



Poissonian Image Restoration Via the L_1/L_2 -Based Minimization

Mujibur Rahman Chowdhury¹ · Chao Wang^{2,3} · Yifei Lou⁴

Received: 7 January 2024 / Revised: 11 August 2024 / Accepted: 12 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

This study investigates the Poissonian image restoration problems. In particular, we propose a novel model that incorporates L_1/L_2 minimization on the gradient as a regularization term combined with a box constraint and a nonlinear data fidelity term, specifically crafted to address the challenges caused by Poisson noise. We employ a splitting strategy, followed by the alternating direction method of multipliers (ADMM) to find a model solution. Furthermore, we show that under mild conditions, the sequence generated by ADMM has a sub-sequence that converges to a stationary point of the proposed model. Through numerical experiments on image deconvolution, super-resolution, and magnetic resonance imaging (MRI) reconstruction, we demonstrate superior performance made by the proposed approach over some existing gradient-based methods.

Keywords Image restoration · L_1/L_2 minimization · Alternating direction method of multiplier · Poisson denoising

Mathematics Subject Classification 49N45 · 65K10 · 90C05 · 90C26

1 Introduction

Achieving a superior digital image quality holds significant importance across diverse domains, e.g., medical diagnosis, remote sensing, and astronomical imaging. However, the

✉ Mujibur Rahman Chowdhury
mujib.chowdhury2020@gmail.com; mujibur.chowdhury@McKesson.com

Chao Wang
wangc6@sustech.edu.cn

Yifei Lou
yflou@unc.edu

¹ McKesson Corporation, Irving, TX, USA

² Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518005, Guangdong Province, China

³ National Centre for Applied Mathematics Shenzhen, Shenzhen 518055, China

⁴ Department of Mathematics, School of Data Science and Society, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

recorded images are often corrupted by disparate artifacts such as undersampling, blurring, and noise [1]. For instance, super-resolution (SR) [2–5] involves reconstructing a high-quality image from its under-sampled low-resolution data. This process is challenging due to its inherently ill-posed nature, arising from an insufficient number of measurements. In a parallel vein, magnetic resonance imaging (MRI) undergoes reconstruction from under-sampled data in the Fourier space. Both SR and MRI applications highlight a common focus on addressing challenges associated with data degradation and scarcity [6, 7].

Many image recovery approaches assume the linearity of the degradation process, whether it involves convolutions with a specific point spread function (PSF), a collection of tomographic projections, or the straightforward identity operator for denoising. To measure the data misfit, minimizing the least-squares (LS) term is commonly used and defined by

$$\int_{\Omega} (\mathbf{A}\mathbf{u}(x) - \mathbf{f})^2 dx, \quad (1)$$

where $\mathbf{A}$ is a linear degradation operator, $\mathbf{u}$ is an image defined on a bounded domain of $\Omega \subset \mathbb{R}^2$, and $\mathbf{f}$ denotes the degraded measurements. Note that the gradient of the least-squares (1) is a linear function for $\mathbf{u}$, thus computationally favorable to minimize. However, the LS formulation, as a data fidelity term, is only effective for Gaussian noise.

It is challenging to apply algorithms designed for linear operators and Gaussian noise to tackle other nonlinear inverse problems due to optimization difficulties caused by nonlinearity [8, 9]. This paper focuses on the Poissonian image restoration, which entails a nonlinear data fidelity term. Specifically, the data obtained by photon-count devices is often corrupted by Poisson noise with applications to applied optics [10], astronomical [11] and biomedical fields [12, 13]. Unlike Gaussian noise, Poisson noise is signal-dependent, thus making it challenging to remove. It follows from statistical properties of the Poisson distribution [14, 15] that the data fidelity term for Poissonian image restoration can be formulated by

$$\int_{\Omega} (\mathbf{A}\mathbf{u}(x) - \mathbf{f} \ln(\mathbf{A}\mathbf{u}(x))) dx, \quad (2)$$

with the same setup as in (1).

Aside from machine learning and deep learning methodologies that require a large number of training data, a standalone approach for image recovery tasks often considers certain regularizers to impose prior knowledge and constraints for guiding the reconstruction process from degraded images. For example, total variation (TV) [16] is a celebrated regularization model in image processing. While its primary advantage lies in edge preservation, it can induce stair-casing artifacts. To mitigate these artifacts, fractional-order total variation (FOTV) [17, 18] is formulated to generate piece-wise smooth outputs. A family of the sum or the difference of two norms on the image gradient has been popular for image applications, including $L_2 + L_p$ with $p < 1$ [19], $L_1 - \alpha L_2$ [20, 21], and $L_1 - \beta L_q$ with $q > 1$ [22]. In addition, nonlocal operators [23] based on patch similarity have emerged to preserve fine structures. Other regularizations in the image-processing literature include sparsity-promoting regularizations [24–26], minimax concave penalty (MCP) regularization [27], hyperbolic tangent function [28], and a weighted Schatten norm [29] that exploits low-rank structures among similar patches. Additionally, there are some hybrid models, such as TV and wavelet frame [30], TV and tensor sparse representation [31], joint low-rank and nonlocal prior [32], and low-rank and sparse decomposition method [33]. Specifically for Poissonian image restoration, the FOTV regularization is combined with the data fidelity term (2) for denoising [34] and nonblind deconvolution [35, 36]. The frame-based analysis approach was considered

in [37]. Moreover, Kumar et al. [38] proposed low-rank Poisson denoising, while a non-local means approach was presented in [39]. Recently, advanced Poisson denoising methods have taken advantage of deep learning [40] and self-supervised learning [41]. Zha et al. [42] considered both low-rank regularization and deep priors to get rid of Poisson noise.

The recent success of the L_1/L_2 regularization in sparse signal processing [43, 44] has drawn significant attention to this scale-invariant model. An image, specifically with a piece-wise constant structure, is sparse after taking the operator, suggesting that L_1/L_2 on the gradient could serve as an effective prior for imaging applications. This L_1/L_2 regularization is applied to the CT reconstruction problem in [45–47], where the noise is assumed to be additive Gaussian distribution. In this paper, we introduce an L_1/L_2 -based image restoration model tailored for Poisson statistics and adapt the alternating direction method of multipliers (ADMM) [48] to solve the optimization problem. The convergence analysis is particularly intricate due to the nonlinearity, nonconvexity, and nonseparability in L_1/L_2 regularization. The nonquadratic and non-separable characteristics of the Poisson log-likelihood further complicate challenges in this model. Notably, the absence of a Lipschitz-continuous gradient in the data fidelity term (2) for Poisson statistics becomes a major impediment when analyzing the convergence of gradient-based algorithms. In light of these challenges, we conduct theoretical analysis to establish a subsequential convergence under certain assumptions. Furthermore, our numerical experiments yield promising results, showcasing the proposed approach's exceptional performance in diverse applications, including image deconvolution, super-resolution, and MRI reconstruction. These experiments emphasize the model's superiority in recovering images with intricate piece-wise structures, providing empirical validation to complement the theoretical advancements.

The remaining sections of the paper are outlined below. In Sect. 2, we present the notations and examine the L_1/L_2 model along with the derivation for the Poisson data fidelity term. Our proposed image restoration method and algorithm are described in Sect. 3. In Sect. 4, we establish the convergence analysis of the algorithm under some mild conditions. Section 5 is devoted to experimental results. Finally, Sect. 6 concludes the paper.

2 Preliminaries

We adopt a discrete formulation for our problem setting. A two-dimensional image with dimensions $m \times n$ can be expressed as a vector, where the $((i-1)m+j)$ -th component represents the intensity value at pixel (i, j) using a conventional linear index. We define the discrete gradient operator as

$$D\mathbf{u} := \begin{bmatrix} D_x \\ D_y \end{bmatrix} \mathbf{u}, \quad (3)$$

where the operators D_x and D_y represent finite forward differences with a periodic boundary condition in x , y -directions, respectively. Define $N := mn$, $\mathcal{X} := \mathbb{R}^N$, and $\mathcal{Y} := \mathbb{R}^{2N}$. Throughout the paper, we employ bold letters to represent vectors, capital letters to denote matrices or linear operators, and calligraphic letters to indicate vector spaces. We have the flexibility to employ standard norms on vectors $\mathbf{u} \in \mathcal{X}$ and $D\mathbf{u} \in \mathcal{Y}$. For example, the L_1 applied on $\mathbf{u}$ promotes the sparsity of $\mathbf{u}$, while the L_1 norm applied to $D\mathbf{u}$, i.e., $\|D\mathbf{u}\|_1$, is so-called (anisotropic) TV regularization [49] that encourages the image gradient to be sparse.

Wang et al. [45] proposed the L_1/L_2 penalty on the gradient,

$$\min_{\mathbf{u} \in \mathcal{X}} \frac{\|D\mathbf{u}\|_1}{\|D\mathbf{u}\|_2} + \frac{\beta}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2, \quad (4)$$

where $\mathbf{f}$ is the observed data in a vector form, A denotes a linear degradation operator, and $\beta > 0$ is a weighting parameter. The work [45] focused on the problem of limited-angle CT reconstruction, where the linear operator A is a truncated (due to limited angle) Radon transform. The least-squares fit in (4) indicates implicitly that the noise in the data $\mathbf{f}$ adheres to a Gaussian distribution.

In this paper, we are interested in the Poisson noise distribution. If the noise, as a random variable, follows a Poisson distribution, its expected value is given by $(\mathbf{A}\mathbf{u})_i$ for the i -th entry. Thus, the probability distribution of the measured data can be written by:

$$p(\mathbf{f} | \mathbf{u}) = \prod_i \frac{(\mathbf{A}\mathbf{u})_i^{f_i}}{f_i!} e^{-(\mathbf{A}\mathbf{u})_i}.$$

The maximum a posteriori (MAP) estimation, achieved through the negative log-likelihood function, asymptotically leads to the following data fidelity term

$$\min_{\mathbf{u} \in \mathcal{X}} D_{KL}(\mathbf{A}\mathbf{u}, \mathbf{f}), \quad (5)$$

where $D_{KL}(\mathbf{x}, \mathbf{w})$ refers to the generalized Kullback-Leibler divergence [50]:

$$D_{KL}(\mathbf{x}, \mathbf{w}) = \left\langle \mathbf{w}, \ln \frac{\mathbf{w}}{\mathbf{x}} \right\rangle + \langle \mathbf{1}, \mathbf{x} - \mathbf{w} \rangle, \quad (6)$$

with the all-one vector $\mathbf{1}$. The division of two vectors and the logarithmic function in (6) are applied elementwise. Note that the objective in (5) can be rewritten as $\langle \mathbf{1}, \mathbf{A}\mathbf{u} \rangle - \langle \mathbf{f}, \ln(\mathbf{A}\mathbf{u}) \rangle$.

3 Proposed Method

We propose an L_1/L_2 -regularized Poissonian image restoration model along with a box constraint, i.e.,

$$\min_{\mathbf{u} \in \mathcal{X}} \frac{\|D\mathbf{u}\|_1}{\|D\mathbf{u}\|_2} + \beta \langle \mathbf{1}, \mathbf{A}\mathbf{u} \rangle - \beta \langle \mathbf{f}, \ln(\mathbf{A}\mathbf{u}) \rangle \quad \text{s.t. } \mathbf{u} \in [t_1, t_2]^N. \quad (7)$$

Here $\mathbf{u} \in [t_1, t_2]^N$ signifies that each element of $\mathbf{u}$ falls within the bounds of $[t_1, t_2]$. This is commonly observed in image processing, where pixel values are often constrained to intervals like $[0, 1]$ or $[0, 255]$. Note that the optimal solution of (7), denoted by $\hat{\mathbf{u}}$, satisfies $\mathbf{A}\hat{\mathbf{u}} > \mathbf{0}$ so that $\ln(\mathbf{A}\hat{\mathbf{u}})$ is well-defined, where the notation $>$ means each component of $\mathbf{A}\hat{\mathbf{u}}$ is larger than 0.

We apply the alternating direction method of multipliers (ADMM) [48] to solve the optimization problem (7) by reformulating it into an equivalent form

$$\begin{aligned} \min_{\mathbf{u} \in \mathcal{X}, \mathbf{g} \in \mathcal{Y}} \quad & E(\mathbf{u}, \mathbf{g}) := \frac{\|D\mathbf{u}\|_1}{\|\mathbf{g}\|_2} + \beta \langle \mathbf{1}, \mathbf{A}\mathbf{u} \rangle - \beta \langle \mathbf{f}, \ln(\mathbf{A}\mathbf{u}) \rangle + \delta_{[t_1, t_2]^N}(\mathbf{u}) \\ \text{subject to} \quad & \mathbf{g} = D\mathbf{u}, \end{aligned} \quad (8)$$

where $\delta_Q(\mathbf{x})$ is the indicator function that forces $\mathbf{x}$ to belong to a feasible set Q , i.e.,

$$\delta_Q(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in Q, \\ +\infty & \text{otherwise.} \end{cases} \quad (9)$$

With a dual variable $\mathbf{t}$, the augmented Lagrangian function related to (8) is written as

$$\begin{aligned} \mathcal{L}(\mathbf{u}, \mathbf{g}; \mathbf{t}) = & \frac{\|\mathbf{Du}\|_1}{\|\mathbf{g}\|_2} + \beta \langle \mathbf{1}, \mathbf{Au} \rangle - \beta \langle \mathbf{f}, \ln(\mathbf{Au}) \rangle + \delta_{[t_1, t_2]^N}(\mathbf{u}) \\ & + \langle \rho \mathbf{t}, \mathbf{Du} - \mathbf{g} \rangle + \frac{\rho}{2} \|\mathbf{Du} - \mathbf{g}\|_2^2, \end{aligned} \quad (10)$$

where $\rho > 0$ is a fixed parameter to be tuned. Then the ADMM scheme is given as

$$\begin{cases} \mathbf{u}^{(k+1)} = \arg \min_{\mathbf{u}} \mathcal{L}(\mathbf{u}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) \\ \mathbf{g}^{(k+1)} = \arg \min_{\mathbf{g}} \mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}; \mathbf{t}^{(k)}) \\ \mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + \mathbf{Du}^{(k+1)} - \mathbf{g}^{(k+1)}. \end{cases} \quad (11)$$

The $\mathbf{g}$ -subproblem in (11) has a closed-form solution given by [43, page A3655]

$$\mathbf{g}^{(k+1)} = \begin{cases} \mathbf{e}^{(k)} & \mathbf{Du}^{(k+1)} + \mathbf{t}^{(k)} = \mathbf{0} \\ \tau^{(k)} (\mathbf{Du}^{(k+1)} + \mathbf{t}^{(k)}) & \mathbf{Du}^{(k+1)} + \mathbf{t}^{(k)} \neq \mathbf{0}, \end{cases} \quad (12)$$

where $\mathbf{e}^{(k)}$ is a random vector with its L_2 norm equal to $\sqrt[3]{\frac{\|\mathbf{Du}^{(k+1)}\|_1}{\rho}}$ and $\tau^{(k)}$ is determined by the expression $\frac{1}{3} + \frac{1}{3}(\zeta^{(k)} + \frac{1}{\zeta^{(k)}})$ with

$$\zeta^{(k)} = \sqrt[3]{\frac{27\kappa^{(k)} + 2 + \sqrt{(27\kappa^{(k)} + 2)^2 - 4}}{2}} \quad \text{and} \quad \kappa^{(k)} = \frac{\|\mathbf{Du}^{(k+1)}\|_1}{\rho \|\mathbf{Du}^{(k+1)} + \mathbf{t}^{(k)}\|_2^3}.$$

Regarding the $\mathbf{u}$ -subproblem in (11), we elaborate on it as solving the minimization problem:

$$\min_{\mathbf{u} \in \mathcal{X}} \frac{\|\mathbf{Du}\|_1}{\|\mathbf{g}^{(k)}\|_2} + \beta \langle \mathbf{1}, \mathbf{Au} \rangle - \beta \langle \mathbf{f}, \ln(\mathbf{Au}) \rangle + \frac{\rho}{2} \|\mathbf{Du} - \mathbf{g}^{(k)} + \mathbf{t}^{(k)}\|_2^2 + \delta_{[t_1, t_2]^N}(\mathbf{u}), \quad (13)$$

which requires an iterative process to find the optimal solution. For this purpose, we introduce three variables $\mathbf{v} := \mathbf{u}$, $\mathbf{d} := \mathbf{Du}$, $\mathbf{q} := \mathbf{Au}$. Then the objective function in (13) can be expressed as

$$\frac{\|\mathbf{d}\|_1}{\|\mathbf{h}^{(k)}\|_2} + \beta \langle \mathbf{1}, \mathbf{q} - \mathbf{f} \ln \mathbf{q} \rangle + \frac{\rho}{2} \|\mathbf{Du} - \mathbf{g}^{(k)} + \mathbf{t}^{(k)}\|_2^2 + \delta_{[t_1, t_2]^N}(\mathbf{v}). \quad (14)$$

The corresponding augmented Lagrangian is expressed as,

$$\begin{aligned} \mathcal{L}^{(k)}(\mathbf{u}, \mathbf{q}, \mathbf{d}, \mathbf{v}; \mathbf{y}, \mathbf{w}, \mathbf{z}) = & \frac{\|\mathbf{d}\|_1}{\|\mathbf{h}^{(k)}\|_2} + \beta \langle \mathbf{1}, \mathbf{q} - \mathbf{f} \ln \mathbf{q} \rangle + \frac{\rho}{2} \|\mathbf{Du} - \mathbf{g}^{(k)} + \mathbf{t}^{(k)}\|_2^2 \\ & + \delta_{[t_1, t_2]^N}(\mathbf{v}) + \langle \mu \mathbf{y}, \mathbf{u} - \mathbf{v} \rangle + \frac{\mu}{2} \|\mathbf{v} - \mathbf{u}\|_2^2 \\ & + \langle \gamma \mathbf{w}, \mathbf{Du} - \mathbf{d} \rangle + \frac{\gamma}{2} \|\mathbf{d} - \mathbf{Du}\|_2^2 \\ & + \langle \lambda \mathbf{z}, \mathbf{Au} - \mathbf{q} \rangle + \frac{\lambda}{2} \|\mathbf{q} - \mathbf{Au}\|_2^2, \end{aligned} \quad (15)$$

where $\mathbf{y}$, $\mathbf{w}$, $\mathbf{z}$ serve as dual variables, and λ , γ , μ are positive parameters. The superscript k in $\mathcal{L}$ denotes the Lagrangian for the $\mathbf{u}$ -subproblem in (11) at the k -th iteration. The ADMM framework to minimize (14) is similar to (11) and is summarized in Algorithm 1.

When differentiating $\mathcal{L}^{(k)}$ with respect to $\mathbf{u}$, we obtain the update,

$$\mathbf{u}_{j+1} = \left(\lambda A^T A + (\rho + \gamma) D^T D + \mu I \right)^{-1} \mathbf{c}_{j+1}, \quad (16)$$

with $\mathbf{c}_{j+1} = \rho D^T (\mathbf{g}^{(k)} - \mathbf{t}^{(k)}) + \lambda A^T (\mathbf{q}_j - \mathbf{z}_j) + \gamma D^T (\mathbf{d}_j - \mathbf{w}_j) + \mu (\mathbf{v}_j - \mathbf{y}_j)$ and the identity matrix I . Here the subscript j is the inner loop index, in contrast to the superscript k in (11) for counting outer iterations. Note that when the matrix $A^T A$ can be diagonalized by the Fourier transform, the inversion in (16) can be calculated efficiently by the fast Fourier transform (FFT). Its rationale lies in the fact that $D^T D$ with the gradient operator D defined in (3) can be diagonalized by the Fourier transform. In our experiments, we consider image deblurring, super-resolution, and MRI construction, all of which involve frequency measurements, and hence, the update for $\mathbf{u}_{j+1}$ can be obtained via FFT. We formulate the $\mathbf{q}$ -subproblem as follows,

$$\min_{\mathbf{q}} \left\{ \beta \langle \mathbf{1}, \mathbf{q} - \mathbf{f} \ln \mathbf{q} \rangle + \frac{\lambda}{2} \|\mathbf{A}\mathbf{u} + \mathbf{z}_j - \mathbf{q}\|_2^2 \right\}, \quad (17)$$

which can be solved entry-wise as a standard one-dimensional KL divergence problem [51, Lemma 2.2]. Specifically, the update for $\mathbf{q}$ can be given by the quadratic formula,

$$\mathbf{q}_{j+1} = \frac{(\lambda \mathbf{A}\mathbf{u}_{j+1} + \lambda \mathbf{z}_j - \beta) + \sqrt{(\lambda \mathbf{A}\mathbf{u}_{j+1} + \lambda \mathbf{z}_j - \beta)^2 + 4\beta\lambda\mathbf{f}}}{2\lambda}. \quad (18)$$

Note that we choose the positive solution from the quadratic formula, i.e., $\mathbf{q}_{j+1} > 0$.

Lastly, it is straightforward to get the closed-form solution for $\mathbf{d}$, that is,

$$\mathbf{d}_{j+1} = \text{shrink} \left(D\mathbf{u}_{j+1} + \mathbf{w}_j, \frac{1}{\gamma \|\mathbf{g}^{(k)}\|_2} \right), \quad (19)$$

where $\text{shrink}(\mathbf{x}, \nu)_i := \text{sign}(x_i) \max\{|x_i| - \nu, 0\}$ for the i -th entry.

We summarize the pseudo-code for minimizing the proposed model (7) in Algorithm 1 that consists of (outer) iterations of (11) with (inner) loops for solving the $\mathbf{u}$ -subproblem.

4 Convergence Analysis

In this section, we analyze the convergence of the ADMM iterations (11) under the following assumptions:

- (A1) The linear operator A satisfies $A\mathbf{1} = \mathbf{1}$ and has nonnegative entries.
- (A2) $\mathcal{N}(D) \cap \mathcal{N}(A) = \{0\}$, where the gradient operator D is defined in (3) and $\mathcal{N}$ denotes the null space.
- (A3) The norm of the sequence $\{\mathbf{g}^{(k)}\}$ produced by (11) possesses a (uniform) lower bound. Specifically, there exists a positive constant ϵ such that $\|\mathbf{g}^{(k)}\|_2 \geq \epsilon$ holds for all iterations k .

Note that the operator A for denoising is the identity matrix, while the convolution kernel for deblurring sums to 1. Therefore, the assumption (A1) is trivial for denoising and deblurring. The assumption (A2) is standard in convergence analysis for TV-based regularizations [21, 45, 46, 52]. The last assumption (A3) requires a uniform lower bound of the iterates $\{\mathbf{g}^{(k)}\}$, which is seemingly strong. However, it is inevitably essential in the convergence analysis for an objective function that is not coercive. Fortunately, the assumption (A3) is easy to verify numerically.

Algorithm 1: The L_1/L_2 minimization in the Poisson noise model.

```

1 Input: the observed image  $\mathbf{f}$  and the linear operator  $A$ 
2 Set parameters:  $\beta, \lambda, \gamma, \mu, \epsilon$ , OuterMax, InnerMax
3 Initialization:  $k = 0, \mathbf{u}^{(0)} = \mathbf{0}, \mathbf{v}^0 = \mathbf{0}, \mathbf{d}^{(0)} = \mathbf{0}, \mathbf{g}^{(0)} = \mathbf{0}, \mathbf{t}^{(0)} = \mathbf{0}, \mathbf{y}^{(0)} = \mathbf{0}, \mathbf{w}^{(0)} = \mathbf{0}, \mathbf{z}^{(0)} =$ 
    $\mathbf{0}, \mathbf{n}^{(0)} = \mathbf{0}, \mathbf{q}^{(0)} = \mathbf{f}$ 
4 while  $\frac{\|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\|_2}{\|\mathbf{u}^{(k)}\|_2} > \epsilon$  and  $k \leq \text{OuterMax}$  do
5    $j = 0$ .
6   while  $j \leq \text{InnerMax}$  do
7      $\mathbf{c}_{j+1} = \mathbf{n}^{(k)} + \lambda A^T(\mathbf{q}_j - \mathbf{z}_j) + \gamma D^T(\mathbf{d}_j - \mathbf{w}_j) + \mu(\mathbf{v}_j - \mathbf{y}_j)$ 
      $\mathbf{u}_{j+1} = \left( \lambda A^T A + (\rho + \gamma) D^T D + \mu I \right)^{-1} \mathbf{c}_{j+1}$ 
      $\mathbf{q}_{j+1} = \frac{(\lambda A \mathbf{u}_{j+1} + \lambda \mathbf{z}_j - \beta) + \sqrt{(\lambda A \mathbf{u}_{j+1} + \lambda \mathbf{z}_j - \beta)^2 + 4\beta\lambda\mathbf{f}}}{2\lambda}$ 
      $\mathbf{d}_{j+1} = \text{shrink}(D \mathbf{u}_{j+1} + \mathbf{w}_j, \frac{1}{\gamma \|\mathbf{g}^{(k)}\|_2})$ 
      $\mathbf{y}_{j+1} = \mathbf{y}_j + \mathbf{u}_{j+1} - \mathbf{v}_{j+1}$ 
      $\mathbf{w}_{j+1} = \mathbf{w}_j + D \mathbf{u}_{j+1} - \mathbf{d}_{j+1}$ 
      $\mathbf{z}_{j+1} = \mathbf{z}_j + A \mathbf{u}_{j+1} - \mathbf{q}_{j+1}$ 
      $j = j + 1$ 
8   end
9   return  $\mathbf{u}^{(k+1)} = \mathbf{u}_j$ 
10   $\mathbf{g}^{(k+1)} = \begin{cases} \mathbf{e}^{(k)} & D \mathbf{u}^{(k+1)} + \mathbf{t}^{(k)} = \mathbf{0} \\ \tau^{(k)} (D \mathbf{u}^{(k+1)} + \mathbf{t}^{(k)}) & D \mathbf{u}^{(k+1)} + \mathbf{t}^{(k)} \neq \mathbf{0} \end{cases}$ 
11   $\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + D \mathbf{u}^{(k+1)} - \mathbf{g}^{(k+1)}$ 
12   $\mathbf{n}^{(k+1)} = \rho D^T(\mathbf{g}^{(k)} - \mathbf{t}^{(k)})$ 
13   $k = k + 1$ 
14 end
15 Output:  $\mathbf{u}^{(k)}$ 

```

As the sequence $\{\mathbf{u}^{(k)}\}$ is bounded (by $[t_1, t_2]$), then $\{\nabla \mathbf{u}^{(k)}\}$ is also bounded and we denote $M = \sup_k \{\|\nabla \mathbf{u}^{(k)}\|_1\}$. To estimate the bounds of $\mathbf{u}$ -subproblem, we rely on a property of strongly convex functions characterized in Lemma 1.

Lemma 1 [53] *A function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is strongly convex with $\mu > 0$ if and only if the following inequality*

$$f(\mathbf{w}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{w} - \mathbf{x} \rangle + \frac{\mu}{2} \|\mathbf{w} - \mathbf{x}\|_2^2,$$

holds for any vectors $\mathbf{x}, \mathbf{w} \in \mathbb{R}^N$.

Lemma 2 (sufficient decreasing) *Let $\{(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}, \mathbf{t}^{(k)})\}$ be the sequence of iterates defined by (11). Under the assumptions (A1)–(A3) and a sufficiently large $\rho > 0$, there exist two positive parameters c_1, c_2 such that*

$$\begin{aligned} & \mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k+1)}) - \mathcal{L}(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) \\ & \leq -c_1 \|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\|_2^2 - c_2 \|\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\|_2^2. \end{aligned} \quad (20)$$

Proof We start by showing that there exists a positive parameter c such that

$$\mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) - \mathcal{L}(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) \leq -c \|\mathbf{u}^{(k)} - \mathbf{u}^{(k+1)}\|_2^2. \quad (21)$$

As $\frac{\|D\mathbf{u}\|_1}{\|\mathbf{g}^{(k)}\|_2}$ and the indicator function $\delta_{[t_1, t_2]^N}(\mathbf{u})$ are convex, it is sufficient to show that

$$F(\mathbf{u}) := \beta \langle \mathbf{1}, A\mathbf{u} \rangle - \beta \langle \mathbf{f}, \ln(A\mathbf{u}) \rangle + \langle \rho \mathbf{t}^{(k)}, D\mathbf{u} - \mathbf{g}^{(k)} \rangle + \frac{\rho}{2} \|D\mathbf{u} - \mathbf{g}^{(k)}\|_2^2, \quad (22)$$

is a strongly convex function with a positive constant c , as the desired inequality (21) follows from Lemma 1. For this purpose, let $\phi(A\mathbf{u}) := \langle \mathbf{1}, A\mathbf{u} \rangle - \langle \mathbf{f}, \ln(A\mathbf{u}) \rangle$ and $\mathbf{q} = A\mathbf{u}$. It follows from the Chain rule that

$$\frac{\partial^2 \phi}{\partial \mathbf{u}^2} = A^T \frac{\partial^2 \phi}{\partial \mathbf{q}^2} A = A^T \text{diag} \left(\frac{\mathbf{f}}{(A\mathbf{u})^2} \right) A, \quad (23)$$

where the division and the square operations are conducted elementwise and $\text{diag}(\cdot)$ takes a vector into a diagonal matrix. Due to the presence of the logarithm in (22), the functional F is defined on the domain of $\Omega := \{\mathbf{u} \mid A\mathbf{u} > \mathbf{0}\}$ and thus differentiable on Ω . Specifically, the Hessian matrix of F can be expressed by

$$H = \beta A^T \text{diag} \left(\frac{\mathbf{f}}{(A\mathbf{u})^2} \right) A + \rho D^T D. \quad (24)$$

Using the assumption (A1) and the upper bound of $\mathbf{u}$, we have $A\mathbf{u} \leq t_2 \mathbf{1}$, thus leading to a lower bound of the vector $\frac{\mathbf{f}}{(A\mathbf{u})^2}$ as $\eta := \frac{\min \mathbf{f}}{t_2^2} > 0$. Denote by μ the smallest eigenvalue of the matrix $\beta \eta A^T A + \rho D^T D$. We show that μ is strictly positive. Otherwise, $\mu = 0$, which implies that there exists a nonzero vector $\mathbf{v}$ such that $\mathbf{v}^T (\beta \eta A^T A + \rho D^T D) \mathbf{v} = 0$. It is straightforward that $\mathbf{v}^T A^T A \mathbf{v} \geq 0$ and $\mathbf{v}^T D^T D \mathbf{v} \geq 0$. Therefore, one shall have $\mathbf{v}^T A^T A \mathbf{v} = 0$ and $\mathbf{v}^T D^T D \mathbf{v} = 0$, which contradicts with the assumption (A2) that $\mathcal{N}(D) \cap \mathcal{N}(A) = \mathbf{0}$.

Applying Lemma 1 on F yields

$$F(\mathbf{u}^{(k)}) \geq F(\mathbf{u}^{(k+1)}) + \langle \nabla F(\mathbf{u}^{(k+1)}), \mathbf{u}^{(k)} - \mathbf{u}^{(k+1)} \rangle + \frac{\mu}{2} \|\mathbf{u}^{(k)} - \mathbf{u}^{(k+1)}\|_2^2. \quad (25)$$

Let $\mathbf{p} \in \partial \|D \cdot\|_1(\mathbf{u}^{(k+1)})$ and $\mathbf{r} \in \partial \delta_{[t_1, t_2]^N}(\mathbf{u}^{(k+1)})$ such that $\frac{\mathbf{p}}{\|\mathbf{g}^{(k)}\|_2} + \nabla F(\mathbf{u}^{(k+1)}) + \mathbf{r} = \mathbf{0}$ due to the optimality condition for $\mathbf{u}^{(k+1)}$. By the definition of subgradient, one has

$$\|D\mathbf{u}^{(k)}\|_1 \geq \|D\mathbf{u}^{(k+1)}\|_1 + \langle \mathbf{p}, \mathbf{u}^{(k)} - \mathbf{u}^{(k+1)} \rangle, \quad (26)$$

$$\delta_{[t_1, t_2]^N}(\mathbf{u}^{(k)}) \geq \delta_{[t_1, t_2]^N}(\mathbf{u}^{(k+1)}) + \langle \mathbf{r}, \mathbf{u}^{(k)} - \mathbf{u}^{(k+1)} \rangle. \quad (27)$$

Combining (25)–(27) along with the definition of $\mathcal{L}(\mathbf{u}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)})$ in (10), we get

$$\mathcal{L}(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) \geq \mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) + \frac{\mu}{2} \|\mathbf{u}^{(k)} - \mathbf{u}^{(k+1)}\|_2^2. \quad (28)$$

By taking $c = 2\mu > 0$, (28) is equivalent to (21).

As our model (7) only differs to the one in [45] by the data fidelity, we can borrow the error bounds of $\mathbf{t}, \mathbf{g}$, summarized as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k)}) - \mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) &\leq -\frac{\rho \epsilon^3 - 6M}{2\epsilon^3} \|\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\|_2^2, \\ \mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k+1)}) - \mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k)}) &\leq \left(\frac{16N}{\rho \epsilon^4} \right) \|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\|_2^2 \\ &\quad + \left(\frac{4M^2}{\rho \epsilon^6} \right) \|\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\|_2^2. \end{aligned} \quad (29)$$

Please refer to [45, Lemmas 4.2–4.3] for the proofs, specifically (A.10) and Lemma 4.2 in [45]. Combining the inequalities (28)–(29), we get the inequality (20) holds with $c_1 = c - \frac{16N}{\rho \epsilon^4}$

and $c_2 = \frac{\rho}{2} - \frac{3M}{\epsilon} - \frac{4M^2}{\rho\epsilon^6}$. If ρ is sufficiently large, we can have both c_1 and c_2 are strictly positive. $\square$

Lemma 3 (subgradient bound) *Let $\{(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}, \mathbf{t}^{(k)})\}$ be the sequence of iterates defined by (11), then there exist a vector $\boldsymbol{\eta}^{(k+1)} \in \partial\mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k+1)})$ and two positive constants c_3, c_4 such that*

$$\|\boldsymbol{\eta}^{(k+1)}\|_2 \leq c_3 \|\mathbf{t}^{(k+1)} - \mathbf{t}^{(k)}\|_2 + c_4 \|\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\|_2. \quad (30)$$

Proof We express the optimality conditions of (11) by,

$$\begin{cases} \frac{\mathbf{p}^{(k+1)}}{\|\mathbf{g}^{(k+1)}\|_2} + \beta A^T (\mathbf{1} - \frac{\mathbf{f}}{A\mathbf{u}^{(k+1)}}) + \mathbf{r}^{(k+1)} + \rho D^T (\mathbf{t}^{(k)} + D\mathbf{u}^{(k+1)} - \mathbf{g}^{(k)}) = \mathbf{0} \\ -\frac{\|D\mathbf{u}^{(k+1)}\|_1}{\|\mathbf{g}^{(k+1)}\|_2^3} \mathbf{g}^{(k+1)} + \rho (\mathbf{g}^{(k+1)} - D\mathbf{u}^{(k+1)} - \mathbf{t}^{(k)}) = \mathbf{0} \\ \mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + D\mathbf{u}^{(k+1)} - \mathbf{g}^{(k+1)}, \end{cases} \quad (31)$$

where $\mathbf{p}^{(k+1)} \in \partial\|D\|_1(\mathbf{u}^{(k+1)})$ and $\mathbf{r}^{(k+1)} \in \partial\delta_{[t_1, t_2]^N}(\mathbf{u}^{(k+1)})$. In addition, we define $\boldsymbol{\eta}_1^{(k+1)}$, $\boldsymbol{\eta}_2^{(k+1)}$, and $\boldsymbol{\eta}_3^{(k+1)}$ as follows,

$$\begin{cases} \boldsymbol{\eta}_1^{(k+1)} := \frac{\mathbf{p}^{(k+1)}}{\|\mathbf{g}^{(k+1)}\|_2} + \beta A^T (\mathbf{1} - \frac{\mathbf{f}}{A\mathbf{u}^{(k+1)}}) + \mathbf{r}^{(k+1)} \\ \quad + \rho D^T (\mathbf{t}^{(k+1)} + D\mathbf{u}^{(k+1)} - \mathbf{g}^{(k+1)}) \\ \boldsymbol{\eta}_2^{(k+1)} := -\frac{\|D\mathbf{u}^{(k+1)}\|_1}{\|\mathbf{g}^{(k+1)}\|_2^3} \mathbf{g}^{(k+1)} + \rho (\mathbf{g}^{(k+1)} - D\mathbf{u}^{(k+1)} - \mathbf{t}^{(k+1)}) \\ \boldsymbol{\eta}_3^{(k+1)} := D\mathbf{u}^{(k+1)} - \mathbf{g}^{(k+1)}. \end{cases} \quad (32)$$

Clearly, we have

$$\begin{aligned} \boldsymbol{\eta}_1^{(k+1)} &\in \partial_{\mathbf{u}}\mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k+1)}) \\ \boldsymbol{\eta}_2^{(k+1)} &\in \partial_{\mathbf{g}}\mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k+1)}) \\ \boldsymbol{\eta}_3^{(k+1)} &\in \partial_{\mathbf{t}}\mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k+1)}; \mathbf{t}^{(k+1)}), \end{aligned}$$

Combining both (31) and (32), we get

$$\begin{cases} \boldsymbol{\eta}_1^{(k+1)} = \frac{\mathbf{p}^{(k+1)}}{\|\mathbf{g}^{(k+1)}\|_2} - \frac{\mathbf{p}^{(k+1)}}{\|\mathbf{g}^{(k)}\|_2} + \rho D^T (\mathbf{t}^{(k+1)} - \mathbf{t}^{(k)}) + \rho D^T (\mathbf{g}^{(k)} - \mathbf{g}^{(k+1)}) \\ \boldsymbol{\eta}_2^{(k+1)} = \rho (\mathbf{t}^{(k)} - \mathbf{t}^{(k+1)}) \\ \boldsymbol{\eta}_3^{(k+1)} = \rho (\mathbf{t}^{(k+1)} - \mathbf{t}^{(k)}), \end{cases} \quad (33)$$

By triangle inequality and the lower bound of $\|\mathbf{g}^{(k)}\|_2$, we have

$$\left| \frac{1}{\|\mathbf{g}^{(k+1)}\|_2} - \frac{1}{\|\mathbf{g}^{(k)}\|_2} \right| = \frac{\|\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\|_2}{\|\mathbf{g}^{(k)}\|_2 \|\mathbf{g}^{(k+1)}\|_2} \leq \frac{\|\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\|_2}{\epsilon^2}.$$

As $\|\mathbf{p}^{(k+1)}\|_2 \leq 2\sqrt{2N}$ and $\|D^T\|_2 \leq 2\sqrt{2}$, we obtain the desired inequality (30) with $\boldsymbol{\eta}^{(k+1)} = (\boldsymbol{\eta}_1^{(k+1)}, \boldsymbol{\eta}_2^{(k+1)}, \boldsymbol{\eta}_3^{(k+1)})$, $c_3 = 2\rho + 2\sqrt{2}\rho$, and $c_4 = \frac{2\sqrt{2N}}{\epsilon^2} + 2\sqrt{2}\rho$. $\square$

Theorem 1 *Under the assumptions (A1)–(A3) and a sufficiently large $\rho > 0$, the sequence $\{(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)})\}$ defined by (11) has a sub-sequence that converges to a stationary point $(\mathbf{u}^*, \mathbf{g}^*; \mathbf{t}^*)$ of (8), i.e., satisfying $\mathbf{0} \in \partial E(\mathbf{u}^*, \mathbf{g}^*)$ and $\mathbf{g}^* = D\mathbf{u}^*$.*

Proof As $\mathbf{u}^{(k)}$ is bounded by $[t_1, t_2]$, then $\|D\mathbf{u}^{(k)}\|_1$ is bounded, denoted by M . Combining the optimality conditions for the $\mathbf{g}$ -subproblem and the $\mathbf{t}$ -subproblem in (31), we obtain

$$\|\mathbf{t}^{(k+1)}\|_2 = \left\| \frac{\|D\mathbf{u}^{(k+1)}\|_1}{\rho} \frac{\mathbf{g}^{(k+1)}}{\|\mathbf{g}^{(k+1)}\|_2^3} \right\|_2 \leq \frac{M}{\rho\epsilon^2}, \quad (34)$$

where we use the assumption (A3) that $\|\mathbf{g}^{(k)}\|_2 \geq \epsilon$. Therefore, $\{\mathbf{t}^{(k)}\}$ is bounded. As the $\mathbf{g}$ -update (12) depends on $D\mathbf{u}$ and $\mathbf{t}$, then $\|\mathbf{g}^{(k)}\|_2$ is also bounded. Thus, the sequence $\{(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}, \mathbf{t}^{(k)})\}$ is bounded, and it has an accumulation point by the Bolzano-Weierstrass theorem. Let $(\mathbf{u}^*, \mathbf{g}^*, \mathbf{t}^*)$ be any accumulation point of $\{(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}, \mathbf{t}^{(k)})\}$ and let $\{(\mathbf{u}^{(k_j)}, \mathbf{g}^{(k_j)}, \mathbf{t}^{(k_j)})\}$ be a subsequence with

$$\lim_{k_j \rightarrow \infty} (\mathbf{u}^{(k_j)}, \mathbf{g}^{(k_j)}, \mathbf{t}^{(k_j)}) = (\mathbf{u}^*, \mathbf{g}^*, \mathbf{t}^*).$$

Taking the telescopic summation of (20) leads to

$$\mathcal{L}(\mathbf{u}^{(K)}, \mathbf{g}^{(K)}; \mathbf{t}^{(K)}) \leq \mathcal{L}(\mathbf{u}^{(0)}, \mathbf{g}^{(0)}; \mathbf{t}^{(0)}) - c_1 \sum_{k=0}^{K-1} \|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\|_2^2 - c_2 \sum_{k=0}^{K-1} \|\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\|_2^2. \quad (35)$$

Note that

$$\begin{aligned} \langle \mathbf{1}, A\mathbf{u}^{(k)} \rangle - \langle \mathbf{f}, \ln(A\mathbf{u}^{(k)}) \rangle &= \langle \mathbf{1}, A\mathbf{u}^{(k)} - \mathbf{f} \ln(A\mathbf{u}^{(k)}) \rangle \\ &= \left\langle \mathbf{1}, A\mathbf{u}^{(k)} + \mathbf{f} \ln\left(\frac{\mathbf{f}}{A\mathbf{u}^{(k)}}\right) - \mathbf{f} \ln \mathbf{f} \right\rangle \\ &\geq \langle \mathbf{1}, \mathbf{f} - \mathbf{f} \ln \mathbf{f} \rangle. \end{aligned} \quad (36)$$

The last inequality comes from the fact $x \ln x \geq x - 1$, $\forall x > 0$ and $A\mathbf{u}^{(k)} > 0$. Combining (34) and (36), we can estimate a lower bound of $\mathcal{L}(\mathbf{u}^{(k+1)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)})$

$$\begin{aligned} \mathcal{L}(\mathbf{u}^{(k)}, \mathbf{g}^{(k)}; \mathbf{t}^{(k)}) &= \frac{\|D\mathbf{u}^{(k)}\|_1}{\|\mathbf{g}^{(k)}\|_2} + \beta \langle \mathbf{1}, A\mathbf{u}^{(k)} \rangle - \beta \langle \mathbf{f}, \ln(A\mathbf{u}^{(k)}) \rangle \\ &\quad + \frac{\rho}{2} \|D\mathbf{u}^{(k)} - \mathbf{g}^{(k)} + \mathbf{t}^{(k)}\|_2^2 - \frac{\rho}{2} \|\mathbf{t}^{(k)}\|_2^2 \\ &\geq \beta \langle \mathbf{1}, \mathbf{f} - \mathbf{f} \ln \mathbf{f} \rangle - \frac{M^2}{\rho \epsilon^4}. \end{aligned}$$

Therefore, it follows from (35) that $\|\mathbf{u}^{(k)} - \mathbf{u}^{(k+1)}\|_2 \rightarrow 0$, $\|\mathbf{g}^{(k)} - \mathbf{g}^{(k+1)}\|_2 \rightarrow 0$, and, using [45, Lemma 4.2], $\|\mathbf{t}^{(k)} - \mathbf{t}^{(k+1)}\|_2 \rightarrow 0$ as $k \rightarrow \infty$. The limit points are denoted as $(\mathbf{u}^*, \mathbf{g}^*, \mathbf{t}^*)$, and further convergence results are derived, including $(\mathbf{u}^{(k_j+1)}, \mathbf{g}^{(k_j+1)}, \mathbf{t}^{(k_j+1)}) \rightarrow (\mathbf{u}^*, \mathbf{g}^*, \mathbf{t}^*)$ and $\nabla \mathbf{u}^* = \mathbf{g}^*$. By invoking Lemma 3, it is concluded that $\mathbf{0} \in \partial \mathcal{L}(\mathbf{u}^*, \mathbf{g}^*, \mathbf{t}^*)$, establishing that $(\mathbf{u}^*, \mathbf{g}^*)$ is a stationary point of (8). $\square$

Remark: If one applies ADMM to decompose all the variables, it requires the introduction of two Lagrangian multipliers, leading to a so-called three-block ADMM [54] that lacks a convergence guarantee. Current research in the field of general optimization (8) often requires an accompanying function, such as an objective function, merit function, or augmented Lagrangian function, that possesses properties such as coerciveness, separability, or Lipschitz differentiability within a specified domain. As none of these properties holds for the L_1/L_2 formulation, we adopt the splitting scheme (8) that has a convergence guarantee. On the other hand, there is a gap in the convergence analysis, as the $\mathbf{u}$ subproblem cannot be solved precisely in practice. Instead, we utilize another iterative scheme to solve the subproblem inexactly. We empirically show that such an inexact solver leads to convergence (see Sect. 5.1), but rigorous proof will be left to future work.

5 Experimental Results

In this section, we present extensive experiments to demonstrate the performance of the proposed model (7) for restoring images corrupted by Poisson noise. We show three applications of our method: image deconvolution, super-resolution, and MRI reconstruction in Sects. 5.2–5.4, respectively. All numerical results are conducted under Windows 10 and MATLAB version 9.9 (R2020b) running on a Desktop with Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz and 24GB RAM. Our MATLAB source code will be available at GitHub¹ after publication.

To vary the Poisson noise level, we scale the pixel value to $[0, P]$ for the original image before applying the linear operator A , where the value P controls the noise level. The smaller P is, the more noisy the image looks, and hence, it becomes more challenging for image restoration. We then set the box constraint to $[0, P]$, i.e., $t_1 = 0$ and $t_2 = P$. As indicated in Algorithm 1, $\mathbf{u}$ is initialized by the zero vector. We set $\frac{\|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\|_2}{\|\mathbf{u}^{(k)}\|_2} < 10^{-5}$. We re-scale the final solution to get back in the same range.

We use two quantifiers, i.e., peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index, to evaluate the performance for comparison. The PSNR is defined as

$$\text{PSNR}(\mathbf{x}, \mathbf{w}) = 10 \ln_{10} \frac{mnP^2}{\|\mathbf{x} - \mathbf{w}\|_2^2}, \quad (37)$$

where P denotes the maximum intensity value (peak) of the original image $\mathbf{x}$, and $\mathbf{w}$ represents the restored image. The definition of SSIM relies on the local similarity index on two small patches x and w , that is,

$$\text{ssim}(x, w) = \frac{(2\mu_x\mu_w + c_1)(2\sigma_{xw} + c_2)}{(\mu_x^2 + \mu_w^2 + c_1)(\sigma_x^2 + \sigma_w^2 + c_2)}, \quad (38)$$

where μ_x and μ_w are their respective means, σ_x^2 and σ_w^2 represent their respective variances, σ_{xw} is the co-variance of x and w , and $c_1, c_2 > 0$ are constants which prevent having zero in the denominator. Then the overall SSIM is the mean of local similarity indexes, i.e.,

$$\text{SSIM}(\mathbf{x}, \mathbf{w}) = \frac{1}{K} \sum_{i=1}^{(k)} \text{ssim}(x_i, w_i), \quad (39)$$

where x_i and w_i are the corresponding patches indexed by i in the two images $\mathbf{x}$ and $\mathbf{w}$, respectively, and K is the number of patches. The PSNR serves as a reliable metric for gauging human subjective sensation, with a higher PSNR indicating superior quality in the restored image. Conversely, the SSIM aligns more closely with the quality perception of the human visual system. An SSIM value approaching 1 signifies that the restored image's characteristics, including edges and textures, closely resemble those of the original image.

5.1 Algorithmic Behavior

In this subsection, we present some computational aspects of our algorithm's settings. We start by the impact of the maximum number of iterations for the inner loop, denoted as InnerMax, on the super-resolution (SR) problem discussed in [55]. This problem involves reconstructing a high-resolution (HR) image from a low-resolution (LR) version. We test

¹ Once the paper is published, the codes will be available at <https://github.com/mujib2020>.

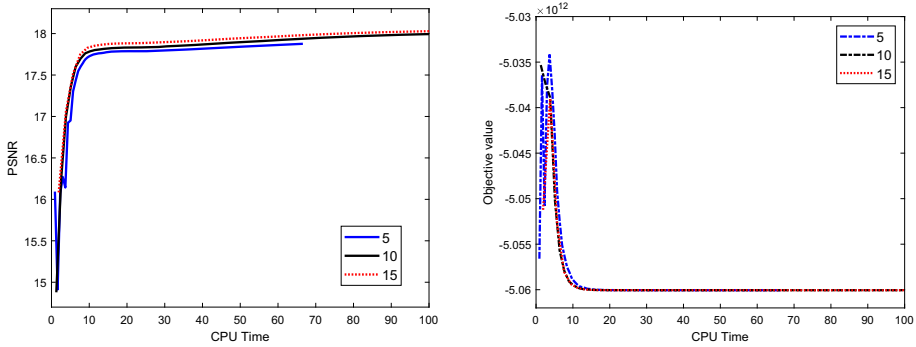


Fig. 1 The effects of the maximum number of the inner loops in terms of the PSNR (left) and objective value (right) with respect to CPU time

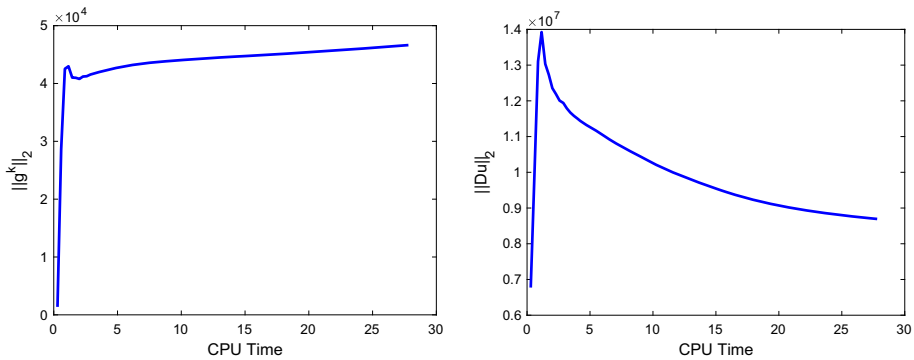


Fig. 2 Empirical verification on the boundedness of $\|\mathbf{g}^{(k)}\|_2$ and $\|\mathbf{Du}^{(k)}\|_2$ using the SR problem

InnerMax values of 5, 10, and 15, and present the corresponding PSNR and objective values against CPU time in Fig. 1. We observe that with a small InnerMax, the performance curves are erratic. However, setting InnerMax to 10 or more stabilizes these curves. Based on these findings, we choose InnerMax = 10 and OuterMax = 500 for the maximum iterations in the inner and outer loops, respectively, for subsequent experiments.

Next, we turn our focus to the assumptions underpinning the convergence analysis. Specifically, assumption (A3) posits a lower bound for the value of $\|\mathbf{g}^{(k)}\|_2$. In Fig. 2, we empirically verify this assumption by plotting the values of $\|\mathbf{g}^{(k)}\|_2$ in terms of CPU time. In addition, we also numerically show the boundedness of $\|\mathbf{Du}^{(k)}\|_2$. These findings empirically affirm that our algorithm meets all the specified assumptions.

5.2 Image Deblurring/Deconvolution

We use two standard testing images, labeled by Shape and Satellite as shown in Fig. 3. We define the blurring kernel as a 9×9 Gaussian Kernel with a standard deviation of $\sqrt{3}$, represented as `fspecial('gaussian', 9, sqrt(3))` in MATLAB. Before convolution, we set the peak value to P , $P/2$, $P/5$, $P/10$ with $P = 255$, as four specific levels of Poisson noise.

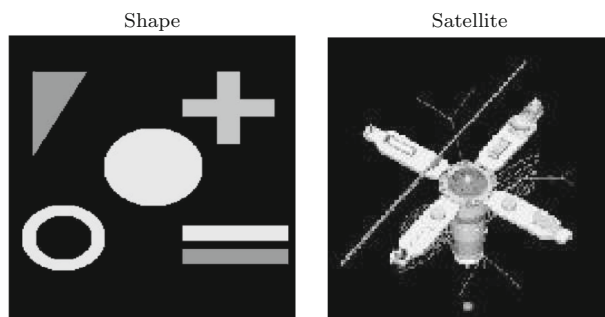


Fig. 3 Testing images: shape (left) and satellite (right)

Table 1 Comparison of image deconvolution results with each entry presenting PSNR and SSIM values

Test image	Peak ($P = 255$)	Input	NPTool	BM3D	FOTV	L_1/L_2
Shape	P	25.52/0.88	33.34/0.98	30.14/0.97	34.32/0.99	34.84/0.99
	$P/2$	24.61/0.85	29.97/0.94	29.53/0.94	33.50/0.99	34.08/0.99
	$P/5$	22.80/0.83	30.90/0.96	28.63/0.95	31.77/0.98	33.42/0.99
	$P/10$	20.88/0.81	29.38/0.96	28.02/0.94	30.07/0.96	32.69/0.99
Satellite	P	20.44/0.79	22.44/0.86	22.25/0.86	22.85/0.87	22.98/0.88
	$P/2$	20.20/0.78	22.00/0.85	22.00/0.85	22.51/0.86	22.62/0.87
	$P/5$	19.61/0.77	21.85/0.85	21.60/0.84	21.79/0.83	22.04/0.85
	$P/10$	18.78/0.74	21.52/0.83	21.22/0.83	21.59/0.82	21.59/0.83

The most favorable outcomes are emphasized in bold

We compare Algorithm 1 with some existing Poisson deblurring methods, including NPTool [56], BM3D [57], and FOTV [35]. We select the parameters that yield the highest PSNR value within their respective sets: $\beta \in \{10^{-1}, 10^{-2}\}$, $\rho \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}\}$, $\lambda \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$, $\gamma \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}\}$, and $\mu \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. The only tuning parameter for NPTool is selected among $\{4 \times 10^{-6}, 2 \times 10^{-5}, 7 \times 10^{-5}, 1 \times 10^{-4}, 2 \times 10^{-4}, 3 \times 10^{-4}, 5 \times 10^{-4}, 9 \times 10^{-4}\}$. There is no tuning parameter for BM3D. both NPTool and BM3D are initialized by the input data f .

In Table 1, the quantitative comparison in terms of PSNR and SSIM are listed for different peak levels. The proposed method achieves the best performance in all the cases of Shape and Satellite images. Visual comparison results are depicted in Figs. 4, 5 for Shape and Satellite, respectively. It is obvious in Fig. 4 that only the proposed method returns a piece-wise constant output. Additionally in Fig. 5, the proposed method preserves the most of the features in the wing of the satellite, while getting rid of Poisson noise.

5.3 Image Super Resolution

The setup of super-resolution (SR) considered in [55] is to reconstruct a high-resolution (HR) image from its low-resolution (LR) counterpart. In other words, SR recovers an image from low-frequency measurements, *i.e.*, the data is restricted within a square in the center of the frequency domain, corresponding to the low-frequency regime.

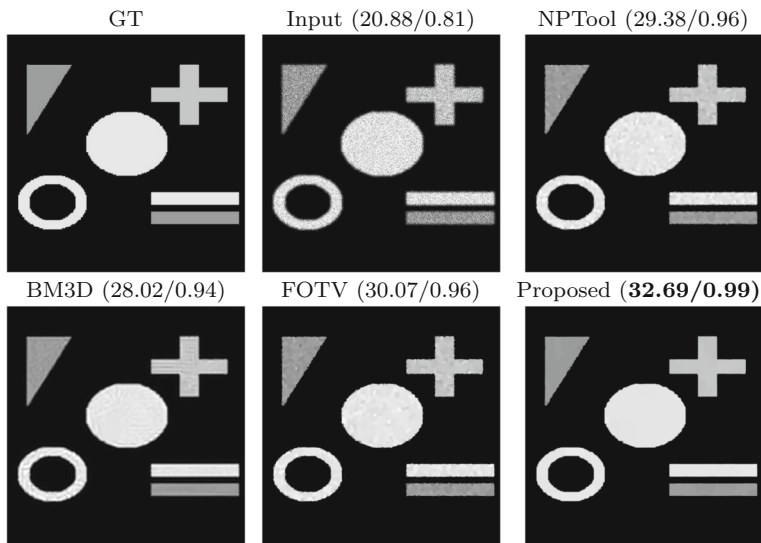


Fig. 4 Deconvolution results of shape with peak at $P/10 = 25.5$

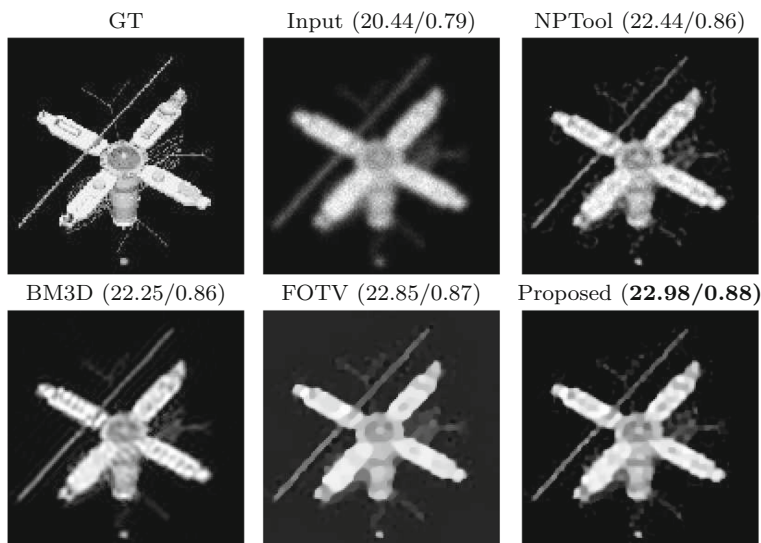


Fig. 5 Deconvolution results of satellite with peak at $P = 255$

To evaluate the performance of our method for SR, we use an original image from [58]; see Fig. 6. The size of the image is 688×688 . We restrict the data within a square in the center of the frequency domain. So, the sampling ratios depend on the size of the square. We consider two cases: 1% (box size 70×70) and 2% (box size 100×100) low-frequency measurements compared to the entire dimension. We set the peak values to 255 and 10^3 , each corresponding to a level of Poisson noise. We compare the performance of L_1/L_2 with L_1 norm on the gradient and minimax concave penalty (MCP) regularization [27], including a direct method called zero-filling (ZF) [4, 46], where the unknown frequency is filled by zero

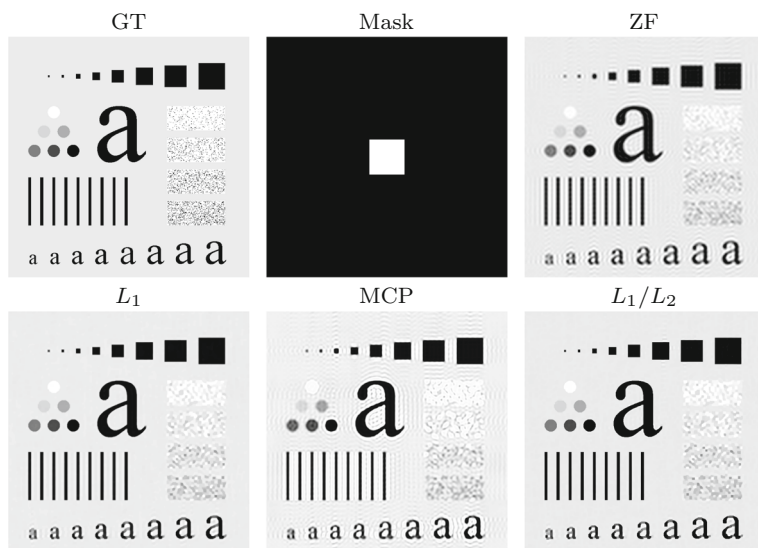


Fig. 6 Super resolution (2%) with peak $P = 10^3$

Table 2 Super resolution

Frequency	Peak	ZF	L_1	MCP	L_1/L_2
2%	255	17.52	18.96	18.24	19.24
	10^3	17.53	18.78	18.16	19.31
1%	255	16.19	17.48	17.56	17.91
	10^3	16.20	17.53	17.47	17.96

Bold values indicate the best results

Each entry contains PSNR

and followed by inverse Fourier Transform. The quantitative comparison in terms of PSNR and SSIM is presented in Table 2, showing significant improvements made by the proposed L_1/L_2 over the competing methods. The visual results of SR from 2% frequency data are pictured in Figs. 6 and 7 for the peak value to be 1000 and 255, respectively. Both MCP and ZF have severe ringing artifacts. The L_1 norm on the gradient is equivalent to the TV regularization, which has a known drawback of losing contrast.

5.4 MRI Reconstruction

In this subsection, we investigate the MRI reconstruction. The phantom of Shepp-Logan (SL) generated by MATLAB command `phantom(256)` is used as a testing image; see Fig. 8. We use two different masks to get the approximately 4% frequency measurements, called radial and Cartesian masks shown in Figs. 8 and 9, respectively. The “Cartesian mask” in MRI is named for its alignment with the Cartesian coordinate system, following a grid-like pattern in the frequency space (also known as k -space) for systematic data sampling. This alignment allows for easy and efficient sampling of data. The horizontal lines represent the frequency encoding lines in k -space. These lines in the Cartesian mask indicate that data is sampled along these lines. The density and distribution of these lines can affect the

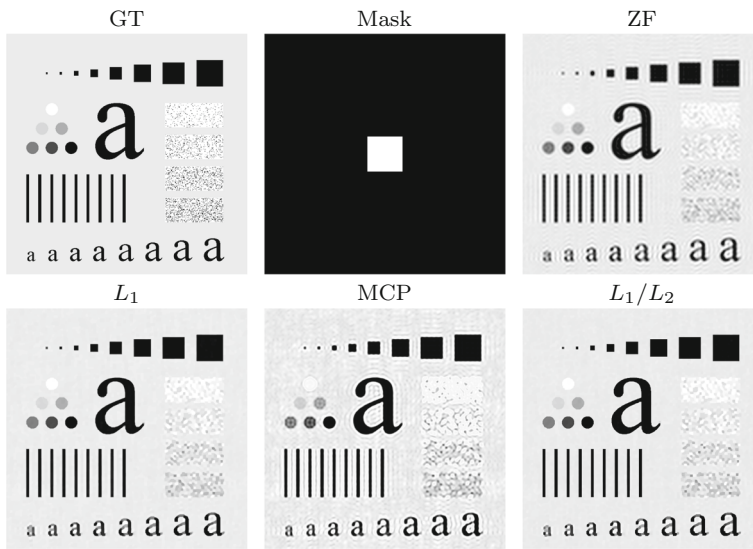


Fig. 7 Super resolution (2%) with peak $P = 255$

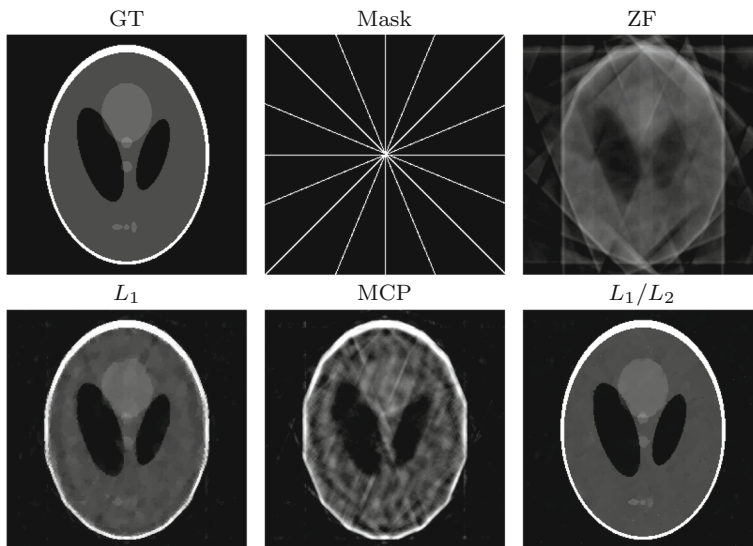


Fig. 8 MRI Reconstruction from approximately 4% frequency measurements along 8 radial lines

resolution and quality of the reconstructed MRI image. To speed up the imaging process, we sample fewer lines which is known as under-sampling. The central square in a Cartesian mask signifies a higher sampling density in the middle of k-space, capturing crucial signal information to maintain image contrast and quality, thereby mitigating aliasing artifacts from under-sampling elsewhere.

Similar to the super-resolution we include the results obtained by zero-filling. The numerical comparison in terms of PSNR and SSIM is given in Table 3, while the visual results

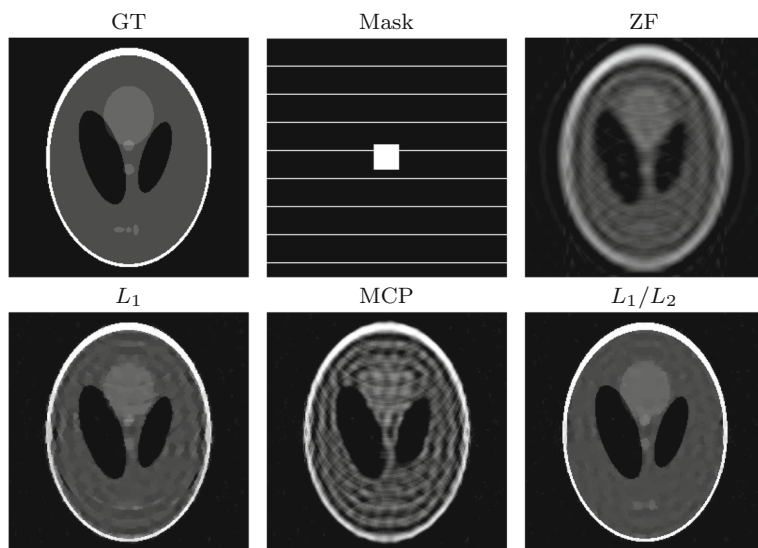


Fig. 9 MRI reconstruction from approximately 4% frequency measurement with Cartesian mask

Table 3 MRI reconstruction

Mask	ZF	L_1	MCP	L_1/L_2
Radial	16.20/0.53	25.28/0.93	20.43/0.75	41.47/0.99
Cartesian	17.98/0.66	26.09/0.92	20.34/0.73	32.07/0.97

Bold values indicate the best results

Each entry contains PSNR/SSIM

are presented in Figs. 8, 9. ZF fails to reconstruct the Shepp-Logan image due to the limited number of low-frequency measurements and Poisson noise. The L_1/L_2 regularization method gives a significant improvement over other regularization methods (L_1 and MCP), quantitatively and visually for producing sharper results.

6 Conclusion

In this paper, we proposed a new method to restore images that are subject to Poisson noise. Specifically, We formulated an unconstrained model that incorporates the L_1/L_2 regularization, the data fidelity that deals with the Poisson noise, and a box constraint. We applied a variable splitting algorithm called ADMM to find the model solution and established its subsequential convergence. We conducted extensive experiments on image deconvolution, super-resolution, and MRI reconstruction to demonstrate the superior performance of the proposed method over some existing methods. To maintain the efficiency of our algorithm, we implement an approximate method in the inner loop, although our current analysis of convergence is based on the assumption that this loop is solved precisely. In our future work, we aim to address this discrepancy by building on analysis from [59, 60].

Funding YL acknowledges the support from NSF CAREER DMS-2414705. CW acknowledges the support from the Natural Science Foundation of China (No. 12201286), the Shenzhen Science and Technology

Program 20231115165836001, HKRGC Grant No. CityU11301120, National Key R&D Program of China (2023YFA1011400), and the Shenzhen Fundamental Research Program JCYJ20220818100602005.

Data Availability The codes will be available at <https://github.com/mujib2020>.

Declarations

Competing interests The authors declare that they have no competing interests.

References

1. He, L., Gao, S., Deng, L.-J., Wang, Y., Wang, C.: Denoiser-guided image deconvolution with arbitrary boundaries and incomplete observations. *Signal Process.* **214**, 109226 (2024)
2. Hardie, R., Barnard, K., Armstrong, E.: Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.* **6**(12), 1621–1633 (1997)
3. Tipping, M., Bishop, C.: Bayesian image super-resolution. In: *Advances in Neural Information Processing System*, pp. 1303–1310 (2003)
4. Farsiu, S., Robinson, M., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution. *IEEE Trans. Image Process.* **13**(10), 1327–1344 (2004)
5. Chang, K., Ding, P.L.K., Li, B.: Single image super resolution using joint regularization. *IEEE Signal Process. Lett.* **25**(4), 596–600 (2018)
6. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**(6), 1182–1195 (2007)
7. Marques, E.C., Maciel, N., Naviner, L., Cai, H., Yang, J.: A review of sparse recovery algorithms. *IEEE Access* **7**, 1300–1322 (2018)
8. Luenberger, D.G., Ye, Y., et al.: *Linear and Nonlinear Programming*, vol. 2. Springer, Berlin (1984)
9. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer, Berlin (1999)
10. Wang, C., Chan, R., Nikolova, M., Plemmons, R., Prasad, S.: Nonconvex optimization for 3-dimensional point source localization using a rotating point spread function. *SIAM J. Imag. Sci.* **12**(1), 259–286 (2019)
11. Dai, L., Lu, M., Wang, C., Prasad, S., Chan, R.: Locnet: deep learning-based localization on a rotating point spread function with applications to telescope imaging. *Opt. Express* **31**(24), 39341–39355 (2023)
12. Nowak, R., Kolaczyk, E.D.: A Bayesian multiscale framework for poisson inverse problems. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, IEEE, pp. 1741–1744 (1999)
13. Sarder, P., Nehorai, A.: Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Process. Mag.* **23**(3), 32–45 (2006)
14. Figueiredo, M.A., Bioucas-Dias, J.M.: Restoration of Poissonian images using alternating direction optimization. *IEEE Trans. Image Process.* **19**(12), 3133–3145 (2010)
15. Zhang, X., Ng, M.K., Bai, M.: A fast algorithm for deconvolution and Poisson noise removal. *J. Sci. Comput.* **75**, 1535–1554 (2018)
16. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**(1–4), 259–268 (1992)
17. Chen, D., Chen, Y., Xue, D.: Fractional-order total variation image denoising based on proximity algorithm. *Appl. Math. Comput.* **257**, 537–545 (2015)
18. Zhang, J., Chen, K.: A total fractional-order variation model for image restoration with nonhomogeneous boundary conditions and its numerical solution. *SIAM J. Imag. Sci.* **8**(4), 2487–2518 (2015)
19. Wu, T., Shao, J., Gu, X., Ng, M.K., Zeng, T.: Two-stage image segmentation based on nonconvex l2-lp approximation and thresholding. *Appl. Math. Comput.* **403**, 126168 (2021)
20. Lou, Y., Zeng, T., Osher, S., Xin, J.: A weighted difference of anisotropic and isotropic total variation model for image processing. *SIAM J. Imag. Sci.* **8**(3), 1798–1823 (2015)
21. Bui, K., Park, F., Lou, Y., Xin, J.: A weighted difference of anisotropic and isotropic total variation for relaxed mumford-shah color and multiphase image segmentation. *SIAM J. Imag. Sci.* **14**(3), 1078–1113 (2021)
22. Huo, L., Chen, W., Ge, H., Ng, M.K.: L1-beta lq minimization for signal and image recovery. *SIAM J. Imag. Sci.* **16**(4), 1886–1928 (2023)
23. Lou, Y., Zhang, X., Osher, S., Bertozzi, A.: Image recovery via nonlocal operators. *J. Sci. Comput.* **42**(2), 185–197 (2010)

24. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010)
25. Zhang, X., Lu, Y., Chan, T.: A novel sparsity reconstruction method from poisson data for 3d bioluminescence tomography. *J. Sci. Comput.* **50**, 519–535 (2012)
26. Shen, Z., Chen, Q., Yang, F.: A convex relaxation framework consisting of a primal-dual alternative algorithm for solving l0 sparsity-induced optimization problems with application to signal recovery based image restoration. *J. Comput. Appl. Math.* **421**, 114878 (2023)
27. You, J., Jiao, Y., Lu, X., Zeng, T.: A nonconvex model with minimax concave penalty for image restoration. *J. Sci. Comput.* **78**(2), 1063–1086 (2019)
28. Li, F., Lv, X.-G.: A nonconvex nonsmooth image prior based on the hyperbolic tangent function. *J. Sci. Comput.* **97**(3), 55 (2023)
29. Xie, Y., Gu, S., Liu, Y., Zuo, W., Zhang, W., Zhang, L.: Weighted Schatten p -norm minimization for image denoising and background subtraction. *IEEE Trans. Image Process.* **25**(10), 4842–4857 (2016)
30. Lv, X.-G., Song, Y.-Z., Li, F.: An efficient nonconvex regularization for wavelet frame and total variation based image restoration. *J. Comput. Appl. Math.* **290**, 553–566 (2015)
31. Deng, K., Wen, Y.W., Li, K., Zhang, J.: Hybrid model of tensor sparse representation and total variation regularization for image denoising. *Signal Process.* 109352 (2023)
32. Yuan, W., Liu, H., Liang, L., Wang, W., Liu, D.: Image restoration via joint low-rank and external nonlocal self-similarity prior. *Signal Process.* **215**, 109284 (2024)
33. Fan, R., Jing, M., Shi, J., Li, L., Wang, Z.: TVRPCA+: low-rank and sparse decomposition based on spectral norm and structural sparsity-inducing norm. *Signal Process.* **217**, 109319 (2024)
34. Chowdhury, M.R., Zhang, J., Qin, J., Lou, Y.: Poisson image denoising based on fractional-order total variation. *Inverse Probl. Imaging* **14**(1)
35. Chowdhury, M.R., Qin, J., Lou, Y.: Non-blind and blind deconvolution under Poisson noise using fractional-order total variation. *J. Math. Imag. Vis.* **62**(9), 1238–1255 (2020)
36. Chowdhury, M.M.R.: Fractional-Order Total Variation Based Image Denoising, Deconvolution, and CT Reconstruction Under Poisson Statistics, The University of Texas at Dallas (2020)
37. Huang, J., Huang, T.-Z.: A nonstationary accelerating alternating direction method for frame-based poissonian image deblurring. *J. Comput. Appl. Math.* **352**, 181–193 (2019)
38. Kumar, P.G., Sahay, R., Ranjan: Low rank Poisson denoising (lrpd): a low rank approach using split bregman algorithm for Poisson noise removal from images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2019)
39. Evangelista, R.C., Salvadeo, D.H., Mascarenhas, N.D.: A new bayesian Poisson denoising algorithm based on nonlocal means and stochastic distances. *Pattern Recognit.* **122**, 108363 (2022)
40. Liang, H., Liu, R., Wang, Z., Ma, J., Tian, X.: Variational bayesian deep network for blind Poisson denoising. *Pattern Recognit.* **143**, 109810 (2023)
41. Khademi, W., Rao, S., Minnerath, C., Hagen, G., Ventura, J.: Self-supervised Poisson-Gaussian denoising. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2131–2139 (2021)
42. Zha, Z., Wen, B., Yuan, X., Zhou, J., Zhu, C.: Simultaneous nonlocal low-rank and deep priors for Poisson denoising. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), IEEE, pp. 2320–2324 (2022)
43. Rahimi, Y., Wang, C., Dong, H., Lou, Y.: A scale-invariant approach for sparse signal recovery. *SIAM J. Sci. Comput.* **41**(6), A3649–A3672 (2019)
44. Wang, C., Yan, M., Rahimi, Y., Lou, Y.: Accelerated schemes for the L_1/L_2 minimization. *IEEE Trans. Signal Process.* **68**, 2660–2669 (2020)
45. Wang, C., Tao, M., Nagy, J.G., Lou, Y.: Limited-angle ct reconstruction via the L_1/L_2 minimization. *SIAM J. Imag. Sci.* **14**(2), 749–777 (2021)
46. Wang, C., Tao, M., Chuah, C.-N., Nagy, J., Lou, Y.: Minimizing L_1 over L_2 norms on the gradient. *Inverse Prob.* **38**(6), 065011 (2022)
47. Wu, T., Mao, Z., Li, Z., Zeng, Y., Zeng, T.: Efficient color image segmentation via quaternion-based l1/l2 regularization. *J. Sci. Comput.* **93**(1), 9 (2022)
48. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
49. Osher, S.J., Esedoglu, S.: Decomposition of images by the anisotropic Rudin–Osher–Fatemi model. *Commun. Pure Appl. Math.* **57**, 1609–1626 (2003)
50. Le, T., Chartrand, R., Asaki, T.J.: A variational approach to reconstructing images corrupted by Poisson noise. *J. Math. Imag. Vis.* **27**(3), 257–263 (2007)
51. Teuber, T., Steidl, G., Chan, R.H.: Minimization and parameter estimation for seminorm regularization models with I-divergence constraints. *Inverse Prob.* **29**(3), 035007 (2013)

52. Chan, T.F., Shen, J.: Image processing and analysis: variational, PDE, wavelet, and stochastic methods. SIAM (2005)
53. Boyd, S.P., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
54. Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Math. Program.* **155**(1), 57–79 (2016)
55. Candès, E.J., Fernandez-Granda, C.: Super-resolution from noisy data. *J. Fourier Anal. Appl.* **19**(6), 1229–1254 (2013)
56. Landi, G., Piccolomini, E.L.: NPTool: a matlab software for nonnegative image restoration with newton projection methods. *Numer. Algor.* **62**(3), 487–504 (2013)
57. Azzari, L., Foi, A.: Variance stabilization in Poisson image deblurring, In: IEEE 14th International Symposium on Biomedicine and Imaging, IEEE, pp. 728–731 (2017)
58. Gonzalez, C.R., Woods, R., Eddins, S.: Digital Image Processing Using Matlab, Gatesmark (2009)
59. Chen, L., Li, X., Sun, D., Toh, K.-C.: On the equivalence of inexact proximal alm and admm for a class of convex composite programming. *Math. Program.* **185**(1), 111–161 (2021)
60. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. *Math. Program.* **137**(1), 91–129 (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.