Inverse Problems and Imaging doi:10.3934/ipi.2020071

BILINEAR CONSTRAINT BASED ADMM FOR MIXED POISSON-GAUSSIAN NOISE REMOVAL

Jie Zhang

School of Mathematical Sciences,Tianjin Normal University
Tianjin, 300387, China
Department of Mathematics, The University of Hong Kong
Pokfulam, Hong Kong

Yuping Duan

Center for Applied Mathematics, Tianjin University
Tianjin, 300072, China

Yue Lu

School of Mathematical Sciences,Tianjin Normal University
Tianjin, 300387, China

Michael K. Ng

Department of Mathematics, The University of Hong Kong
Pokfulam, Hong Kong

Huibin Chang*

School of Mathematical Sciences,Tianjin Normal University
Tianjin, 300387, China

(Communicated by Yunmei Chen)

Abstract. In this paper, we propose new operator-splitting algorithms for the total variation regularized infimal convolution (TV-IC) model [6] in order to remove mixed Poisson-Gaussian (MPG) noise. In the existing splitting algorithm for TV-IC, an inner loop by Newton method had to be adopted for one nonlinear optimization subproblem, which increased the computation cost per outer loop. By introducing a new bilinear constraint and applying the alternating direction method of multipliers (ADMM), all subproblems of the proposed algorithms named as BCA (short for **B**ilinear **C**onstraint based **A**DMM algorithm) and BCA$_f$ (short for a variant of BCA with **f**ully splitting form) can be very efficiently solved. Especially for the proposed BCA$_f$, they can be calculated without any inner iterations. The convergence of the proposed algorithms are investigated, where particularly, a Huber type TV regularizer is adopted to guarantee the convergence of BCA$_f$. Numerically, compared to existing primal-dual algorithms for the TV-IC model, the proposed algorithms, with fewer tunable parameters, converge much faster and produce comparable results meanwhile.

2020 *Mathematics Subject Classification.* Primary: 65K10, 65F22; Secondary: 94A08, 46N10.
Key words and phrases. Mixed Poisson-Gaussian noise; total variation; alternating direction method of multipliers; bilinear constraint; convergence.

* Corresponding author: Huibin Chang.

1. **Introduction.** As the result of photon counting and thermal noise to the detectors, it is very common that the observed image is corrupted by the mixed Poisson and Gaussian (MPG) noise. The MPG denoising has been extensively studied in [21, 36, 28, 2, 1, 15] and references therein. Generally speaking, the idea of MPG noise removal is based on the maximum a posteriori (MAP) following the Bayes' law. Chakrabarti and Zickler [7] approximated the MPG noise with a shifted Poisson likelihood. A generalized Anscombe transformation was proposed for MPG noise removal in [35, 30], while its unbiased inversion was given in [27]. In order to choose a correct MPG noise model, Reyes and Schönlieb [33] proposed a nonsmooth PDE-constrained optimization strategy. A reweighted L^2 method was proposed by Li et al. [24], which approximated the Poisson component noise with weighted Gaussian noise. The convexity and Lipschitz differentiability of Poisson-Gaussian negative log-likelihood was proven by Chouzenoux et al. in [11], where a convergent primal-dual algorithm was given in the case of approximation of the infinite sum for data discrepancy. More recent works considered the general joint MAP formulation, showing that Gaussian noise model [34] and Poisson noise model [23] can be combined together in order to remove the MPG noise. Lanza et al. [22] proposed a primal-dual based iterative algorithm for total variation (TV) regularized model (TV-PD), where one subproblem required additional inner loop by Newton method. In practice, in order to reduce computational cost, the TV-PD algorithm ran with very few Newton iterations and the corresponding convergence guarantee with such inexact inner solver was unknown. Calatroni et al. [6] proposed the TV regularized infimal convolution (TV-IC) model, consisting of infimal convolution combination of standard data fidelities classically associated to one single-noise distribution, and the TV regularization term. A Semi-Smooth Newton algorithm (SSN) was designed [6] to solve the TV-IC model. Combining with an outer BFGS-type algorithm, an optimality system was also proposed to find the optimal regularization parameters [6]. Although no inner iteration is required by using the SSN method, one has to compute some second-order information, which is computationally challenging.

In order to solve the TV-IC model more efficiently, we will introduce a new bilinear constraint to reformulate the model, which essentially helps to establish the iterative algorithms without Newton iteration in the inner loop. Then we apply alternating direction method of multipliers (ADMM) [17, 16, 14, 3, 40] to the reformulated model, leading to the proposed **B**ilinear **C**onstraint based **A**DMM algorithm (BCA). Due to the nonconvex term in the augmented Lagrangian caused by the bilinear constraint, it seems quite difficult to study the global convergence. Instead, by assuming the iterative sequences have a uniform and strictly positive lower bounds, we prove the local convergence of proposed BCA, in the sense that corresponding iterative sequences are bounded and any limit point is a stationary point of the saddle problem of the augmented Lagrangian functional. Due to the existence of total variation term of the original variable, inner loop is still needed for the proposed BCA. In order to reduce such extra computational cost, we further develop a variant of **BCA** with **f**ully splitting form (BCA$_f$). However, it seems much more difficult to guarantee the convergence for BCA$_f$ algorithm for traditional TV-IC model, which is non-differentiable. Inspired by [6], we are able to prove the local convergence of this fully splitting algorithm by using the Huber type TV regularizer. Extensive numerical experiments further verify the faster convergence of the proposed algorithms while producing comparable recovery results. Especially, it

demonstrates that the proposed algorithms with fewer tunable parameters converge much faster than TV-PD [22] for the TV-IC model.

This paper is organized as follows. Section 2 reviews briefly the TV-IC model and ADMM algorithm. Section 3 introduces the proposed BCA and BCA$_f$ algorithms as well as the related convergence analysis. Section 4 presents the numerical experiments to show the performances of proposed algorithms in terms of convergence and recovery quality as well as the robustness with respect to the parameters and the number of inner iterations of BCA. Finally, conclusions and future work are given in section 5.

2. Review of TV-IC model and ADMM.

In this section, we will briefly review the TV-IC model and the ADMM algorithm.

2.1. Review of TV-IC model.

Let $u, f \in \mathbb{R}^n$ be the ground truth and observed images corrupted by MPG noise respectively, satisfying

$$f = v + \mathbf{n},$$

where $v \sim \text{Poisson}(u)$, $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$. The general joint MAP estimation [22, 6] is given

$$\begin{aligned}(u^\star, v^\star) &= \arg\max_{(u,v)} \prod_i P(v_i, u_i | f_i) \\ &= \arg\max_{(u,v)} \prod_i P(f_i|v_i) P(v_i|u_i) P(u_i).\end{aligned} \tag{1}$$

Incorporating the density function of Poisson and Gaussian distributions, and further taking the negative logarithm in (1), one has

$$\begin{aligned}(u^\star, v^\star) &= \arg\min_{(u,v)} -\ln\Big(\prod_i P(f_i|v_i) P(v_i|u_i) P(u_i)\Big) \\ &= \arg\min_{(u,v)} \left\{ \frac{\lambda_1}{2} \sum_i (f_i - v_i)^2 + \sum_i |\nabla u_i| + \lambda_2 \sum_i (u_i - v_i \ln u_i + \ln v_i!) \right\},\end{aligned}$$

with ∇ denoting the discrete gradient (finite difference) operator and $|\cdot|$ denotes the L^2 norm of a vector, where a Gibbs prior distribution of $P(u_i) = \exp(-|\nabla u_i|)$ is considered.

Using the standard Stirling approximation of the logarithm of the factorial function, the following TV-IC model was established [22, 6]

$$\min_{u,v} H(u, v), \tag{2}$$

where $H(u, v)$ is:

$$H(u,v) = \frac{\lambda_1}{2} \sum_i (f_i - v_i)^2 + \lambda_2 \sum_i (u_i - v_i \ln \frac{u_i}{v_i} - v_i) + \sum_i |\nabla u_i| + \chi_{\mathcal{V}}(v),$$

$\chi_{\mathcal{V}}$ is the characteristic function of the positivity constraint set $\mathcal{V} = \{v : v_i \geq \epsilon > 0 \ \forall i\}$

$$\chi_{\mathcal{V}}(v) = \begin{cases} 0 & v \in \mathcal{V}, \\ +\infty & \text{otherwise}, \end{cases}$$

which is derived by the property of v ($v \sim \text{Poisson}(u)$). The notation $\sum_i |\nabla u_i|$ denotes the standard discrete TV regularization. Here we remark that we require

that v is lower bounded by a positive scalar ϵ, which is introduced for the purpose of studying convergence guarantee of proposed algorithms.

2.2. **Review of ADMM.** The ADMM [17, 16, 14, 3, 40, 37, 29] is one of the popular first-order operator-splitting algorithm in image processing, which can handle complex constraints and non-smooth and non-convex objective functional. Compared with the gradient descent algorithm, it is more stable since it gets rid of directly calculating the derivative of the objective functional and therefore allows for big stepsize. Hence, we apply the ADMM to solve the TV-IC model. In this part, we will give a brief introduction of the ADMM. Consider the optimization problem below

$$\begin{aligned} \min_{\mathbf{x},\mathbf{z}} \quad & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{s.t.} \quad & A\mathbf{x} + B\mathbf{z} = \mathbf{m} \end{aligned} \tag{3}$$

with variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^m$, where $A \in \mathbb{R}^{p\times n}$, $B \in \mathbb{R}^{p\times m}$, and $\mathbf{m} \in \mathbb{R}^p$. The augmented Lagrangian is given below

$$L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T(A\mathbf{x} + B\mathbf{z} - \mathbf{m}) + \frac{\rho}{2}\left\|A\mathbf{x} + B\mathbf{z} - \mathbf{m}\right\|_2^2,$$

with the parameter $\rho > 0$ and the multiplier $\mathbf{y} \in \mathbb{R}^p$. In order to solve the saddle point problem

$$\max_{\mathbf{y}} \min_{\mathbf{x},\mathbf{z}} L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}),$$

The ADMM consists of the following iterations to determine $(k+1)^{th}$ solutions as

$$\begin{aligned} \mathbf{x}^{k+1} &:= \arg\min_{\mathbf{x}} L_\rho\left(\mathbf{x}, \mathbf{z}^k, \mathbf{y}^k\right) \\ \mathbf{z}^{k+1} &:= \arg\min_{\mathbf{z}} L_\rho\left(\mathbf{x}^{k+1}, \mathbf{z}, \mathbf{y}^k\right) \\ \mathbf{y}^{k+1} &:= \mathbf{y}^k + \rho\left(A\mathbf{x}^{k+1} + B\mathbf{z}^{k+1} - \mathbf{m}\right). \end{aligned}$$

given the previous iteration solutions $(\mathbf{x}^k, \mathbf{z}^k, \mathbf{y}^k)$.

3. **Proposed algorithms.** In this section, we will consider how to design more efficient operator-splitting algorithms based on ADMM to solve the TV-IC model (2). If decoupling the problem by introducing an auxiliary variable to replace the original variable u following [22], it will have a subproblem without closed-form solution, due to the existence of the term $v_i \ln v_i$. In order to solve $v-$subproblem, an inner loop by Newton method is needed, which is time-consuming and lack of convergence guarantee with few inner iterations. To further speed up the convergence, a new bilinear constraint ($u_i = v_i w_i$) will be introduced such that the resulting BCA algorithm consists of standard TV-L2 denoising for variable u, and simple closed form solutions for v and w. Its convergence is further derived under the assumption that the iterative sequence of w is uniformly bounded below by a positive number. In order to get a fully splitting scheme, i.e., all subproblems have closed form solutions, a typical constraint $p_i = \nabla u_i$ is introduced additionally such that the BCA$_f$ is obtained within the framework of ADMM. Following the technique developed for BCA, we can prove its theoretical convergence, using the Huber type TV regularizer in the TV-IC model instead of the traditional TV.

3.1. **BCA.** By introducing the bilinear constraint $u_i = v_i w_i$, we rewrite (2) as the following equivalent constrained optimization problem:

$$\begin{aligned} (4)\quad & \min_{(u,v,w)} \left\{ \frac{\lambda_1}{2} \sum_i (f_i - v_i)^2 + \lambda_2 \sum_i (u_i - v_i \ln w_i - v_i) + \sum_i |\nabla u_i| + \chi_{\mathcal{V}}(v) \right\}, \\ & s.t. \qquad u_i = v_i w_i, \ \forall 1 \le i \le n. \end{aligned}$$

Readily one sees that the term $v_i \ln v_i$ disappears, hence we can design a fast algorithm with v and w subproblems all having closed form solutions. As reviewed in subsection 2.2, one has to establish the augmented Lagrangian of the above constrained optimization problem with the penalization parameter $\alpha > 0$ and the multiplier $\Lambda \in \mathbb{R}^n$, is given below

$$\begin{aligned} L_\alpha(u, v, w, \Lambda) =& \frac{\lambda_1}{2} \sum_i (f_i - v_i)^2 + \sum_i |\nabla u_i| + \chi_{\mathcal{V}}(v) \\ & + \lambda_2 \sum_i (u_i - v_i \ln w_i - v_i) \\ & + \langle \Lambda, v \circ w - u \rangle + \frac{\alpha}{2} \|v \circ w - u\|^2, \end{aligned} \tag{5}$$

where $\langle \cdot \rangle$ and $\|\cdot\|$ denote the inner product and norm in L^2 space respectively, and $\circ$ denotes the element-wise multiplication. Note that all the vector multiplications and divisions in this paper are element-wise.

Given the previous iterative solution (u^k, v^k, w^k), the ADMM updates the sequence $(u^{k+1}, v^{k+1}, w^{k+1})$ by solving three subproblems w.r.t. u, v, w, and multiplier update, which is given below:

$$\begin{cases} u^{k+1} = \arg\min_u L_\alpha(u, v^k, w^k, \Lambda^k), & \text{(6a)} \\ v^{k+1} = \arg\min_v L_\alpha(u^{k+1}, v, w^k, \Lambda^k), & \text{(6b)} \\ w^{k+1} = \arg\min_w L_\alpha(u^{k+1}, v^{k+1}, w, \Lambda^k), & \text{(6c)} \\ \Lambda^{k+1} = \Lambda^k + \alpha(v^{k+1} \circ w^{k+1} - u^{k+1}). & \text{(6d)} \end{cases}$$

We will show how to solve these subproblems in the rest of this part.

First, we consider the u-subproblem as

$$\begin{aligned} u^{k+1} =& \arg\min_u \{ \lambda_2 \sum_i u_i + \langle \Lambda^k, v^k \circ w^k - u \rangle \\ & + \frac{\alpha}{2} \left\| v^k \circ w^k - u \right\|^2 + \sum_i |\nabla u_i| \} \\ =& \arg\min_u \left\{ \frac{\alpha}{2} \left\| v^k \circ w^k + \frac{\Lambda^k}{\alpha} - \frac{\lambda_2 \mathbf{1}}{\alpha} - u \right\|^2 + \sum_i |\nabla u_i| \right\}, \end{aligned} \tag{7}$$

where $\mathbf{1} \in \mathbb{R}^n$ is a vector whose elements are all equal to one. This is a standard TV-L2 [34] optimization problem, and one can adopt the gradient projection algorithm for the pre-dual form of total variation minimization [8].

As the update rule of w-subproblem can simplify the calculation of v-subproblem, we consider w-subproblem first.

$$w^{k+1} = \arg\min_w \sum_i \left[-\lambda_2 v_i^{k+1} \ln w_i + \frac{\alpha}{2} (v_i^{k+1} w_i + \frac{\Lambda_i^k}{\alpha} - u_i^{k+1})^2 \right].$$

Obviously it is a convex optimization problem. One can readily get the scalar optimization problem of this convex optimization problem

$$w_i^{k+1} = \arg\min_{w_i} \left\{ -\lambda_2 v_i^{k+1} \ln w_i + \frac{\alpha}{2}(v_i^{k+1} w_i + \frac{\Lambda_i^k}{\alpha} - u_i^{k+1})^2 \right\}.$$

The optimality condition of the above problem is

$$\alpha(v_i^{k+1})^2 w_i^2 + \left(\Lambda_i^k v_i^{k+1} - \alpha v_i^{k+1} u_i^{k+1}\right) w_i - \lambda_2 v_i^{k+1} = 0.$$

We can obtain a closed-form solution (also the global minimizer) of this problem

$$w_i^{k+1} = \frac{1}{2v_i^{k+1}} \left[(u_i^{k+1} - \frac{\Lambda_i^k}{\alpha}) + \sqrt{(u_i^{k+1} - \frac{\Lambda_i^k}{\alpha})^2 + \frac{4\lambda_2 v_i^{k+1}}{\alpha}} \right]. \tag{8}$$

Finally, we consider the v-subproblem as

$$\begin{aligned} v^{k+1} = \arg\min_{v_i \geq \epsilon} \sum_i [& \frac{\lambda_1}{2}(f_i - v_i)^2 - \lambda_2(v_i \ln w_i^k + v_i) \\ & + \frac{\alpha}{2}(v_i w_i^k + \frac{\Lambda_i^k}{\alpha} - u_i^{k+1})^2]. \end{aligned}$$

This optimization problem can be computed independently with respect to each component of v, therefore, we can consider the scalar optimization problem

$$v_i^{k+1} = \arg\min_{v_i \geq \epsilon} \left\{ \frac{\lambda_1}{2}(f_i - v_i)^2 - \lambda_2(v_i \ln w_i^k + v_i) + \frac{\alpha}{2}(v_i w_i^k + \frac{\Lambda_i^k}{\alpha} - u_i^{k+1})^2 \right\}.$$

One easily obtains the optimal solution of the above problem below

$$v^{k+1} = \max\left(\epsilon \mathbf{1}, \tilde{v}^{k+1}\right),$$

with the notation $\max(\cdot,\cdot)$ taking the element-wise maximum of two vectors, where $\tilde{v}^{k+1}$ is defined as follows

$$\tilde{v}^{k+1} = \frac{\mathbf{1}}{\lambda_1 \mathbf{1} + \alpha(w^k)^2} \circ \left(\lambda_1 f + \lambda_2 \ln w^k + \lambda_2 - w^k \circ \Lambda^k + \alpha w^k \circ u^{k+1}\right), \tag{9}$$

which corresponds to the unconstrained optimal solution. Note that $\frac{\mathbf{a}}{\mathbf{b}}$ denotes the element-wise division of two vectors $\mathbf{a}$ and $\mathbf{b}$.

In order to further simplify the calculation of $v-$subproblem, one can obtain the following lemma.

Lemma 3.1. *Letting Λ^{k+1}, w^{k+1} be generated by (6c)-(6d), then we have*

$$\Lambda^{k+1} \circ w^{k+1} = \lambda_2 \mathbf{1}. \tag{10}$$

Proof. Considering the update rule of w-subproblem in (6c) and multiplier update in (6d), we have

$$\begin{aligned} 0 &= \alpha(v^{k+1})^2 \circ (w^{k+1})^2 + (\Lambda^k \circ v^{k+1} - \alpha v^{k+1} \circ u^{k+1}) \circ w^{k+1} - \lambda_2 v^{k+1} \\ &= v^{k+1} \circ w^{k+1} \circ (\Lambda^k + \alpha v^{k+1} \circ w^{k+1} - \alpha u^{k+1}) - \lambda_2 v^{k+1} \\ &= v^{k+1} \circ w^{k+1} \circ \Lambda^{k+1} - \lambda_2 v^{k+1}. \end{aligned}$$

Further due to $v_i^{k+1} \geq \epsilon > 0 \ \ \forall\, i$, we can prove this lemma. □

Remark 1. By Lemma 3.1, The equation (9) can be simplified below

$$\begin{aligned}\tilde{v}^{k+1} &= \frac{\mathbf{1}}{\lambda_1\mathbf{1}+\alpha(w^k)^2}\circ\left(\lambda_1 f+\lambda_2\ln w^k+\lambda_2-w^k\circ\Lambda^k+\alpha w^k\circ u^{k+1}\right),\\ &= \frac{\mathbf{1}}{\lambda_1\mathbf{1}+\alpha(w^k)^2}\circ\left(\lambda_1 f+\lambda_2\ln w^k+\alpha w^k\circ u^{k+1}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}v^{k+1} &= \max\left(\epsilon\mathbf{1},\tilde{v}^{k+1}\right),\\ &= \max\left(\epsilon\mathbf{1},\frac{\mathbf{1}}{\lambda_1\mathbf{1}+\alpha(w^k)^2}\circ\left(\lambda_1 f+\lambda_2\ln w^k+\alpha w^k\circ u^{k+1}\right)\right).\end{aligned}\tag{11}$$

Algorithm 1 summarizes the overall BCA algorithm.

Algorithm 1 BCA

Input: Noisy data f and parameters λ_1, λ_2, α
Initialization: $u^0=f, v^0=f, w^0=\mathbf{1}, \Lambda^0=\mathbf{0}, k=0$.
1: **while** Stopping criteria is not satisfied **do**
2: Find u^{k+1} by (7)
3: Find v^{k+1} by (11)
4: Find w^{k+1} by (8)
5: Update the multipliers by

$$\Lambda^{k+1}=\Lambda^k+\alpha(v^{k+1}\circ w^{k+1}-u^{k+1}).$$

6: $k\leftarrow k+1$.
7: **end while**

3.2. **Convergence analysis of BCA.** Readily one knows that the reformulated optimization problem (4) cannot be interpreted as a (two-block) problem (as reviewed in subsection 2.2), whose objective function is of sum of two functions without coupled variables. Due to the bilinear constraint and coupled term $v_i\ln w_i$ in the objective function, it does not also belong to the problems considered either for convex optimization problem [25, 10, 12] with objective functions of sum of no less than three functions or nonconvex optimization problem [18, 9] with bilinear constraint. The current algorithm introduces similar bilinear constraint as for the blind ptychography problem in [9]. However, the proof technique for [9] cannot directly apply to the current BCA, since the linear relation between the iterative multipliers and the auxiliary variable does not hold for BCA, and more specifically, their relation is bilinear (See Lemma 3.1). Moreover, the coercivity of the objective function is not trivial. Therefore, one has to develop a new technique for convergence guarantee.

To guarantee the sufficient decrease and boundedness of the iterative sequence, we make the following assumption. Although limited by current analysis technique we cannot remove it, it can be verified numerically (See Fig. 13 in the numerical part of this paper).

Assumption 1. The iterative sequence $\{w^k\}$ generated by BCA algorithm has a uniformly positive lower bound, i.e., $w_i^k\geq c>0,\forall i$, where c is a positive constant which is independent to k.

Lemma 3.2. *Let $T(x) = \frac{1}{2}\|Ax - b\|^2 + M(x)$, with convex function M. Letting x^* be a stationary point of $T(x)$ (also a global minimizer), i.e. $0 \in \partial T(x^*)$, where $\partial T(x)$ denotes the subdifferential of $T(x)$ in the convex analysis sense, then we have*

$$T(x) - T(x^*) \geq \|A(x - x^*)\|^2.$$

Proof. Let $H(x) = \frac{1}{2}\|Ax - b\|^2$. Since x^* is a stationary point of T, one has

$$-\nabla H(x^*) \in \partial M(x^*).$$

Readily one has

$$M(x) - M(x^*) \geq \langle -\nabla H(x^*), x - x^* \rangle \quad \forall\ x.$$

Then we have

$$T(x) - T(x^*) \geq H(x) - H(x^*) - \langle \nabla H(x^*), x - x^* \rangle = \frac{1}{2}\|A(x - x^*)\|^2,$$

that immediately concludes this lemma. □

In the following, within the framework in [37, 9, 19, 26, 29, 18] developed for the analysis of ADMM for nonconvex nonsmooth optimization problem, we will first prove that the iterative sequence satisfies the sufficient descent condition. Then, the relative error condition for the iterative sequence will be derived. Finally one can derive the subsequence convergence of the proposed BCA.

Lemma 3.3. *Letting $(u^k, v^k, w^k, \Lambda^k)$ be the sequence generated by BCA in Algorithm 1, and $\alpha > \frac{\sqrt{2}\lambda_2}{c^2\epsilon}$, then under Assumption 1 we have*

$$\begin{aligned} &L_\alpha(u^k, v^k, w^k, \Lambda^k) - L_\alpha(u^{k+1}, v^{k+1}, w^{k+1}, \Lambda^{k+1}) \geq \frac{\alpha}{2}\|u^{k+1} - u^k\|^2 \\ &+ \frac{\lambda_1}{2}\|v^{k+1} - v^k\|^2 + \frac{\alpha}{2}\|w^k \circ (v^{k+1} - v^k)\|^2 + C_1\|v^{k+1} \circ (w^{k+1} - w^k)\|^2, \end{aligned} \tag{12}$$

where C_1 is a positive constant which is independent to k.

Proof. For u-subproblem, by Lemma 3.2, one readily has

$$L_\alpha(u^k, v^k, w^k, \Lambda^k) - L_\alpha(u^{k+1}, v^k, w^k, \Lambda^k) \geq \frac{\alpha}{2}\|u^{k+1} - u^k\|^2. \tag{13}$$

Similarly using Lemma 3.2, for v-subproblem, one can obtain

$$\begin{aligned} &L_\alpha(u^{k+1}, v^k, w^k, \Lambda^k) - L_\alpha(u^{k+1}, v^{k+1}, w^k, \Lambda^k) \\ \geq &\frac{\lambda_1}{2}\|v^{k+1} - v^k\|^2 + \frac{\alpha}{2}\|w^k \circ (v^{k+1} - v^k)\|^2. \end{aligned} \tag{14}$$

For w-subproblem, one gets

$$\begin{aligned} &L_\alpha(u^{k+1}, v^{k+1}, w^k, \Lambda^k) - L_\alpha(u^{k+1}, v^{k+1}, w^{k+1}, \Lambda^k) \\ \geq &\frac{\alpha}{2}\|v^{k+1} \circ (w^{k+1} - w^k)\|^2. \end{aligned} \tag{15}$$

By (6d), Lemma 3.1 and Assumption 1 one has

$$
\begin{aligned}
&L_\alpha(u^{k+1},v^{k+1},w^{k+1},\Lambda^k)-L_\alpha(u^{k+1},v^{k+1},w^{k+1},\Lambda^{k+1})\\
&=-\frac{1}{\alpha}\left\|\Lambda^{k+1}-\Lambda^k\right\|^2=-\frac{1}{\alpha}\left\|\frac{\lambda_2\mathbf{1}}{w^{k+1}}-\frac{\lambda_2\mathbf{1}}{w^k}\right\|^2\\
&=-\frac{\lambda_2^2}{\alpha}\left\|\frac{v^{k+1}\circ(w^{k+1}-w^k)}{v^{k+1}\circ w^{k+1}\circ w^k}\right\|^2\\
&\geq-\frac{\lambda_2^2}{\alpha c^4\epsilon^2}\left\|v^{k+1}\circ(w^{k+1}-w^k)\right\|^2.
\end{aligned}
\tag{16}
$$

Since $\alpha>\frac{\sqrt{2}\lambda_2}{c^2\epsilon}$, further by (13)-(16), one can conclude to this lemma. □

Lemma 3.4. *Denote $\mathcal{G}:\Omega\to\mathbb{R}$ by*

$$
\mathcal{G}(u,v,w)=\frac{\lambda_1}{2}\left\|f-v\right\|^2+\lambda_2\left\langle(\mathbf{1}-\frac{\mathbf{1}}{w})\circ u-v\circ\ln w,\mathbf{1}\right\rangle+\frac{\alpha}{2}\left\|v\circ w-u\right\|^2,
$$

with $\alpha>\max\left(\lambda_2,\lambda_2\left(\frac{1}{c}-1\right)^2\right)$, where $\Omega:=\{(u,v,w)\mid v_i\geq\epsilon>0,w_i\geq c>0\ \forall\ i;u,v,w\in\mathbb{R}^n\}$. If $\|(u,v,w)\|_\Omega:=\max\{\|u\|_\infty,\|v\|_\infty,\|w\|_\infty\}\to+\infty$, then we have $\mathcal{G}(u,v,w)\to+\infty$.

Proof. For all $(u,v,w)\in\Omega$, one readily has

$$
\begin{aligned}
&\mathcal{G}(u,v,w)\geq\frac{\lambda_1}{2}\|f-v\|^2+\lambda_2\left\langle(\mathbf{1}-\frac{\mathbf{1}}{w})\circ u-v\circ w,\mathbf{1}\right\rangle+\frac{\alpha}{2}\|v\circ w-u\|^2\\
&=\frac{\lambda_1}{2}\|f-v\|^2+\lambda_2\left\langle(\mathbf{1}-\frac{\mathbf{1}}{w})\circ(u-v\circ w)-v,\mathbf{1}\right\rangle+\frac{\alpha}{2}\|v\circ w-u\|^2\\
&\geq\frac{\lambda_1}{2}\|f-v\|^2-\lambda_2\left\langle|\mathbf{1}-\frac{\mathbf{1}}{w}|\circ|u-v\circ w|+v,\mathbf{1}\right\rangle+\frac{\alpha}{2}\|v\circ w-u\|^2\\
&=\frac{\lambda_1}{2}\|f-v\|^2-\lambda_2\langle v,\mathbf{1}\rangle+\frac{\lambda_2}{2}\left\||\mathbf{1}-\frac{\mathbf{1}}{w}|\circ|u-v\circ w|-\mathbf{1}\right\|^2\\
&\quad-\frac{\lambda_2}{2}\left\||\mathbf{1}-\frac{\mathbf{1}}{w}|\circ|u-v\circ w|\right\|^2-\frac{\lambda_2}{2}\|\mathbf{1}\|^2+\frac{\alpha}{2}\|v\circ w-u\|^2\\
&\overset{\text{Assumption 1}}{\geq}\frac{\lambda_1}{2}\|f-v\|^2-\lambda_2\langle v,\mathbf{1}\rangle+\frac{\lambda_2}{2}\left\||\mathbf{1}-\frac{\mathbf{1}}{w}|\circ|u-v\circ w|-\mathbf{1}\right\|^2\\
&\quad-\frac{\lambda_2}{2}M^2\|u-v\circ w\|^2-\frac{\lambda_2 n}{2}+\frac{\alpha}{2}\|v\circ w-u\|^2\\
&=\frac{\lambda_1}{2}\|f-v\|^2-\lambda_2\langle v,\mathbf{1}\rangle+\frac{\lambda_2}{2}\left\||\mathbf{1}-\frac{\mathbf{1}}{w}|\circ|u-v\circ w|-\mathbf{1}\right\|^2\\
&\quad+\left[\frac{\alpha}{2}-\frac{\lambda_2}{2}M^2\right]\|v\circ w-u\|^2-\frac{\lambda_2 n}{2},
\end{aligned}
\tag{17}
$$

where the first inequality is derived by $-\ln w_i\geq-w_i$ if $w_i>0$, $M=\max\left(1,\left(\frac{1}{c}-1\right)\right)$.

In the following part, we consider the following two cases for $\|(u,v,w)\|_\Omega\to+\infty$.

Case 1: $\|v\|_\infty\to+\infty$ or $\|v\circ w-u\|_\infty\to+\infty$. Since $\alpha>\max\left(\lambda_2,\lambda_2\left(\frac{1}{c}-1\right)^2\right)$, one can readily get that $\mathcal{G}(u,v,w)\to+\infty$.

Case 2: There exists two constants $C_2, C_3 > 0$, such that $\|v\|_\infty \leq C_2 < +\infty$, $\|u\|_\infty \to +\infty$, $\|w\|_\infty \to +\infty$, and $\|v \circ w - u\|_\infty \leq C_3 < +\infty$. Then we have

$$\mathcal{G}(u,v,w) = \sum_i [\frac{\lambda_1}{2}(f_i - v_i)^2 + \lambda_2(1 - \frac{1}{w_i})u_i - v_i \ln w_i + \frac{\alpha}{2}(v_i w_i - u_i)^2]. \tag{18}$$

There must exist some i where $u_i \to +\infty$, $w_i \to +\infty$, and $|v_i w_i - u_i| \leq C_3$. Thus, we have

$$\epsilon w_i - u_i \leq |v_i w_i - u_i| \leq C_3.$$

Then we get the lower bound estimate of u_i as

$$u_i \geq \epsilon w_i + C_3. \tag{19}$$

Therefore,

$$\begin{aligned} &\frac{\lambda_1}{2}(f_i - v_i)^2 + \lambda_2(1 - \frac{1}{w_i})u_i - v_i \ln w_i + \frac{\alpha}{2}(v_i w_i - u_i)^2 \\ &\geq \lambda_2(1 - \frac{1}{w_i})u_i - v_i \ln w_i \\ &\geq \lambda_2(1 - \frac{1}{w_i})(\epsilon w_i + C_3) - C_2 \ln w_i \\ &= \lambda_2 \epsilon w_i - \lambda_2 \epsilon + \lambda_2 C_3 - \frac{\lambda_2 C_3}{w_i} - C_2 \ln w_i. \end{aligned} \tag{20}$$

Since $\lim\limits_{w \to +\infty} \frac{\lambda_2 \epsilon w - \lambda_2 \epsilon + \lambda_2 C_3 - \frac{\lambda_2 C_3}{w} - C_2 \ln w}{w} = \lambda_2 \epsilon > 0$, we can readily get $(\lambda_2 \epsilon w - \lambda_2 \epsilon + \lambda_2 C_3 - \frac{\lambda_2 C_3}{w} - C_2 \ln w) \to +\infty$ as $w \to +\infty$. Thus, we can derive that

$$\frac{\lambda_1}{2}(f_i - v_i)^2 + \lambda_2(1 - \frac{1}{w_i})u_i - v_i \ln w_i + \frac{\alpha}{2}(v_i w_i - u_i)^2 \to +\infty.$$

If the variables u_j, v_j, w_j do not satisfy the above two cases, they cannot tend to the infinity. Hence, in summary, we can conclude that $\mathcal{G}(u,v,w) \to +\infty$ as $\|(u,v,w)\|_\Omega \to +\infty$. □

Theorem 3.5. *Letting $\alpha > \max(\frac{\sqrt{2}\lambda_2}{c^2 \epsilon}, \lambda_2, \lambda_2(\frac{1}{c} - 1)^2)$, under Assumption 1, we have*

(1) The sequence $(u^k, v^k, w^k, \Lambda^k)$ generated by proposed BCA is bounded and has at least one limit point.
(2) The successive errors $u^{k+1} - u^k \to 0$, $v^{k+1} - v^k \to 0$, $w^{k+1} - w^k \to 0$, and $\Lambda^{k+1} - \Lambda^k \to 0$ as $k \to +\infty$.
(3) Each limit point $(u^*, v^*, w^*, \Lambda^*)$ is a stationary point of $L_\alpha(u,v,w,\Lambda)$, and (u^*, v^*) is a stationary point of $H(u,v)$.

Proof. (1) If $\alpha > \frac{\sqrt{2}\lambda_2}{c^2 \epsilon}$, by Lemma 3.3, we get

$$\begin{aligned} &L_\alpha(u^k, v^k, w^k, \Lambda^k) - L_\alpha(u^{k+1}, v^{k+1}, w^{k+1}, \Lambda^{k+1}) \\ &\geq \frac{\alpha}{2}\|u^{k+1} - u^k\|^2 + \frac{\lambda_1}{2}\|v^{k+1} - v^k\|^2 + \frac{\alpha}{2}\|w^k \circ (v^{k+1} - v^k)\|^2 \\ &\qquad + C_1 \|v^{k+1} \circ (w^{k+1} - w^k)\|^2 \\ &\geq \frac{\alpha}{2}\|u^{k+1} - u^k\|^2 + \frac{\lambda_1 + \alpha c^2}{2}\|v^{k+1} - v^k\|^2 + C_1 \epsilon^2 \|w^{k+1} - w^k\|^2. \end{aligned} \tag{21}$$

Next, we will show that $L_\alpha(u^k, v^k, w^k, \Lambda^k)$ is lower bounded. Readily one knows that

$$\begin{aligned} & L_\alpha(u^k, v^k, w^k, \Lambda^k) \\ & \geq \frac{\lambda_1}{2}\left\|f - v^k\right\|^2 + \lambda_2\left\langle (\mathbf{1} - \frac{\mathbf{1}}{w^k}) \circ u^k - v^k \circ \ln w^k, \mathbf{1}\right\rangle + \frac{\alpha}{2}\left\|v^k \circ w^k - u^k\right\|^2 \\ & = \mathcal{G}(u^k, v^k, w^k). \end{aligned} \tag{22}$$

Then, following (22) and Lemma 3.4, the sequences $\{u^k\}$, $\{v^k\}$, $\{w^k\}$ and $\{\mathcal{G}(u^k, v^k, w^k)\}$ are all bounded as well as the boundedness of $\{\Lambda^k\}$ due to Lemma 3.1.

Due to the boundedness of $(u^k, v^k, w^k, \Lambda^k)$, there exists a convergent subsequence $(u^{k_i}, v^{k_i}, w^{k_i}; \Lambda^{k_i})$, i.e., $(u^{k_i}, v^{k_i}, w^{k_i}, \Lambda^{k_i}) \to (u^*, v^*, w^*, \Lambda^*)$.

(2) By (22), one readily knows that the sequence $L_\alpha(u^k, v^k, w^k, \Lambda^k)$ is bounded below. Therefore, further by summing up (21) from $k = 1$ to ∞ implies that

$$\sum_{k=1}^{\infty}\left\|u^{k+1} - u^k\right\|^2 + \left\|v^{k+1} - v^k\right\|^2 + \left\|w^{k+1} - w^k\right\|^2 < \infty.$$

That immediately implies that $u^{k+1} - u^k \to 0$, $v^{k+1} - v^k \to 0$, $w^{k+1} - w^k \to 0$. By Lemma 3.1, one can also knows that $\Lambda^{k+1} - \Lambda^k \to 0$.

(3) It follows from the optimality condition of u-subproblem that there exists $q_1^{k+1} \in \partial \sum_i |\nabla u_i^{k+1}|$ such that

$$q_1^{k+1} + \lambda_2 \mathbf{1} - \Lambda^k - \alpha(v^k \circ w^k - u^{k+1}) = 0.$$

Letting $p_1^{k+1} = q_1^{k+1} + \lambda_2 - \Lambda^{k+1} - \alpha(v^{k+1} \circ w^{k+1} - u^{k+1}) \in \partial_u L_\alpha(u^{k+1}, v^{k+1}, w^{k+1}; \Lambda^{k+1})$, then we have

$$\begin{aligned} \|p_1^{k+1}\| &= \|q_1^{k+1} + \lambda_2 \mathbf{1} - \Lambda^{k+1} - \alpha(v^{k+1} \circ w^{k+1} - u^{k+1})\| \\ &= \|\Lambda^k - \Lambda^{k+1} + \alpha(v^k \circ w^k - v^{k+1} \circ w^{k+1})\| \\ &\leq \|\Lambda^{k+1} - \Lambda^k\| + \alpha\|w^k \circ (v^{k+1} - v^k)\| + \alpha\|v^{k+1} \circ (w^{k+1} - w^k)\|. \end{aligned} \tag{23}$$

The optimality condition of v-subproblem implies that there exists $q_2^{k+1} \in \partial\chi_{\mathcal{V}}(v^{k+1})$ such that

$$\begin{aligned} & q_2^{k+1} + \lambda_1 v^{k+1} + \alpha(w^k)^2 \circ v^{k+1} - \lambda_1 f \\ & \quad - \lambda_2 \ln w^k - \lambda_2 + w^k \circ \Lambda^k - \alpha w^k \circ u^{k+1} = 0. \end{aligned}$$

Letting $p_2^{k+1} = q_2^{k+1} + \lambda_1 v^{k+1} + \alpha(w^{k+1})^2 \circ v^{k+1} - \lambda_1 f - \lambda_2 \ln w^{k+1} - \lambda_2 + w^{k+1} \circ \Lambda^{k+1} - \alpha w^{k+1} \circ u^{k+1} \in \partial_u L_\alpha(u^{k+1}, v^{k+1}, w^{k+1}; \Lambda^{k+1})$, then we have

$$
\begin{aligned}
&\|p_2^{k+1}\| \\
&= \|q_2^{k+1} + \lambda_1 v^{k+1} + \alpha(w^{k+1})^2 \circ v^{k+1} - \lambda_1 f - \lambda_2 \ln w^{k+1} - \lambda_2 \\
&\qquad + w^{k+1} \circ \Lambda^{k+1} - \alpha w^{k+1} \circ u^{k+1}\| \\
&= \|\alpha v^{k+1} \circ [(w^{k+1})^2 - (w^k)^2] - \lambda_2(\ln w^{k+1} - \ln w^k) + \Lambda^{k+1} \circ w^{k+1} \\
&\qquad - \Lambda^k \circ w^k - \alpha u^{k+1}(w^{k+1} - w^k)\| \\
&= \|\alpha v^{k+1} \circ (w^{k+1} + w^k) \circ (w^{k+1} - w^k) - \lambda_2(\ln w^{k+1} - \ln w^k) \\
&\qquad - \alpha u^{k+1} \circ (w^{k+1} - w^k)\| \\
&= \|(\Lambda^{k+1} - \Lambda^k) \circ (w^{k+1} - w^k) + \alpha v^{k+1} \circ w^k \circ (w^{k+1} - w^k) \\
&\qquad - \lambda_2(\ln w^{k+1} - \ln w^k)\| \\
&\le \|(\Lambda^{k+1} - \Lambda^k)\|\|(w^{k+1} - w^k)\| + \alpha\|v^{k+1} \circ w^k \circ (w^{k+1} - w^k)\| \\
&\qquad + \lambda_2\|\ln w^{k+1} - \ln w^k\|.
\end{aligned} \tag{24}
$$

By the optimality condition of w-subproblem and (6d), we have

$$
\begin{aligned}
&\|\nabla_w L_\alpha\left(u^{k+1}, v^{k+1}, w^{k+1}; \Lambda^{k+1}\right)\| \\
&= \|\frac{\alpha(v^{k+1})^2 \circ (w^{k+1})^2 + (\Lambda^{k+1} \circ v^{k+1} - \alpha v^{k+1} \circ u^{k+1}) \circ w^{k+1} - \lambda_2 v^{k+1}}{w^{k+1}}\| \\
&= \|v^{k+1} \circ (\Lambda^{k+1} - \Lambda^k)\|,
\end{aligned} \tag{25}
$$

and

$$
\begin{aligned}
&\|\nabla_\Lambda L_\alpha\left(u^{k+1}, v^{k+1}, w^{k+1}; \Lambda^{k+1}\right)\| \\
&= \|v^{k+1} \circ w^{k+1} - u^{k+1}\| = \frac{1}{\alpha}\|\Lambda^{k+1} - \Lambda^k\|.
\end{aligned} \tag{26}
$$

Finally, (23)-(26) and Item (1) in this theorem suggest that $(u^*, v^*, w^*; \Lambda^*)$ is a stationary point of $L_\alpha(u, v, w, \Lambda)$. Since $(u^*, v^*, w^*; \Lambda^*)$ is a stationary point, we have $u^* = v^* w^*$ from (26), then (23) and (24) imply that $0 \in \partial_u H(u^*, v^*)$ and $0 \in \partial_v H(u^*, v^*)$, i.e., (u^*, v^*) is a stationary point of $H(u, v)$. □

We remark that in order to prove the theoretical convergence of the proposed BCA algorithm, we assume that ϵ is a positive constant. Simulation results reported in the experimental section of this paper will not be affected if ϵ is selected appropriately. As for the case $\epsilon = 0$, we will investigate the theoretical convergence in the future.

3.3. **BCA$_f$.** The proposed BCA algorithm has a subproblem in (7) w.r.t. total variation minimization problem, which requires inner loop. To get a fully splitting scheme, we propose the following BCA$_f$ algorithm.

We introduce one more auxiliary variable p satisfying the constraint $p_i = \nabla u_i$ ($p \in \mathbb{R}^{n,2}$ with $p_i \in \mathbb{R}^2$ as its i^{th} row), in additional to the constraint $u_i = v_i w_i$ in proposed BCA, and then rewrite (2) as the following equivalent constrained

optimization problem:

$$\begin{aligned}(27)\quad &\min_{(u,v,w,p)} \left\{ \frac{\lambda_1}{2}\sum_i (f_i - v_i)^2 + \sum_i |p_i| + \chi_{\mathcal{V}}(v) + \lambda_2 \sum_i (u_i - v_i \ln w_i - v_i) \right\}, \\ &s.t. \quad u_i = v_i w_i, \quad p_i = \nabla u_i.\end{aligned}$$

Then we can easily get the augmented Lagrangian of the above constrained optimization problem by introducing the multipliers Λ_w and Λ_p:

$$\begin{aligned}(28)\quad &L_{\alpha_w,\alpha_p}(u,v,w,p,\Lambda_w,\Lambda_p) = \frac{\lambda_1}{2}\sum_i (f_i - v_i)^2 + \sum_i |p_i| \\ &+ \lambda_2 \sum_i (u_i - v_i \ln w_i - v_i) + \langle \Lambda_w, v\circ w - u\rangle + \langle \Lambda_p, p - \nabla u\rangle \\ &+ \frac{\alpha_w}{2}\|v\circ w - u\|^2 + \frac{\alpha_p}{2}\|p - \nabla u\|^2 + \chi_{\mathcal{V}}(v),\end{aligned}$$

where $\alpha_w > 0$ and $\alpha_p > 0$ are the penalization parameters.

Given the previous iterative solution $(u^k, v^k, w^k, p^k, \Lambda_w^k, \Lambda_p^k)$, the ADMM consists of the following iterations

$$\begin{cases} u^{k+1} = \arg\min_u L_{\alpha_w,\alpha_p}(u, v^k, w^k, p^k, \Lambda_w^k, \Lambda_p^k), & (29a)\\ v^{k+1} = \arg\min_v L_{\alpha_w,\alpha_p}(u^{k+1}, v, w^k, p^k, \Lambda_w^k, \Lambda_p^k), & (29b)\\ w^{k+1} = \arg\min_w L_{\alpha_w,\alpha_p}(u^{k+1}, v^{k+1}, w, p^k, \Lambda_w^k, \Lambda_p^k), & (29c)\\ p^{k+1} = \arg\min_p L_{\alpha_w,\alpha_p}(u^{k+1}, v^{k+1}, w^{k+1}, p, \Lambda_w^k, \Lambda_p^k), & (29d)\\ \Lambda_w^{k+1} = \Lambda_w^k + \alpha_w(v^{k+1}\circ w^{k+1} - u^{k+1}), & (29e)\\ \Lambda_p^{k+1} = \Lambda_p^k + \alpha_p(p^{k+1} - \nabla u^{k+1}). & (29f)\end{cases}$$

We will show how to solve the subproblems w.r.t. (u, v, w, p) in the rest of this part. We consider the u-subproblem below:

$$u^{k+1} = \arg\min_u \left\{ \lambda_2 \sum_i u_i + \frac{\alpha_w}{2}\left\| v^k \circ w^k + \frac{\Lambda_w^k}{\alpha_w} - u \right\|^2 + \frac{\alpha_p}{2}\left\| p^k + \frac{\Lambda_p^k}{\alpha_p} - \nabla u \right\|^2 \right\}.$$

The first-order optimality condition of this subproblem is directly given below:

$$(30)\qquad \alpha_w u - \alpha_p \triangle u = -\lambda_2 + \Lambda_w^k - \nabla\cdot\Lambda_p^k + \alpha_w v^k \circ w^k - \alpha_p \nabla\cdot p^k,$$

where $\nabla\cdot$ denotes the divergence operator (conjugate operator of negative gradient $-\nabla$). We can readily solve the above equations by using conjugate gradient (CG) method or fast Fourier transform.

For v, w-subproblems (same to BCA, but with different notations), one can readily obtain that

$$(31)\qquad v^{k+1} = \max\left(\epsilon\mathbf{1}, \tilde{v}^{k+1}\right),$$

where

$$\tilde{v}^{k+1} = \frac{\mathbf{1}}{\lambda_1\mathbf{1} + \alpha_w (w^k)^2} \circ \left(\lambda_1 f + \lambda_2 \ln w^k + \lambda_2 - w^k \circ \Lambda_w^k + \alpha_w w^k \circ u^{k+1}\right).$$

and

$$w_i^{k+1} = \frac{1}{2v_i^{k+1}} \left[(u_i^{k+1} - \frac{\Lambda_{w(i)}^k}{\alpha_w}) + \sqrt{(u_i^{k+1} - \frac{\Lambda_{w(i)}^k}{\alpha_w})^2 + \frac{4\lambda_2 v_i^{k+1}}{\alpha_w}} \right]. \tag{32}$$

For the p-subproblem, one has

$$p^{k+1} = \arg\min_p \left\{ \sum_i |p_i| + \frac{\alpha_p}{2} \left\| p + \frac{\Lambda_p^k}{\alpha_p} - \nabla u^{k+1} \right\|^2 \right\}.$$

The solution is exactly the soft thresholding of $\nabla u^{k+1} - \frac{\Lambda_p^k}{\alpha_p}$:

$$p^{k+1} = \text{Thresh}_{\frac{1}{\alpha_p}}(\nabla u^{k+1} - \frac{\Lambda_p^k}{\alpha_p}), \tag{33}$$

where $\text{Thresh}_\eta(p) := \max\{0, |p| - \eta\} \circ \text{sign}(p)$, and $\text{sign}(p) := (\frac{p^{(1)}}{|p|}, \frac{p^{(2)}}{|p|})$, $|p| = \sqrt{|p^{(1)}|^2 + |p^{(2)}|^2}$. Note that all the operations here are element-wise.

Finally, the overall algorithm summarizing the above analysis is given below:

Algorithm 2 BCA$_f$

Input: Noisy data f and parameters λ_1, λ_2, α_w, α_p.
Initialization: $u^0 = f, v^0 = f, w^0 = \mathbf{1}, p^0 = \mathbf{0}, \Lambda_w^0 = \mathbf{0}, \Lambda_p^0 = \mathbf{0}, k = 0$.

1: **while** Stopping criteria is not satisfied **do**
2: Solve u^{k+1} by using CG for (30)
3: Find v^{k+1} by (31)
4: Find w^{k+1} by (32)
5: Find p^{k+1} by (33)
6: Update the multipliers by

$$\Lambda_w^{k+1} = \Lambda_w^k + \alpha_w(v^{k+1} \circ w^{k+1} - u^{k+1});$$
$$\Lambda_p^{k+1} = \Lambda_p^k + \alpha_p(p^{k+1} - \bigtriangledown u^{k+1}).$$

7: $k \leftarrow k + 1$.
8: **end while**

Here we remark that the convergence study of BCA$_f$ seems more difficult. If directly following the technique for BCA in the last subsection, since the subproblem for p is non-differentiable, the successive errors of Λ_p cannot be controlled by the successive errors of p such that it seems impossible to guarantee the sufficient decrease of the whole iterative sequences. In order to investigate the convergence of BCA$_f$ algorithm, we adopt the Huber type TV (HTV) regularization [20, 5] instead of the traditional total variation for the TV-IC model, which is Lipschitz differentiable. This change will not affect the solutions to the subproblems w.r.t. u, v, w. Hence only the p-subproblem of BCA$_f$ algorithm with HTV is given below.

$$p^{k+1} = \arg\min_p \left\{ \sum_i H_\tau(p_i) + \frac{\alpha_p}{2} \left\| p + \frac{\Lambda_p^k}{\alpha_p} - \nabla u^{k+1} \right\|^2 \right\}, \tag{34}$$

where

$$H_\tau(p) = \begin{cases} \frac{\tau}{2}|p|^2 & |p| \le \frac{1}{\tau}, \\ |p| - \frac{1}{2\tau} & \text{otherwise}, \end{cases}$$

is Huber function, τ is a positive constant. By using an operator splitting technique [31], we can readily get that

$$p^{k+1} = \max\left(\frac{\alpha_p}{\tau+\alpha_p}, 1 - \frac{1}{\alpha_p|\nabla u^{k+1} - \frac{\Lambda_p^k}{\alpha_p}|}\right) \circ \left(\nabla u^{k+1} - \frac{\Lambda_p^k}{\alpha_p}\right) \tag{35}$$

Following the same framework in section 3.2, we can prove the convergence of algorithm BCA$_f$-HTV (BCA$_f$ with Huber type TV regularization). Assumption 1 is still required to ensure that the iterative sequence is sufficiently decreased and bounded. Since the procedure of BCA$_f$-HTV's convergence analysis is similar to BCA, here we only give relevant theorems and no longer give detailed proofs.

Lemma 3.6. *Letting $w^{k+1}, p^{k+1}, \Lambda_w^{k+1}, \Lambda_p^{k+1}$ be generated by (29c), (34), (29e) and (29f), then we have*

$$\begin{aligned} \Lambda_w^{k+1} \circ w^{k+1} &= \lambda_2 \mathbf{1}. \\ \Lambda_p^{k+1} &= -\nabla H_\tau(p^{k+1}). \end{aligned} \tag{36}$$

Lemma 3.7. *Letting $(u^k, v^k, w^k, p^k, \Lambda_w^k, \Lambda_p^k)$ be the sequence generated by algorithm BCA$_f$-HTV, and $\alpha_w > \frac{\sqrt{2}\lambda_2}{c^2\epsilon}, \alpha_p > \sqrt{2}L$, then under Assumption 1 we have*

(37)

$$\begin{aligned} &L_{\alpha_w,\alpha_p}(u^k, v^k, w^k, p^k, \Lambda_w^k, \Lambda_p^k) - L_{\alpha_w,\alpha_p}(u^{k+1}, v^{k+1}, w^{k+1}, p^{k+1}, \Lambda_w^{k+1}, \Lambda_p^{k+1}) \\ \geq\ & \frac{\alpha_w}{2}\|u^{k+1} - u^k\|^2 + \frac{\lambda_1}{2}\|v^{k+1} - v^k\|^2 + \frac{\alpha_w}{2}\|w^k \circ (v^{k+1} - v^k)\|^2 \\ &+ D_1\|v^{k+1} \circ (w^{k+1} - w^k)\|^2 + D_2\|p^{k+1} - p^k\|^2, \end{aligned}$$

where L is the Lipschitz coefficient for the gradient of the Huber function $H_\tau(p)$, D_1 and D_2 are positive constants which are independent to k.

Lemma 3.8. *Denote $\mathcal{G}_1 : \Omega_1 \to \mathbb{R}$ by*

$$\begin{aligned} \mathcal{G}_1(u, v, w, p) = \frac{\lambda_1}{2}\|f - v\|^2 + \lambda_2\left\langle (\mathbf{1} - \frac{\mathbf{1}}{w}) \circ u - v \circ \ln w, \mathbf{1}\right\rangle \\ + \frac{\alpha_w}{2}\|v \circ w - u\|^2 + \frac{\alpha_p}{2}\|p - \nabla u\|^2, \end{aligned}$$

with $\alpha_w > \max\left(\lambda_2, \lambda_2\left(\frac{1}{c} - 1\right)^2\right)$, where $\Omega_1 := \{(u, v, w, p) \mid v_i \geq \epsilon > 0, w_i \geq c > 0 \ \ \forall\ i; u, v, w \in \mathbb{R}^n, p \in \mathbb{R}^{n\times 2}\}$. If $\|(u, v, w, p)\|_{\Omega_1} := \max\{\|u\|_\infty, \|v\|_\infty, \|w\|_\infty, \|p\|_\infty\} \to +\infty$, then we have $\mathcal{G}_1(u, v, w, p) \to +\infty$.

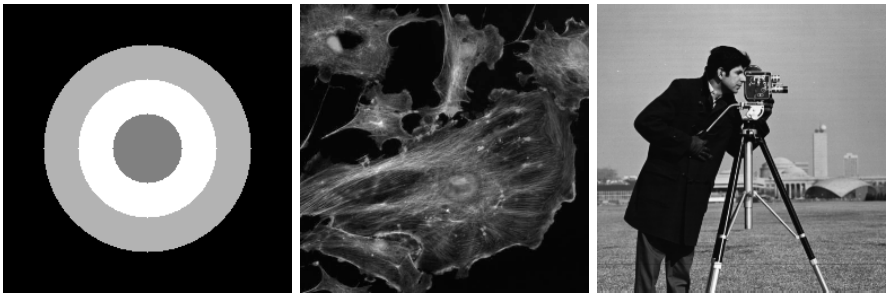
Now, let us consider the convergence of algorithm BCA$_f$-HTV.

Theorem 3.9. *Letting $\alpha_w > \max(\frac{\sqrt{2}\lambda_2}{c^2\epsilon}, \lambda_2, \lambda_2(\frac{1}{c} - 1)^2), \alpha_p > \sqrt{2}L$, under Assumption 1, we have*

1. *The sequence $(u^k, v^k, w^k, p^k, \Lambda_w^k, \Lambda_p^k)$ generated by proposed BCA$_f$-HTV is bounded and has at least one limit point.*
2. *The successive errors $u^{k+1} - u^k \to 0$, $v^{k+1} - v^k \to 0$, $w^{k+1} - w^k \to 0$, $p^{k+1} - p^k \to 0$, $\Lambda_w^{k+1} - \Lambda_w^k \to 0$ and $\Lambda_p^{k+1} - \Lambda_p^k \to 0$ as $k \to +\infty$.*
3. *Each limit point $(u^*, v^*, w^*, p^*, \Lambda_w^*, \Lambda_p^*)$ is a stationary point of $L_\alpha(u, v, w, p, \Lambda_w, \Lambda_p)$, and (u^*, v^*) is a stationary point of $H(u, v)$.*

4. **Numerical experiments.** Since Poisson noise is data-dependent, the noise level of the observed images depends on the pixel value, and therefore we introduce a scale factor $\eta \in (0, \infty)$ to control the scale of the image (simulating different number of photons detector received), which is inversely proportional to the amount of noise added to the data, i.e. $v_i \sim \text{Poisson}(\eta u)/\eta$. Meanwhile, we add Gaussian noise with different variances σ^2.

The Signal-to-Noise Ratio (SNR) in dB is used to measure the quality of the recovery result, defined as: $\text{SNR}(u, u_g) = -10 \log_{10} \frac{\sum_i |u_i - (u_g)_i|^2}{\sum_i |u_i|^2}$, where u_g is the ground-truth image (See Fig. 1) and u is the reconstructed image. The structural similarity (SSIM) index [38] is also provided to measure the quality of restored results (Bigger values mean the better quality).



(A) (B) (C)

FIGURE 1. (A) Circles; (B) Fluorescent Cells; (C) Cameraman

We set the stopping criterion as the successive error $\text{SE} := \frac{\|u_{k+1} - u_k\|}{\|u_k\|} \leq \xi$ or the iteration reaches 1000, where ξ is a desired tolerance. To evaluate the performance of the proposed algorithms, we compare them with other operator-splitting algorithms, including the L^2 data fitting (TV+L^2) method [39], KL-divergence (TV+KL) method [40], the exact Poisson-Gaussian (TV+EPG) likelihood model [11], Shifted-Poisson (TV+SP) method [7], the combination of L^2 data fitting and KL-divergence (TV+KL+L^2) method [32, 4] and the primal-dual (TV+PD) method for (2) following [22]. The last three algorithms were specially designed for the MPG noise. All the parameters for the compared algorithms are tuned heuristically to gain optimal image quality. For fair comparison, we set $\xi = 5 \times 10^{-4}$ for all compared algorithms and set initialization of the variable w.r.t. reconstructed output to noisy images. In order to further evaluate the performance of BCA$_f$-HTV, we compare it with the SSN algorithm [6] which is also designed for Huber-regularized TV-IC model.

All compared algorithms are implemented in Matlab, and performed using a Laptop with Intel Core i5 processor and 8GB RAM.

4.1. **Performances and convergence.** We first show how to determine the optimal inner iteration number for the proposed BCA, where we employ the gradient descent method proposed by Chambolle [8] to solve the u-subproblem. To find the optimal inner iteration number (More inner iterations will increase the overall computational cost), we show the SNR values of different inner iteration numbers

$(1, 2, 5, 10, 20, 100)$ in Table 1. Obviously one can see that when the inner iteration number is greater than 10, the SNR value is almost unchanged. Hence, in the latter tests, the inner loop number sets to 10 for the proposed BCA.

TABLE 1. SNR changes w.r.t. the number of inner iterations using gradient descent method [8] for proposed BCA Algorithm.

η	σ	1	2	5	10	20	100
1	10^{-1}	15.20	18.08	17.57	18.16	18.17	18.17
1	10^{-4}	16.20	18.39	16.35	18.40	18.40	18.40
4	10^{-1}	19.92	21.92	21.90	21.98	21.97	21.98
4	10^{-4}	20.05	22.37	22.36	22.47	22.48	22.49
16	10^{-1}	22.50	25.17	25.27	25.28	25.26	25.28
16	10^{-4}	23.59	26.17	26.40	26.37	26.38	26.40

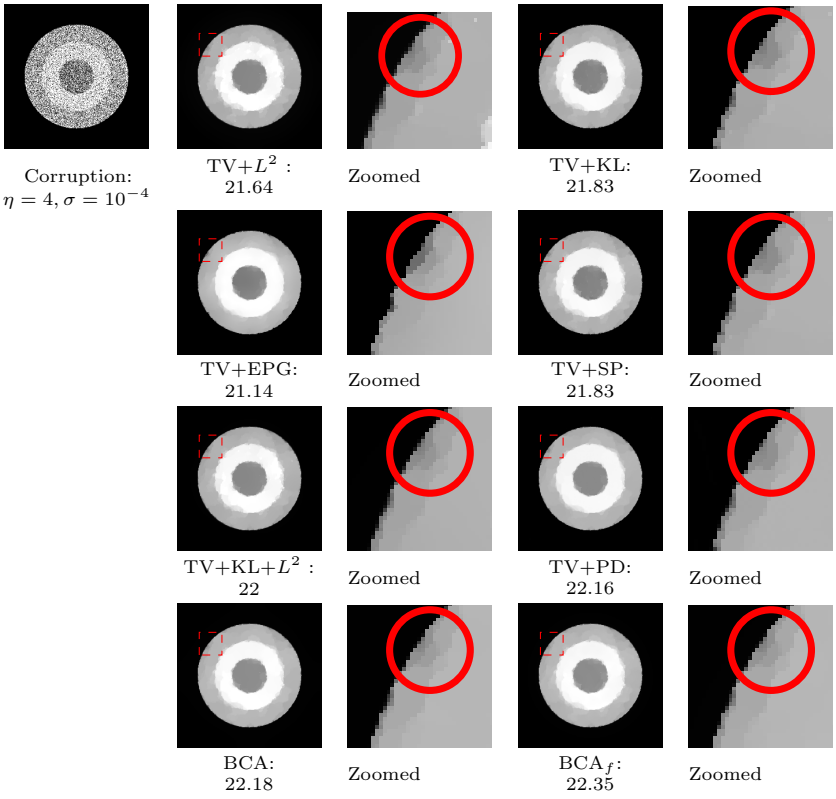


FIGURE 2. Recovery results by proposed algorithms and other compared algorithms (with SNRs(dB) below the figures) for the image "Circles" in the case of MPG noises which are generated with $\eta = 4, \sigma = 10^{-4}$.

We evaluate the performances of our proposed methods – BCA and BCA$_f$, compared with six other methods. The recovery results with zoomed regions are put in Figs. 2-4, where noises are generated with $\eta = 4, \sigma = 10^{-4}$, $\eta = 16, \sigma = 10^{-4}$, and $\eta = 64, \sigma = 10^{-1}$ for the three different images respectively. Generally speaking, one readily sees that the proposed BCA and BCA$_f$ generate better results compared with denoising methods including TV+L^2, TV+KL and TV+SP. In Fig. 2, one can observe that the region located at the red circles in the recovery results by the proposed BCA and BCA$_f$, especially the part below the edges, appears more flat than other compared algorithms. The recovery accuracy of recovery results by proposed algorithms with higher SNRs is also better than other compared algorithms, inferred from Fig. 2.

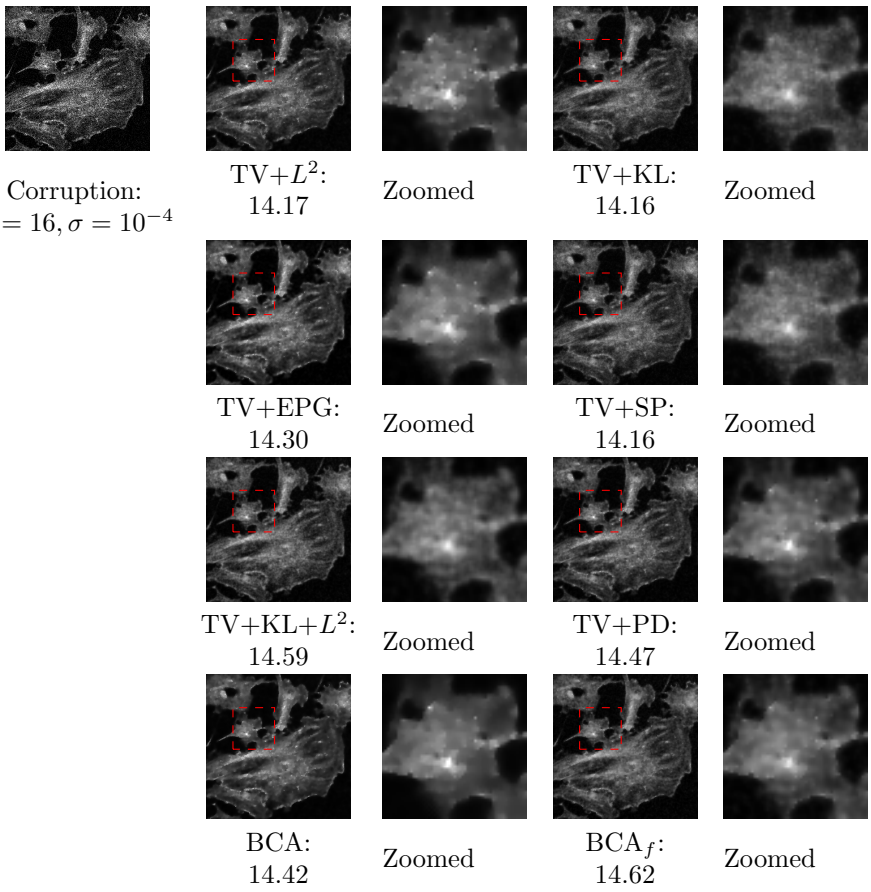


FIGURE 3. Recovery results by proposed algorithms and other compared algorithms (with SNRs(dB) below the figures) for the image "Fluorescent Cells" in the case of MPG noises which are generated with $\eta = 16, \sigma = 10^{-4}$.

In Fig. 4, there are less grey blocks in the restoration images of our proposed algorithms. We plot some line profiles to show it much clearer in Fig.5. It is obvious that the line profiles of clean image and our proposed methods are much smoother

than the others from 150 to 256 (x-axis) in Fig. 5, which indicates there are less grey blocks in our recovery images.Table 2 reports the SNRs and SSIMs of recovery images for all compared algorithms with more different noisy levels ($\eta = 1, 4, 16$ and $\sigma = 10^{-1}, 10^{-4}$), that demonstrates that the proposed algorithms gain highest SNRs and SSIMs averagely.

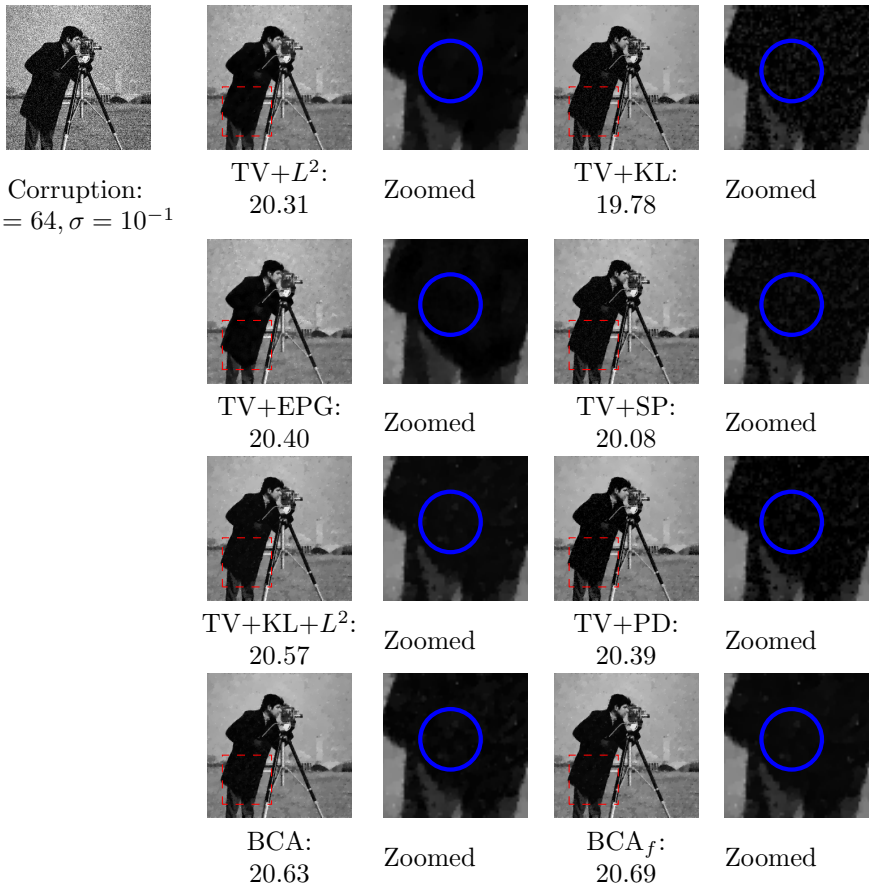


FIGURE 4. Recovery results by proposed algorithms and other compared algorithms (with SNRs(dB) below the figures) for the image "Cameraman" in the case of MPG noises which are generated with $\eta = 64, \sigma = 10^{-1}$. Comparing with other results, there are less grey blocks in the restoration images of proposed algorithms on the blue circle, which can show the better performance of our proposed algorithms.

In order to further show the advantage in term of speed for the proposed algorithms compared with TV+PD for the same model, we report the SNRs changes w.r.t. the elapsed CPU time in Fig. 6. Readily one can see the proposed algorithm converges much faster than TV+PD[1]. Table 3 reports the computational time of

[1]The iteration number for subproblems solved by Newton method affect the convergence speed of TV+PD, and 5 iterations are adopted heuristically to gain best speed.

the proposed algorithms and TV+PD. It is obvious that the proposed algorithms have higher speed than TV+PD. The fully splitting algorithm BCA_f computes much faster than BCA. We also remark that our proposed BCA and BCA_f have fewer parameters (three parameters for BCA, and four parameters for BCA_f), while compared TV+PD has six parameters.

TABLE 2. Denoising performance (First row: SNR in dB. Second row: SSIM.) with Poisson-Gaussian Noise

Image	η	σ	Noisy	TV+L^2	TV+KL	TV+EPG	TV+SP	TV+KL+L^2	TV+PD	BCA	BCA_f
Circle	1	10^{-1}	2.50	17.21	16.68	17.07	16.76	17.28	17.44	17.84	**17.97**
			0.0452	0.6147	0.4044	0.6580	0.4059	0.3975	0.7544	0.7125	**0.9007**
	1	10^{-4}	2.57	17.34	17.16	17.11	17.24	17.76	17.00	18.02	**18.10**
			0.5494	0.6087	0.7997	0.7740	0.8678	0.8251	0.8711	0.8913	**0.9029**
	4	10^{-1}	5.92	21.48	20.44	20.74	20.77	21.29	21.64	**22.01**	21.78
			0.0687	0.7149	0.4763	0.6260	0.5305	0.5259	0.9216	0.7153	**0.9357**
	4	10^{-4}	6.32	21.64	21.83	21.14	21.83	22.00	22.16	22.18	**22.35**
			0.5793	0.7397	0.9385	0.9395	0.9385	0.9299	0.9466	**0.9472**	0.9180
	16	10^{-1}	9.88	24.64	22.33	22.54	22.62	23.89	23.54	**25.28**	25.04
			0.0971	0.8411	0.5088	0.5315	0.5436	0.5438	0.8141	**0.9697**	0.8079
	16	10^{-4}	11.55	25.46	26.41	25.14	26.41	26.46	26.47	26.37	**27.12**
			0.6149	0.8816	0.9500	0.9447	0.9500	0.9594	0.9651	**0.9676**	0.9168
	Average		6.46	21.30	20.81	20.62	20.94	21.45	21.38	21.95	**22.06**
			0.3258	0.7335	0.6796	0.745	0.7061	0.6969	0.8788	0.8673	**0.8970**
Fluorescent Cells	1	10^{-1}	1.16	9.88	9.72	9.29	9.72	9.96	9.48	**10.37**	10.33
			0.0402	0.4861	0.4508	0.3149	0.4532	0.4512	0.4572	**0.5026**	0.4971
	1	10^{-4}	1.22	9.97	9.83	9.46	9.83	9.98	9.79	**10.43**	10.41
			0.0598	0.5058	0.4954	0.3289	0.4954	0.5003	0.4471	**0.5108**	0.5014
	4	10^{-1}	3.14	11.10	11.58	11.12	11.54	11.75	11.62	11.66	**12.06**
			0.1181	0.5554	0.5369	0.5093	0.5239	0.5588	0.5674	0.5753	**0.5801**
	4	10^{-4}	3.59	11.25	11.88	11.32	11.88	12.17	11.88	11.96	**12.38**
			0.1680	0.5765	0.6078	0.4593	0.6078	0.6133	0.5531	**0.6160**	0.6139
	16	10^{-1}	5.77	13.43	12.66	12.52	12.64	13.27	13.45	13.37	**13.50**
			0.2282	0.6424	0.5998	0.5271	0.5989	0.6383	0.6669	0.6557	**0.6685**
	16	10^{-4}	7.87	14.17	14.16	14.30	14.16	14.59	14.47	14.42	**14.62**
			0.4003	0.7360	0.7228	0.6895	0.7228	0.7388	0.7309	**0.7379**	0.7368
	Average		3.79	11.63	11.64	11.34	11.63	11.95	11.78	12.04	**12.22**
			0.1691	0.5837	0.5689	0.4710	0.5670	0.5835	0.5704	**0.5997**	0.5996
Cameraman	1	10^{-1}	1.97	13.06	14.25	13.13	14.19	14.30	14.30	**14.59**	14.56
			0.0496	0.4167	0.5498	0.3452	0.5322	0.5432	0.5524	0.5633	**0.5644**
	1	10^{-4}	2.00	13.09	14.36	13.04	14.36	14.40	14.33	**14.57**	14.52
			0.0628	0.4238	0.4602	0.3342	0.4602	0.4760	0.4362	**0.5854**	0.5659
	4	10^{-1}	5.02	15.66	16.46	15.5	16.43	16.56	16.17	16.79	**16.83**
			0.1178	0.5729	0.6552	0.4863	0.6540	0.6720	0.6422	0.6417	**0.6655**
	4	10^{-4}	5.28	15.81	16.27	15.64	16.27	16.47	16.14	16.99	**17.00**
			0.1514	0.5879	0.6301	0.5027	0.6301	0.6160	0.6047	0.6697	**0.6770**
	16	10^{-1}	9.06	18.25	18.60	18.24	18.60	18.81	18.18	18.99	**19.02**
			0.1992	0.6393	0.7126	0.6355	0.7181	0.7163	0.6527	0.7112	**0.7214**
	16	10^{-4}	10.24	18.88	19.44	18.83	19.44	19.64	19.59	**19.81**	19.53
			0.2920	0.6980	0.7321	0.6616	0.7321	0.7503	0.7317	0.7614	**0.7764**
	Average		5.60	15.79	16.56	15.73	16.55	16.70	16.45	**16.96**	16.91
			0.1455	0.5564	0.6233	0.4940	0.6211	0.6290	0.6033	0.6555	**0.6618**

Since SSN and BCA$_f$-HTV both considered the HTV regularization, we compare their performance, with the same Huber regularization parameter $\tau = 1e3$. Both algorithms are implemented in Matlab and performed using a Laptop with Intel Core i5 processor and 8G RAM. Fig. 7 shows that BCA$_f$-HTV can produce comparable results with SSN with less computational complexity.

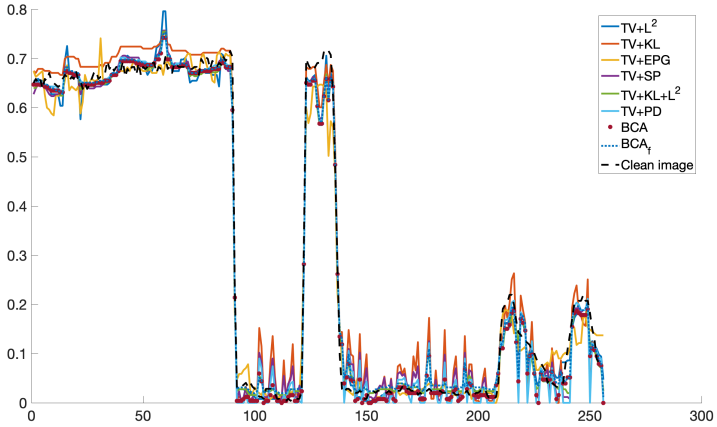


FIGURE 5. Line profiles of recovery results showed in Fig. 4

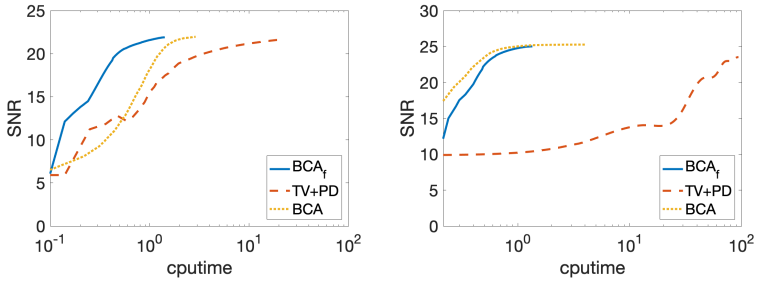
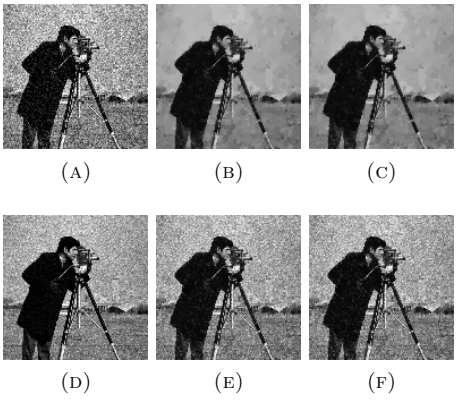


FIGURE 6. Histories of SNR changes for proposed algorithms and TV+PD w.r.t. the elapsed CPU time (in log scale). Left: $\eta = 4, \sigma = 10^{-1}$. Right: $\eta = 16, \sigma = 10^{-1}$.

To show the convergence of BCA, BCA$_f$ and BCA$_f$-HTV, we plot the convergence curves in Fig. 8 ($\eta = 1$ and $\sigma = 10^{-1}, 10^{-4}$). One can readily see that the errors (SE) are quite steadily decreasing, demonstrating the convergence of the proposed algorithms.

TABLE 3. Computational time of the proposed algorithms and TV+PD (in seconds)

Image	η	σ	TV+PD	BCA	BCA$_f$
Circle	1	10^{-1}	40.4971	4.6052	**1.0314**
	1	10^{-4}	18.5340	4.6922	**1.2650**
	4	10^{-1}	7.3436	0.7641	**0.7001**
	4	10^{-4}	39.6196	1.1891	**0.6030**
	16	10^{-1}	36.3221	1.2652	**0.4643**
	16	10^{-4}	39.5773	0.6123	**0.3070**
	Average		30.3156	2.1880	**0.7285**
Fluorescent Cells	1	10^{-1}	21.7999	8.3016	**2.1735**
	1	10^{-4}	41.6892	8.9243	**2.5687**
	4	10^{-1}	40.1406	4.1263	**2.4025**
	4	10^{-4}	38.6224	3.8464	**2.0889**
	16	10^{-1}	37.7174	7.3265	**6.1742**
	16	10^{-4}	7.7788	7.6185	**1.1258**
	Average		31.2914	6.6906	**2.7556**
Cameraman	1	10^{-1}	9.3735	1.3062	**1.2272**
	1	10^{-4}	40.9848	3.2183	**0.9053**
	4	10^{-1}	36.8623	**0.6872**	1.5084
	4	10^{-4}	36.4358	**0.6179**	1.4209
	16	10^{-1}	3.3691	**0.5151**	0.9925
	16	10^{-4}	11.0407	**0.3607**	0.5260
	Average		23.0140	1.1176	**1.0967**



(A) (B) (C)

(D) (E) (F)

FIGURE 7. Rcovery results by SSN and BCA$_f$-HTV. (A): Noisy image, SNR=12.79. (B): Recovery result by SSN, $SNR = 18.86, t = 6.763386s$. (C): Recovery result by BCA$_f$-HTV, $SNR = 18.88, t = 2.377923s$. (D): Poisson residuums of noisy images. (E): Poisson residuums by SSN. (F): Poisson residuums by BCA$_f$-HTV.

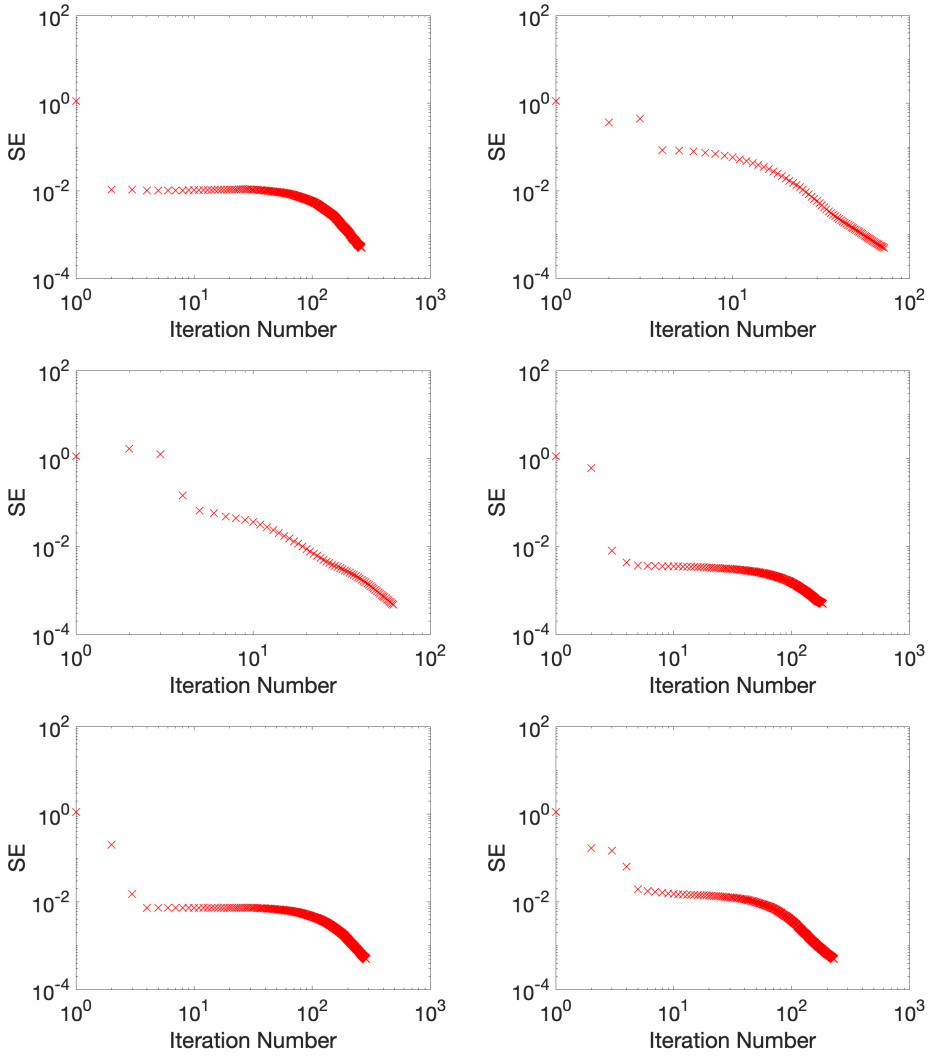


FIGURE 8. SE ($\frac{\|u_{k+1}-u_k\|}{\|u_k\|}$) changes v.s. iteration number (both in log scale): Top: BCA; Medium:BCA$_f$; Bottom: BCA$_f$-HTV; Left: $\eta = 1, \sigma = 10^{-1}$; Right: $\eta = 1, \sigma = 10^{-4}$. The test image is Circles(Fig1(A)).

4.2. **Robustness w.r.t. the parameters.** One readily knows that the model parameters λ_1 and λ_2 are critical to recovery results, which essentially balance the data fitting terms and the regularization terms. Here we only study how the algorithm parameters α, α_w and α_p affect the performance of proposed BCA and BCA$_f$ respectively. In Fig. 9, the SNR changes w.r.t. α for BCA algorithm is illustrated, in which λ_1 and λ_2 are fixed, and α varies from 12 to 500. Readily one can see that BCA is robust w.r.t. the parameter α.

As for BCA$_f$, the SNR changes w.r.t. different parameters are illustrated in Fig. 10(A), in which λ_1, λ_2 and α_p are fixed, and α_w varies from 18 to 700. The SNR changes are put in Fig. 10(B), where α_p varies from 1 to 512. Similar tests

are conducted for BCA$_f$-HTV, and related results are put in Fig. 11(A) and Fig. 11(B). All these results demonstrate that BCA$_f$ and BCA$_f$-HTV are quite robust w.r.t. the parameters α_w and α_p.

We have introduced a positive scalar ϵ for convergence guarantee of proposed BCA algorithm. To demonstrate its reasonableness, we show the performances changes w.r.t. this parameter, and put the SNRs changes of recovery results by BCA algorithm in Fig. 12. One can readily observe that when ϵ lies in the range from 10^{-10} to 10^{-1}, the SNR value is quite stable.

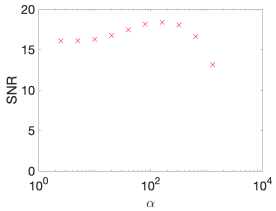
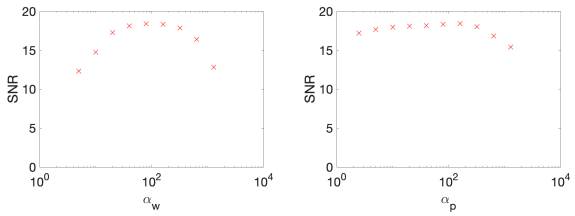
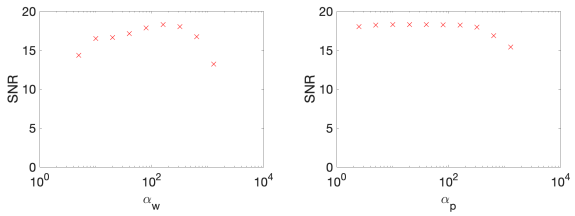


FIGURE 9. The performance of BCA w.r.t. α (in log scale), where $\eta = 1, \sigma = 10^{-4}$, using test image Circles(Fig1(A)).



(A) (B)

FIGURE 10. The performance of BCA$_f$ w.r.t. α_w (in log scale), α_p (in log scale), where $\eta = 1, \sigma = 10^{-4}$, using test image Circles(Fig1(A))



(A) (B)

FIGURE 11. The performance of BCA$_f$-HTV w.r.t. α_w (in log scale), α_p (in log scale), where $\eta = 1, \sigma = 10^{-4}$, using test image Circles(Fig1(A))

4.3. **Numerical validations.** To validate the Assumption 1 numerically, which is used for the convergence analysis of the proposed BCA, we plot the minimum value curves of the iterative solutions $\{w^k\}$ (see Fig. 13), which clearly show that the minimum value of w during iteration are almost bigger than 0.2, which show the reasonableness of this assumption.

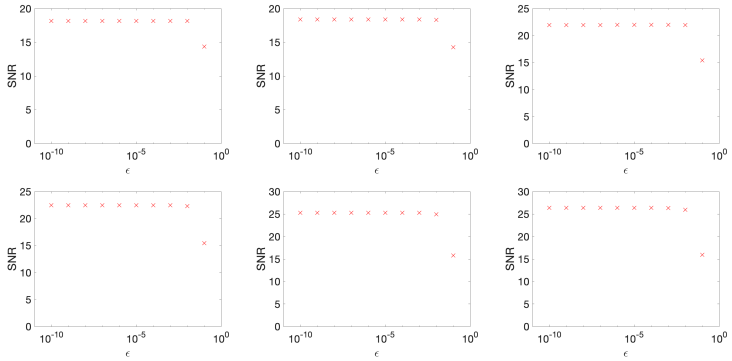


FIGURE 12. SNR changes w.r.t. the parameter ϵ for BCA Algorithm, using test image Circles(Fig1(A)).

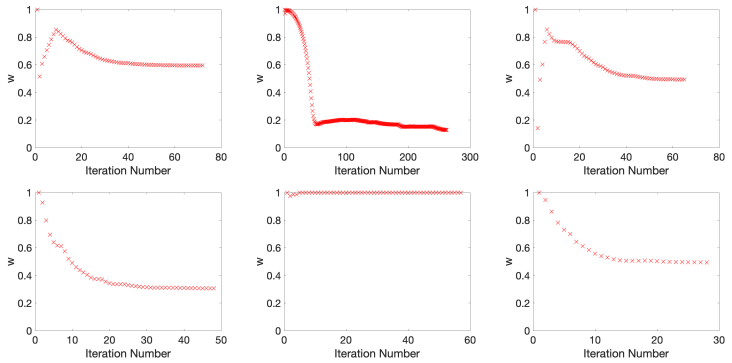


FIGURE 13. Minimum value curves of w for BCA Algorithm. The test image is Circles(Fig1(A)).

5. **Conclusion.** In this paper, we proposed new operator-splitting algorithms for MPG noise removal. A new bilinear constraint was introduced to ensure that all corresponding subproblems can be very efficiently solved. Numerical experiments showed that the proposed algorithms produced comparable results visually. Especially, compared with the TV+PD solving the same variational model, the proposed algorithms with fewer tunable parameters produced better recovery results at much faster speed. In the future, we aim to analyze the convergence of the proposed algorithm BCA_f with TV regularizer, particularly without Assumption 1, although it is empirically verifiable. Inspired by the outlook on parameter learning in [6], we plan to explore the optimal noise model learning as well. In addition, we are also

interested in extending the proposed algorithms to more general image reconstruction problem [13] as well as deblurring with background noise, and we also leave it as future work.

Acknowledgments. The authors would like to thank the editor and referees for their valuable comments. We are also grateful to Professor Luca Calatroni for providing the numerical results of SSN for Fig. 7.

J. Zhang and H. Chang were partially supported by National Natural Science Foundation of China (11871372, 11501413) and Natural Science Foundation of Tianjin (18JCYBJC16600). Y. Duan was partially supported by National Natural Science Foundation of China (12071345, 11701418), Major Science and Technology Project of Tianjin (18ZXRHSY00160) and Recruitment Program of Global Young Expert. Y. Lu was supported in part by Natural Science Foundation of China (11601389) and Doctoral Foundation of Tianjin Normal University (52XB1513). M. Ng was partially supported by HKRGC GRF 12200317, 12300218, 12300519 and 17201020.

REFERENCES

[1] B. Begovic, V. Stankovic and L. Stankovic, Contrast enhancement and denoising of Poisson and Gaussian mixture noise for solar images, *18th IEEE International Conference on Image Processing (ICIP),* Brussels, Belgium, 2011, 185–188.

[2] F. Benvenuto, A. L. Camera, C. Theys, A. Ferrari, H. Lantéri and M. Bertero, The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise, *Inverse Problems*, **24** (2008), 035016, 20pp.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Found. Trends Mach. Learn.*, **3** (2011), 1–122.

[4] L. Calatroni, C. Cao, J. D. L. Reyes, C. Schönlieb and T. Valkonen, Bilevel approaches for learning of variational imaging models, *Variational Methods,* 252–290, *Radon Ser. Comput. Appl. Math.*, 18, De Gruyter, Berlin, 2017. arXiv:1505.02120.

[5] L. Calatroni and K. Papafitsoros, Analysis and automatic parameter selection of a variational model for mixed Gaussian and salt-and-pepper noise removal, *Inverse Problems*, **35** (2019), 114001, 37pp.

[6] L. Calatroni, J. D. L. Reyes and C. Schönlieb, Infimal convolution of data discrepancies for mixed noise removal, *SIAM Journal on Imaging Sciences*, **10** (2017), 1196–1233.

[7] A. Chakrabarti and T. E. Zickler, Image restoration with signal-dependent Camera noise, preprint, arXiv:1204.2994.

[8] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vis.*, **20** (2004), 89–97.

[9] H. Chang, P. Enfedaque and S. Marchesini, Blind ptychographic phase retrieval via convergent alternating direction method of multipliers, *SIAM Journal on Imaging Sciences*, **12** (2019), 153–185.

[10] C. Chen, B. He, Y. Ye and X. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, *Mathematical Programming*, **155** (2016), 57–79.

[11] E. Chouzenoux, A. Jezierska, J. Pesquet and H. Talbot, A convex approach for image restoration with exact poisson-gaussian likelihood, *SIAM J. Imaging Sciences*, **8** (2015), 2662–2682.

[12] W. Deng, M. Lai, Z. Peng and W. Yin, Parallel multi-block ADMM with O (1/k) convergence, *Journal of Scientific Computing*, **71** (2017), 712–736.

[13] Q. Ding and Y. Long and X. Zhang and J. A. Fessler, Statistical image reconstruction using mixed poisson-gaussian noise model for X-ray CT, preprint, arXiv:1801.09533.

[14] J. Eckstein and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Mathematical Programming*, **55** (1992), 293–318.

[15] A. Foi, M. Trimeche, V. Katkovnik and K. Egiazarian, Practical poissonian-gaussian noise modeling and fitting for single-image raw-data, *IEEE Transactions on Image Processing*, **17** (2008), 1737–1754.
[16] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, *Computers & Mathematics with Applications*, **2** (1976), 17–40.
[17] R. Glowinski and A. Marroco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires, *Revue Française D'automatique, Informatique, Recherche Opérationnelle. Analyse Numérique*, **9** (1975), 41–76.
[18] D. Hajinezhad and Q. Shi, Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications, *Journal of Global Optimization*, **70** (2018), 261–288.
[19] M. Hong, Z. Luo and M. Razaviyayn, Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems, *SIAM Journal on Optimization*, **26** (2016), 337–364.
[20] P. J. Huber, Robust Estimation of a Location Parameter, *The Annals of Mathematical Statistics*, **35** (1964), 73–101.
[21] A. Jezierska, C. Chaux, J. Pesquet and H. Talbot, An EM approach for Poisson-Gaussian noise modeling, 19th European Signal Processing Conference (EUSIPCO), Barcelona, Spain, 2011, 2244–2248.
[22] A. Lanza, S. Morigi, F. Sgallari and Y. Wen, Image restoration with Poisson-Gaussian mixed noise, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, **2** (2014), 12–24.
[23] T. Le, R. Chartrand and T. Asaki, A variational approach to reconstructing images corrupted by poisson noise, *Journal of Mathematical Imaging and Vision*, **27** (2007), 257–263.
[24] J. Li and Z. Shen and R. Yin and X. Zhang, A reweighted L^2 method for image restoration with Poisson and mixed Poisson-Gaussian noise, *Inverse Problems & Imaging*, **9** (2015), 875–894.
[25] T. Lin, S. Ma and S. Zhang, On the global linear convergence of the admm with multiblock variables, *SIAM Journal on Optimization*, **25** (2015), 1478–1497.
[26] Y. Lou and M. Yan, Fast L1-L2 minimization via a proximal operator, *Journal of Scientific Computing*, **74** (2018), 767–785.
[27] M. Mäkitalo and A. Foi, Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise, *IEEE Transactions on Image Processing*, **22** (2013), 91–103.
[28] Y. Marnissi, Y. Zheng and J. Pesquet, Fast variational Bayesian signal recovery in the presence of Poisson-Gaussian noise, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Shanghai, China, 2016, 3964–3968.
[29] J. Mei, Y. Dong, T. Huang and W. Yin, Cauchy noise removal by nonconvex ADMM with convergence guarantees, *Journal of Scientific Computing*, **74** (2018), 743–766.
[30] F. Murtagh, J.-L Starck and A. Bijaoui, Image restoration with noise suppression using a multiresolution support, *Astronomy & Astrophysics, Suppl. Ser*, **112** (1995), 179–189.
[31] B. O' Donoghue, G. Stathopoulos and S. Boyd, A splitting method for optimal control, *IEEE Transactions on Control Systems Technology*, **21** (2013), 2432–2442.
[32] C. T. Pham, G. Gamard, A. Kopylov and T. Tran, An algorithm for image restoration with mixed noise using total variation regularization, *Turkish Journal of Electrical Engineering & Computer Sciences*, **26** (2018), 2832–2846.
[33] J. D. L. Reyes and C. Schönlieb, Image denoising: Learning the noise model via nonsmooth PDE-constrained optimization, *Inverse Problems & Imaging*, **7** (2013), 1183–1214.
[34] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica. D*, **60** (1992), 259–268.
[35] J.-L. Starck, F. Murtagh and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, New York, USA, 1998.
[36] D. N. H. Thanh and S. D. Dvoenko, A method of total variation to remove the mixed Poisson-Gaussian noise, *Pattern Recognition and Image Analysis*, **26** (2016), 285–293.
[37] Y. Wang, W. Yin and J. Zeng, Global convergence of ADMM in nonconvex nonsmooth optimization, *Journal of Scientific Computing*, **78** (2019), 29–63.

[38] 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*, **13** (2004), 600–612.
[39] C. Wu and X. Tai, Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models, *SIAM Journal on Imaging Sciences*, **3** (2010), 300–339.
[40] C. Wu, J. Zhang and X. Tai, Augmented Lagrangian method for total variation restoration with non-quadratic fidelity, *Inverse Problems & Imaging*, **5** (2011), 237–261.

Received January 2020; revised September 2020.

E-mail address: zj199607@connect.hku.hk
E-mail address: yuping.duan@tju.edu.cn
E-mail address: jinjin403@sina.com
E-mail address: mng@maths.hku.hk
E-mail address: changhuibin@gmail.com