PAPER

A New DY Conjugate Gradient Method and Applications to Image Denoising

Wei XUE[†a)], **Junhong REN**[††b)], **Xiao ZHENG**[†c)], ***Nonmembers***, **Zhi LIU**[†††d)], ***Member***, ***and*** **Yueyong LIANG**[††††e)], ***Nonmember***

SUMMARY Dai-Yuan (DY) conjugate gradient method is an effective method for solving large-scale unconstrained optimization problems. In this paper, a new DY method, possessing a spectral conjugate parameter β_k, is presented. An attractive property of the proposed method is that the search direction generated at each iteration is descent, which is independent of the line search. Global convergence of the proposed method is also established when strong Wolfe conditions are employed. Finally, comparison experiments on impulse noise removal are reported to demonstrate the effectiveness of the proposed method.

key words: *unconstrained optimization, conjugate gradient method, line search, image denoising*

1. Introduction

We consider the unconstrained optimization problem

$$\min \{f(x) \mid x \in R^n\}, \tag{1}$$

where $f : R^n \to R$ is continuously differentiable, and its gradient ∇f is denoted by g. Usually, iterative methods designed for solving (1) are of the form

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

where α_k is a positive stepsize, and $d_k \in R^n$ is a search direction. Numerous methods have been proposed to solve (1) in the past decades. Due to the simplicity of iteration and the low memory requirements, conjugate gradient (CG) methods are very popular [19], [29], [32], [33]. The search direction d_k is usually defined as

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

where g_k is the gradient of f at x_k, and $\beta_k \in R$ is the conjugate parameter that characterizes the method. Different β_k results in different numerical performance. In [17], Hestenes and Stiefel first proposed a CG method for minimizing a convex quadratic function. Later on, Fletcher and Reeves [14] applied the CG method to general unconstrained optimization problems. Well-known formulas for β_k are the Hestenes-Stiefel (HS) [17], Fletcher-Reeves (FR) [14], Polak-Ribière-Polyak (PRP) [27], Conjugate Descent (CD) [6], Liu-Storey (LS) [22], and DY [13]. In the six methods mentioned above, FR, CD, and DY methods usually have global convergence, but their numerical performance is unattractive. The methods of HS, PR, and LS are very practical, but may not globally converge. In this paper, we mainly focus on the DY method with $\beta_k^{DY} = \|g_k\|^2/(d_{k-1}^T y_{k-1})$, where y_{k-1} is defined as $y_{k-1} = g_k - g_{k-1}$, and $\|\cdot\|$ denotes the Euclidean norm.

DY method has been studied by many researchers over the past decade. In 2008, Andrei [1] proposed a modified DY method which satisfies both sufficient descent and conjugacy conditions, independently of the line search, and in [2], Andrei proposed a hybrid conjugate gradient algorithm in which the conjugate parameter β_k is computed as a convex combination of β_k^{HS} and β_k^{DY}. In 2009, Zhang [36] proposed two modified versions of the DY formula. One is based on the modified BFGS method [21] with the other on the ideas of [31], [37]. In 2013, Babaie-Kafaki [3] suggested a hybridization of HS and DY using a quadratic relaxation. In 2016, Sato [28] presented a generalization of DY's Euclidean CG method to a Riemannian algorithm that requires only the weak Wolfe conditions. Other related work can be found in [4], [23], [24], etc.

Recently, there has been growing interest in the descent CG method. Motivated by the work of [16], Yu et al. [33] proposed a *descent DY* (DDY) CG method, which is given by

$$\beta_k^{DDY} = \beta_k^{DY} - \frac{C\|g_k\|^2}{(y_{k-1}^T d_{k-1})^2} g_k^T d_{k-1}, \tag{4}$$

where C is a positive parameter. If $C \geq 1/4$, then the search direction generated by the DDY method can always satisfy the (sufficient) descent condition, i.e., $g_k^T d_k \leq -(1 - \frac{1}{4C})\|g_k\|^2 < 0$. Furthermore, the authors presented a *spectral DDY* (SDDY) method, in which, the conjugate parameter takes the following form

Manuscript received June 14, 2018.
Manuscript revised August 30, 2018.
Manuscript publicized September 14, 2018.

[†]The authors are with the School of Computer Science and Technology, Anhui University of Technology, Maanshan 243032, China.
[††]The author is with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.
[†††]The author is with the Department of Mathematical and Systems Engineering, Shizuoka University, Hamamatsu-shi, 432–8561 Japan.
[††††]The author is with PHIMA Intelligence Technology Co., Ltd., Maanshan 243000, China.
a) E-mail: cswxue@ahut.edu.cn (Corresponding author)
b) E-mail: junhong.ren@ia.ac.cn
c) E-mail: xzheng@ahut.edu.cn (Corresponding author)
d) E-mail: liu@shizuoka.ac.jp
e) E-mail: lyy@magang.com.cn

DOI: 10.1587/transinf.2018EDP7210

$$\beta_k^{SDDY} = \beta_k^{SDY} - \frac{C\|g_k\|^2}{\delta_k(y_{k-1}^T d_{k-1})^2} g_k^T d_{k-1}, \tag{5}$$

where $\beta_k^{SDY} = \beta_k^{DY}/\delta_k$, and $\delta_k = y_{k-1}^T s_{k-1}/\|s_{k-1}\|^2$. Further details of δ_k can be found in [5]. Similarly, if $C \geq 1/4$, then (5) satisfies the (sufficient) descent condition. In [12], Cheng proposed two-term PRP-based CG method, where the search direction is defined as

$$d_k = \begin{cases} -g_k & \text{for k = 0,} \\ -g_k + \beta_k^{PRP}(I - \frac{g_k g_k^T}{g_k^T g_k})d_{k-1} & \text{for k} \geq 1. \end{cases} \tag{6}$$

The search directions generated by (6) always satisfy $g_k^T d_k = -\|g_k\|^2 < 0$, which implies the sufficient descent condition. Simulation results show the potential advantages of this method.

In this paper, we give a new formula for β_k^{SDDY} and focus on the d_k defined in (6). The resulting search direction is a descent direction, and the proposed CG method is globally convergent.

The remainder of this paper is organized as follows. In Sect. 2, we present a modification of spectral DDY formula, which possesses the sufficient descent property independent of line search conditions. Global convergence analysis is given in Sect. 3. In Sect. 4, we apply the proposed method to image denoising. Experimental results are given in Sect. 5.

2. New Formula for β_k

In this paper, we focus on the direction d_k defined in (6) with the conjugate parameter β_k^{SDDY} defined in (5). In order to establish the global convergence result for general objective functions, we reformulate β_k^{SDDY} as

$$\beta_k^{SDDY} = \beta_k^{SDY} - \min\{\beta_k^{SDY}, \frac{C\|g_k\|^2 g_k^T d_{k-1}}{\delta_k(y_{k-1}^T d_{k-1})^2}\}, \tag{7}$$

which implies that $\beta_k^{SDDY} \geq 0$ for all k. By substituting (7) into (6), we obtain the following search direction

$$d_k = \begin{cases} -g_k & \text{for k = 0,} \\ -g_k + \beta_k^{SDDY}(I - \frac{g_k g_k^T}{g_k^T g_k})d_{k-1} & \text{for k} \geq 1. \end{cases} \tag{8}$$

It follows from (8) that this d_k is a descent direction.

In the following, we call our method the *new spectral descent DY* (hereinafter referred as NsdDY) conjugate gradient method. We state the steps of NsdDY in Algorithm 1.

3. Convergence Analysis

In this section, we establish a convergence theorem for Algorithm 1. We assume that $g_k \neq 0$ for all k; otherwise, a stationary point has been found.

To analyse the global convergence of Algorithm 1, we first make the following assumptions on the objective function f.

Assumption 1: The level set $\mathcal{L} = \{x \in R^n : f(x) \leq f(x_0)\}$

Algorithm 1: NsdDY

Data. Given the parameters $0 < \sigma_1 \leq \sigma_2 < 1$, $C > 1/4$, relaxation factor $\tau > 0$ and $x_0 \in R^n$. Set $k := 0$.
While "not converged", **Do**
Compute d_k by (8), where β_k^{SDDY} is determined by (7).
Find $\alpha_k > 0$ such that the following Wolfe conditions hold

$$\begin{aligned} & f(x_k + \alpha_k d_k) - f(x_k) \leq \sigma_1 \alpha_k g_k^T d_k, \\ & |g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma_2 g_k^T d_k. \end{aligned} \tag{9}$$

Compute $\tilde{x}_k = x_k + \alpha_k d_k$.
Compute $x_{k+1} = (1-\tau)x_k + \tau\tilde{x}_k$.
Set $k := k + 1$.
End Do

---

is bounded, namely, there exists a positive constant B, such that $\|x\| \leq B, \forall x \in \mathcal{L}$.

Assumption 2: In some neighborhood $\mathcal{L}_N$ of $\mathcal{L}$, f is continuously differentiable, and its gradient g is Lipschitz continuous with Lipschitz constant $L > 0$, that is, $\|g(x)-g(y)\| \leq L\|x-y\|$, $\forall\ x, y \in \mathcal{L}_N$.

Under Assumptions 1 and 2, we have the following lemma called Zoutendijk condition [26] without proof, which is often used to prove global convergence results of conjugate gradient-based methods.

Lemma 1: [Zoutendijk condition] Suppose Assumptions 1 and 2 hold. Consider any iterative method of the form (2), where d_k is a descent direction and α_k satisfies (9), then $\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty$.

The next theorem establishes the global convergence of the Algorithm 1 when f is strongly convex, that is, there exists constant $\xi > 0$ such that $\xi\|x-y\|^2 \leq (g(x)-g(y))^T(x-y)$ for all $x, y \in \mathcal{L}_N$.

Theorem 1: Let $\{x_k\}$ be the sequence generated by Algorithm 1. If Assumptions 1 and 2 hold and f is strongly convex, then $\liminf_{k\to\infty} \|g_k\| = 0$.

Proof 1: Suppose that $\liminf_{k\to\infty} \|g_k\| > 0$, and define $\gamma = \inf\{\|g_k\| : k > 0\}$. Since $g_k \neq 0$, it follows that $\gamma > 0$. By the second equation of (9), we have

$$y_{k-1}^T d_{k-1} = (g_k - g_{k-1})^T d_{k-1} \geq (\sigma_2 - 1) g_{k-1}^T d_{k-1}.$$

Combining this with $g_{k-1}^T d_{k-1} = -\|g_{k-1}\|^2$, we have

$$y_{k-1}^T d_{k-1} \geq (1-\sigma_2)\gamma^2.$$

By the second equation in (9), it holds that

$$g_k^T d_{k-1} \geq \sigma_2 g_{k-1}^T d_{k-1} = \sigma_2 (g_k - y_{k-1})^T d_{k-1}.$$

Since $\sigma_2 < 1$, we have $g_k^T d_{k-1} \geq (\frac{-\sigma_2}{1-\sigma_2}) y_{k-1}^T d_{k-1}$, i.e.,

$$|\frac{g_k^T d_{k-1}}{y_{k-1}^T d_{k-1}}| \leq \max\{1, \frac{\sigma_2}{1-\sigma_2}\}.$$

By the strong convexity assumption of f, it holds that $\xi \leq \delta_k \leq L$. From all above derivations, we obtain

$$\begin{aligned}|\beta_k^{SDDY}| &= |\beta_k^{SDY} - \min\{\beta_k^{SDY}, \tfrac{C\|g_k\|^2}{\delta_k(y_{k-1}^T d_{k-1})^2} g_k^T d_{k-1}\}| \\ &\le |\beta_k^{SDY}| + \tfrac{C\|g_k\|^2}{\delta_k|y_{k-1}^T d_{k-1}|} \tfrac{|g_k^T d_{k-1}|}{|y_{k-1}^T d_{k-1}|} \\ &= \tfrac{1}{\delta_k} \tfrac{1}{|y_{k-1}^T d_{k-1}|}(\|g_k\|^2 + C\|g_k\|^2 \tfrac{|g_k^T d_{k-1}|}{|y_{k-1}^T d_{k-1}|}) \\ &\le \tfrac{\Gamma^2}{\xi\gamma^2(1-\sigma_2)}(1 + C\max\{1, \tfrac{\sigma_2}{1-\sigma_2}\}) \triangleq \Phi\Psi,\end{aligned}$$

where $\Phi = \frac{\Gamma^2}{\xi\gamma^2(1-\sigma_2)}$, $\Psi = 1 + C\max\{1, \frac{\sigma_2}{1-\sigma_2}\}$, $\Gamma = \max_{x\in\mathcal{L}_\mathcal{N}\|g(x)\|}$. Therefore,

$$\begin{aligned}\|d_k\| &= \| - g_k + \beta_k^{SDDY}(I - \tfrac{g_k g_k^T}{g_k^T g_k})d_{k-1}\| \\ &= \| - (1 + \beta_k^{SDDY} \tfrac{g_k^T d_{k-1}}{\|g_k\|^2})g_k + \beta_k^{SDDY} d_{k-1}\| \\ &\le \|g_k\| + 2|\beta_k^{SDDY}|\|d_{k-1}\| \\ &\le \Gamma + 2\Phi\Psi\|d_{k-1}\| \\ &\le \cdots \\ &\le \Gamma + \Gamma(2\Phi\Psi) + \Gamma(2\Phi\Psi)^2 + \cdots + \Gamma(2\Phi\Psi)^k \\ &= \Gamma\tfrac{(2\Phi\Psi)^k - 1}{2\Phi\Psi - 1}.\end{aligned}$$

Let $\Theta = \max_k\{\Gamma\frac{(2\Phi\Psi)^k-1}{2\Phi\Psi-1}\}$, then we have $\frac{\|g_k\|^4}{\|d_k\|^2} \ge \frac{\gamma^4}{\Theta^2}$. Further,

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = +\infty,$$

which contradicts to the Zoutendijk condition. This completes the proof.

4. Applications to Image Denoising

Based on the analysis above, the proposed algorithm NsdDY can be used to handle unconstrained optimization problems. Here, we apply it to impulse noise denoising. There are two main kinds of impulse noise: salt-and-pepper noise and random-valued noise. For images corrupted by salt-and-pepper noise, the noisy pixels only take the maximum number and the minimum in a dynamic range, while for random-valued noise, the noisy pixels can be any random numbers. The goal of image denoising is to remove the noise while preserving image details. The median-based filter was once the most popular method for removing impulse noise [11], [18], [20]. The filter first locates possible noisy pixels, and then replaces them with the median values or their variants. However, the main drawback is that this replacement technique can blur the details of image features such as the possible presence of edges, especially when the noise ratio is high.

In order to overcome the drawback, Chan et al. [9], [10] proposed a two-phase scheme which combines the advantages of adaptive median filter and variational method. For salt-and-pepper noise, the first phase is to identify the noisy pixels by using the adaptive median filter [18], while for random-valued noise, it is accomplished by using the adaptive center-weighted median filter [11]. Let X be the true image of size M-by-N, and $\mathcal{A} = \{1, 2, \cdots, M\} \times \{1, 2, \cdots, N\}$ be the index set of X. Let the set of indices of the noisy pixels detected in the first phase denote by $\mathcal{N}$, where $\mathcal{N} \subset \mathcal{A}$. Let $u = [u_{i,j}]_{(i,j)\in\mathcal{N}}$ be a column vector of length l ordered lexicographically (l is the number of elements of $\mathcal{N}$), and $y_{i,j}$ denote the observed pixel value at position (i, j). Then, the second phase it to recover the noisy pixels by minimizing the following edge-preserving regularization function

$$F_\alpha(u) = \sum_{(i,j)\in\mathcal{N}} |u_{i,j} - y_{i,j}| + \frac{\mu}{2} \sum_{(i,j)\in\mathcal{N}} (2 \cdot S_{i,j} + T_{i,j}), \quad (10)$$

where $S_{i,j} = \sum_{(m,n)\in\mathcal{V}_{i,j}\setminus\mathcal{N}} \varphi_\alpha(u_{i,j} - y_{m,n})$, $T_{i,j} = \sum_{(m,n)\in\mathcal{V}_{i,j}\bigcap\mathcal{N}} \varphi_\alpha(u_{i,j} - u_{m,n})$, and $\mathcal{V}_{i,j}$ denotes the set of the four closest neighbors of the pixel at position $(i, j) \in \mathcal{A}$. φ is an edge-preserving function, and $\alpha > 0$ is a parameter. Examples of φ_α are $\sqrt{\alpha + u^2}$ and $|u|^\alpha$.

The function $F_\alpha(u)$ only applies to the selected noisy pixels, i.e., the uncorrupted pixels are unchanged. The function $F_\alpha(u)$ in (10) is nonsmooth because of the 1-norm data-fitting term $|u_{i,j} - y_{i,j}|$, it is expensive to get the minimizer. To improve the computational efficiency, Cai et. al suggested to remove the data-fitting term in [7]. This operation can transform $F_\alpha(u)$ into a smooth function which can be minimized

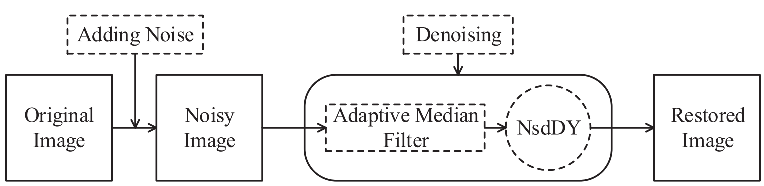

Fig. 1 Flow chart of our experiments.

Fig. 2 Original test images 512 × 512: Lena, Barbara, Cameraman, Boat

Table 1 Summary of restoration results for PRPCG/DSCG/NsdDY methods (the bold font represents the optimal value).

Test Images	Noise Level	PRPCG				DSCG				NsdDY			
		NI	CPU Time	SNR	SSIM	NI	CPU Time	SNR	SSIM	NI	CPU Time	SNR	SSIM
Lena	10%	296	28.01	55.9018	**0.9978**	141	13.15	55.9312	**0.9978**	**79**	**6.07**	**55.9805**	**0.9978**
	30%	360	56.83	49.7132	**0.9912**	256	40.25	49.7144	**0.9912**	**107**	**14.13**	**49.7270**	**0.9912**
	50%	323	74.11	45.2247	**0.9794**	236	56.29	45.2249	**0.9794**	**132**	**26.55**	**45.2268**	**0.9794**
	70%	308	72.58	40.9010	**0.9515**	255	60.00	**40.9018**	**0.9515**	**110**	**23.25**	40.9014	**0.9515**
	90%	349	81.14	34.0608	0.8527	368	94.41	34.0774	**0.8528**	**203**	**48.16**	**34.1207**	**0.8528**
Barbara	10%	328	31.03	**46.2025**	**0.9914**	248	23.90	46.1898	**0.9914**	**107**	**8.61**	46.1042	**0.9914**
	30%	291	45.61	**39.4577**	**0.9648**	256	40.12	39.4551	**0.9648**	**110**	**14.66**	39.4213	0.9647
	50%	255	57.96	36.3925	**0.9290**	215	49.71	36.3963	0.9289	**117**	**23.55**	**36.4037**	0.9289
	70%	184	42.72	**33.8104**	**0.8764**	162	38.76	33.8103	**0.8764**	**104**	**22.42**	33.8101	**0.8764**
	90%	320	69.77	31.5994	0.7643	303	69.34	31.6026	0.7644	**191**	**42.62**	**31.6074**	**0.7645**
Cameraman	10%	222	21.00	52.4619	**0.9993**	176	16.57	52.4758	**0.9993**	**55**	**4.18**	**52.5340**	**0.9993**
	30%	283	45.81	**49.2107**	**0.9962**	158	25.59	49.2068	**0.9962**	**161**	**22.87**	49.2061	**0.9962**
	50%	311	73.27	**43.7880**	**0.9872**	253	59.76	43.7758	**0.9872**	**138**	**29.28**	43.7412	**0.9872**
	70%	301	70.85	**38.1212**	0.9609	235	56.20	38.1201	0.9609	**133**	**30.01**	38.0989	**0.9610**
	90%	398	88.28	29.8083	**0.8483**	343	77.84	29.8122	**0.8483**	**231**	**50.85**	**29.8179**	**0.8483**
Boat	10%	357	34.84	53.2215	**0.9955**	247	24.02	53.2218	**0.9955**	**60**	**4.82**	**53.2564**	**0.9955**
	30%	376	60.73	**46.9295**	**0.9836**	254	43.76	46.9234	**0.9836**	**95**	**14.24**	46.9032	0.9835
	50%	344	79.26	42.7586	**0.9615**	259	59.94	**42.7619**	0.9614	**122**	**24.60**	42.7607	0.9614
	70%	288	66.36	38.3880	**0.9141**	234	54.67	**38.3896**	**0.9141**	**116**	**24.13**	38.3865	**0.9141**
	90%	313	67.42	31.9907	0.7535	317	70.90	31.9961	0.7536	**162**	**35.48**	**32.0405**	**0.7538**

Fig. 3 Restoration results when noise level = 30%.

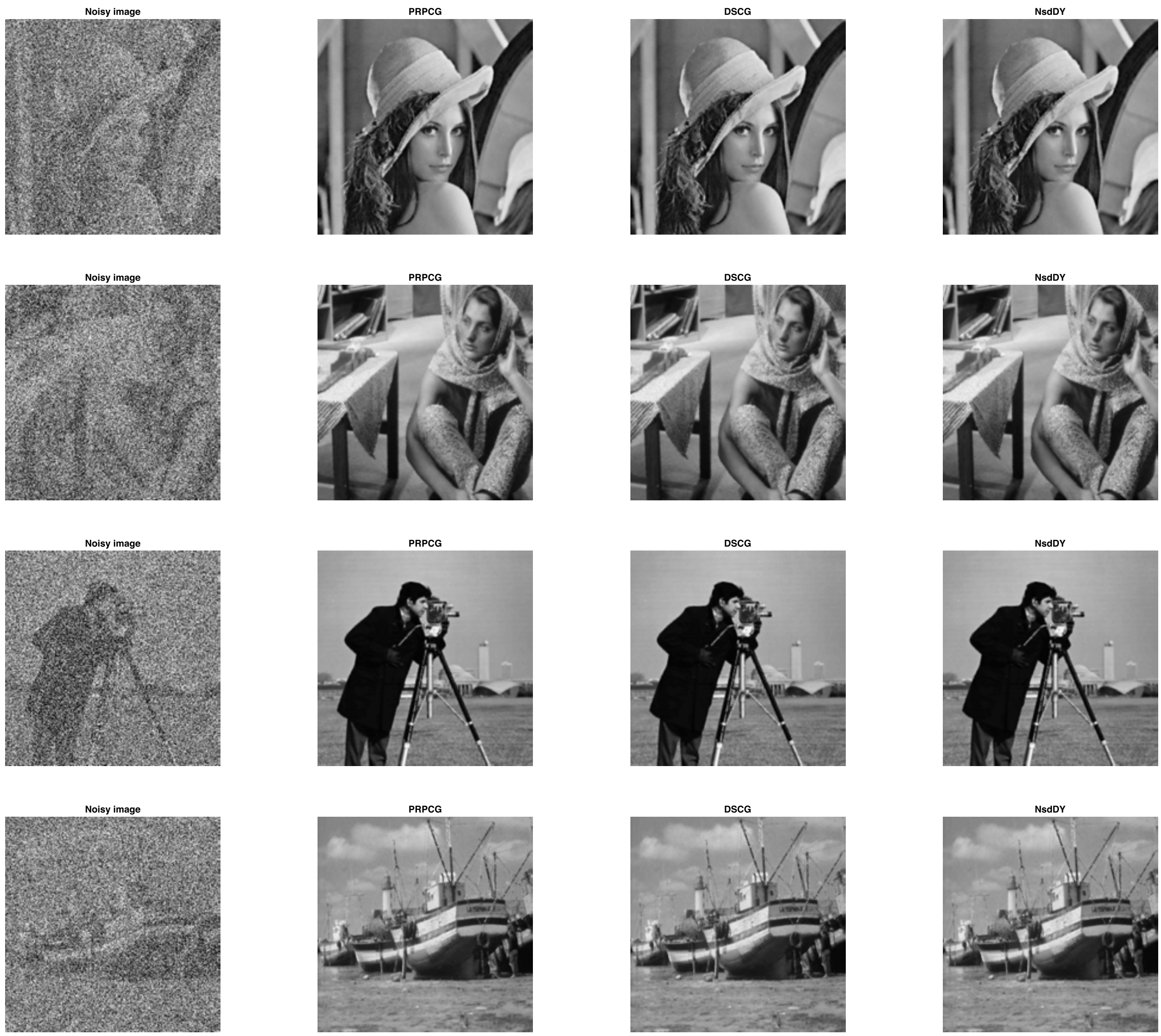

Fig. 4 Restoration results when noise level = 70%.

efficiently. Therefore, the objective function that we are going to minimize in this paper takes the following form

$$\mathcal{F}_\alpha(u) = \sum_{(i,j)\in\mathcal{N}} (2 \cdot S_{i,j} + T_{i,j}). \tag{11}$$

From the results in [7], we see that by minimizing $\mathcal{F}_\alpha(u)$ instead of $F_\alpha(u)$ in the second phase, the denoising quality is not affected. And according to the numerical results, the *PRP conjugate gradient* (PRPCG) method is the most efficient method among the given methods. Later on, Yu et al. [34] proposed a *descent spectral conjugate gradient* (DSCG) method for solving (11). Compared to PRPCG, the DSCG method can reduce the computing time while obtaining the same restored image quality. Total variation regularization-based image denoising methods are also popular, interested readers may refer to [8], [25], [35]. This is not the point of this paper, we do not elaborate here.

5. Numerical Results

In this section, we report some numerical results to demonstrate the performance of NsdDY algorithm for salt-and-pepper impulse noise removal by minimizing (11). Our line search subroutine chooses the initial guess for step size such as $\alpha_k = (g_k^T d_k)/\|d_k\|^2$, and further computes it such that the Wolfe conditions (9) hold. The parameters in the NsdDY method are specified as follows: $\sigma_1 = 10^{-4}$, $\sigma_2 = 0.1$, $C = 0.5$, $\tau = 1.8$. Figure 1 shows the flow chart of our experiments. Figure 2 displays the original test images. All simulations are implemented by MATLAB 2015a on a PC.

To show the performance of NsdDY method, we compare it with the PRPCG method and the DSCG method. It should be emphasized in this paper that we are mainly concerned with the speed of solving the minimization of (11), in which the potential function is $\varphi_\alpha(u) = \sqrt{\alpha + u^2}$ with

$\alpha = 100$. We use Signal-to-Noise Ratio (SNR)† and Structural Similarity (SSIM)†† to measure the quality of the restored images. Stopping criterions of both methods are

$$\frac{|f(u_k) - f(u_{k-1})|}{|f(u_k)|} \le 10^{-6} \text{ and } \frac{\|u_k - u_{k-1}\|}{\|u_k\|} \le 10^{-6}.$$

Experimental results are given in Table 1. We report the number of iterations (NI), the computing time (CPU Time) (in seconds) required for the whole denoising process, the SNR (in dB) of the restored image, and the SSIM values. From Table 1, we can see that the NsdDY method is the fastest of the three methods, and it requires fewer iterations. Moreover, we note that the SNR and SSIM values attained by these three methods are very similar. Figures 3 and 4 show the restoration results obtained by the PRPCG, DSCG, and NsdDY methods, respectively. These results show that the proposed NsdDY method can restore corrupted image quite well in an efficient manner.

In conclusion, we proposed a new modified Dai-Yuan conjugate gradient formula and introduced a new conjugate gradient method called NsdDY. We established its global convergence under the strong Wolfe line search conditions. Simulation experiments verified that NsdDY can significantly reduce the computing time while obtaining the same restored image quality.

It is worth mentioning that in this paper we mainly focus on the image denoising problem, the proposed approach can also be applied to other fields, such as task assignment and route planning [15], [38], which is our research direction in future.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61806004, and in part by the Major Technologies R&D Special Program of Anhui, China (No. 16030901060).

References

[1] N. Andrei, "A Dai-Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstraind optimization," Applied Mathematics Letters, vol.21, no.2, pp.165–171, 2008.
[2] N. Andrei, "Another hybrid conjugate gradient algorithm for unconstrained optimization," Numerical Algorithms, vol.47, no.2, pp.143–156, 2008.
[3] S. Babaie-Kafaki, "A hybrid conjugate gradient method based on a quadratic relaxation of the Dai-Yuan hybrid conjugate gradient parameter," Optimization, vol.62, no.7, pp.929–941, 2013.
[4] S. Babaie-Kafaki and R. Ghanbari, "A hybridization of the Hestenes-Stiefel and Dai-Yuan conjugate gradient methods based on a least-squares approach," Optimization Methods and Software, vol.30, no.4, pp.673–681, 2015.

† $\text{SNR} = 20\log_{10}(\|u\|^2/\|\bar{u} - u\|^2)$, where u and $\bar{u}$ are the original image and the restored image respectively.

†† $\text{SSIM} = [l(x,y)]^{\rho} \cdot [c(x,y)]^{\varsigma} \cdot [s(x,y)]^{\zeta}$, where $\rho > 0$, $\varsigma > 0$ and $\zeta > 0$ are parameters used to adjust the relative importance of the three components. SSIM is proposed to capture the loss of image structure [30].

[5] E.G. Birgin and J.M. Martínez, "A spectral conjugate gradient method for unconstrained optimization," Applied Mathematics and Optimization, vol.43, no.2, pp.117–128, 2001.
[6] R. Fletcher, Practical Methods of Optimization, Volume 1: Unconstrained Optimization, Wiley, New York, 1987.
[7] J.-F. Cai, R. Chan, and B. Morini, "Minimization of an edge-preserving regularization functional by conjugate gradient type methods," in: Image Processing Based on Partial Differential Equations, in: Mathematics and Visualization, Springer, Berlin Heidelberg, pp.109–122, 2007.
[8] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," Journal of Mathematical Imaging and Vision, vol.20, no.1-2, pp.89–97, 2004.
[9] R.H. Chan, C.-W. Ho, and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detector and edge-preserving regularization," IEEE Trans. Image Process., vol.14, no.10, pp.1479–1485, 2005.
[10] R.H. Chan, C. Hu, and M. Nikolova, "An iterative procedure for removing random-valued impulse noise," IEEE Signal Process. Lett., vol.11, no.12, pp.921–924, 2004.
[11] T. Chen and H.R. Wu, "Adaptive impulse detection using center-weighted median filters," IEEE Signal Process. Lett., vol.8, no.1, pp.1–3, 2001.
[12] W. Cheng, "A two-term PRP-based descent method," Numerical Functional Analysis and Optimization, vol.28, no.11-12, pp.1217–1230, 2007.
[13] Y.H. Dai and Y. Yuan, "A nonlinear conjugate gradient with a strong global convergence property," SIAM Journal on Optimization, vol.10, no.1, pp.177–182, 1999.
[14] R. Fletcher and C.M. Reeves, "Function minimization by conjugate gradients," The Computer Journal, vol.7, no.2, pp.149–154, 1964.
[15] B. Gu, Y. Chen, H. Liao, Z. Zhou, and D. Zhang, "A distributed and context-aware task assignment mechanism for collaborative mobile edge computing," Sensors, vol.18, no.8, 2018.
[16] W.W. Hager and H. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search," SIAM Journal on Optimization, vol.16, no.1, pp.170–192, 2005.
[17] M.R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," Journal of Research of the National Bureau of Standards, vol.49, no.6, pp.409–436, 1952.
[18] H. Hwang and R.A. Haddad, "Adaptive median filters: New algorithms and results," IEEE Trans. Image Process., vol.4, no.4, pp.499–502, 1995.
[19] H. Ji and Y. Li, "Block conjugate gradient algorithms for least squares problems," Journal of Computational and Applied Mathematics, vol.317, pp.203–217, 2017.
[20] S.-J. Ko and Y.H. Lee, "Center weighted median filters and their applications to image enhancement," IEEE Trans. Circuits Syst., vol.38, no.9, pp.984–993, 1991.
[21] D.-H. Li and M. Fukushima, "A modified BFGS method and its global convergence in nonconvex minimization," Journal of Computational and Applied Mathematics, vol.129, no.1-2, pp.15–35, 2001.
[22] Y. Liu and C. Storey, "Efficient generalized conjugate gradient algorithms, part 1: theory," Journal of Optimization Theory and Applications, vol.69, no.1, pp.129–137, 1991.
[23] J. Liu and S. Li, "Multivariate spectral DY-type projection method for convex constrained nonlinear monotone equations," Journal of Industrial & Management Optimization, vol.13, no.1, pp.283–295, 2017.
[24] I.E. Livieris, V. Tampakas, and P. Pintelas, "A descent hybrid conjugate gradient method based on the memoryless BFGS update," Numerical Algorithms, https://doi.org/10.1007/s11075-018-0479-1, 2018.
[25] S. Niu, S. Zhang, J. Huang, Z. Bian, W. Chen, G. Yu, Z. Liang, and J. Ma, "Low-dose cerebral perfusion computed tomography image restoration via low-rank and total variation regularizations," Neurocomputing, vol.197, pp.143–160, 2016.

[26] J. Nocedal and S.J. Wright, Numerical Optimization, 2nd edition, Springer, New York, 2006.
[27] E. Polak and G. Ribière, "Note sur la convergence de méthodes de directions conjuguées," Revue Française d'Informatique et de Recherche Opérationnelle, Série rouge, vol.3, no.16, pp.35–43, 1969.
[28] H. Sato, "A Dai-Yuan-type Riemannian conjugate gradient method with the weak Wolfe conditions," Computational Optimization and Application, vol.64, no.1, pp.101–118, 2016.
[29] D.G. Skariah and M. Arigovindan, "Nested conjugate gradient algorithm with nested preconditioning for non-linear image restoration," IEEE Trans. Image Process., vol.26, no.9, pp.4471–4482, 2017.
[30] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error measurement to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.
[31] Z. Wei, S. Yao, and L. Liu, "The convergence properties of some new conjugate gradient methods," Applied Mathematics and Computation, vol.183, no.2, pp.1341–1350, 2006.
[32] S. Yao and L. Ning, "An adaptive three-term conjugate gradient method based on self-scaling memoryless BFGS matrix," Journal of Computational and Applied Mathematics, vol.332, pp.72–85, 2018.
[33] G. Yu, L. Guan, and W. Chen, "Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization," Optimization Methods and Software, vol.23, no.2, pp.275–293, 2008.
[34] G. Yu, J. Huang, and Y. Zhou, "A descent spectral conjugate gradient method for impulse noise removal," Applied Mathematics Letters, vol.23, no.5, pp.555–560, 2010.
[35] G. Yu, W. Xue, and Y. Zhou, "A nonmonotone adaptive projected gradient method for primal-dual total variation image restoration," Signal Processing, vol.103, pp.242–249, 2014.
[36] L. Zhang, "Two modified Dai-Yuan nonlinear conjugate gradient methods," Numerical Algorithms, vol.50, no.1, pp.1–16, 2009.
[37] L. Zhang, W. Zhou, and D. Li, "Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search," Numerische Mathematik, vol.104, no.4, pp.561–572, 2006.
[38] Z. Zhou, J. Feng, B. Gu, B. Ai, S. Mumtaz, J. Rodriguez, and M. Guizani, "When mobile crowd sensing mMeets UAV: Energy-efficient task assignment and route planning," IEEE Trans. Commun., DOI: 10.1109/TCOMM.2018.2857461, 2018.

Wei Xue received the Ph.D. degree in computer science and technology from Nanjing University of Science and Technology, Nanjing, China, in 2017. He is now a lecturer at the School of Computer Science and Technology, Anhui University of Technology. His current research interests include machine learning, computer version, statistical analysis and nonlinear optimization.

Junhong Ren received the M.S. degrees in Software Engineering from Beijing University of Aeronautics and Astronautics, Beijing, China, in 2017. He is now a software engineer at Institute of automation, Chinese academy of sciences. His current research interests include computer version and software testing.

Xiao Zheng received the B.S degree from Anhui University, PR China, in 1997, and the Ph.D. degree in computer science and technology from Southeast University in 2014. He is currently a professor at the School of Computer Science and Technology, Anhui University of Technology. His research interests include service computing, mobile cloud computing and machine learning.

Zhi Liu received the B.E. from University of Science and Technology of China and the Ph.D. degree in informatics from National Institute of Informatics. He is currently an assistant professor at Shizuoka University. He was a Junior Researcher (Assistant Professor) at Waseda University and a JSPS research fellow in National Institute of Informatics. His research interest includes video network transmission, vehicular networks and mobile edge computing. He was the recipient of the IEEE StreamComm2011 best student paper award, 2015 IEICE Young Researcher Award and ICOIN2018 best paper award. He has been a Guest Editor of journals including Wireless Communications and Mobile Computing, Sensors and IEICE Transactions on Information and Systems. He has been serving as the chair for number of international conference and workshops. He is a member of IEEE and IEICE.

Yueyong Liang received the B.E. from Tsinghua University, Beijing, China, in 1992. He is currently a senior engineer at PHIMA Intelligence Technology Co., Ltd., China. His research interests include intelligent manufacturing, cloud computing and internet of things.