# Image deblurring via adaptive proximal conjugate gradient method

**Han Pan[1], Zhongliang Jing[1], Minzhe Li[1], Peng Dong[1]**
[1] School of Aeronautics and Astronautics, Shanghai Jiao Tong University
No. 800, Dong Chuan Road, Min Hang District, Shanghai, China
[e-mail: hanpan@sjtu.edu.cn, zljing@sjtu.edu.cn]
*Corresponding author: Han Pan

## Abstract

It is not easy to reconstruct the geometrical characteristics of the distorted images captured by the devices. One of the most popular optimization methods is fast iterative shrinkage/ thresholding algorithm. In this paper, to deal with its approximation error and the turbulence of the decrease process, an adaptive proximal conjugate gradient (APCG) framework is proposed. It contains three stages. At first stage, a series of adaptive penalty matrices are generated iterate-to-iterate. Second, to trade off the reconstruction accuracy and the computational complexity of the resulting sub-problem, a practical solution is presented, which is characterized by solving the variable ellipsoidal-norm based sub-problem through exploiting the structure of the problem. Third, a correction step is introduced to improve the estimated accuracy. The numerical experiments of the proposed algorithm, in comparison to the favorable state-of-the-art methods, demonstrate the advantages of the proposed method and its potential.

*Keywords:* Image restoration; optimization method; total variation minimization

## 1. Introduction

**I**mage restoration has been studied for a long time [1][2], which is motivated by the spatial (blur) and point degradations of the original scene. This problem is ill-posed problem, varies in different situations. Each elements of the imaging system contributes to the degradations of the system. In the past years, this subject received an increasing amount of interest [42][43] [44][45][46]. The main goal of image restoration is to restore the original image from the degraded scene. A common imaging model consists of three components, i.e. the observed image $s$, the original scene $v$ and the presented noise $\varepsilon$. This model can be summarized as follows

$$s = K \otimes v + \varepsilon \tag{1}$$

where $K$ denotes the convolution operator on clean image $v$. And the noise level $\varepsilon$ varies in different conditions, which includes Gaussian, salt-and-pepper noise, speckle noise [3] and other distortion situations [4][5][6][7]. For recovering more edges and textures or improving the stability of the restoration procedure, some regularization terms [10] are developed to obtain a favorable solution. This procedure requires prior acknowledge necessary about the sampling process or the image data.

A popular regularization term is total variation (TV) proposed by Rudin-Osher and Fatemi [8] (ROF), which can extract the image gradient and preserve edges. And its formation can be presented as follows

$$F(v) = \arg \min_{v \in D} f(v) + \lambda TV(v) = \arg \min_{v \in D} \left\{ \frac{1}{2} \| K \otimes v - s \|_F^2 + \lambda TV(v) \right\}, \tag{2}$$

where $\| \cdot \|_F$ denotes Frobenius norm on the matrix space $D \in R^{m \times n}$, the parameter $\lambda$ is a positive coefficient. It can be noted that TV regularization term would introduce some difficulties, such as non-smooth of objective function and enormous storage requirement. In the past decades, numerous researchers presented vast majority of optimization methods to tackle this problem. On the one hand, some improvements on this term were presented, e.g. non-local based total variation model [9]. On the other hand, various optimization techniques were presented for accelerating the reconstruction procedure, such as gradient projection-like methods, projected Newton method (PN), gradient projection-reduced Newton method (GPRN) and projection-conjugate gradient method [10], which can be viewed as a unified framework of steeped descent algorithm.

Despite of existed optimization schemes, recently, some efficient optimization methods have been developed. These methods consist of proximal splitting schemes [14][15][17], iterative shrinkage/thresholding algorithm (IST) and its variants [11][12][23], sparse reconstruction by separable approximation (SpaRSA) [13]. Among these methods, IST-like algorithms were examined to be simple and efficient minimization methods. Proximal splitting scheme is a key procedure to the success of these methods. Its key idea is to solve the resulting sub-problem. Another mathematical programming method for this problem is the framework of augmented Lagrange method combined with the variable splitting techniques [18]. It has been devoted to image restoration, such as compressive magnetic resonance (CMR) image reconstruction [31], a hot air turbulence removal and object detection method [25], sparse representation (SR) based image restoration [26] and compressed sensing [27][28].

In this paper, an effective image deblurring method, inspired by the woks in [19][30], is proposed. This scheme can provide a generalization representation of proximal conjugate algorithm in [19] for image de-blurring problem. The motivation of the proposed method is to handle the practical limitations in [19][30], e.g. the inaccuracy estimate of the sub-problem and the dimension limitation. Accordingly, the framework of adaptive proximal conjugate gradient (APCG) is proposed. The proposed method can divide into three stages mainly. First, an improved procedure for approximating inverse Hessian matrix is presented for the resulting sub-problem. Second, the sub-problem is solved by an adaptive regularization gradient projection method, characterized by an iterate-varying hyper-ellipsoidal based norm. Then, an efficient solution to this sub-problem is offered, which can be viewed as an incremental method on the problem and backtracking adaptively. Finally, a conjugate gradient method is applied to adjust the approximate error of the second stage. The presented method can exploit the special structure of the problem and improve the estimated accuracy. The convergence properties, presented in [19], guaranteed the effectiveness of the proposed method. At last, the simulation experiments and the performance evaluation, compared with the state-of-the-art methods, demonstrate the effectiveness and efficiency of the proposed method.

The rest of the paper is organized as follows. In section 2, a review on the methods of IST and FISTA is presented. Section 3 demonstrates our framework and a solution to the resulting sub-problem. In Section 4, extensive experiments and discussion in details under several degradation situations are performed. At last, the conclusion on the proposed method is carried out.

## 2. Preliminaries

### 2.1 An overview of proximal splitting methods

In this section, a short description of the schemes, i.e. the basic definitions of the proximal gradient scheme, the proximal Newton splitting scheme [19] and fast iterative/thresholding algorithm [12], are presented.

**Definition 2.1**: proximal gradient scheme is expressed as follows

$$v = pro_\rho^{TV}(x) = \arg\min_{v \in E} \left\{ \lambda TV(v) + \frac{\rho}{2} \|v - x\|_2^2 \right\}, \tag{3}$$

where $\rho$ is a positive scale parameter, viewed as penalty parameter, and sometimes can be configured with a constant parameter, such as Lipschitz constant. And $E$ denotes the matrix space $E \in R^{m \times n}$. Its theoretical results have been discussed and analyzed in [10].

**Definition 2.2**: proximal Newton splitting scheme [19], which can be defined as follows

$$v = pro_H^{TV}(x) = \arg\min_{v \in E} \left\{ \lambda TV(v) + \frac{1}{2} \|v - x\|_H^2 \right\}, \tag{4}$$

where $H$ is a generalization representation to the parameter $\rho$ and assumed to be symmetric and positive definiteness matrix. And $x$ can be viewed as approaching point. The symbol $H$ can be chosen to be a proper matrix. Solving the problem (4) can be considered as a procedure to weight different structures locally, and may bring in an alternate way toward the optimal solution. A general solution to this problem may be difficult because of $TV(x)$ and the determination of the weighting matrix $H$. It can be observed that $H$ may have an influence on the storage requirement and computation complexity, which have not been investigated

fully in the work [11][12]. A general scheme for minimizing TV based prlobem is presented in **Algorithm 1**.

**Algorithm 1.** Proximal splitting scheme for minimizing TV

1. Input: $v_0$, the gradient of $F(v): \nabla F(v_0)$, the step size $t_i$ and the iteration times $K$
2. **for** $i = 1$ to $K$ **do**
3. $\quad x_i = v_{i-1} - t_i \nabla f(v_{i-1})$
4. $\quad v_i = \Theta_{\lambda,i}(x_i) = pro_H^{TV}(x)$
5. **end for**

However, the performance of proximal splitting scheme depends on the implementation, including step size $t_i$, stopping condition, and initialization parameters despite of its apparently simplicity. A simple solution to $\ell_1$-based minimization problem by proximal splitting scheme is IST. The key procedure of this scheme can be given by

$$\Theta_{\lambda,i}(x_i) = (|x_i| - \rho)\,\mathrm{sgn}(x_i), \tag{5}$$

where $\rho$ is a chosen threshold for discriminating different coefficients. It can be seen that the procedure is piece-wise. However, this scheme suffers from low convergence rate and approximate accuracy.

To accelerate the convergence rate of IST, a fast and simple numerical scheme, named fast iterative shrinkage/thresholding algorithm (FISTA) [12], is proposed. It should be noted that the proximal splitting method and IST-like algorithms involve with the operation on the approximation point $v_{i-1} - t_i \nabla f(x_{i-1})$ and the mechanism to obtain the step size $t_i$.

## 2.2 Conjugate gradient method

The conjugate gradient method is popular for solving large scale optimization problems. In this subsection, this scheme is intruded to remedy the local increment in FISTA [12]. Then, the backward step is employed for the proposed algorithm. Many of the special effects have been devoted to the conjugate gradient method [32]. A detailed discussion of the theoretical and performance picture of conjugate gradient method have been presented. The convergence results of the proposed method are similar to the work in [7]. So, the proofs of these results are ignored. A general procedure of conjugate gradient method, presented in algorithm 2, is introduced to the proposed algorithm.

**Algorithm 2.** General congujate gradient method for non-smooth function

1. Input: $v_k$, the gradient of $F(v): \nabla F(v_k)$, the step size $\beta_k$
2. Repeat untile convegence
3. $p_k = -\nabla F(v_k)$
4. $d_k = p_k + \beta_k d_{k-1}$
5. % choose a step length $\alpha_k > 0$
6. $\quad \alpha_i = C_i(p_k, v_k)$
7. $\quad v_{k+1} = v_k - \alpha_k d_k$

The symbol $C(\cdot)$ denotes the procedure for obtaining the conjugate gradient update parameter that belongs to a real number.

$$C_i^{FR}(\hat{v}_{i+1}, v_i) = \frac{\left\| \nabla F(\hat{v}_{i+1}) \right\|^2}{\left\| \nabla F(v_i) \right\|^2} \tag{6}$$

$$C_i^{PRP}(\hat{v}_{i+1}, v_i) = \frac{\left\langle \nabla F(\hat{v}_{i+1}), \nabla F(\hat{v}_{i+1}) - \nabla F(v_{i+1}) \right\rangle}{\left\| \nabla F(v_i) \right\|^2} \tag{7}$$

$C_i^{FR}$ was developed by Fletcher and Reeves [35]. $C_i^{PRP}$, proposed by Polak and Ribiere and Polyak [36][37], behaves a better way in a practical way. Although there are some classical methods for choosing this parameter, Eq.(7) is mainly employed in this paper. We extended the algorithm by combining classical general descent algorithm and line search by choosing adaptive step size. The details of these steps are presented in **algorithm 3**.

## 3. Proposed method

### 3.2 Adaptive proximal-conjugate gradient method

The principle of the proposed framework is to solve the problem Eq.(2) through striking a balance between the effort expended at each iteration to solve the corresponding sub-problem and the number of iterations required for convergence. The main challenges of the proposed framework have three aspects. On the one hand, when the dimensions growing, especially, the calculation of the approximate inverse Hessian matrix $H_i^{-1}$ is required to be a suitable formation. Unfavorably, FISTA has not been explored this problem. The accuracy of the proposed framework is completely determined by the approximate accuracy of the sub-problem. Furthermore, the relative cost of the steps of many optimization methods changes when the problem becomes large. Then, certain difficulties arise.

The proposed method for total variation minimization problem, including three stages, is presented in **algorithm 3**. First stage in the step 3 is provided for overcoming the limitations in [19][30]. It should be noted that this matrix is assumed to be symmetric and positive matrix. Second stage in the step 4 can convert Eq.(3) into a general formation Eq.(4). Or more precisely, the resulting problem can be viewed as hyper-ellipsoidal norm based sub-problem defined in Eq.(8). It should be noted that this generalized representation can product ellipsoidal-norm changes locally. The third stage, including the steps from 5 to 8, is employed for adjusting the approximate error through introducing conjugate gradient method.

<div align="center"><strong>Algorithm 3.</strong> The framework of APCG</div>

| |
|---|
| 1.  Input: the observation data $x$, the stopping criterion $tol$, the step size $t_i$, the conjugate parameter $\alpha$ and the iteration itmes $K$ |
| 2.  **for** $i = 1$ to $K$ **do** <br>       % first stage |
| 3.       $z_{i+1} = v_{i-1} - H_i^{-1} \nabla f(v_{i-1})$ <br>       % second stage |
| 4.       $\hat{v}_{i+1} = pro_{H_i}^{TV}(z_{i+1})$ <br>       % third stage |

5.          $\varphi_{i+1} = v_i - \hat{v}_{i+1}$
            % Calculate the conjugate parameter $\alpha_i$
6.          $\alpha_i = C_i(\hat{v}_{i+1}, v_i)$
7.          $d_{i+1} = -\varphi_{i+1} + \alpha_i d_i$
            % a suitable $\alpha_i$ step size $t_i$
8.          $v_{i+1} = v_i + t_i d_{i+1}$
9.   **end for**
10.  until terminated by the suitable stopping conditions

The matrix $H_i^{-1}$ is not presented to approximate the toeplitz matrix simply. Its goal is to provide a more accurate and positive definiteness penalty parameter matrix to the sub-problem in step 4. The practical importance of $H_i^{-1}$ can be presented in many ways. First, based the presentation in the literature [39], the matrix $H_i^{-1}$ generalizes the representation of penalty parameter and dominates the progress toward the feasible solution. Second, empirical experiments, presented in FISTA and alternative direction multiplier method (ADMM) [40], indicate that the selection of $H_i^{-1}$ is very important. Some theoretical analyses were presented in [19][39][41]. $H_i^{-1}$ can be calculated in **Algorithm** 4. It can be noted that this matrix is updated iterate-to-iterate. This algorithm can provide various formations of approximate inverse Hessian matrix.

**Remark 1:** An approximate solution of the step 4 is presented in **algorithm** 4. This solution of the sub-problem is improved from the work in [12], which can be expressed as follows

$$pro_{H_i}^{TV}(z_{i+1}) = \arg\min_{v \in E}\left\{\lambda TV(v) + \frac{1}{2}\|v - z_{i+1}\|_{H_i}^2\right\}. \tag{8}$$

We provide this algorithm without proofs. The advantage of this algorithm can exploit the structure of $H_i^{-1}$ matrix directly, indicates the computation of the true Hessian matrix is bypassed. All the key aspects of **algorithm 4** is the main difference to the work in [19][30].

**Algorithm 4.** A general approximation method to $H_i^{-1}$

1.   Input: the iterate $i, v_{i-1}, v_i$, $0 < \pi < 1$, $0 < \psi_{\min} < \psi_{\max}$ and
     stopping criteria $tol$
2.   $\varphi_i = v_i - v_{i-1}$
3.   $\theta_i = \nabla f(v_i) - \nabla f(v_{i-1})$
4.   **if** $i == 1$ **then**
5.       $H_{tmp}^{-1} = \psi_{\min} I$
6.       $e_i = 0$
7.   **else**
8.       $\psi_{BB} = \dfrac{\langle \varphi_i, \theta_i \rangle}{\|\theta_i\|^2}$
9.       $\psi_{BB} \mapsto [\psi_{\min}, \psi_{\max}]$

$$10. \quad H_{tmp}^{-1} = \pi \psi_{BB} I$$

$$11. \quad \delta = \varphi_i - H_{tmp}^{-1} \theta_i$$

12. **if** $\langle \delta, \theta_i \rangle \le tol \times \|\theta_i\|_F \|\delta\|_F$ **then**

$$13. \quad e_i = 0$$

14. **else**

$$15. \quad e_i = (\varphi_i - H_{tmp}^{-1} \theta_i) / \sqrt{\langle \varphi_i - H_{tmp}^{-1} \theta_i, \theta_i \rangle}$$

16. **end if**

17. **end if**

$$18. \quad H_i^{-1} = H_{tmp}^{-1} + O(e_i e_i^T)$$

**Remark 2**: The proposed method is improved from the work in [12]. It is extended to a more feasible solution by saving the storage requirement. This weighting scheme can improve the performance and reliability of the algorithm through searching the descent direction adaptively.

**Remark 3**: The symbol $O(\cdot)$ stands for the operation on $e_k e_k^T$, which can produce some special structure matrices [20], such as tri-diagonal matrix, up-arrowhead and down-arrowhead. In this paper, the tri-diagonal matrix is taken mainly.

**Remark 4**: The proposed framework is different from the works [19][31] in several ways. First, the proposed framework extends the original scheme [31] by introducing the conjugate gradient method, which can reduce the approximate error in some degree. Second, the proposed method deals with the resulting problem by taking account into the special pattern of the matrix. A further benefit of this modification is that lower storage requirement and computational cost is required. Third, the computation of exact Hessian matrix is avoided, which can reduce the computational demand significantly when the dimension growing.

**Remark 5**: For ADMM [40] and FISTA [12] based methods, there are many differences in many ways. First, the main difference lies in the update and the form of the penalty parameter [12]. In contrast, the proposed algorithm generalizes the penalty parameter into a full matrix penalty formation. Second, the solution to $\ell_1$-based sub-problem [40] is different to the proposed method. The application of IST in [40] may smooth local structures or texture components in some extent. The geometric illustration, offered in the reference [39], indicates the choice and the form of the penalty parameter has an important influence on the convergence process. In summary, the essential features of the proposed method are the bounded and hyper-ellipsoidal adaption algorithm. The convergence analysis of the proposed framework confirmed that it should be a novel algorithm for this problem.

### 3.4 Some theoretical results of APCG

**Some notations:** An iterative sequence $v_i$ converges to $v^*$. Based on the most effective technique by comparing the improvement at each step to the improvement at the previous step. An algorithm converges $q$-quadratically is denoted as follows:

$$\|v_{i+1} - v^*\| = Q(v_i - v^*), \tag{9}$$

and the function $f(v)$ assumed to be locally convex and $\nabla^2 f$ to be locally Lipschitz continuous. Similarly, $q$-superlinearly is defined as follows

$$\left\|v_{i+1}-v^*\right\|=Qs(v_i-v^*),\tag{10}$$

The symbols $Q$ and $Qs$ are related to the rates of convergence of the presented method. And a method would be practicable if iterative process is the closeness to the required point with some rapidity.

**Convergence analysis of APCG:** some assumptions and theoretical results against the iterative sequence generated by the APCG are presented.

**Lemma 1** [19]. Suppose $f(v)$ is twice continuously differentiable and $\nabla^2 f$ is locally Lipschitz continuous with constant $\Phi$. If $H_i^{-1}$ satisfies Dennis-More criterion and their eigenvalues are bounded, then the unit step length satisfies the sufficient descent condition after sufficiently many iterations.

**Theorem 2.** Suppose (1) $f(v)$ is locally strongly convex with constant $m$, (2) $\nabla^2 f$ is locally Lipschitz continuous with constant $\Phi$, and (3) then the hyper-ellipsoidal adaption algorithm converges $q$-superlinearly to $v^*$.

**Proof.** The assumptions of Lemma 1 are satisfied; then the stage of the conjugate gradient satisfy the sufficient descent condition after many iterations:

$$v_{i+1}=v_i+t_i d_{i+1}\tag{11}$$

where $t_i$ assumed to be equal to 1. Based on the convergence results (c.f. Theorem 3.3 in [19]), i.e. converges $q$-quadratically

$$\left\|v_{i+1}-v^*\right\|\le\frac{\Phi}{m}\left\|v_i^{phe}-v^*\right\|+\left\|d_{i+1}-\nabla v_i^{phe}\right\|,\tag{12}$$

where $\nabla v_i^{phe}=\varphi_{i+1}$ denotes the proximal hyper-ellipsoids direction. It should be noted that the eigenvalues of hyper-ellipsoids matrix provided by Algorithm 3 is box-constrained, then the last term can be reformatted as follows

$$\left\|d_{i+1}-\nabla v_i^{phe}\right\|\le\sqrt{\frac{1+\theta}{m}}\left\|(\nabla^2 f(v_i)-H_i)d_{i+1}\right\|^{0.5}\left\|d_{i+1}\right\|^{0.5}.\tag{13}$$

The details of Eq. (13) referred the Proposition 3.5 in [19]. It is notable that the $\nabla^2 f$ and $d_{i+1}$, then it can be observed that

$$\left\|(\nabla v(x_i)-H_i)d_{i+1}\right\|\le\Phi\left\|v_i-v^*\right\|\left\|d_{i+1}\right\|+o(\left\|d_{i+1}\right\|).\tag{14}$$

Due to the results (Lemma 2 in [38]) and $q$-quadratically convergence (Theorem 3.3 [19]), then we can get

$$\left\|d_{i+1}\right\|\le Q(\left\|v_i-v^*\right\|)+\theta\left\|v_i-v^*\right\|.\tag{15}$$

putting these expressions into Eq.(13), then we can get

$$\left\|d_{i+1}-\nabla v_i^{phe}\right\|\le Qs(\left\|v_i-v^*\right\|),\tag{16}$$

considering the Eq.(16) and Eq.(12), then we can obtain

$$\left\|v_{i+1}-v^*\right\|\le\frac{\Phi}{m}\left\|v_i^{phe}-v^*\right\|^2+o(v_i-v^*),\tag{17}$$

at last, it can be seen that $v_k$ converges to $v^*$ $q$-superlinearly.

## 3.5 An approximate solution to the problem (2)

The fundamental issues of the framework of APCG are the presence of $H_i^{-1}$ and the corresponding sub-problem. Moreover, the computational expenses of this framework are referred to two estimated procedures mainly, i.e. the solution to the sub-problem Eq.(6) and the conjugate gradient method. Based on $H_i^{-1}$ provided by **algorithm** 4, an improved scheme from the work [12], is proposed in **algorithm** 5.

<div>

**Algorithm 5.** A hyper-ellipsoids regularization gradient projection method

1.  Input: the variable $v$, the obtaned $H_i^{-1}$, $\beta_{\max} = \max(diag(H_i^{-1}))$,
    the regularization parameter $\lambda$, inner iterate number $N$
2.  Output: the optimal solution $v^*$
3.  Initialization:
4.  $(u_1, r_1) = (x_0, y_0) = (0^{(m-1) \times n}, 0^{m \times (n-1)}), t_1 = 1$
5.  **for** $i = 1$ to $K$ **do**
6.  $\quad z_i = \dfrac{1}{8\lambda\beta_{\max}} L^T (P_B(v - \lambda H_i^{-1} L(u_k, r_k)))$
7.  $\quad (x_k, y_k) = P_D[(u_k, r_k) + z_k]$
8.  $\quad t_k = \dfrac{1 + \sqrt{1 + 4t_k^2}}{2}$
9.  $\quad (u_{k+1}, r_{k+1}) = (x_k, y_k) + \dfrac{t_k + 1}{t_{k+1}}(x_k - x_{k-1}, y_k - y_{k-1})$
10. **end for**
11. $v^* = P_B[v - \lambda H_i^{-1} L(x_N, y_N)]$ % up to the stopping point

</div>

**Remark 6**: The operation $P$ stands for the orthogonal projection operator on the convex set $C$. In this paper, the set $C$ is defined as a bound condition. For the symbol $D$, it denotes similar convex projection on the dual space. The real definition of the $L$ and the adjoint operator $L^T$ can be found in [12].

## 4. Numerical experiments and discussion

### 4.1 The experiment settings

In this subsection, the experimental settings for evaluating the referred methods are presented. The IST [23], the FISTA [12], FTVD based method [40] (i.e. ADMM) and the proposed method are implemented in Matlab. The assessment procedures are performed on a work station shipped with an Intel processor Core2 Quad 2.66 GHz and 3G RAM. Some synthetic examples are applied to the proposed method, i.e. Gaussian (Gau) blur, uniform (Uni) blur and disk based blur. The Gaussian noise with zero mean and a variance $\sigma^2 = 10^{-3}$ is added to the blurred image. In this paper, two criteria are used to measure the performance numerically, i.e. signal-to-noise (SNR) and structural similarity (SSIM) index [34].

The settings of the experiments are described in **Table 1**. All experiments are designed and performed for some goals, such as numerical behavior of the sub-problem, sensitivity analyses of penalty parameter and noise-levels, a wide range of datasets and blurring operations. In order to have a fair comparison between FTVD based method and our method, the regularization parameter $\lambda$ is chosen to be $= 5 \times 10^{-4}$ from the row 3 to 5.

All the required methods are performed with default parameters. Meanwhile, all the pixel intensity values are rearranged to the box bound from 0 to 1. And the inner iterative numbers for the sub-problem are chosen to 10. The image dataset can be downloaded from http://www.cipr.rpi.edu/resource/stills/index.html. CIPR datasets, contain about 101 images, including aerial (7), brodatz (53), canon (18) and Kodak (23). The code of FTVD is obtained from the web site: http://www.caam.rice.edu/~optimization/L1/ftvd.

**Table 1.** The general configuration of the experiments

| No. | Name | Blur kernel | Kernel Size | Variance | $\lambda$ | $\sigma^2$ |
|---|---|---|---|---|---|---|
| 1 | Airplane | Gau | 9 | 4 | 0.0010 | 0.001 |
| 2 | Lena | Gau | 9 | 4 | 0.0010 | 0.001-0.01 |
| 3 | Lena | Gau | 9 | 4 | 0.0005 | 0.001 |
| 4 | Lena | Gau | 9 | 4 | 0.0005 | 0.001 |
| 5 | CIPR | Gau, disk, Uni | 9 | 4 or / | 0.0005 | 0.001 |

## 4.2 The first experiment on Airplane

In this subsection, the visual perception and numerical performance of the experiment on the airplane ($512 \times 512$) are demonstrated. The second row of **Table 2** exhibits the maximum scores of the SNR and the SSIM, which are obtained by performing 100 iterates. It can be notable that the proposed method is better than other methods separately, which is more than 1dB in term of the SNR. In another viewpoint on these objective results, the proposed method performs a better behavior. **Fig. 1** demonstrates the observed data and the visual results of these methods. The restoration image is more similar to the original image. It can be noted that the proposed algorithm restores more local features, for example the textures, edges and shapes. As an evaluation of the reconstruction process, **Fig.2** displays the curves of the SNR and SSIM at every iteration. The quantitative comparison of these methods indicates that our method is better than the other methods.



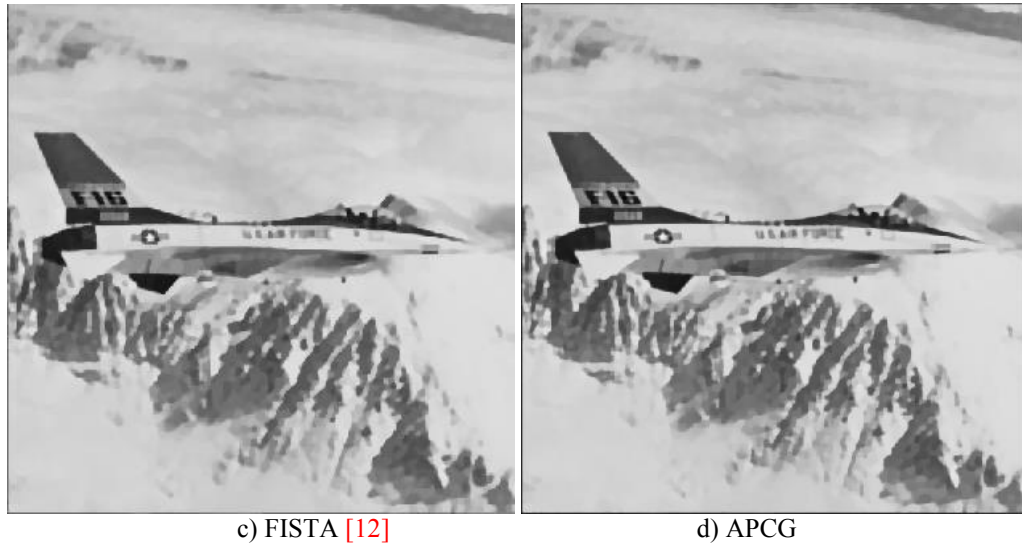a) The blurry and noisy image                    b) IST [23]

c) FISTA [12]                                        d) APCG
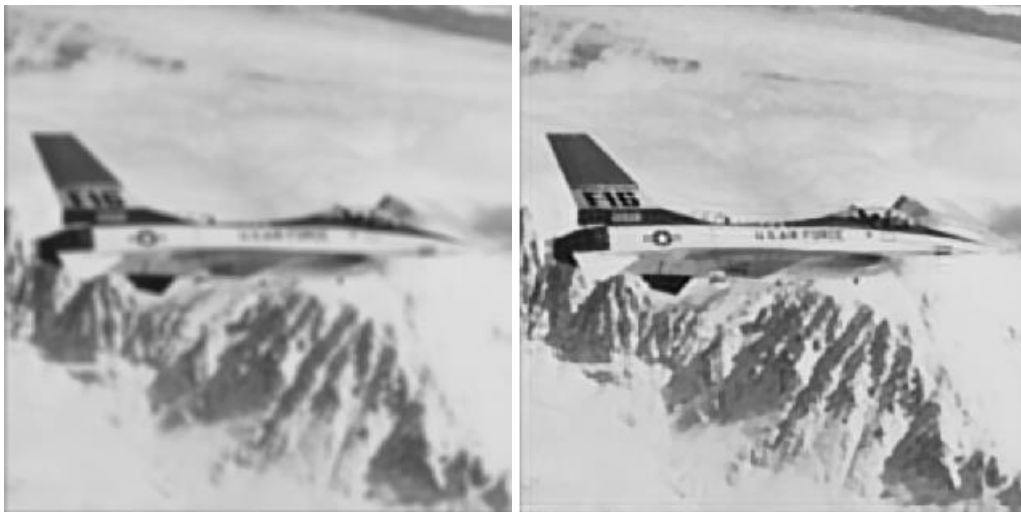
**Fig. 1.** The reconstruction results of the first experiment ($\lambda = 1 \times 10^{-3}$). (a) The observation image, (b) for IST, (c) for FISTA and (d) for APCG.
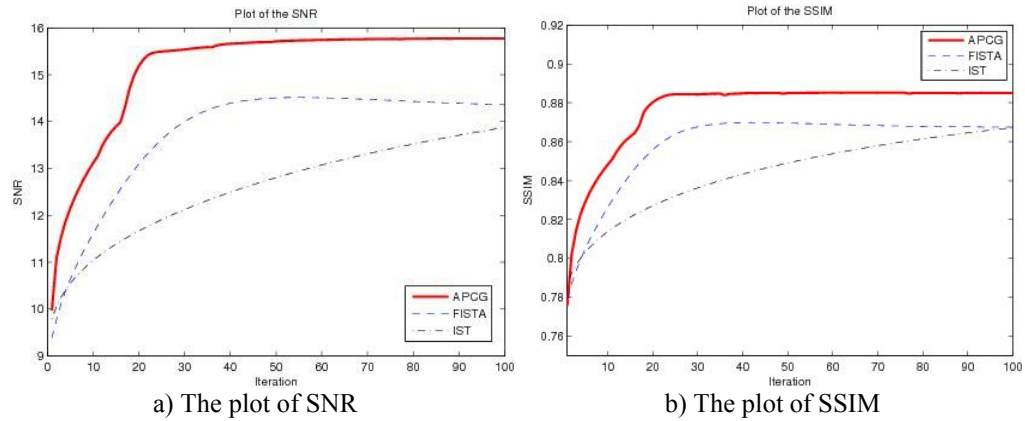


a) The plot of SNR                                        b) The plot of SSIM

**Fig. 2.** The SNR and the SSIM of the first experiment ($\lambda = 1 \times 10^{-3}$). (a) for the SNR, (b) for the SSIM.

**Table 2.** The numerical results of the 1st and 2nd experiments ($\lambda = 10^{-3}$)

| Image | Algorithm | SNR(dB) | SSIM | $I_{tol}$ |
|---|---|---|---|---|
| Airplane | IST [23] | 13.8592 | 0.8670 | **100** |
| | FISTA [12] | 14.5131 | 0.8699 | 537 |
| | APCG | **15.7729** | **0.8852** | 770 |
| Lena | IST [23] | 12.6211 | 0.7923 | **100** |
| | FISTA [12] | 13.2072 | 0.7992 | 918 |
| | APCG | **14.0288** | **0.8177** | 776 |

|  |  |
|---|---|
| a) The blurry and noisy image | b) IST [23] |
| c) FISTA [12] | d) APCG |

**Fig. 3.** The reconstruction results of the 2$^{nd}$ experiment. (a) The observation image, (b) for IST, (c) for FISTA and (d) for APCG.
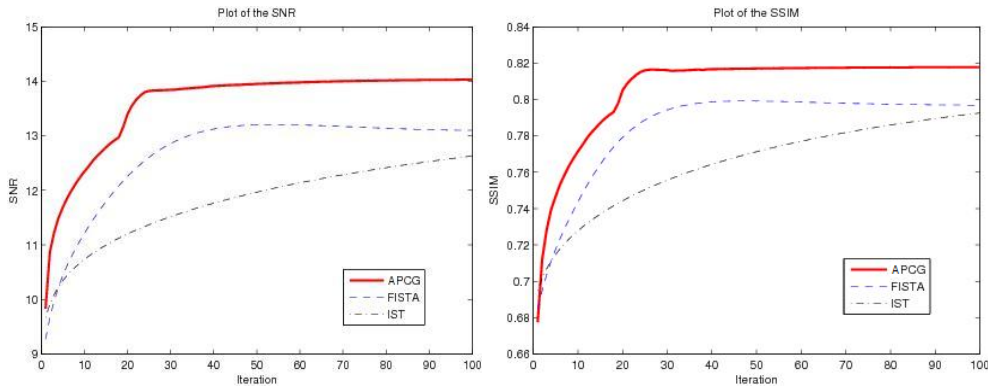


**Fig. 4.** The SNR and SSIM of the 2$^{nd}$ experiment ( $\lambda = 1 \times 10^{-3}$ ). (a) for the SNR, (b) for the SSIM.

## 4.3 An evaluation on the solutions to the sub-problem (8)

In this subsection, an assessment on the sub-problem $pro_{H_i}^{TV}$ was performed. It can be noted that is a identify matrix in FISTA based method. First, the reconstruction results were examined in Lena image. Second, the times of inner iteration for solving $pro_{H_i}^{TV}$ was measured and evaluated. These experiments can provide a comprehensive view on the numerical behavior of the proposed method.

The visual outcomes are presented in **Fig.3**. In comparison to the other methods, the proposed method can restore a sharper and clear latent image. As shown in **Table 2**, it is notable that the numerical results, based on the indices of the SNR and SSIM, imply that the proposed method is better than the other methods again. The total iterate numbers of APCG based method are presented in five column of **Table 2**.
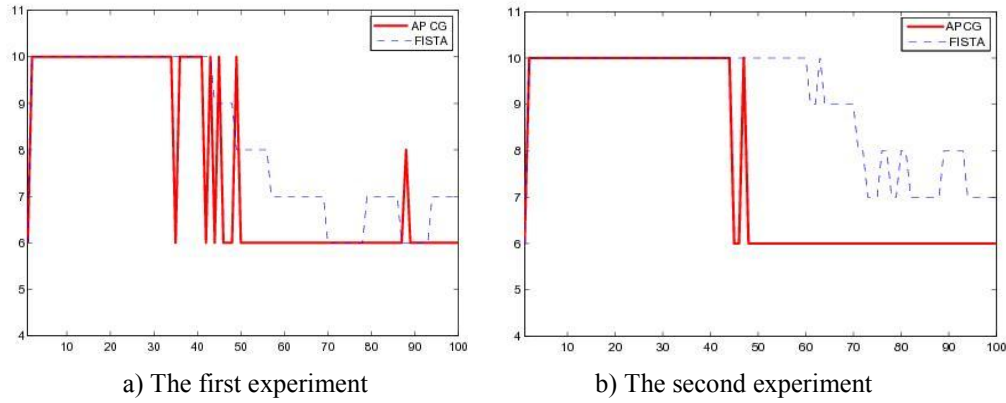


a) The first experiment                          b) The second experiment

**Fig. 5.** The inner iteration numbers of the first and second experiments ( $\lambda = 1 \times 10^{-3}$ ).

The influence on the resulting sub-problem based is presented in **Fig. 5**. The five column of **Table 2** presents the total iterate numbers $I_{tol} = \sum_{i=1}^{100} I_i$ by summarizing all the minor iterates $I_i$ when solving Eq.(6). It is evident that the IST conducts the lowest iterate numbers. However, compared to the FISTA ( $I_{tol} = 837$ ), the proposed method achieves a more performance improvement and involves with a lower total iterate numbers ( $I_{tol} = 770$ ). This phenomenon can be observed in **Fig. 5** carefully. Although the curve of FISTA decreases from 10 to 6 slowly, it can be noted that the outcome of the APCG is mainly equal to 6 and hold it for many times, which can be observed in **Fig. 5** about $I_{i=50-100}$.

## 4.4 Sensitivity analysis of penalty parameters

In this subsection, some experiments are performed for observing the numerical behavior of FISTA and FTVD in various penalty parameters. In another view on these numerical results, it can be seen that the update process may not be effective in any iterate. In **Table 3**, the symbol * denotes the fail of the algorithm. The symbol $\rho$ defined in FISTA is penalty parameter, and $\beta$ for FTVD. It is notable that the generalization representation of the proposed framework suggests an alternative approach to this problem. When $\beta$ in FTVD decreases, the behavior of the SNR history would act a different way.

**Table 3.** The evaluation of the FISTA( $\rho$ ) and FTVD( $\beta$ ) on 'Lena' image( $\lambda = 5 \times 10^{-4}$ )

| **FISTA( $\rho$ )** | $\rho = 1$ | $\rho = 2$ | $\rho = 4$ | $\rho = 6$ | $\rho = 8$ | APCG |
|---|---|---|---|---|---|---|
| SNR(dB) | * | 13.8268 | 13.8261 | 13.7816 | 13.6977 | **14.8286** |
| SSIM | * | 0.8183 | 0.8183 | 0.8183 | 0.8176 | **0.8429** |
| FTVD( $\beta$ ) | $\beta = 1$ | $\beta = 2$ | $\beta = 4$ | $\beta = 6$ | $\beta = 8$ | APCG |
| SNR(dB) | 14.7884 | 14.2542 | 13.9982 | 14.0300 | 14.0204 | **14.8286** |
| SSIM | 0.8372 | 0.8289 | 0.8184 | 0.8190 | 0.8194 | **0.8429** |

## 4.5 Sensitivity analysis of noise-levels

**Fig. 6** demonstrates the influence of the noise-levels on these optimization methods. It also can be noted that the IST generates a more stable curve, which implies that it is not sensitive to the parameter $\sigma^2$. These appearances may relate to the hard thresholding on the latent image. This operation may keep regional features away. For the APCG, although it behaves quite more sensitive to noise-levels increasingly, but it should be noted that its reconstruction results, evaluated by the indices of SNR and SSIM, are better than other methods roughly.
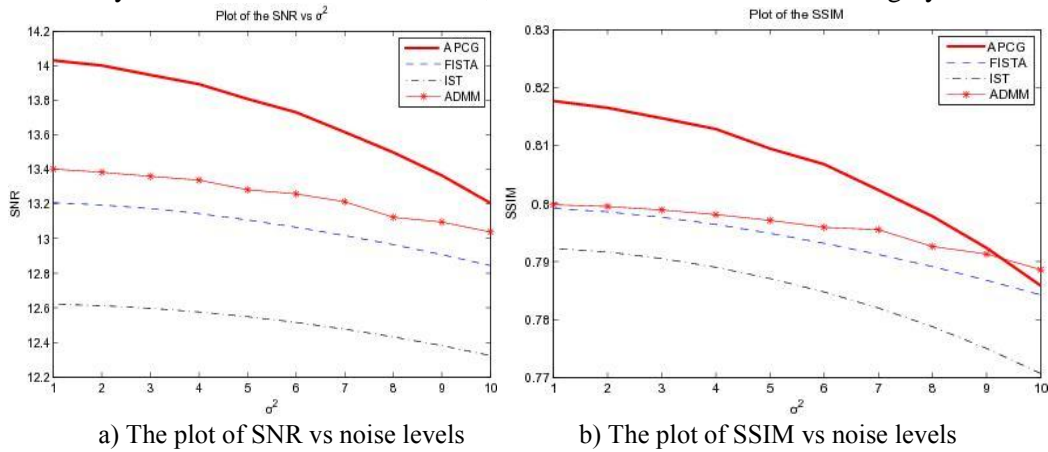


|                     a) The plot of SNR vs noise levels                     |                    b) The plot of SSIM vs noise levels                    |

**Fig. 6.** Sensitivity analysis of noise-levels on Lena ($\lambda = 1 \times 10^{-3}$), (a) for the SNR, (b) for the SSIM.

## 4.6 The numerical results of CIPR still images

The performance of the proposed method is verified in many extended experiments. These experiments are performed in three different blur operations. Similar to the previous experiments, the better numerical results of the referred methods are displayed in bold.

The outcomes of all these methods for typical degradation conditions are presented in **Table 4** and **Table 5**, separately about SNR and SSIM. It can be seen that our method fulfills a substantial performance improvement in comparison to the other methods. The last two rows present the mean and variance of the numerical results of these blurring conditions. The outcomes of these experiments about these four datasets indicate that our method can be applied to a wider variety of images. Meanwhile, it can be seen that the means and variances presented in **Table 4** and **Table 5** demonstrate that the proposed method behaves more stable than other methods. For the outcomes of the variance of SSIM, FTVD based method is better than other methods. Our approach is a natural extension in the context of proximal splitting scheme. In part, this is due to the introduction of APCG framework.

**Table 4.** The average SNR of CIPR still images($\lambda = 5 \times 10^{-4}$)

| Datasets | Blur | IST | FISTA | FTVD | APCG |
|----------|------|-----|-------|------|------|
| Aerial | Gaussian | 7.7633 | 8.2151 | 8.8933 | **9.4915** |
|  | Uniform | 8.1361 | 8.6773 | 9.3417 | **10.1482** |
|  | Disk | 8.9940 | 9.4742 | 9.9633 | **10.9053** |
| Brodatz | Gaussian | 5.7726 | 6.2479 | 7.3100 | **8.4392** |
|  | Uniform | 5.8361 | 6.4789 | 7.6429 | **8.6181** |
|  | Disk | 7.6894 | 8.1613 | 8.9471 | **10.3089** |
| Canon | Gaussian | 17.3769 | 18.4123 | 18.5343 | **19.3459** |
|  | Uniform | 17.6414 | 18.8424 | 19.3171 | **19.8065** |
|  | Disk | 18.9354 | 19.9118 | 19.7843 | **21.1592** |

| Image | Blur | | | | |
|---|---|---|---|---|---|
| Kodak | Gaussian | 9.8484 | 10.4914 | 10.9186 | **11.6067** |
| | Uniform | 9.8957 | 10.5536 | 11.0471 | **11.7211** |
| | Disk | 11.3457 | 11.9445 | 12.1814 | **13.2898** |
| Average | Gaussian | 10.1903 | 10.841675 | 11.4140 | **12.2208** |
| | Uniform | 10.3773 | 11.1380 | 11.8372 | **12.5734** |
| | Disk | 11.7411 | 12.3729 | 12.7190 | **13.9158** |
| Variance | Gaussian | 5.0718 | 5.3366 | 4.9712 | **4.9292** |
| | Uniform | 5.1200 | 5.3994 | 5.1766 | **4.9856** |
| | Disk | 5.0291 | 5.2649 | **4.8999** | 4.9977 |

**Table 5.** The average SSIM of CIPR still images ( $\lambda = 5 \times 10^{-4}$ )

| Image | Blur | IST | FISTA | FTVD | APCG |
|---|---|---|---|---|---|
| Aerial | Gaussian | 0.6662 | 0.6763 | 0.6956 | **0.7100** |
| | Uniform | 0.6875 | 0.6988 | 0.7168 | **0.7360** |
| | Disk | 0.7404 | 0.7451 | 0.7672 | **0.7878** |
| Brodatz | Gaussian | 0.6737 | 0.6919 | 0.7162 | **0.7376** |
| | Uniform | 0.6816 | 0.7076 | 0.7330 | **0.7538** |
| | Disk | 0.7630 | 0.7750 | 0.7985 | **0.8219** |
| Canon | Gaussian | 0.9077 | 0.9162 | 0.8798 | **0.9266** |
| | Uniform | 0.9078 | 0.9217 | 0.8883 | **0.9322** |
| | Disk | 0.9304 | 0.9343 | 0.9047 | **0.9455** |
| Kodak | Gaussian | 0.7320 | 0.7464 | 0.7606 | **0.7737** |
| | Uniform | 0.7512 | 0.7661 | 0.7814 | **0.7975** |
| | Disk | 0.7830 | 0.7902 | 0.8089 | **0.8245** |
| Average | Gaussian | 0.7449 | 0.7577 | 0.7631 | **0.7870** |
| | Uniform | 0.7570 | 0.7736 | 0.7780 | **0.8049** |
| | Disk | 0.8042 | 0.8112 | 0.8198 | **0.8449** |
| Variance | Gaussian | 0.1123 | 0.1098 | **0.0824** | 0.0966 |
| | Uniform | 0.1053 | 0.1031 | **0.0773** | 0.0887 |
| | Disk | 0.0858 | 0.0842 | **0.0592** | 0.0691 |

## 4.7 Discussion

In this paper, we have show that the efficiency of APCG, which involves the unified framework and its approximate solution. It is obvious that the presented method acts a better way than other methods in term of the improvement on the SNR and SSIM, the significant decrement in total iterate numbers $I_{tol}$ , sensitivity analysis of penalty parameters and noise-levels. The numerical results and the visual inspection of the restoration pictures demonstrate an involvement of the generalization representation of the proximal term and an efficient strategy to linear search effectively, leading to the success of the proposed framework. In another viewpoint on this condition, the association of the generalized proximal term with the conjugate gradient restores more high frequencies information from the degraded image. Taken together, these results indicate that the combination of these elements is essential to the achievement of the reconstruction performance. In term of the indices of the SNR and the SSIM in CIPR datasets and blurring conditions, the outcomes of the proposed method are available for a variety of datasets. In the meanwhile, an insight on the inner iterate numbers $I_i$ at a single iterate of the APCG, presented in **Fig. 5(a)** and **Fig. 5(b)**, implies that the approximate error at every iterate can be alleviated in some degree. The summary of total iterations exhibited in **Table 2** verifies the effectiveness of the proposed method again. In

contrast to other methods, this difference may relate to the method used. It is possible that the composite framework with the hyper-ellipsoidal norm based proximal term and the backtracking line search mechanism lead to the success of the APCG.

However, there are some limitations on our method. First, the FTVD based method, implemented by ADMM, provides a well balance between the complexity and the run-time of the algorithm. However, we argued that the proposed framework can provide a unified representation of penalty parameter. Second, the proposed method is required to calculate the **Algorithm 4**. Although this may introduce some computational expense, the reduction in the total iterate numbers of the solution to the sub-problem compensates this extra computational cost to some extent. Third, the storage of the matrix $H_k^{-1}$ may be required to be huge when the size of the problem growing.

In this paper, an alternate approach is proposed to handle the existed problems of the work in [19][30]. Despite of the extra issues of APCG based method, the evaluation and outcomes of the experiments confirm us that the proposed method is superior to the other methods and is suitable for large-scale optimization problems. Thanks to the accurate estimation of the iterate variable, the proposed method behaves a better way in these experiments.

# 5. Conclusion

In this paper, an efficient optimization framework for the image de-blurring problem, named adaptive proximal conjugate gradient method, is presented. This framework can be viewed as a variable hyper-ellipsoidal norm based method, which is formulated by generating weighting matrix adaptively and offering a good linear direction. The proposed framework may be of functional significance to the other large scale problems. To deal with the practical limitations of the proposed framework, an approximation solution to the sub-problem is presented. Observed from the experiments, the proposed algorithm can provide an alternate approach to total variation minimization problem. The experiment results based on the pictorial and numerical results of the referred methods, indicate that the proposed method is superior to the other methods.

# Acknowledgment

# References

[1] H. Andrews and B. Hunt, Digital image restoration, Upper Saddle River, NJ: Prentice-Hall, 1977. http://dl.acm.org/citation.cfm?id=2481044

[2] D. Kundur and D. Hatzinakos, "Blind Image Deconvolution," *IEEE Signal Processing Magazine*, Vol.13, pp.43-64, 1996. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=489268&tag=1

[3] Fujieda, I., Kosugi, T., Inaba, Y., "Speckle Noise Evaluation and Reduction of an Edge-Lit Backlight System Utilizing Laser Diodes and an Optical Fiber," *Journal of Display Technology*, vol.5, no.11, pp. 414-417, Nov. 2009. Article (CrossRef Link)

[4] J. Song and S. Park, "Assessment of Image Quality Degraded by Tone Rendering Distortion," *J. Display Technol*, Vol. 7, pp. 365-372, 2011. Article (CrossRef Link)

[5] E. Buckley, "Real-Time Error Diffusion for Signal-to-Noise Ratio Improvement in a Holographic Projection System," *J. Display Technol*., Vol. 7, pp. 70-76, 2011. Article (CrossRef Link)

[6] Sung In Cho; Suk-Ju Kang; Young Hwan Kim, "Image Quality-Aware Backlight Dimming With

Color and Detail Enhancement Techniques," *Journal of Display Technology*, Vol.9, No.2, pp. 112-121, Feb. 2013. Article (CrossRef Link)

[7] Ruomei Yan, Ling Shao, Cvetkovic, S.D., Klijn, J., "Improved Nonlocal Means Based on Pre-Classification and Invariant Block Matching," *Journal of Display Technology*, Vol.8, No.4, pp. 212-218, April 2012. Article (CrossRef Link)

[8] L. I. Rudin, S. J. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms," *Physica D*, Vol.60, pp. 259-268, 1992. Article (CrossRef Link)

[9] G. Peyre, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Proc. of European Conference on Computer Vision*, pp. 57-68, 2008. Article (CrossRef Link)

[10] C. Vogel, "Computational Methods for Inverse problems," *SIAM*, 2002. Article (CrossRef Link)

[11] Beck A., Teboulle M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci*, Vol. 2, pp. 183-202, 2009. Article (CrossRef Link)

[12] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. On Image Process*, Vol. 18, No. 11, pp. 2419-2434, Nov. 2009. Article (CrossRef Link)

[13] Wright, S.J., Nowak, R.D., Figueiredo, M.A.T., "Sparse reconstruction by separable approximation," *Acoustics, IEEE International Conference on Speech and Signal Processing*, (ICASSP), pp.3373-3376, March 31 2008-April 4 2008. Article (CrossRef Link)

[14] P. Combettes and V.Wajs. "Signal recovery by proximal forward backward splitting. Multi scale Modeling and Simulation," Vol.4, No.4, pp.1168-1200, 2005. Article (CrossRef Link)

[15] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185-212. Springer, New York, 2011. Article (CrossRef Link)

[16] Combettes, P.L., "Solving monotone inclusions via compositions of non-expansive averaged operators," *Optimization*, Vol. 53, pp. 475-504, 2004. Article (CrossRef Link)

[17] Combettes, P.L. "Iterative construction of the resolvent of a sum of maximal monotone operators," *J. Convex Anal.* Vol.16, pp. 727-748, 2009. Article (CrossRef Link)

[18] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345-2356, September, 2010. Article (CrossRef Link)

[19] J. D. Lee, Y. Sun, and M. A. Saunders. "Proximal Newton-type methods for minimizing functions in composite form," *Advances in Neural Information Processing Systems*, Vol.25, pp. 827-835, 2012. Article (CrossRef Link)

[20] Jorge Nocedal, Stephen J. Wright, "Numerical Optimization," *2nd ed., Springer*, 2006. Article (CrossRef Link)

[21] J. Bioucas-Dias and M. Figueiredo. "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. On Image Processing*, Vol.16, pp. 2992-3004, 2007. Article (CrossRef Link)

[22] A. Chambolle. "An algorithm for total variation minimization and applications," *J. Math. Imaging Vision*, Vol. 20, No. 1-2, pp. 89-97, 2004. Article (CrossRef Link)

[23] I. Daubechies, M. Defrise, and C. De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, Vol. 57, No. 11, pp. 1413-1457, 2004. Article (CrossRef Link)

[24] C. Vonesch, S. Ramani, and M. Unser, "Recursive risk estimation for nonlinear image deconvolution with a wavelet-domain sparsity constraint," in *Proc. of 15th IEEE Int. Conf. Image Processing (ICIP)*, pp. 665- 668, 2008. Article (CrossRef Link)

[25] Omar Oreifej, Xin Li, Mubarak Shah, "Simultaneous Video Stabilization and Moving Object Detection in Turbulence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, No. 2, pp. 450-462, 2013. Article (CrossRef Link)

[26] Haichao Zhang, Yanning Zhang, "Sparse representation based iterative incremental image deblurring", in *Proc. of 16th IEEE International Conference on Image Processing (ICIP)*, pp. 1293-1296, 7-10 Nov. 2009. Article (CrossRef Link)

[27] Rostami, M., Michailovich, O., Zhou Wang, "Image Deblurring Using Derivative Compressed

Sensing for Optical Imaging Application," *IEEE Transactions on Image Processing*, Vol.21, No.7, pp. 3139-3149, July 2012. Article (CrossRef Link)

[28] Jianwei Ma, Le Dimet F.-X., "Deblurring From Highly Incomplete Measurements for Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, Vol.47, No.3, pp. 792-802, March 2009. Article (CrossRef Link)

[29] Radoslaw Pytlak. "Conjugate Gradient Algorithms in Nonconvex Optimization," *Springer*, 2009. Article (CrossRef Link)
Kowalski, M., "Proximal algorithm meets a conjugate descent", *Tech. rep.*, 2010.

[30] Junzhou Huang, Shaoting Zhang and Dimitris Metaxas. "Fast Optimization for Mixture Prior Models," in *Proc. of the 11th European Conference on Computer Vision*, ECCV'10, Crete, Greece, September 2010. Article (CrossRef Link)

[31] R. Tomioka and K.Aihara. "Classifying matrices with a spectral regularization," in *Proc. of the 24th international conference on machine learning*, pp. 895-902. ACM press, 2007. Article (CrossRef Link)

[32] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images," *IEEE Trans. Image Process.*, Vol. 16, No. 5, pp. 1395-1411, May 2007. Article (CrossRef Link)

[33] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600-612, Apr. 2004. Article (CrossRef Link)

[34] R. Fletcher and C. Reeves. "Function minimization by conjugate gradients," *Comput. Journal*, Vol. 7, pp. 149-154, 1964. Article (CrossRef Link)

[35] E. Polak and Ribire. "Note sur la convergence de directions conjugues," *Revue Fran caise dInformatique et de Recherche Oprationelle*, 3(16):3543, 1969.

[36] B.T. Polyak. "The conjugate gradient method in extreme problems", *USSR Comp. Math. Math. Phys.*, Vol. 9, pp. 94-112, 1969.

[37] P. Tseng and S.Yun. "A coordinate gradient descent method for nonsmooth separable minimization," *Math. Prog. Ser. B*, Vol. 117, No. 1, pp. 387-423, 2009. Article (CrossRef Link)

[38] D. P. Bertsekas, "Convex optimization theory" *Athena Scientific*, 2009.

[39] Min Tao and Junfeng Yang, "Alternating direction algorithms for total variation deconvolution in image reconstruction," *Tech. Rep. TR0918*, Nanjing University, China, 2009. Article (CrossRef Link)

[40] M. Schmidt, N. Le Roux, F. Bach. "Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization," *Advances in Neural Information Processing Systems (NIPS)*, 2011. Article (CrossRef Link)

[41] K. Y. Shin, B. J. Kang, K. R. Park, "Super-Resolution Iris Image Restoration using Single Image for Iris Recognition," *KSII Transactions on Internet and Information Systems (TIIS)*, Vol. 4, No. 2, pp. 117-137, 2010. Article (CrossRef Link)

[42] J. J. Hwang and H. R. Wu, "Stereo Image Quality Assessment Using Visual Attention and Distortion Predictors," *KSII Transactions on Internet and Information Systems*, Vol. 5, No. 9, pp. 1613-1631, September 2011. Article (CrossRef Link)

[43] J. Yu, D. C. Tao, M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Transactions on Image Processing*, Vol. 21, No. 7, pp. 3262-3272, 2012. Article (CrossRef Link)

[44] J. Yu, R. Hong, M. Wang and J. You, "Image clustering based on sparse patch alignment framework," *Pattern Recognition*, Vol. 47, No. 11, pp. 3512-3519, 2014. Article (CrossRef Link)

[45] J. Yu, D. C. Tao, M. Wang and Y. Rui, "Learning to Rank using User Clicks and Visual Features for Image Retrieval," *IEEE Transactions on Cybernetics*, Vol. 45, No. 4, pp. 767-779, 2015. Article (CrossRef Link)

**Han Pan** was born in GuangXi, PR. China. He received his Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2014. He is currently a postdoctoral fellow at Shanghai Jiao Tong University, Shanghai, China. His research interests include image restoration, information fusion, and convex optimization.

**Zhongliang Jing** received his B.S., M.S., and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 1983, 1988, and 1994, respectively, all in Electronics and Information Technology. Currently, he is Cheung Kong professor, and executive dean at the School of Aeronautics and Astronautics, Shanghai Jiao Tong University, China. Prof. Jing is an editorial board member of the Science China: Information Sciences, Chinese Optics Letters as well as International Journal of Space Science and Engineering. His major research interests include multi-source information acquisition, processing and fusion, target tracking, and aerospace control.

**Minzhe Li** was born in JiangSu, PR China. He received his BS, MS degree from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009 and 2012. He is currently pursuing a Ph.D. degree at Shanghai Jiao Tong University, Shanghai, China. His research interests include tracking, information fusion.
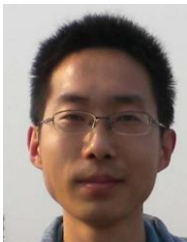
**Peng Dong** received the Ph.D. degree in armament launch theory and technology from the Northwestern Polytechnical University, Xi'an, in 2013. Currently, he is a postdoctoral fellow at Shanghai Jiao Tong University. His research interests include target tracking and information fusion.