Development of a New Numerical Conjugate Gradient Technique for Image Processing

HAWRAZ N. JABBAR¹, YOKSAL A. LAYLANI¹, ISSAM A.R. MOGHRABI², BASIM A. HASSAN³

¹Department of Mathematics, College of Sciences, University of Kirkuk, IRAQ
²Department of Computer Science, College of Arts and Sciences, University Central Asia, Naryn, KYRGYZ REPUBLIC
³Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, IRAQ

Abstract: - We present a new iterative conjugate gradient technique for image processing. The technique is based on a new derivation of the conjugacy coefficient and develops a variant of the classical Fletcher-Reeves conjugate gradient method. The derivation exploits a quadratic function model. The new method is intended to minimize the presence of noise by utilizing the adaptive median filter (AMF) to reduce salt-and-pepper noise, while the adaptive center-weighted median filter (ACWMF) is used to reduce random-valued noise. The theoretical convergence properties of the method are proven and then tested on a basic set of images using MATLAB. The results show that the proposed algorithm is more efficient than the classical Fletcher-Reeves (FR) method, as measured by the signal-to-noise ratio (PSNR). The number of iterations and the number of function evaluations are also lower for the proposed method. The favorable performance of the new algorithm provides promise for deriving similar techniques that enhance the speed and efficiency of image-processing libraries.

Key-Words: - Conjugate Gradient methods, image restoration, impulse noise reduction.


1 Introduction

Extensive research and practical applications have been devoted to the field of image restoration across various domains in scientific and engineering. This field focuses on the restoration of an image file from a degraded observation. For instance, pictures captured by telescopes and satellites often suffer from degradation caused by air turbulence. Moreover, images frequently encounter noise originating from environmental effects, transmission channels, and other associated elements throughout the processes of acquisition, resizing, and communication. Consequently, these factors adversely affect the image quality, resulting in distortion and loss of valuable information. Moreover, noise can have a detrimental impact on subsequent image-processing tasks, including image analysis, image tracking, and video processing.

Therefore, image cleansing plays a pivotal role in contemporary image processing systems.

The objective of image cleansing is to restore the original image quality by reducing the presence of noise in noisy images. Nevertheless, this task presents a challenge due to the difficulty in distinguishing between noise, edges, and textures, as these elements often possess high-frequency characteristics. Consequently, throughout the cleansing process, restored images may inadvertently lose certain significant details. In essence, the primary challenge faced by image processing systems lies in recovering relevant information from noisy images while effectively removing noise, ultimately leading to the generation of high-quality images. In certain scenarios, it becomes necessary to recover stellar images not necessarily observed directly within the Earth's atmosphere. The main objective of this research is...
to devise a set of optimization methodologies that can effectively handle edge-preserving regularization (EPR) objective functions.

When it comes to image processing methods, a comparative analysis reveals the distinct strengths and characteristics of various approaches, with conjugate gradient methods standing out in specific contexts. Classical methods, such as Fourier Transform-based techniques, excel in capturing global frequency information but may fall short when dealing with localized features. Meanwhile, wavelet-based methods offer a compromise by combining both global and local information, making them versatile for various applications. Machine learning-based approaches, particularly deep learning models like convolutional neural networks (CNNs), have gained immense popularity for their ability to learn complex hierarchical features directly from data, showcasing remarkable performance in tasks like image recognition. Conjugate gradient methods have proven their viability in image processing, especially when it comes to noise reduction. The conjugate gradient algorithm’s ability to converge rapidly, particularly in cases of ill-conditioned systems, makes it well-suited for large-scale optimization tasks in image processing. This iterative optimization technique ensures that each iteration provides a substantial reduction in the objective function, contributing to the overall enhancement of image quality. The methods can achieve robust and computationally efficient solutions, aligning with the demands of real-world applications where both accuracy and speed are paramount.

To address impulse noise reduction, a recent advancement was made in the form of a two-phase technique described in, [1]. This technique utilizes the adaptive median filter (AMF) to mitigate salt-and-pepper noise, while for random-valued noise, the adaptive method of center-weighted median filter (ACWMF) is employed. The ACWMF is further enhanced by implementing the variable window technique, which enhances its ability to detect and address severe damage in images, [1]. For this study, we exclusively focus on handling salt-and-pepper noise.

Let \( X \) represent the actual picture and \( A = \{1,2,3,\ldots,M\} \times \{1,2,3,\ldots,N\} \) represent the index set of \( X \) and \( N \subset A \) refer to the set of noisy pixel indices detected throughout the first phase. Also, let \( P_{ij} \) denote the set of the four nearest neighbors at position \((i,j) \in A, y_{ij} \). In addition, we use \( u_{ij} = [u_{ij}]_{(i,j)\in N} \) to indicate a lexicographically organized column vector of length \( c \), where \( c \) represents the size of \( N \). Therefore, the minimization of the following function will restore the noisy pixels:

\[
f_u(u) = \sum_{(i,j)\in N} \left[ |u_{ij} - y_{ij}| + \frac{\beta}{2} (2 \times S_{ij}^1 + S_{ij}^2) \right],
\]

where \( \beta \) is the regularization parameter,

\[
S_{ij}^1 = 2 \sum_{(m,n)\in P_{ij}\cap N} \phi_u(u_{ij} - y_{mn})
\]

and

\[
S_{ij}^2 = \sum_{(m,n)\in P_{ij}\cap N} \phi_u(u_{ij} - y_{mn}).
\]

Function (1) ensures the preservation of all the edges \( \phi_u = \sqrt{\alpha + x^2}, \alpha > 0 \). Generally, impulsive noise can be described with this function. The slavish AMF introduced in, [3] is fundamentally based on minimizing (1). In practical applications, the non-smooth data-fitting term can be omitted since it is not necessary for the next phase, which specifically aims to recover only the poor-quality pixels after noise reduction. Consequently, various optimization strategies can be employed to minimize the following smooth EPR (such as, [2]):

\[
f_u(u) = \sum_{(i,j)\in N} \left[ (2 \times S_{ij}^1 + S_{ij}^2) \right],
\]

Since the Conjugate Gradient (CG) methods have low storage requirements, they prove to be highly effective in tackling unconstrained minimization problems expressed as:

\[
\text{Min} \ f(x) \ , \ x \in \mathbb{R}^n,
\]

(see, [3]). To solve (1), subsequent solution estimates are generated using

\[
x_{k+1} = x_k + \alpha_k d_k,
\]

where the step length \( \alpha_k \) is traditionally approximated by performing a one-dimensional line search. The approximation suffices since finding the exact solution is time-consuming and may even be not possible to obtain. However, for quadratic functions, the step length \( \alpha_k \) can be expressed exactly as, [4], [5], [6], [8]

\[
\alpha_k = -\frac{g_k^T d_k}{d_k^T d_k}.
\]

For non-quadratic problems, \( \alpha_k \) is determined to guarantee that the computed search direction is sufficiently descent through enforcing the strong Wolfe conditions, [8]

\[
f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \quad (6a)
\]

and

\[
d_k^T g(x_k + \alpha_k d_k) \geq \sigma d_k^T g_k, \quad (6b)
\]

where \( 0 < \delta < \sigma < 1 \). The search directions for CG methods are obtained using

\[
d_{k+1} = -g_{k+1} + \beta_k s_k,
\]

\[
\beta_k = \frac{\alpha_k}{\alpha_{k-1}}\quad \text{or} \quad \beta_k = \frac{\alpha_k}{\alpha_{k-1}}\frac{\alpha_{k-1}}{\alpha_{k-2}} \ldots \frac{\alpha_2}{\alpha_1}.
\]
where $\beta_k$ is taken as a conjugacy parameter. Both $d_k$ and $d_{k+1}$ satisfy the condition for conjugacy
\[ d_i^TQd_j = 0, \forall i \neq j, \]
for a symmetric matrix $Q \in R^{n \times n}$.

Particularly intriguing are the global convergence characteristics of CG algorithms. According to, [9], the Fletcher and Reeves (FR) formula for $\beta_k$ has the best convergence results. On the other hand, the Hestenes-Stiefel (HS) method, a highly recognized CG technique, fails to meet the global convergence criterion under inexact line search, [10]. The two choices of $\beta_k$ are given, respectively, as:
\[ \beta_k^{FR} = \frac{g_k^Tg_{k+1}}{g_k^Tg_k} \quad \text{and} \quad \beta_k^{HS} = \frac{y_k^Tg_{k+1}}{d_k^Ty_k}, \]
where $y_k = g_{k+1} - g_k$. An attractive property of the Hestenes-Stiefel formula is the fact that it satisfies the conjugacy criteria.

Numerous alternative approaches have been investigated to upgrade the numerical behavior of CG methods, considering their advantageous storage demands (refer to, [11], [12], [13], [14], for further details). A wide range of problems can be addressed with CG methods, including machine learning, mechanics, nonlinear and differential equations, and many others. Furthermore, an additional potential domain for their application lies within Human Performance Technology (HPT). HPT heavily relies on the numerical Performance Improvement (PI) attributes of computer systems, which are facilitated by specialized algorithms enabling logical assessments, [15]. An empirical study found that CG methods improve the performance and efficiency of mobile users and help them adopt mobile Electronic Performance Support Systems (EPSS), [16].

One useful approach that has proven viability in improving the performance of the CG methods, is based on the incorporation of the quasi-Newton idea in developing better converging CG methods, [7]. This is achieved by rewriting (7) as
\[ -Q_{k+1}^{-1}Q_{k+1} = -g_{k+1} + \beta_k s_k, \]
where $Q_{k+1}$ is the Hessian matrix of the function being minimized, [13].

Distinguishing itself from conventional CG algorithms, the aforementioned approach exhibits a distinctive quality of consistently generating improved downhill directions while adhering to the conjugacy properties, as demonstrated by the reported outcomes. Subsequently, in the subsequent section, a quadratic model is utilized to derive novel conjugacy parameters $\beta_k$, leading to the development of a new CG algorithm.

## 2 Deriving the New Parameter
The key idea of the derivation of the new CG parameter is the utilization of a classical quadratic model given by
\[ f(u) = f(u_{k+1}) + g_{k+1}^T(u - u_{k+1}) + \frac{1}{2}(u - u_{k+1})^TA(u - u_{k+1}), \]
where $Q$ is the constant Hessian of the quadratic function. The gradient of the model is expressed as
\[ g_{k+1} = g_k + Q(u_k)s_k, \]
for $s_k = u_{k+1} - u_k$.

From (10) and (11), the second-order curvature is given by
\[ s_k^TQ(u_k)s_k = 2(f_k - f_{k+1}) + 2g_{k+1}^Ts_k, \]
or, equivalently,
\[ s_k^TQ(u_k)s_k = 2(f_k - f_{k+1}) + 2y_k^Ts_k + 2g_k^Ts_k, \]
Equation (13) leads to the following matrix definition
\[ Q(u_k) = \frac{2(f_k - f_{k+1}) + 2y_k^Ts_k + 2g_k^Ts_k}{s_k^Ts_k}I_n, \]
where $I_n$ is the $n \times n$ Identity matrix. Substituting (14) in (9) yields a new conjugacy parameter as follows:
\[ \beta_k^{BBD} = \frac{1 - \frac{s_k^Ts_k}{2(f_k - f_{k+1}) + 2y_k^Ts_k + 2g_k^Ts_k}}{s_k^Ts_k}. \]

The algorithmic framework is given next.

**BBD Algorithm:**

i) Start with the initial solution point, $x_1 \in R^n$. Set $k = 1$ and $d_1 = -g_1$. If $\|g_1\| \leq 10^{-6}$, then terminate.

ii) Find $\alpha_k > 0$ that satisfies conditions (6).

iii) Calculate $x_{k+1} = x_k + \alpha_k d_k$ and the corresponding gradient $g_{k+1} = g(x_{k+1})$. If $\|g_{k+1}\| \leq 10^{-6}$, then halt.

iv) Calculate $\beta_k$ using (15) and construct $d_{k+1}$ from (7).

v) $k = k + 1$ and go to (ii).

## 3 Convergence Analysis
To ensure the global convergence of the BBD algorithm on uniformly convex problems, it is necessary to rely on the following assumptions.

i. The level set $\Omega = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded.

ii. There exists a constant $L > 0$ such that the gradient $g$ of the objective function is Lipschitz continuous in some neighborhood $\Delta$ of $\Omega$, such that
\[ \| g(\alpha) - g(\tau) \| \leq L \| \alpha - \tau \|, \forall \tau, \alpha \in \Lambda \quad (16) \]

(see, [14], for more details).

In this particular situation, there is a stable \( \Gamma \geq 0 \) such that, provided that certain function assumptions are met, \( \| \nabla f(x) \| \leq \Gamma \).

**Theorem 1.**

If \( s_k^T y_k \neq 0 \), the search directions given by (7), using (15), are descent directions.

**Proof.** We have \( g_0^T d_0 = -\|g_0\|^2 < 0 \) since \( d_0 = -g_0 \). Consider \( d_k^T g_k \leq 0 \) to be true. Multiplying (6) by \( g_{k+1} \) results in

\[
d_{k+1}^T g_{k+1} - g_k^T g_{k+1} + (1 - \frac{s_k^T y_k}{2(f_k - f_{k+1}) + 2y_k^T s_k + 2g_k^T s_k}) y_k^T g_{k+1} s_k^T g_{k+1}.
\]

Let \( \alpha_k s_k^T y_k = \hat{s}_k^T y_k(2(f_k - f_{k+1}) + 2y_k^T s_k + 2g_k^T s_k) \), then it is easy to show that

\[
d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + (\frac{\alpha_k s_k^T y_k - \hat{s}_k^T y_k}{\alpha_k s_k^T y_k}) y_k^T g_{k+1} s_k^T g_{k+1}.
\]

Now using the Lipschitz condition leads to \( y_k^T g_{k+1} \leq L \hat{s}_k^T g_{k+1} \) and \( s_k^T y_k \leq L \hat{s}_k^T s_k \). Thus, it can be deduced that

\[
d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + (\frac{\alpha_k L \hat{s}_k^T s_k - \hat{s}_k^T s_k}{\alpha_k s_k^T y_k}) \frac{L(s_k^T g_{k+1})^2}{s_k^T y_k}.
\]

Because \( L \) and \( \alpha_k \) are very small, it is clear that

\[
d_{k+1}^T g_{k+1} \leq 0.
\]

The proof is established.

For any conjugate gradient approach, employing the strong Wolfe conditions (6), the general convergence results in, [15], apply and are stated in Lemma 1 below.

**Lemma 1.**

If assumptions (i) and (ii) are true, then for any conjugate gradient method using \( d_{k+1} = -g_{k+1} + \beta_k d_k \), for \( \alpha_k \) selected to satisfy the strong Wolfe conditions (6) the following applies:

\[ \sum_{k>1}^{1/\|d_{k+1}\|^2} = \infty, \quad (21) \]

then

\[ \lim_{k \to \infty} (\inf \| g_{k+1} \|) = 0. \quad (22) \]

The same results were used in [13], [15], [16], [17], [18].

We now utilize the results in Lemma 1 to prove the same for our method.

**Theorem 2.**

If a constant \( \mu > 0 \) exists such that it satisfies, for any \( k \):

\[ (\nabla f(u) - \nabla f(w))^T (u - w) \geq \mu \|u - w\|^2, \quad \forall u, w \in R^n, \]

then by Lemma 1, the following holds:

\[ \lim_{k \to \infty} (\inf \| g_{k+1} \|) = 0. \quad (23) \]

**Proof.** It is clear from (12) that:

\[ \|d_{k+1}\| = \|g_{k+1} + (1 - \omega) \frac{2g_{k+1}y_k}{s_k^T y_k} \| s_k \|, \]

where \( \omega = \frac{s_k^T y_k}{s_k^T s_k} \). Using Cauchy's inequality:

\[ \|d_{k+1}\| \leq \|g_{k+1}\| + |1 - \omega| \frac{\|g_{k+1}\| \|y_k\|}{\|s_k\| \|y_k\|} \|s_k\| \leq (2 - \omega) \|g_{k+1}\|. \]

Therefore, \( \| \nabla f(x) \| \leq \Gamma \) implies that:

\[ \sum_{k \geq 1} \frac{1}{\|d_k\|^2} - \frac{1}{2 - \omega} \frac{1}{\Gamma} \sum_{k \geq 1} \frac{1}{\|d_k\|^2} = \infty. \quad (27) \]

It follows that \( \lim_{k \to \infty} \inf \| g_k \| = 0 \) using Lemma 1.\( \square \)

**4 Numerical Results**

We evaluate the new algorithm's performance in the context of reducing salt-and-pepper impulse noise (3). The test images utilized in this evaluation are presented in Table 1. Additionally, Table 1 provides the numerical results obtained from comparing the classical CG FR method with the newly derived algorithm. The comparison is based on parameters that include the number of function/gradient evaluations, count of iterations, and Peak signal-to-noise ratio (PSNR), [2], [19], [20]. We use MATLAB 2015a for all simulations. This study focuses on developing an efficient and fast way to reduce carbon emissions in (3). We use the PSNR value, [21], [22], to assess the corrected images' pixel quality:

\[ PSNR = 10 \log_{10} \frac{255^2}{\sum_{i,j \in M,N}(f_0(i,j) - f_1(i,j))^2}, \quad (28) \]
where $u_{i,j}^r$ and $u_{i,j}^*$ refer to the pixel values of the denoised and initial images, respectively. The termination conditions for both procedures are as follows:

$$\frac{|f(u_k) - f(u_{k-1})|}{|f(u_k)|} \leq 10^{-4}$$

and

$$\|f(u_k)\| \leq 10^{-4}(1 + |f(u_k)|).$$

Figure 1, Figure 2, Figure 3, and Figure 4 showcase the outcomes achieved by implementing the algorithms on noisy images. Specifically, the first image in Figures 1, 2, 3, and 4 depicts the images corrupted with 70% salt-and-pepper noise. The results obtained from the FR method are represented in the second image of each figure. The third image in Figure 1, Figure 2, Figure 3, and Figure 4 exhibits the outcomes of the BBD method. The proposed BBD image correction method is demonstrated to be effective and efficient evidenced by the visual representations.

<table>
<thead>
<tr>
<th>Image</th>
<th>Noise level r (%)</th>
<th>FR-Method</th>
<th>BBD-Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NI</td>
<td>NF</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td>Le</td>
<td>50</td>
<td>82</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>81</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>108</td>
<td>211</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>52</td>
<td>53</td>
</tr>
<tr>
<td>Ho</td>
<td>70</td>
<td>63</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>111</td>
<td>214</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>El</td>
<td>70</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>65</td>
<td>114</td>
</tr>
<tr>
<td>c512</td>
<td>50</td>
<td>59</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>78</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>121</td>
<td>236</td>
</tr>
</tbody>
</table>

Table 1. Numerical results of FR and BBD algorithms
Fig. 1: Demonstrates the results of algorithms FR and BBD of 256 * 256 Lena image

Fig. 2: Demonstrates the results of algorithms FR and BBD of 256 * 256 house images.
Fig. 3: Demonstrates the results of algorithms FR and BBD of 256 * 256 Elaine image
5 Conclusion
In this paper, the primary objective was to develop innovative and modified conjugate gradient formulae that supersede the performance of the conventional Fletcher-Reeves conjugate gradient (FR) approach, specifically in the context of picture restoration. Through a comprehensive analysis, the experimental results validate the global convergence of the proposed novel techniques, particularly when subjected to the strong Wolfe line search conditions. The application of the Wolfe conditions ensures both sufficient decrease and curvature conditions in the optimization process. The convergence analysis reveals that, even in the presence of complex, ill-conditioned systems inherent in image processing tasks, the proposed method consistently converges globally. The experimental results consistently demonstrate that the newly introduced algorithm, referred to as BBD, consistently achieves remarkable reductions in iteration counts and function evaluations. Remarkably, these efficiency improvements are achieved without compromising the quality of picture restoration. Further research may focus on looking at other possibilities that utilize more of the quasi-Newton methods within CG algorithms, such as the ones proposed in, [23].

References:


**Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)**
- Hawraz Jabbar contributed to the mathematical derivations.
- Yoksal Laylani carried out the numerical tests on the new method.
- Issam Moghrabi drafted the document and contributed to the derivation.
- Basim Hassan did the coding necessary for carrying out the tests.

**Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself**
No funding was received for conducting this study.

**Conflict of Interest**
The authors have no conflicts of interest to declare.

**Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)**
This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US