A reweighted $\ell^2$ method for image restoration with Poisson and mixed Poisson-Gaussian noise

Jia Li*  Zuowei Shen*  Rujie Yin†  Xiaoqun Zhang‡

Abstract

We study weighted $\ell^2$ fidelity in variational models for Poisson noise related image restoration problems. Gaussian approximation to Poisson noise statistic is adopted to deduce weighted $\ell^2$ fidelity. Different from usual weighted $\ell^2$ approximation, we propose a reweighted $\ell^2$ fidelity with sparse regularization by framelets. Based on the split Bregman algorithm introduced in [22], the proposed numerical scheme is composed of three easy subproblems that involve quadratic minimization, soft shrinkage and matrix vector multiplications. Unlike usual least square approximation of Poisson noise, we dynamically update the underlying noise variance from previous estimate. The solution of the proposed algorithm is shown to be the same as the one obtained by minimizing Kullback-Leibler divergence fidelity with the same regularization. This reweighted $\ell^2$ formulation can be easily extended to mixed Poisson-Gaussian noise case. Finally, the efficiency and quality of the proposed algorithm compared to other Poisson noise removal methods are demonstrated through denoising and deblurring examples. Moreover, mixed Poisson-Gaussian noise tests are performed on both simulated and real digital images for further illustration of the performance of the proposed method.

Keywords: weighted $\ell^2$ fidelity, Gaussian approximation, split Bregman algorithm, framelets

1 Introduction

We consider an imaging system whose output data is a vector $f \in \mathbb{R}^M$ and the true underlying image is $u \in \mathbb{R}^N$. The observation model is often described by

$$f = Au$$

where $A$ denotes a linear operator from $\mathbb{R}^N$ to $\mathbb{R}^M$. In this paper, we focus on variational image restoration from observation $f$ contaminated by Poisson or mixed Poisson-Gaussian noise. Typically, variational models for image restoration are composed of two terms, one is a data fidelity term and the other is a regularization term for modeling a priori knowledge on unknown images. In general, the data fidelity term keeps the inline true image $u$ close enough to the input data $f$, so that the solution is meaningful. According to different noise statistics, the fidelity term takes different forms. It is well known that the least square fidelity is used for additive white Gaussian noise (AWGN). Such fidelity is mostly considered in literature for its good characterization of system noise. On the other hand, non-Gaussian types of noise are also encountered in real images. An important variant is Poisson noise, which is generally

---

*J. Li and Z. Shen are with the department of mathematics, National University of Singapore. Block S17, 10 Lower Kent Ridge Road, Singapore 119076; {lijia,matzuows}@nus.edu.sg
†R. Yin is with Zhiyuan College, Shanghai Jiao Tong University, 800, Dongchuan Road, Shanghai, China, 200240. ry27@duke.edu
‡X. Zhang is with the department of mathematics, MOE-LSC, and institute of natural sciences, Shanghai Jiao Tong University. 800, Dongchuan Road, Shanghai, China, 200240. xqzhang@sjtu.edu.cn
observed in photon-limited images, such as fluorescence microscopy, emission tomography and images have low exposure region like photos taken at night with high ISO setting. Due to its importance in medical imaging and the wide commercial application, a vast amount of literature in image restoration is devoted to problems encountering Poisson noise (see [2] and the references therein). The poisson noise data, i.e. the probability of receiving $y$ particles is given by

$$P(y) = \frac{e^{-\tau \tau^y y!}}{y!}, y = 0, 1, 2, \cdots$$

(2)

where $\tau$ is the expected value and the variance of random counts. Based on the statistics of Poisson noise, a generalized Kullback-Leibler (KL)-divergence [11] fidelity can be derived (see section 2.3.1 below) which is usually used in Poisson image restoration. However, difficulties in computational stability and efficiency arise due to the logarithm function that appears in the KL-divergence function. Moreover, in the case of image acquired with CCD camera, the mixture of Poisson and read-out noise (modeled as AWGN) is a more appropriate model, although it is seldom considered in literature. In [25], a Stein’s unbiased risk estimator (SURE) based on Poisson-Gaussian statistics is constructed in wavelet transform for mixed Poisson-Gaussian noise denoising, but the construction of estimators is so complicated that it cannot be easily extended for more general image restoration problem such as deblurring. Recently, Gong et al. proposed in [23] a universal $\ell^1 + \ell^2$ fidelity term for mixed or unknown noise and achieve encouraging numerical results, however the statistical analysis of it remains unclear.

Besides the fidelity term, we need a regularization term to control noise and artifacts in reconstructed images. A simple but effective idea is to urge a sparse representation in some transform domain for the recovered image. For a large class of applications [15, 8, 36], penalizing the $\ell^1$ norm of the transform coefficients of a signal leads to such a sparse solution, as largely illustrated in compressed sensing community in recent few years. The choice of transform is crucial to obtain a reasonable solution, and one popular choice is total variation proposed by Rudin-Osher-Fatemi [29]. Another choice are framelets, which have been shown useful and superior to total variation for different restoration tasks, see [13, 10] and the references therein. Recently, the relation of sparse framelets representation and total variation has been revealed in [10], and theoretically total variation can be interpreted as a special form of $\ell^1$ framelets regularization. For these reasons, we consider framelets regularization in this paper.

The main contribution of this paper is in introducing a least square fidelity term, a reweighted $\ell^2$ fidelity, that well approximates KL-divergence and developing an efficient iterative algorithm solving the optimization problem with framelets regularization. In fact, the idea of approximating KL-divergence by a fixed weighted least-square has been used for a long time in medical imaging reconstruction and image deblurring [2]. Recently, such a technique is restudied by Stagliano et al. [34], where the KL-divergence fidelity term is approximated by a weighted least square involving the unknown image. Combining with regularization, a scaled gradient projection (SGP) method with line search technique has been applied to solve the resulted problem. However, the efficiency of the algorithm involves the decomposition of a specific smoothed total variation regularization and line searching, which can not be easily generalized to other sparse regularization. We have a similar approximation, but viewed it as a Gaussian approximation to the Poisson distribution of the same variance and the variance is dynamically estimated using an iterative algorithm based on split Bregman iteration [22]. The original split Bregman iteration is ideal for solving $\ell^2$ fidelity based $\ell^1$ regularization, where the $\ell^1$ minimization is executed by a soft shrinkage, promising the efficiency of our modified version of it. In addition, the solution obtained by our algorithm has the same reweighted $\ell^2$ energy and framelets sparsity as that of the minimizer of the KL-divergence fidelity using the same regularization. Furthermore, the model immediately gives a reasonable extension to the mixed Poisson-Gaussian noise case with a small modification on the weight.
This paper is organized in three consecutive parts. In the first part, we review the background of variational restoration models, the split Bregman method and the existing Poisson fidelities. Then we present the reweighted $\ell^2$ fidelity for Poisson noise through Gaussian approximation and maximizing likelihood (ML) and then build the corresponding image restoration models and algorithms. We also provide an analysis of the proposed models and algorithms which reveals the connection to the classical KL-divergence model. The models are further extended to the mixed noise case, where both Poisson noise and Gaussian white noise present. The last part is dedicated to numerical simulation where other fidelity terms and models are compared numerically to ours for Poisson image denoising/deblurring and the mixed Poisson-Gaussian denoising/deblurring cases.

2 Background

2.1 Variational image restoration model

The observation model in consideration takes the following generic form

$$f = Au + c + \epsilon$$

where $c$ is a fixed background image vector in $\mathbb{R}^M$ and possibly 0, and $\epsilon$ is the noise perturbation. Generally, for additive white Gaussian noise setting, it is assumed that $c = 0$ and $\epsilon$ has independent normal distribution of mean 0 and variance $\sigma^2$ for each pixel. The operator $A$ may come from the imaging system such as the point spread function (PSF), the exterior disturbance such as motion or a combination of the two sources. In this paper, we mainly consider denoising and deblurring, where $A$ is either an identity or a convolution operator.

Image restoration by variational model usually takes the following form:

$$\min_u F(u) + \lambda G(u)$$

where $F(u)$ and $G(u)$ are the fidelity term and the regularization term respectively, and $\lambda$ is a positive scaling factor. The fidelity $F(u)$ is derived using classical maximum a posterior probability (MAP) $P(u|f)$ estimation, while the regularization $G(u)$ is designed based on a priori assumptions on $u$. As introduced previously, it is nowadays standard technique to penalize the $\ell^1$ norm of representation coefficients in a transform domain. Therefore, the following variational model is largely studied for image restoration from AWGN:

$$\min_u \frac{1}{2}\|Au - f\|_2^2 + \lambda\|Du\|_1$$

where $\| \cdot \|_1$ denotes the usual $\ell^1$ vector norm and $D$ is a linear transform, such as discrete gradient used in total variation [29], Fourier transform, local cosine transforms, wavelet or framelets [9].

2.2 Split Bregman algorithm

Split Bregman method introduced in [22] is designed to solve the variational model taking the form as (5). More generally, we consider the minimization problem

$$\min_u F(u) + \lambda\|Du\|_1$$

where $F(u)$ is a convex functional representing fidelity, such as $F(u) = \frac{1}{2}\|Au - f\|_2^2$ as in the case of AWGN. By introducing an auxiliary variable $d = Du$, the following alternating split
Bregman scheme solves (6) for $\mu > 0$,

$$
\begin{align*}
&u_{k+1} = \arg\min_u F(u) + \frac{\mu}{2} \|Du - d_k + b_k\|^2_2 \\
d_{k+1} = \arg\min_d \lambda \|d\|_1 + \frac{\mu}{2} \|Du_{k+1} - d + b_k\|^2_2 \\
b_{k+1} = b_k + Du_{k+1} - d_{k+1}
\end{align*}
$$

(7)

The efficiency of this algorithm relies on the closed form solution of the second subproblem. In fact, $d_{k+1}$ is given by the so-called soft-shrinkage operator

$$
d_{k+1} = \text{sign}(Du_{k+1} + b_k) \cdot \max(|Du_{k+1} + b_k| - \lambda/\mu, 0)
$$

(8)

where each operation is componentwisely performed. The nonsmooth optimization problem (6) is thus decomposed into three subproblems; the first subproblem is simple to solve when $F(u)$ is in quadratic form. This method has been demonstrated to be very efficient for $\ell^1$ type of minimization in a large variant of applications. In a more general setting, it is shown to be equivalent to classical Douglas-Racheford and alternating direction multiplier method (ADMM) [16, 30], and the convergence analysis is given in this framework. In [9], Cai et al. studied the application of such an algorithm for framelets based image restoration and provide a detailed convergence proof. Further approximation of the quadratic subproblem was considered in [37] in order to maximally decouple the subproblems. In this paper, we intend to take advantage of the efficiency of split Bregman and Poisson noise formulation to develop an efficient algorithm for Poisson noise related image restoration problem. Therefore, a reweighted $\ell^2$ fidelity term which is closely related to the problem (6) is considered.

### 2.3 Poisson noise related data fidelities

We assume that the observation $f$ is corrupted by Poisson noise (see (2)), i.e.

$$f \sim P(Au + c)
$$

(9)

Since both $f$ and $u$ denote photon counts, their elements are nonnegative; the following assumptions are made for the linear operator $A$ and $c$ as in [32, 3, 7]:

**Assumption 1**

- The observation data $f > 0$, and the underlying true $u \geq 0$.
  - The linear operator $A$ satisfies the following conditions:
    $$A_{ij} \geq 0, \quad \sum_i A_{ij} > 0, \forall j, \quad \sum_j A_{ij} > 0, \forall i.$$
    where $A_{ij}$ is the $(i, j)$ element of the imaging matrix $A$.
  - The fixed background image $c > 0$.

For most image restoration models, $A$ is the convolution or line integral operator, and these conditions can be easily fulfilled without loss of generality. The first assumption is used to avoid model deficiency as in KL-divergence and reweighted $\ell^2$. The second assumption implies that for $u \geq 0$, $Au = 0$ if and only if $u = 0$.

Given $Au$ and $c$, we have the likelihood of observing $f$

$$P(f|Au + c) = \prod_{i=1}^{M} \frac{(Au + c)_i^{f_i} e^{-(Au+c)_i}}{f_i!}
$$

(10)
where \((Au + c)_i\) denotes the \(i^{th}\) element of \(Au + c\). By the property of the Poisson distribution, we have the expectation (mean) and the variance \(f\) are

\[
E(f|Au + c) = \text{Var}(f|Au + c) = Au + c
\]  

Before presenting the proposed reweighted \(\ell^2\) fidelity, we first review two existing fidelities for Poisson statistics for comparison.

### 2.3.1 KL-divergence

The most popular fidelity for Poisson noise is the generalized Kullback-Leibler (KL)-divergence fidelity [11], which can be derived directly by MAP method. Maximizing a posterior probability is equivalent to minimizing the negative log likelihood of (10), i.e.

\[
\min_{u \geq 0} -\log P(f|Au + c) = \min_{u \geq 0} 1^T (Au + c - \log(\text{factorial})f) - f^T \log(Au + c),
\]  

where \(1\) is the vector with all 1 entries.

If we neglect the constant term \(\log(\text{factorial})\) which is unrelated to the unknown \(u\), we obtain the following fidelity term

\[
F(u) = 1^T (Au + c) - f^T \log(Au + c)
\]  

Combining with sparse regularization and nonnegativity constraint on photon counts \(u\), we get the restoration model as follows,

\[
\min_{u \geq 0} 1^T (Au + c) - f^T \log(Au + c) + \lambda \|Du\|_1
\]  

Generally, (14) is a difficult optimization problem because of the nonsmooth regularization term and the KL-divergence term. Optimization of KL-divergence fidelity with nonnegative constraint are typically solved by Expectation-Maximization (EM) algorithm [24]. It is known that the convergence of the EM algorithm is slow and it may introduce so-called “checkboard effect” [3, 34, 7]. In [7], the authors proposed a two-step iteration method called EM-TV for solving (14) when \(D\) is a discrete differential operator. We change the TV regularization to framelets and rename it as EM+\(\ell^1\) algorithm for later discussion and numerical comparison with our scheme. The algorithms is described as follows,

\[
\begin{aligned}
  u_{k+\frac{1}{2}} &= u_k A^+(\frac{f}{Au_k + c}) \\
  u_{k+1} &= \arg\min_{u \geq 0} \frac{1}{2} \left\| \frac{u - u_{k+\frac{1}{2}}}{\sqrt{u_k}} \right\|_2^2 + \lambda \|Du\|_1 
\end{aligned}
\]  

This algorithm and its variants have been proved to be efficient for Poisson noise removal in PET and nasoscopy image deconvolution [7]. Under adequate conditions, the convergence of the algorithm with a damped parameter is provided in [6]. However, the conditions are also hard to verify in practice and we also need a subproblem solver for the second step when \(D\) is not an orthogonal operator.

The other kind of method considered in [31, 18] is to apply directly the split Bregman method (7) by introducing a variable \(d = Du\) on the model (14). Here, for the reason of comparison, we present the algorithm proposed in [31] for solving (14) by inducing three extra
variables to represent $d^{(1)} = Au$, $d^{(2)} = D u$ and $d^{(3)} = u$ to handle the nonnegativity constraint. The overall scheme can be written as

$$
\begin{align}
  u_{k+1} &= \text{argmin}_u ||Au - d^{(1)}_k + b^{(1)}_k||_2^2 + ||Du - d^{(2)}_k + b^{(2)}_k||_2^2 + ||u - d^{(3)}_k + b^{(3)}_k||_2^2 \\
  d^{(1)}_{k+1} &= \text{argmin}_{d^{(1)}} 1^T (d^{(1)} + c) - f^T \log(d^{(1)} + c) + \frac{1}{2\gamma} ||Au_{k+1} - d^{(1)} + b^{(1)}||_2^2 \\
  d^{(2)}_{k+1} &= \text{argmin}_{d^{(2)}} \lambda ||d^{(2)}||_1 + \frac{1}{2\gamma} ||Du_{k+1} - d^{(2)} + b^{(2)}||_2^2 \\
  d^{(3)}_{k+1} &= \text{argmin}_{d^{(3)} \geq 0} \frac{1}{2\gamma} ||u_{k+1} - d^{(3)} + b^{(3)}||_2^2 \\
  b^{(1)}_{k+1} &= b^{(1)}_k + Au_{k+1} - d^{(1)}_k \\
  b^{(2)}_{k+1} &= b^{(2)}_k + Du_{k+1} - d^{(2)}_k \\
  b^{(3)}_{k+1} &= b^{(3)}_k + u_{k+1} - d^{(3)}_{k+1}
\end{align}
$$

(16)

Several auxiliary variables have been introduced and we will draw into comparison with our proposed algorithm in the numerical implementation.

2.3.2 Anscombe transform

Another well-known technique of Poisson denoising is the Anscombe transform [1] used in image denoising, when the linear operator $A$ in (9) is the identity. The Anscombe transform is defined as the following non-linear transform

$$
A : x \mapsto 2\sqrt{x + \frac{3}{8}}
$$

(17)

If $x$ is a random variable that obeys the Poisson distribution with mean and variance $\tau$, the transformed random variable $Ax$ follows an approximated standard Gaussian distribution $\mathcal{N}(\tau, 1)$. Apply the Anscombe transform on $f \sim P(u + c)$, and let $\tilde{f} := Af$, $\tilde{u} := A(u + c)$, then $\tilde{f}$ follows approximately a normal distribution $\mathcal{N}(\tilde{u}, 1)$. Hence, the usual least square fidelity term can be applied for $\tilde{f}$ and we obtain the following denoising model

$$
\tilde{u}^* = \text{argmin}_{\tilde{u} \geq 0} \frac{1}{2} ||\tilde{u} - \tilde{f}||_2^2 + \lambda ||D\tilde{u}||_1
$$

(18)

and the final image is given by $u^* = A^{-1}\tilde{u}^* - c$. The regularization used in the above model (18) forces $\tilde{u}$ to be sparse in the transformed domain. This is roughly equivalent to requiring sparsity on $u$ in the transform domain since the Anscombe transform (17) is monotone increasing and keeps the order of magnitude of elements in $u$. We note that this model is generally applied for denoising model, and the adaption to image deblurring is not easy since it involves inverse of $A$, which is non-linear.

3 Weighted least square for Poisson noise image restoration

3.1 Framelets regularization

A countable set $X \subset L_2(\mathbb{R})$ is called a tight frame of $L_2(\mathbb{R})$ if

$$
f = \sum_{h \in X} \langle f, h \rangle h \quad \forall f \in L_2(\mathbb{R}),
$$

(19)

where $\langle \cdot, \cdot \rangle$ is the inner product of $L_2(\mathbb{R})$. Given a finite collection of functions $\Psi = \{\psi_1, \psi_2, ..., \psi_m\}$, define $X = \{\psi_{n,k,l} = 2^n/2\psi_l(2^n \cdot k), 1 \leq l \leq m, \ n, k \in \mathbb{Z}\}$. If $X$ satisfies the condition of tight
frame, then \( X \) is called a wavelet tight frame and \( \Psi \) is called the wavelet. We choose tight frame basis as a representation basis of images due to its multi-resolution property that allows fast algorithm implementation; its redundancy encourages sparse representation of images, as studied in [12],[13]. Generally, multi-resolution analysis (MRA) based wavelets can be generated by the unitary extension principle (UEP) in [28]. The wavelet tight frame generated through UEP is based on a refinable function \( \phi = \psi_0 \) and the wavelet functions in \( \Psi \) satisfies:

\[
\hat{\psi}_l(2^l \cdot) = \hat{h}_l \hat{\phi}, l \in \{0, 1, 2, \ldots, m\}
\]

and

\[
\sum_{l=0}^{m} |\hat{h}_l(\xi)|^2 = 1
\]

and

\[
\sum_{l=0}^{m} \hat{h}_l(\xi)\hat{h}_l(\xi + \pi) = 0
\]

where the sequence \( h_l \) is called the refinement mask of the wavelet tight frame system. By collecting all the refinement masks of wavelet tight frame system, we can generate the fast tight frame transform or decomposition operator \( W \). The matrix \( W \) is consisted of \( J+1 \) sub-filtering operators \( W_0, W_1, \ldots W_J \). Among them, \( W_0 \) is the low-pass filtering operator and the rest are high-pass filtering operators. Correspondingly, by unitary extension principle [28], the operator \( W^T \) is the fast tight frame reconstruction operator and we have \( W^T W = I \), i.e., \( W^T W u = u \) for any image \( u \). More details on discrete algorithms of framelet transforms can be found in [13].

3.2 Model and algorithm

We consider Poisson noise for the observation model (9). Let

\[
\epsilon = f - Au - c,
\]

which can be interpreted as an additive perturbation noise. Given \( Au \) and \( c \), the conditional mean of \( \epsilon \) is

\[
E(\epsilon|Au + c) = E(f|Au + c) - Au - c = 0
\]

and the conditional variance is

\[
\text{Var}(\epsilon|Au + c) = \text{Var}(f|Au + c) = Au + c.
\]

We approximate \( \epsilon \) by additive Gaussian noise, following the normal distribution \( \mathcal{N}(0, Au + c\Sigma) \), i.e.

\[
P(\epsilon|Au + c) \simeq \exp\left\{ -\frac{1}{2} (f - Au - c)^T \Sigma^{-1} (f - Au - c) \right\}
\]

s.t. \( \text{diag}(\Sigma) = Au + c \). By further imposing the independence of noise at each pixel, \( \Sigma \) is a diagonal matrix. We take the negative log of the normal distribution (21),

\[
- \log P(\epsilon|Au + c) \propto \frac{1}{2} (f - Au - c)^T \Sigma^{-1} (f - Au - c),
\]

and set (22) as the fidelity term, which is equivalent to maximum likelihood. Let \( \|x\|_Q^2 = x^T Q x \) be the weighted \( \ell^2 \) norm of a vector \( x \in \mathbb{R}^N \) with respect to a symmetric positive definite matrix \( Q \). Then (22) can be reformulated as

\[
F(u) = \| Au + c - f \|_{\Sigma^{-1}}^2 = \|Au + c - f\|_{\text{diag}(Au+c)-1}^2 = \left\| \frac{Au + c - f}{\sqrt{Au + c}} \right\|_2^2
\]
Notice that both the division and the square root operator in the \( \ell^2 \) norm on the right hand side of (23) are element wise. The positive definitiveness of \( \Sigma^{-1} \) is guaranteed by the assumptions that \( c > 0 \) and \( Au \geq 0 \) for \( u \geq 0 \).

Combining the fidelity \( F(u) \) with sparse framelets regularization and nonnegativity constraint, we have the following restoration model
\[
\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au+c-f}{\sqrt{Au+c}} \right\|^2_2 + \lambda \|Wu\|_1
\]  
(24)

This formulation is also considered in [34] and a scale gradient projection method is applied to (24), directly solving the nonlinear objective function, where a smoothed version of the \( \ell^1 \) regularization term was adopted for calculating the gradient.

To solve (24), we are interested in taking advantage of the weighted least square structure and utilizing recently proposed efficient sparse regularization scheme, such as split Bregman presented in Section 2.2. However, the fidelity term \( F(u) \) is not quadratic, because \( u \) appears both in the denominator and in the numerator of (23). Therefore, in order to solve the first sub optimization problem in split Bregman algorithm (7), we need to either approximate or directly solve the nonlinear square term.

A reasonable approximation of the unknown weight \( Au+c \), for example, can be the observed data \( f \), so that (23) is approximated by
\[
F(u) = \frac{1}{2} \left\| \frac{Au+c-f}{\sqrt{f}} \right\|^2_2
\]  
(25)

Using this simplification, efficient least square based methods can be applied to solve the subproblem.

However, such an approximation is not accurate, especially when the observed data \( f \) is severely corrupted by noise. Hence we need a better approximation, which may be derived adaptively in an iterative algorithm. For any iterative algorithm, whose sequence of solution \( u_k \) converges to \( u^* \), when \( k \) is big enough, as \( u_k \) becomes stable, \( u_k \) serves as a more precise approximation of \( u^* \) than \( f \). Hence a better approximation to the formulation (24) is given as
\[
\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au+c-f}{\sqrt{Au^*+c}} \right\|^2_2 + \lambda \|Wu\|_1,
\]  
(26)

with sufficiently large number of iterations \( k \). We may approximate the weight \( \frac{1}{Au^*+c} \) by \( \frac{1}{Au_k+c} \) in the algorithm. With such an idea in mind, we combine with the popular Split Bregman iteration and derives a new algorithm called the reweighted \( \ell^2 \) algorithm with split Bregman for Poisson noise:

\[
\begin{align*}
    u_{k+1} &= \arg\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au+c-f}{\sqrt{Au_k+c}} \right\|^2_2 + \frac{\mu}{2} \|Wu - d_k + b_k\|^2_2 \\
    d_{k+1} &= \arg\min_{d} \lambda \|d\|_1 + \frac{\mu}{2} \|d - Wu_{k+1} - b_k\|^2_2 \\
    b_{k+1} &= b_k + (Wu_{k+1} - d_{k+1})
\end{align*}
\]  
(27)

We describe the general method (27) in more detail in Algorithm 1. Note that the first step can be solved by a gradient method with a projection onto the nonnegative orthant. In practice, often a few iterations are sufficient to get a reasonable result.

In the following, we justify the proposed algorithm by some intuitive analysis. Let \((u_k, d_k, b_k)\) be the sequence generated by the algorithm (27). If \((u_k, d_k, b_k)\) converges to \((\bar{u}, \bar{d}, \bar{b})\), then \( \bar{u} \) is
Algorithm 1: Reweighted $\ell^2$ with Split Bregman for Poisson noise restoration Algorithm

Step 0. Set the initial value, $u_0 = f; d_0 = Wf; b_0 = 0; k = 1; \Sigma_0 = Au_0 + c$.

Step 1. while $\|u_{k-2} - u_{k-1}\|_2 > \delta$ or $k = 1$ do

\[ u_k = \arg\min_{u \geq 0} \frac{1}{2} \|Au + c - f\|^2_{\Sigma_{k-1}^{-1}} + \frac{\mu}{2} \|Wu - d_{k-1} + b_{k-1}\|_2^2 \]

\[ d_k = \text{sign}(Wu_k + b_{k-1}) \cdot \max(|Wu_k + b_{k-1}| - \lambda/\mu, 0) \]

\[ b_k = b_{k-1} + Wu_k - d_k \]

\[ \Sigma_k = \text{diag}(Au_k + c) \]

\[ k = k + 1 \]
end while

A minimizer of (14), shown as follows. If $(u_k, d_k, b_k)$ generated by (27) converges to $(\tilde{u}, \tilde{d}, \tilde{b})$, we immediately get

\[ AT \tilde{\Sigma}^{-1}(A\tilde{u} + c - f) + \mu W^T \tilde{b} - \tilde{q} = 0 \]

\[ \lambda \tilde{p} - \mu \tilde{b} = 0 \text{ with } \tilde{p} \in \partial\|d\|_1 \]

\[ W\tilde{u} = \tilde{d} \]

where $\tilde{q} \geq 0$ satisfies $\langle \tilde{q}, \tilde{u} \rangle = 0$, and $\tilde{\Sigma} = \text{diag}(A\tilde{u} + c)$. Therefore, substituting $\tilde{b}$ by $\lambda/\mu \tilde{p}$, we have

\[ AT \tilde{\Sigma}^{-1}(A\tilde{u} + c - f) + \lambda W^T \tilde{p} - \tilde{q} = 0 \]

On the other hand, the first order condition of the model (14) gives

\[ 0 = AT \left( \frac{A\tilde{u} + c - f}{A\tilde{u} + c} \right) + \lambda W^T \tilde{p} - \tilde{q} \]

(29)

where $\tilde{p} \in \partial|\tilde{d}|_1$ with $\tilde{d} = W\tilde{u}$ and $\tilde{q} \geq 0$ is a lagrangian multiplier such that $\langle \tilde{q}, \tilde{u} \rangle = 0$. Since $\tilde{u}$ satisfies exactly this first order condition, it is a minimizer of (14).

A more rigorous analysis for the convergence of the algorithm, with adequate assumptions on the sequences, is presented in the Appendix. Although these conditions are not easy to check in applications where ground truth is unavailable, this analysis may bring some insight for a complete convergence analysis in future.

3.3 Extension to Poisson-Gaussian mixed noise

Previously, we mainly discuss models with Poisson noise only. In real imaging systems, besides Poisson noise which characterizes the fluctuation in counting number of photons, there is other system-inherited noise that can be approximated by AWGN, as considered in [33]. The literature on Poisson-Gaussian mixed noise are limited despite of its importance. We now explore the extended application of the reweighted $\ell^2$ fidelity to the mixed Gaussian-Poisson noise case.

With a small modification of (23), the reweighted $\ell^2$ fidelity can be adapted to the mixed noise case. Let $\sigma^2$ be the variance of AWGN and the observed image $f$ has distribution

\[ f \sim \mathcal{P}(Au + c) + \mathcal{N}(0, \sigma^2) \]

Approximating the probability density function of $f$ by normal distribution (21) with covariance matrix $\Sigma = \text{diag}(Au + c) + \sigma^2 I$, we obtain the new fidelity term as follows for mixed noise:

\[ F(u) = \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au + c + \sigma^2 I}} \right\|_2^2 \]

(30)
Combining with the tight frame regularization, we have the following restoration model

\[
\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au + c + \sigma^2 I}} \right\|^2_2 + \lambda \| \mathcal{W}u \|_1
\]  

(31)

The algorithm solving this model is the same as Algorithm 1 except adding \( \sigma^2 \) in the estimation and updating step of the covariance matrix \( \Sigma \).

4 Numerical results

In this section, numerical experiments on pure Poisson noise model (24) and Poisson-Gaussian mixed noise models (31) are shown for synthesized and real digital photos for demonstrating the performance of the proposed Algorithm 1. Numerical results of denoising/deblurring algorithms discussed in section 2 are also presented for comparison. In all implementations where framelets sparse regularization is used, we choose the piecewise linear B-spline wavelets with 1 level decomposition as the framelets basis. The corresponding filter masks in discrete version are \( h_0 = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right] \), \( h_1 = \left[ \frac{1}{4}, -\frac{1}{4}, -\frac{1}{4} \right] \), \( h_2 = \left[ \sqrt{2}/4, 0, -\sqrt{2}/4 \right] \). The parameter \( \lambda \) is always set to be 0.01 for pure image denoising and 0.001 for image deblurring in presence of noise. The parameter \( \mu \) is simply set to be 1 for all numerical simulations.

In the numerical implementation, we find out that preconditioning on \( \Sigma_k \) is necessary to ensure numerical stability. In particular, we enforce a lower bound \( C_{\sigma} > 0 \) on the diagonal entries of \( \Sigma_k \), i.e. \( \bar{\Sigma}_k = \max\{\Sigma_k, C_{\sigma}\} \) to control its conditional number. To improve the visual quality of the final result, in the postprocessing, we apply a bilateral filter [35] \( B \) is defined as

\[
y[i, j] = B(x) = \frac{1}{\sum_{p,q} w[i, j, p, q]} \sum_{p,q} w[i, j, p, q] x[p, q]
\]

where \([i, j], [p, q]\) are indices of pixels of an image \( x \),

\[
w[i, j, p, q] = G_{\sigma_s}(\sqrt{(i - p)^2 + (j - q)^2}) G_{\sigma_r}(x[i, j] - x[p, q])
\]

and \( G_{\sigma_s}, G_{\sigma_r} \) are Gaussian functions with variance \( \sigma_s \) and \( \sigma_r \) respectively.

Each denoised result \( \tilde{u} \) are evaluated quantitatively by the peak signal-to-noise ratios (PSNR) value defined by

\[
\text{PSNR}(u, \tilde{u}) = 10 \log_{10} \frac{MN(u_{\text{max}} - u_{\text{min}})^2}{\| u - \tilde{u} \|^2_2},
\]

using the ground truth image \( u \), where \( u_{\text{max}} \) and \( u_{\text{min}} \) are its maximal and minimal pixel values respectively and \( M, N \) are the size of the image.

4.1 Poisson denoising

In the following, we compare the visual image quality with the three fidelities through their corresponding denoising methods: the reweighted \( \ell^2 \) method (27) with Algorithm 1, the KL-divergence model (14) with the algorithm (15) as well as algorithm (16) and the Anscombe transform model (18). The noisy images in our test are simulated as follows. The clean images are first rescaled to an intensity range from 0 to 120; then the Poisson noise is added in Matlab using the function \texttt{poissrnd}. The parameters in the first three algorithms are set to the same instead of optimized respectively (especially \( \lambda = 0.05 \)), hence the energy function they minimize are consistent to each other. For the Anscombe transform model, the parameters are tuned (\( \lambda = 0.11 \)) such that the best result is obtained, since the scale of the energy function is changed...
after the Anscombe transform of the initial image. For the denoised results, we do not apply
the bilateral filtering postprocessing to remove artifacts, so that the comparison is clear and
fair.

In Figure 1, the simulation results of the test image cameraman are shown for all the
algorithms considered previously in Section 2. The denoising results of the reweighted $\ell^2$ model
and that of the KL-divergence model are visually the same and their PSNR are comparable to
each other, the difference in pixel value of these two results is shown in Figure 1. The Anscombe
transform model gives a slightly better result than the other two models, whose parameters are
well tuned, mainly in the low pixel value region. The difference of results in pixel value of
the Anscombe transform algorithm (18) and that of the KL-divergence based split Bregman
method (16) is shown as well, where the biggest difference is about 6.

Table 1 shows a quantitative comparison of PSNR of the other test images, fruit, boat and
goldhill. The Anscombe transform algorithm performs substantially better for the test image
goldhill than the other algorithms, mainly because the test image has a large low intensity
region where the Anscombe transform provides a better Gaussian approximation of the low
rate Poisson distribution than a weighted $\ell^2$ formulation. On the other hand, it performs worse
for the test image fruit, which has a small low intensity region but large high intensity region. Figure 2 shows the histogram of absolute residual (error) of reconstruction results in low and
high intensity regions, which further illustrates the limitation of these two methods in different
regions. Therefore, for pure denoising purpose, the Anscombe transform is preferred for low
intensity images while the other three algorithms are preferred for relatively high intensity
images.

<table>
<thead>
<tr>
<th></th>
<th>re-weighted $\ell^2$</th>
<th>KL(SB)</th>
<th>EM</th>
<th>Anscombe</th>
</tr>
</thead>
<tbody>
<tr>
<td># of (outer) iterations</td>
<td>50</td>
<td>400</td>
<td>50(with 5 inner iteration)</td>
<td>100</td>
</tr>
<tr>
<td>cameraman</td>
<td>30.54</td>
<td>30.59</td>
<td>30.57</td>
<td>30.69</td>
</tr>
<tr>
<td>fruit</td>
<td>30.71</td>
<td>30.71</td>
<td>30.71</td>
<td>30.64</td>
</tr>
<tr>
<td>boat</td>
<td>29.19</td>
<td>29.18</td>
<td>29.18</td>
<td>29.19</td>
</tr>
<tr>
<td>goldhill</td>
<td>29.24</td>
<td>29.24</td>
<td>29.24</td>
<td>29.38</td>
</tr>
</tbody>
</table>

Table 1: PSNR value of the denoising result.

The similarity between these algorithms can be explained by the connections between their
convergent solutions. In the Poisson noise only case, the operator $A$ is identity, therefore the
objective function is strictly convex. According to the theorem 5.1, our algorithm (without
bilateral filter processing) converges to the unique minimizer $u^*$ of the KL-divergence fidelity
model with the same framelet sparsity regularization. Consequently, the two solutions should
have the same re-weighted $\ell_2$ energy and $\ell_1$ sparsity of the framelets coefficients, where the
weight in the $\ell_2$ norm is fixed to $u^* + c$ and $u^*$ is approximated by the split Bregman result.
This can be proved numerically as shown in Figure 3; the result of our algorithm with regularized
weight $\tilde{\Sigma}_k$ has weighted $\ell_2$ energy plus $\ell_1$ sparsity close to a minimizer of the KL-divergence
fidelity model calculated by the Split Bregman algorithm (16) as well as the solution given by
the EM algorithm (15). The three solutions are not identical, though, because each algorithm
is embedded with different regularization and generates numerical error differently. In Figure
3, the KL-divergence plus $\ell_1$ sparsity energy is also plotted. Also, our proposed scheme and the
EM algorithm converge faster than the split Bregman method with respect to the number of
iterations, but the EM algorithm involves a considerable number of inner iterations (5 iterations
in our implementation) in solving the second sub-problem; it converges to a solution of slightly
higher energy.
Figure 1: Denoising results of simulated noisy images with peak intensity 120. From left to right, up to down: noisy image, reweighted $\ell^2$ method by Algorithm 1, Anscombe transform (18), KL-divergence model using split Bregman (16) and KL-divergence model with EM-$\ell^1$ algorithm (15). The last image shows the difference in pixel value of Anscombe+$\ell_1$ result and KL(SB) result.
Figure 2: Left: histogram of absolute residual $|u_{recon} - u_{true}|$ in low intensity region of the test image goldhill. Right: histogram of absolute residual $|u_{recon} - u_{true}|$ in high intensity region of the test image fruit.

Figure 3: Left: KL–divergence ($E(u) = 1^T(u + c) - f^T \log(u + c) + \lambda \|Wu\|_1$) energy evolution of Algorithm 1, (15) and (16). Right: reweighted $\ell^2$-norm energy ($E(u) = \frac{1}{2} \left\| \frac{u + c - f}{\sqrt{u^2 + c}} \right\|^2 + \lambda \|Wu\|_1$) evolution of Algorithm 1, (15) and (16).
4.2 Deblurring from Poisson data

The deblurring case is in general more difficult than denoising. We compare our reweighted $\ell^2$ method and the KL-divergence model (14) since Ancombe transform can not be easily applied in this case. For this application, we compare our proposed Algorithm 1 with two different algorithms (EM+$\ell^1$ algorithm (15) and KL-split Bregman algorithm (16)) for the KL-divergence model (14). In this subsection, blurred and noisy images are simulated in the following way. The clean images are first rescaled to an intensity range from 0 to 1200; then they are corrupted by a disk blurring kernel of radius 3 with symmetric boundary condition; finally, the Poisson noise is simulated in the same way as in the denoising case.

The main computation cost of deblurring Algorithm 1 is that $u_{k+1}$ is solved by conjugate gradient method instead of a direct inversion in the first step of the iteration. As shown in Figure 4, the result of the KL-divergence fidelity algorithm is more blurry in contrast to the result of weighted $\ell^2$ fidelity, although they have similar PSNR, see Table 2 for more comparisons. Table 2 shows that our re-weighted method (26) and KL-split Bregman algorithm (16) have higher PSNR value than EM+$\ell^1$ algorithm (15). With the postprocess of bilateral filter, our reweighted method (26) can obtain the best PSNR value among all the algorithms.

![Deblurring results of simulated noisy-blurred "Goldhill" image with peak intensity 1200. First row from left to right: ground truth image, blurry image, EM+$\ell^1$ model (15). Second row from left to right: KL-split Bregman algorithm (16), direct reweighted $\ell^2$ method (26) and reweighted method (26) with bilateral filter post-process.](image)

<table>
<thead>
<tr>
<th></th>
<th>reweighted $\ell^2$ with bilateral filter</th>
<th>reweighted $\ell^2$</th>
<th>KL-Split Bregman</th>
<th>EM+$\ell^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>goldhill</td>
<td>26.82</td>
<td>26.76</td>
<td>26.74</td>
<td>26.62</td>
</tr>
<tr>
<td>cameraman</td>
<td>26.64</td>
<td>26.33</td>
<td>26.41</td>
<td>25.84</td>
</tr>
<tr>
<td>boat</td>
<td>25.70</td>
<td>25.59</td>
<td>25.39</td>
<td>25.08</td>
</tr>
</tbody>
</table>

Table 2: PSNR value of the deblurring result.
Figure 5: Deblurring results of simulated noisy-blurred "Cameraman" image with peak intensity 1200. First row from left to right: ground truth image, blurry image, EM+$\ell_1$ model (15). Second row: zoom-in images of the first row. Third row from left to right: KL-split Bregman algorithm (16), direct reweighted $\ell^2$ method (26) and reweighted method (26) with bilateral filter post-process. Fourth row: zoom-in images of the third row.
In Figure 6, we compare the energy evolution of different algorithms. According to our previous argument, Algorithm 1, if converge, converges to the solution of the model (14). We thus compare the KL-divergence model $E(u) = 1^T(Au + c) - f^T \log(Au + c) + \lambda \|Wu\|_1$ of the three algorithms (EM-$\ell^1$, KL-split Bregman, and the proposed Algorithm 1). We can see that the energy for our proposed Algorithm 1 has the fastest speed of convergence. The evolution of weighted $\ell^2$-norm energy $E(u) = \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au^2 + c}} \right\|_2^2 + \lambda \|Wu\|_1$ is also shown in the right figure of Fig. 6, where $u^*$ is set as the ground truth image. To reach the same PSNR level of the result given by the KL-Split Bregman algorithm after 50 iterations, the proposed Algorithm 1 needs much less iterations (less than 5 iterations), See Fig. 7.

![Figure 6: Top: KL-divergence ($E(u) = 1^T(Au + c) - f^T \log(Au + c) + \lambda \|Wu\|_1$) energy evolution of "Goldhill" image for (15), (16) and Algorithm 1. Bottom: reweighted $\ell^2$-norm energy ($E(u) = \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au^2 + c}} \right\|_2^2 + \lambda \|Wu\|_1$) evolution of "Goldhill" image for (15), (16) and Algorithm 1.](image)

In terms of computational time of deblurring, the reweighted algorithms (1) needs around 100s while the EM-$\ell^1$ around 240s for the same relative error stopping criteria. As a result, our proposed reweighted method (26) can outperform the EM-$\ell^1$ in terms of visual quality, PSNR value, convergence speed and computational time. Moreover, compared with the KL-split Bregman algorithm (16), our proposed Algorithm 1 has faster convergence speed in terms of both the KL-energy and reweighted $\ell^2$-norm energy.

### 4.3 Mixed Poisson-Gaussian noise restoration

In this section, we test the extension of reweighted $\ell^2$ model to mixed Poisson-Gaussian noise for some synthesized data and real photos. The synthesized mixed noised image is simulated by
adding first Poisson noise to a rescaled image with peak intensity 120 and the Gaussian noise of $\sigma = 12$. Fig. 8 shows the denoising result of model (31) with $A$ being identity operator.

In addition, we perform the mixed denoising model on a digital photo taken in a low light environment with high ISO setting. Our denoising result is compared with the embedded noise reduction algorithm in Sony DSLR-A700 camera in Figure 9, and we can see that our result has less noise and better quality.

For deblurring case, we generate blurred image with peak intensity 1200 and then corrupt it with Poisson noise, Gaussian noise of $\sigma = 12$ consecutively. See Fig. 10 for the deblurring result of model (31).

5 Conclusion

In this paper, we studied weighted $\ell^2$ fidelity derived from Gaussian approximation of Poisson statistics. In solving denoising and deblurring models, we proposed a reweighted algorithm based on Split Bregman iteration. The solution of the proposed algorithm is same as the minimizer of KL-divergence fidelity model (14) if the sequence converges. Although we show the convergence of the sequence can be shown under some conditions, but the validity of these
Figure 9: Top: real photo; middle: output of camera using the noise reduction option. bottom: denoised by model (31).
Figure 10: left: input blurred image with mixed Poisson-Gaussian noise, $\sigma = 12$, peak intensity: 1200; right: deblurred by model (31).
conditions is not easy to check. We rely on our numerical experiments to further verify the performance of our algorithm; the proposed algorithm gives competitive result with respect to classical fidelity terms in denoising and deblurring simulations. In addition, the reweighted $\ell^2$ model can be easily extended to the mixed Poisson-Gaussian noise case. As shown by numerical results, the reweighted $\ell^2$ with split Bregman framework is promising in a wide perspective.

Both the algorithm proposed and the theoretical results shown in this paper are an attempt to study and link different variational models in image processing. In this study, we only focused on the fidelity term, using a simple sparsity term as regularization, without utilizing the full image structure for the analysis of Poisson statistics. One possible extension of our current work would be incorporating the image structure and utilizing the spatial correlation of the pixel values to build a more precise statistical model, hence a more precise fidelity term, of the Poisson data. This might substantially improve the result as suggested by the improvement made of a post bilateral filtering process in our deblurring simulation.

Appendix

For the sake of notation simplicity, we drop the constant $c$ in the observation model (3) in the following, but instead substitute the nonnegative constrain $u \geq 0$ by $u \geq c$. Under the assumption that $A > 0, A1 = 1$, the minimizer $u^*$ to this modified minimization problem

$$\min_{u \geq c} 1^T Au - f^T \log(Au) + \lambda \|D u\|_1$$  \hspace{1cm} (32)

can be rewritten as $\tilde{u}^* + c$, where $\tilde{u}^*$ is the minimizer of (14). Therefore, (32) is essentially the same as model (14) up to the constant $c$ that is prefixed. We also consider a variant of the split Bregman algorithm (27) by adding a parameter $\delta_b$ in the third step $b_{k+1} = b_k + \delta_b(Wu_{k+1} - d_{k+1})$. In the following theorem, we consider $1 \geq \delta_b > 0$, which is set to 1 in our algorithm.

**Theorem 5.1** Assume that $u^*$ is an arbitrary minimizer of (14). Assume that $\lambda > 0, \mu > 0, 1 \geq \delta_b > 0, A > 0, A1 = 1$ and $\exists K_0, \delta_1, \delta_2$ s.t. $2\delta_1 + \delta_2 \leq \sqrt{c}$,

$$\left\| \frac{A(u_k - u^*)}{\sqrt{Au^*}} \right\|_\infty \leq \delta_1 \leq \sqrt{c}, \forall k > K_0$$  \hspace{1cm} (33)

and

$$\left\| \frac{f - Au^*}{\sqrt{Au^*}} \right\|_\infty \leq \delta_2$$  \hspace{1cm} (34)

Then we have the following properties:

$$\lim_{k \rightarrow +\infty} Au_k = Au^*$$  \hspace{1cm} (35)

$$\lim_{k \rightarrow +\infty} \frac{1}{2} \left\| \frac{Au_k - f}{\sqrt{Au_k}} \right\|_2^2 = \frac{1}{2} \left\| \frac{Au^* - f}{\sqrt{Au^*}} \right\|_2^2$$

$$\lim_{k \rightarrow +\infty} ||Wu_k||_{\ell_1} = ||Wu^*||_{\ell_1}$$  \hspace{1cm} (36)

In addition, $u^*$ is also an minimizer of (26). Furthermore,

$$\lim_{k \rightarrow +\infty} ||u_k - u^*|| = 0$$  \hspace{1cm} (37)

whenever $u^*$ is the unique solution of (26).
Proof. Let $u^*$ be an solution of (14). By the first order optimality condition, $u^*$ must satisfy

$$0 = A^T \left( 1 - \frac{f}{Au^*} \right) + \lambda \mathcal{W}^T p^* - q^*$$

$$= A^T \left( \frac{Au^* - f}{Au^*} \right) + \lambda \mathcal{W}^T p^* - q^*$$

(38)

where $p^* \in \partial |d^*|$ with $d^* = \mathcal{W} u^*$ and $q^* \geq 0$ is a Lagrange multiplier satisfying

$$\langle q^*, u^* - c \rangle = 0$$

(39)

Let $\Sigma_\ast = \text{diag}(Au^*)$ and $b^* = \frac{\lambda}{\mu} p^*$.

We have

$$\begin{cases}
    0 = A^T \Sigma_\ast^{-1} (Au^* - f) + \mu \mathcal{W}^T (\mathcal{W} u^* - d^* + b^*) - q^* \\
    0 = \lambda p^* + \mu (d^* - \mathcal{W} u^* - b^*), \quad \text{with} \quad p^* \in \partial |d^*| \\
    b^* = b^* + \delta_b (\mathcal{W} u^* - d^*)
\end{cases}$$

(40)

In the rest of this proof, we abuse the notation $\Sigma_\ast$ to either stand for the matrix with diagonal entries $Au^*$ or the vector $Au^*$ when there is no confusion. Therefore, the condition 34 can be rewritten as

$$\left\| \frac{f - \Sigma_\ast}{\sqrt{\Sigma_\ast}} \right\|_{\infty} \leq \delta_2$$

(41)

Since $u^* \geq c$ and by the assumption that $A \geq 0, A1 = 1$, the diagonal elements of $\Sigma_\ast$ are uniformly lower bounded by $c$, i.e. $Au^* \geq Ac1 = c1$ or equivalently

$$\left\| \Sigma_\ast \right\|_{\infty} \geq c.$$  

(42)

Let $\Sigma_k = Au_k$ and rewrite the algorithm (27) into the following form,

$$\begin{cases}
    0 = A^T \Sigma_k^{-1} (Au_{k+1} - f) + \mu \mathcal{W}^T (\mathcal{W} u_{k+1} - d_k + b_k) - q_{k+1} \\
    0 = \lambda p_{k+1} + \mu (d_{k+1} - \mathcal{W} u_{k+1} + b_k), \quad \text{with} \quad p_{k+1} \in \partial |d_{k+1}| \\
    b_{k+1} = b_k + \delta_b (\mathcal{W} u_{k+1} - d_{k+1})
\end{cases}$$

(43)

$$\langle q_k, u_k - c \rangle = 0$$

(44)

Denote the errors by

$$u_k^e = u_k - u^*, \quad d_k^e = d_k - d^*, \quad b_k^e = b_k - b^*.$$  

Then the condition (33) can be rewritten as

$$\left\| \frac{Au_k^e}{\sqrt{\Sigma_\ast}} \right\|_{\infty} \leq \delta_1$$

(45)

Subtracting the first equation of (43) by the first equation of (40), we have

$$0 = A^T (\Sigma_k^{-1} - \Sigma_\ast^{-1}) (Au_{k+1} - f) + A^T \Sigma_k^{-1} Au_k^e + \mu \mathcal{W}^T (\mathcal{W} u_{k+1}^e - d_k^e + b_k^e) - (q_{k+1} - q^*)$$

Take the inner product of the both sides with respect to $u_{k+1}^e$, we obtain

$$0 = \langle u_{k+1}^e, A^T (\Sigma_k^{-1} - \Sigma_\ast^{-1}) (Au_{k+1} - f) \rangle + \| Au_k^e \|_{\Sigma_k^{-1}}^2 + \mu \| \mathcal{W} u_{k+1}^e \|_{\Sigma_\ast}^2$$

$$\quad - \mu \langle u_{k+1}^e, W^T d_k^e \rangle + \mu \langle u_{k+1}^e, W^T b_k^e \rangle - \langle q_k - q^*, u_{k+1} - u^* \rangle$$

(46)
The same manipulation applied to the second equations of (40) and (43) leads to

\[ 0 = \lambda(p_{k+1} - p^*, d_{k+1} - d^*) + \mu\|d_{k+1}^e\|^2_2 - \mu\langle d_{k+1}^e, Wk_{k+1}^e \rangle - \mu\langle d_{k+1}^e, b_k^e \rangle \quad (47) \]

We can also do the similar operation to the third equations of (40) and (43) and obtain

\[ \langle b_k^e, Wu_{k+1}^e - d_{k+1}^e \rangle = \frac{1}{2\delta_b} (\|b_k^e\|^2_2 - \|b_k^e\|^2_2) - \frac{\delta_b}{2} \|Wu_{k+1}^e - d_{k+1}^e\|^2_2. \quad (48) \]

Using (39) and (44), the last term in (46) becomes

\[ -\langle q_k - q^*, u_{k+1} - u^* \rangle = \langle q_k, u^* - c \rangle + \langle q^*, u_{k+1} - c \rangle \quad (49) \]

Adding (47) to (46) and utilizing (49), we get

\[ 0 = \langle u_{k+1}^e, A^T(\Sigma_{k}^{-1} - \Sigma_{*}^{-1})(Au_{k+1} - f) \rangle + \|Au_{k+1}^e\|^2_{\Sigma_{*}^{-1}} + \lambda\langle p_{k+1} - p^*, d_{k+1} - d^* \rangle + \mu\|Wu_{k+1}^e\|^2_2 + \|d_{k+1}^e\|^2_2 \]

\[ -\langle d_{k+1}^e + d_{k}^e, Wu_{k+1}^e \rangle + \langle Wu_{k+1}^e - d_{k+1}^e, b_k^e \rangle + \langle q_k, u^* - c \rangle + \langle q^*, u_{k+1} - c \rangle \quad (50) \]

Substituting (48) into (50), we have

\[ \frac{\mu_2}{2\delta_b} (\|b_k^e\|^2_2 - \|b_k^e\|^2_2) = \]

\[ \langle u_{k+1}^e, A^T(\Sigma_{k}^{-1} - \Sigma_{*}^{-1})(Au_{k+1} - f) \rangle + \|Au_{k+1}^e\|^2_{\Sigma_{*}^{-1}} + \lambda\langle p_{k+1} - p^*, d_{k+1} - d^* \rangle + \mu\|Wu_{k+1}^e\|^2_2 + \|d_{k+1}^e\|^2_2 \]

\[ -\langle d_{k+1}^e + d_{k}^e, Wu_{k+1}^e \rangle - \frac{\delta_b}{2} \|Wu_{k+1}^e - d_{k+1}^e\|^2_2 + \langle q_k, u^* - c \rangle + \langle q^*, u_{k+1} - c \rangle \quad (51) \]

Expanding the first term on the right hand side,

\[ \langle u_{k+1}^e, A^T(\Sigma_{k}^{-1} - \Sigma_{*}^{-1})(Au_{k+1} - f) \rangle = \langle Au_{k+1}^e, \left( \frac{1}{Au_{k}^e} - \frac{1}{Au_{k}^e} \right)(Au_{k+1} - f) \rangle \]

\[ = \langle Au_{k+1}^e, \frac{1}{Au_{k}^e}Au_{k+1} - f \rangle \]

\[ = \langle Au_{k+1}^e, \frac{1}{Au_{k}^e}Au_{k}^e \rangle - \langle f, Au_{k+1}^e \rangle \]

\[ = \langle Au_{k+1}^e, \frac{Au_{k+1}^e}{Au_{k}^e}Au_{k}^e \rangle \Sigma_{*}^{-1} \]

\[ = \langle Au_{k+1}^e, \frac{Au_{k+1}^e}{Au_{k}^e}Au_{k}^e - f \rangle \Sigma_{*}^{-1} \]

\[ = \langle Au_{k+1}^e, \frac{Au_{k+1}^e}{Au_{k}^e}Au_{k}^e - f \rangle \Sigma_{*}^{-1} + \langle Au_{k+1}^e, \frac{Au_{k+1}^e}{Au_{k}^e}Au_{k}^e - f \rangle \Sigma_{*}^{-1} \]

\[ = \langle Au_{k+1}^e, \frac{Au_{k+1}^e}{Au_{k}^e}Au_{k}^e - f \rangle \Sigma_{*}^{-1} + \langle Au_{k+1}^e, \frac{Au_{k+1}^e}{Au_{k}^e}Au_{k}^e - f \rangle \Sigma_{*}^{-1} \]

Note that all the operations between vectors in (52) are element wise except the inner product. Using (42), (41) and (45) we have the following estimation,

\[ \left\| \frac{Au_{k+1}^e}{Au_{k}^e} + \Sigma_{*} \right\|_\infty \leq \left\| \frac{Au_{k+1}^e}{\sqrt{\Sigma_{*}}} \right\|_\infty + \sqrt{\Sigma_{*}} \leq \frac{\delta_1}{\sqrt{\Sigma_{*} - \delta_1}} \leq \frac{\delta_1}{\sqrt{c - \delta_1}}, \quad k \geq K_0 \quad (53) \]

and

\[ \left\| \frac{\Sigma_{*} - f}{Au_{k}^e} \right\|_\infty \leq \left\| \frac{\Sigma_{*} - f}{\sqrt{\Sigma_{*}}} \right\|_\infty \left\| \frac{1}{\sqrt{\Sigma_{*} + \Sigma_{*}}} \right\|_\infty \leq \frac{\delta_2}{\sqrt{c - \delta_1}} \quad (54) \]
\[ \delta = \frac{\delta_1 + \delta_2}{\sqrt{c} - \delta_1} \leq 1, \]

\[ \| (u_{k+1}^e, A^T (\Sigma_k^{-1} - \Sigma_*^{-1}) (Au_{k+1} - f)) \| \]
\[ \leq \| (Au_{k+1}, \frac{Au_{k+1}^e}{Au_k + \Sigma_*}) \| + \| (Au_{k+1}^e, \frac{\Sigma_* - f}{Au_k + \Sigma_*}) \| \]
\[ \leq \frac{\delta_1 + \delta_2}{\sqrt{c} - \delta_1} \| (Au_{k+1}^e, |Au_{k+1}^e|) \|_{\Sigma_*^{-1}} \]
\[ \leq \frac{\delta_2}{2} \| Au_{k+1}^e \|_{\Sigma_*^{-1}}^2 + \| Au_{k+1}^e \|_{\Sigma_*^{-1}}^2. \]

With \( u^* \geq c, u_k \geq c, q^* \geq 0 \) and \( q_k \geq 0 \), we also have
\[ \langle q_k, u^* - c \rangle + \langle q^*, u_{k+1} - c \rangle \geq 0 \]

By substituting (55) and (56) into (51) and summing up from \( k = 0 \) to \( k = K \), we get
\[ \frac{\mu}{2\delta_b} \left( \| b_0^e \|_2^2 - \| b_{K+1}^e \|_2^2 \right) + \sum_{k=0}^{K-1} \left\{ \langle u_{k+1}^e, A^T (\Sigma_k^{-1} - \Sigma_*^{-1}) (Au_{k+1} - f) \rangle + \| Au_{k+1}^e \|_{\Sigma_*^{-1}}^2 \right\} \]
\[ \geq \delta_2 \left( \| Au_{K+1}^e \|_{\Sigma_*^{-1}}^2 - \| Au_{0}^e \|_{\Sigma_*^{-1}}^2 \right) \]
\[ + (1 - \delta) \sum_{k=K_0}^{K} \| Au_{k+1}^e \|_{\Sigma_*^{-1}}^2 + \lambda \sum_{k=0}^{K} \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle - \frac{\mu}{2} \| d_0^e \|_2^2 \]
\[ + \frac{\mu}{2} \| d_{K+1}^e \|_2^2 + \mu \left( \frac{1 - \delta_b}{2} \sum_{k=0}^{K} \| \mathcal{W} u_{k+1}^e - d_{k+1}^e \|_2^2 + \frac{1}{2} \sum_{k=0}^{K} \| \mathcal{W} u_{k+1}^e - d_{k}^e \|_2^2 \right) \]

By rearranging the terms on both sides of (57) and eliminate, we obtain
\[ \frac{\mu}{2\delta_b} \left( \| b_0^e \|_2^2 + \delta \| Au_{0}^e \|_{\Sigma_*^{-1}}^2 + \| d_0^e \|_2^2 \right) \]
\[ \geq (1 - \delta) \sum_{k=K_0}^{K} \| Au_{k+1}^e \|_{\Sigma_*^{-1}}^2 + \lambda \sum_{k=0}^{K} \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle + \frac{\mu}{2} \sum_{k=0}^{K} \| \mathcal{W} u_{k+1}^e - d_{k}^e \|_2^2 \]
\[ + \sum_{k=0}^{K_0-1} \left\{ \langle u_{k+1}^e, A^T (\Sigma_k^{-1} - \Sigma_*^{-1}) (Au_{k+1} - f) \rangle + \| Au_{k+1}^e \|_{\Sigma_*^{-1}}^2 \right\} \]

Since \( p_{k+1} \in \partial |d_{k+1}| \) and \( p^* \in \partial |d^*| \), together with the convexity of \( |\cdot| \), we have
\[ \langle p_{k+1} - p^*, d_{k+1} - d^* \rangle \geq 0 \]

Hence all terms on the right hand side of (58), except the last finite summation, are non-negative with the assumptions of \( \delta < 1, \mu > 0, \lambda > 0 \), and our first observation is as follows
\[ \sum_{k=0}^{\infty} \| Au_{k+1} \|_{\Sigma_*^{-1}}^2 < +\infty \]
which leads to
\[ \lim_{k \to +\infty} \| Au_{k+1} - Au^* \|_{\Sigma_*^{-1}}^2 = 0 \]
Therefore, (35) holds as the weight $\Sigma_*^{-1}$ is positive. This immediately leads to the first part of (36). With (35), the minimization problem (26) is well-defined. Note that the first order optimality condition of (26) is the same as (38) and (39) hence $u^*$ is an minimizer of (26). Likewise we have
\[
\lim_{k \to +\infty} \|Wu_{k+1} - d_k\| = 0
\]
With $Wu^* = d^*$, we further get
\[
\lim_{k \to +\infty} \|Wu_{k+1} - d_k\| = 0
\]
(58) also gives
\[
\lim_{k \to +\infty} \langle p_k - p^*, d_k - d^* \rangle = 0
\]
While notice that $B_{\Omega_k}^p(d_k, d^*) + B_{\Omega_k}^p(d^*, d_k) = \langle p_k - p^*, d_k - d^* \rangle$, where $B_{\Omega_k}^p(u, v)$ with $p \in \partial J(v)$ is the Bregman distance defined on convex function $J$. As the Bregman distance is always positive, (61) together with (60) results in
\[
\lim_{k \to +\infty} |Wu_k| - |Wu^*| - \langle W(u_k - u^*), p^* \rangle = 0
\]
Using (59), we have
\[
\langle W(u_k - u^*), p^* \rangle = \langle W(u_k - u^*), p^* \rangle + \frac{1}{\lambda} \langle A(u_k - u^*), Au^* - f \rangle_{\Sigma_*^{-1}}
\]
\[
= \frac{1}{\lambda} \langle u_k - u^*, A^T \Sigma_*^{-1} (Au^* - f) + \lambda W^T p^* \rangle = 0
\]
The last equality comes from (38). Therefore,
\[
\lim_{k \to +\infty} |Wu_k| - |Wu^*| = 0
\]
which is the second equation in (36).

Next, assume that $u^*$ is the unique solution of 26. Let $E(u) = \frac{1}{2} \|Au - f\|_{\Sigma_*^{-1}}^2 + \lambda \|Wu\|_{\ell_1}$. It is clear that $E(u)$ is convex, lower semi-continuous. Therefore, the uniqueness of the solution leads to (37). 

**Remark 1** In Theorem 5.1, the convergence requires conditions (33) and (34). Although the two conditions are in the form of infinity norm, which enforces a global bound, we can actually restrict ourselves in a local image patch, or in a sub-region of the image domain, because all the element-wise operations and the operator $A$ are local; in other words, these two conditions are local uniform conditions. For Poisson noise $n = Au - f$ of mean $Au$, the standard variation of the noise is of order $O(\sqrt{Au})$, i.e.
\[
Au - f \sim O(\sqrt{Au}).
\]
For rate $Au$ bigger than 10, that is where the pixel value is big enough, the Poisson distribution can be well approximated by a Gaussian distribution with the same mean and variance. Since $P(|x| < 3\sigma) \approx 1$ if $x \sim N(0, \sigma^2)$, $|n| = |Au - f| < 3\sqrt{Au}$ with high probability. This implies that when $c$ is of the order of 10 or bigger, then $Au > c$ is big as well, condition (34) holds for $\delta_2 = 3$ with high probability. The other condition (33) is much harder to check, which involves $u_k$ and we cannot guarantee that $Au_k$ is a better approximation of $Au^*$ than $f$ pointwise. If $c$ is large, that is the pixel value is large, then we can choose $\delta_1$ to be big enough so that (33) holds with high probability. But for the low intensity part, (33) and (34) may fail. This matches our observation in numerical simulations that the algorithm performs worse in the lower intensity region than in the higher one.
Acknowledgments

Xiaoqun Zhang was partially supported by the NNSFC (Grant nos. 11101277, 91330102) and the Shanghai Pujiang Talent Program (Grant no. 11PJ1405900). Zuowei Shen was partially supported by National University of Singapore (Grant no. MOE2011-T2-1-116).

References


