*, **32** (2017), 1299-1312.

[3] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, *IMA J. Numer. Anal.*, **5** (1985), 121-124.

[4] N. Andrei, An unconstrained optimization test functions collection, *Adv. Model. Optim.*, **10** (2008), 147-161.

[5] S. Babaie-Kafaki, N. Mirhoseini and Z. Aminifard, A descent extension of a modified Polak-Ribière-Polyak method with application in image restoration problem, *Optim. Lett.*, **17** (2023), 351-367.

[6] Y. H. Dai, J. Han, G. H. Liu, D. Sun, H. Yin and Y. X. Yuan, Convergence properties of nonlinear conjugate gradient methods, *SIAM J. Optim.*, **10** (2000), 345-358.

[7] Y. H. Dai and L. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Applied Mathematics and Optimization*, **43** (2001), 87-101.

[8] Y. H. Dai and Y. X. Yuan, Convergence properties of the conjugate descent method, *Adv. Math.*, **25** (1996), 552-562.

[9] Y. H. Dai and Y. X. Yuan, Convergence properties of the Fletcher-Reeves method, *IMA J. Numer. Anal.*, **16** (1996), 155-164.

[10] Y. H. Dai and Y. X. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.*, **10** (1999), 177-182.

[11] Y. H. Dai and Y. X. Yuan, Nonlinear conjugate gradient methods, *Shanghai Science and Technology Publisher,* Shanghai.

[12] E. D. Dolan and J. J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.*, **91** (2002), 201-213.

[13] X. L. Dong, H. Liu, Y. L. Xu and X. M. Yang, Some nonlinear conjugate gradient methods with sufficient descent condition and global convergence, *Optim. Lett.*, **9** (2015), 1421-1432.

[14] R. Fletcher, *Practical Methods of Optimization*, Vol. 2. Constrained optimization. A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, 1981.

[15] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, *The Computer Journal*, **7** (1964), 149-154.

[16] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.*, **2** (1992), 21-42.

[17] W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.*, **2** (2006), 35-58.

[18] M. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards*, **49** (1952), 409-436.

[19] W. Hu, J. Wu and G. Yuan, Some modified Hestenes-Stiefel conjugate gradient algorithms with application in image restoration, *Applied Numerical Mathematics*, **158** (2020), 360-376.

[20] M. Li, A family of three-term nonlinear conjugate gradient methods close to the memoryless BFGS method, *Optim. Lett.*, **12** (2018), 1911-1927.

[21] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, part 1: theory, *J. Optim. Theory Appl.*, **69** (1991), 129-137.

[22] G. Ma, H. Lin, W. Jin and D. Han, Two modified conjugate gradient methods for unconstrained optimization with applications in image restoration problems, *Journal of Applied Mathematics and Computing*, **68** (2022), 4733-4758.

[23] H. Mohammad and S. A. Santos, A structured diagonal hessian approximation method with evaluation complexity analysis for nonlinear least squares, *Comput. Appl. Math.*, **37** (2018), 6619-6653.

[24] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Science, 2006.

[25] E. Polak and G. Ribiere, Note sur la convergence de méthodes de directions conjuguées, (French)*Rev. Française Informat. Recherche Opérationnelle*, **3** (1969), 35-43.
[26] B. T. Polyak, The conjugate gradient method in extremal problems, *USSR Comput. Math. & Math. Phys.*, **9** (1969), 94-112.
[27] D. F. Shanno, Globally convergent conjugate gradient algorithms, *Math. Program.*, **33** (1985), 61-67.
[28] W. Sun and Y. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, vol. 1, Springer Science & Business Media, 2006.
[29] P. Wolfe, Convergence conditions for ascent methods, *SIAM Rev.*, **11** (1969), 226-235.
[30] G. Yu, J. Huang and Y. Zhou, A descent spectral conjugate gradient method for impulse noise removal, *Appl. Math. Lett.*, **23** (2010), 555-560.
[31] G. Yuan, T. Li and W. Hu, A conjugate gradient algorithm for large-scale nonlinear equations and image restoration problems, *Applied Numerical Mathematics*, **147** (2020), 129-141.
[32] G. Yuan, J. Lu and Z. Wang, The PRP conjugate gradient algorithm with a modified WWP line search and its application in the image restoration problems, *Applied Numerical Mathematics*, **152** (2020), 1-11.
[33] G. Yuan, J. Lu and Z. Wang, The modified PRP conjugate gradient algorithm under a non-descent line search and its application in the muskingum model and image restoration problems, *Applied Numerical Mathematics*, **25** (2021), 5867-5879.
[34] Y. X. Yuan, Recent advances in numerical methods for nonlinear equations and nonlinear least squares, *Numer. Algebr. Contr. Optim.*, **1** (2011), 15-34.
[35] L. Zhang, W. Zhou and D. Li, A descent modified Polak-Ribiére-Polyak conjugate gradient method and its global convergence, *IMA Journal of Numerical Analysis*, **26** (2006), 629-640.
[36] Z. Zhu, D. Zhang and S. Wang, Two modified DY conjugate gradient methods for unconstrained optimization problems, *Applied Mathematics and Computation*, **373** (2020), 125004.

Received July 2022; 1st revision December 2022; 2nd revision March 2023; early access June 2023.

View publication stats