

A reweighted ℓ^2 method for image restoration with Poisson and mixed Poisson-Gaussian noise*

Jia Li[†] Zuowei Shen[†] Rujie Yin[‡] Xiaoqun Zhang[§]

September 27, 2012

Abstract

We study weighted ℓ^2 fidelity in variational models for Poisson noise related image restoration problems. Gaussian approximation to Poisson noise statistic is adopted to deduce weighted ℓ^2 fidelity. Different from usual weighted ℓ^2 approximation, we propose a reweighted ℓ^2 fidelity with sparse regularization by framelets. Based on Split Bregman algorithm introduced in [20], 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 ℓ^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 ℓ^2 fidelity, Gaussian approximation, split Bregman iteration, framelets

1 Introduction

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

$$f = Au \tag{1}$$

where A denotes a linear operator from \mathbb{R}^N to \mathbb{R}^M . In this paper, we focus on variational image restoration from Poisson and mixed Poisson-Gaussian noisy observation f . Typically, variational models for image restoration are composed of two terms, one is data fidelity term and the other is regularization term for modeling *a priori* knowledge on unknown images. In general, data fidelity term keeps true image u close enough to input data f , otherwise useful information may be discarded in the solution. According to different noise statistics, fidelity term takes different forms. It is well known that the least square fidelity is used for additive white

*Draft on September 27, 2012

[†]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

Gaussian noise (AWGN), and it is mostly considered in literature for its good characterization of system noise. Whereas, non-Gaussian types of noise are also encountered in real images. An important variant is Poisson noise, which is generally observed in images produced by low number of photons, such as fluorescence microscopy, emission tomography, etc. A vast amount of literature in image restoration is devoted to problems encountering Poisson noisy images (see [2] and the references therein). The poisson noise data, i.e. the probability of receiving y particles is given by

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

where τ is the expected value and the variance of random counts. 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 [23], a SURE estimator based on Poisson-Gaussian statistics is constructed in wavelet transform for mixed noise denoising, however the complicated construction of estimators can not be easily extended for more general image restoration problem such as deblurring. Recently, Gong et al. proposed in [21] a universal $\ell^1 + \ell^2$ fidelity term for mixed or unknown noise and achieve encouraging numerical results, however it is hard to derive a meaningful statistical understanding.

Besides 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. Wavelet type of transforms are often chosen to sparsify natural images. For a large class of applications, penalizing the ℓ^1 norm of its transform coefficients leads to sparse representation, as largely illustrated in compressive sensing communality in recent few years. On the other hand, choosing an appropriate transform basis for natural images are also important. One popular choice is total variation proposed by Rudin-Osher-Fatemi [27]. Framelets becomes very useful for different restoration tasks, see [12] and the reference therein. Recently, the relation of sparse framelets representation and total variation has been revealed in [9], and theoretically total variation can be interpreted as a special form of ℓ^1 framelets regularization. For this reason, we will consider framelets regularization in this paper.

The generalized Kullback-Leibler (KL)-divergence [10] fidelity is usually used in Poisson image restoration, and it is directly derived based on exact statistics of Poisson noise (see section 2.3.1 below). However, difficulties on computation stability and efficiency arise due to the logarithm function appeared in the KL-divergence function. We are interested in considering a least square based fidelity by approximating Poisson noise with appropriate Gaussian noise. In fact, the approximation of KL-divergence by a fixed weighted least-square method 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. [32], where the fidelity term is defined as a weighted least square involving the unknown image. Combining with regularization, a scaled gradient method with line search technique has been applied to solve the resulted problem. Here, we consider utilizing the efficiency of ℓ^2 fidelity based regularization algorithm and dynamically estimate Poisson noise variance. More precisely, we propose an iterative algorithm based on split Bregman iteration [20], originally proposed for ℓ^2 fidelity based ℓ^1 regularization, to solve the model with reweighted ℓ^2 fidelity. With this formulation, soft-shrinkage can be naturally incorporated as an efficient ℓ^1 minimization algorithm. Furthermore, the model gives a reasonable extension to the mixed Poisson-Gaussian noise case immediately 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, split Bregman method and existing Poisson fidelities. Then we present the weighted ℓ^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 cases. 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 noise case.

2 Background

2.1 Variational image restoration model

The observation model in consideration takes the following generic form

$$f = Au + c + \epsilon \quad (3)$$

where A denotes a linear operator from $\mathbb{R}^N \rightarrow \mathbb{R}^M$, c is fixed background and possibly 0, and ϵ is noise perturbation. Generally, for additive white Gaussian noise setting, it is assumed that $c = 0$ and ϵ follows independent normal distribution of mean 0 and variance σ^2 for each pixel. The operator A can either come from the imaging system such as the point spread function (PSF) or the disturbance related to the object like its own motion. For image denoising, A is an identity operator; for image deblurring, A is a convolution operator.

Using classical maximum a posterior probability (MAP) $\mathbf{P}(u|f)$ estimation, image restoration by variational model is usually composed of the minimization of the sum of two terms, namely fidelity and regularization:

$$\min_u F(u) + \lambda G(u) \quad (4)$$

where λ is a positive number for balancing the fidelity term $F(u)$ and the regularization term $G(u)$. The fidelity term is related to the noise characteristic and derived by the likelihood function, while the regularization $G(u)$ is designed based on *a priori* assumption on u . As introduced previously, it is nowadays standard technique to penalize the ℓ^1 norm of representation coefficients in transform domain. Therefore, the following variational model is largely studied for image restoration from AWGN:

$$\min_u \frac{1}{2} \|Au - f\|_2^2 + \lambda \|\mathcal{D}u\|_1 \quad (5)$$

where $\|\cdot\|_1$ denotes the usual ℓ^1 vector norm and \mathcal{D} is a linear transform, such as discrete gradient used in total variation [27], Fourier transform, local cosine transforms, wavelet or framelets [8].

2.2 Split Bregman algorithm

Split Bregman method introduced in [20] is designed to solve the variational model taking the form as (5). More generally, we consider the minimization problem

$$\min_u F(u) + \lambda \|\mathcal{D}u\|_1 \quad (6)$$

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 $d = \mathcal{D}u$, the alternating split bregman for solving (6) is given by the following scheme: for $\mu > 0$,

$$\begin{cases} u_{k+1} = \operatorname{argmin}_u F(u) + \frac{\mu}{2} \|\mathcal{D}u - d_k + b_k\|_2^2 \\ d_{k+1} = \operatorname{argmin}_d \lambda \|d\|_1 + \frac{\mu}{2} \|\mathcal{D}u_{k+1} - d + b_k\|_2^2 \\ b_{k+1} = b_k + \mathcal{D}u_{k+1} - d_{k+1} \end{cases} \quad (7)$$

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

$$d_{k+1} = \text{sign}(\mathcal{D}u_{k+1} + b_k) \cdot \max(|\mathcal{D}u_{k+1} + b_k| - \lambda/\mu, 0) \quad (8)$$

where each operation is performed componentwisely. The nonsmooth optimization problem (6) is thus decomposed into three simple subproblems when $F(u)$ is in quadratic form. This method has been demonstrated to be very efficient for ℓ^1 type of minimization in a large variant of applications. In a more general setting, it is shown to be equivalent to classical Douglas-Rachford and alternating direction multiplier method (ADMM) [14, 28], thus the convergence was given in this framework. In [8], Cai et al. studied the application of such algorithm for framelets based image restoration and provide a convergence proof. Further approximation of the quadratic subproblem was considered in [34] 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. For this reason, we will consider a reweighted ℓ^2 fidelity term which is closely related to the problem (6).

2.3 Poisson noise related fidelities

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

$$f \sim \mathbf{P}(Au + c) \quad (9)$$

As in [30, 3, 7], the following assumptions are made for the linear operator A and c :

Assumption 1 • *The observation data $f > 0$, and $u \geq 0$.*

- *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 $c > 0$.*

Since both f and u denote photo counts number, thus their elements are nonnegative. For most of image restoration model, 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 weighted ℓ^2 . The second assumption implies that for $u \geq 0$, $Au = 0$ if and only if $u = 0$, see Lemma 3.1. In practice, if the background can be ignored, we can put $c = \mathbf{1}$ without degrading image quality, where $\mathbf{1}$ is the vector with each entry being 1.

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

$$\mathbf{P}(f|Au + c) = \prod_{i=1}^M \frac{(Au + c)_i^{f_i} e^{-(Au+c)_i}}{f_i!} \quad (10)$$

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

$$\mathbf{E}(f|Au + c) = \text{Var}(f|Au + c) = Au + c \quad (11)$$

Before presenting the proposed reweighted ℓ^2 fidelity, we 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 [10], which can be derived directly by MAP method. By taking a negative log likelihood (10), we obtain

$$-\log \mathbf{P}(f|Au + c) = \mathbf{1}^T(Au + c - \log(f!)) - f^T \log(Au + c), \quad (12)$$

where $\mathbf{1}$ is the vector with each entry being 1.

If we neglect the constant term $\log(f!)$ which is unrelated to the unknown u , we obtain the following fidelity term

$$F(u) = \mathbf{1}^T(Au + c) - f^T \log(Au + c) \quad (13)$$

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

$$\min_{u \geq 0} \mathbf{1}^T(Au + c) - f^T \log(Au + c) + \lambda \|\mathcal{D}u\|_1 \quad (14)$$

Generally, (14) is a difficult optimization problem because of the nonsmooth regularization term and KL-divergence term. Optimization of KL-divergence fidelity with nonnegative constraint are typically solved by a popular iterative algorithm called Expectation-Maximization (EM) algorithm [22]. It is known that the convergence of the EM algorithm is slow and it may introduce so-called "checkboard effect" [3, 32, 7]. In [7], the authors proposed a two-step iteration method called EM-TV for solving (14) when \mathcal{D} is a discrete differential operator. We rename it as EM+ ℓ^1 algorithm for later adoption of framelets regularization. The algorithms is described as

$$\begin{cases} u_{k+\frac{1}{2}} = u_k A^* \left(\frac{f}{Au_k + c} \right) & \text{(EM step)} \\ u_{k+1} = \operatorname{argmin}_{u \geq 0} \frac{1}{2} \left\| \frac{u - u_{k+\frac{1}{2}}}{\sqrt{u_k}} \right\|_2^2 + \lambda \|\mathcal{D}u\|_1 & (\ell^1 \text{ step}) \end{cases} \quad (15)$$

This algorithm and its variants have been proved to be efficient for Poisson noise removal in PET and nasoscopy image deconvolution [7]. The convergence with a damped parameter under adequate condition is provided in [6]. However, the conditions are rather hard to verify since it involves the sequential iterates u_k and u_{k+1} . The other kind of method considered in [29, 16] is to apply directly the split Bregman method (7) by introducing a variable $d = \mathcal{D}u$ on the model (14). Comparing with this kind of method, our proposed algorithm takes a much simpler form and is more convenient for implementation.

2.3.2 Anscombe transform

Another well-known technique of Poisson denoise is the Anscombe transform [1] proposed for image denoising, when the linear operator A in (9) is the identity transform. Anscombe transform is defined as followed

$$\mathcal{A} : x \mapsto 2\sqrt{x + \frac{3}{8}} \quad (16)$$

When x is a random variable that obeys Poisson distribution with mean and variance τ , the transformed random variable $\mathcal{A}x$ follows an approximated standard normal distribution $\mathcal{N}(\tau, 1)$. We now apply the Anscombe transform on the observed data f in the form

$$f \sim \mathbf{P}(u + c)$$

Let $\tilde{f} := \mathcal{A}f$, $\tilde{u} := \mathcal{A}(u + c)$. Since f follows Poisson distribution with mean and variance $u + c$, 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

$$\tilde{u}^* = \operatorname{argmin}_{\tilde{u} \geq 0} \frac{1}{2} \|\tilde{u} - \tilde{f}\|_2^2 + \lambda \|\mathcal{D}\tilde{u}\|_1 \quad (17)$$

and the final image is given by $u^* = \mathcal{A}^{-1}\tilde{u}^* - c$. The regularization used in the above model (17) urges sparsity on \tilde{u} in the transformed domain. This is roughly equivalent to require sparsity on u in Framelet domain since the Anscombe transform (16) 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 nonlinear transform.

3 Weighted least square for Poisson noise image restoration

3.1 Framelets regularization

Here, we choose tight frame basis because of its multi-resolution property and redundancy that is helpful both in algorithm implementation and sparse representation of images [11],[12]. The multi-resolution analysis (MRA) based wavelet can be generated by the unitary extension principle (UEP) of [26]. By collecting all the refinement masks of wavelet tight frame system, we can generate the fast tight frame transform or decomposition operator \mathcal{W} . The matrix \mathcal{W} is consisted of $J + 1$ sub-filtering operators $\mathcal{W}_0, \mathcal{W}_1, \dots, \mathcal{W}_J$. Among them, \mathcal{W}_0 is the low-pass filtering operator and the rest are high-pass filtering operators. Correspondingly, by unitary extension principle [26], the operator \mathcal{W}^T is the fast tight frame reconstruction operator and we have $\mathcal{W}^T \mathcal{W} = \mathcal{I}$, i.e., $\mathcal{W}^T \mathcal{W}u = u$ for any image u . More details on discrete algorithms of framelet transforms can be found in [12].

3.2 Model and algorithm

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

$$\epsilon = f - Au - c. \quad (18)$$

Then the random variable ϵ can be interpreted as an additive perturbation noise. Given Au and c we have the mean of

$$\mathbf{E}(\epsilon|Au + c) = \mathbf{E}(f|Au + c) - Au - c = 0$$

and the variance

$$\operatorname{Var}(\epsilon|Au + c) = \operatorname{Var}(f|Au + c) = Au + c.$$

If we approximate the random difference ϵ by additive Gaussian noise, then ϵ follows normal distribution $\mathcal{N}(0, Au + c)$, i.e.

$$\mathbf{P}(\epsilon|Au + c) \simeq \exp\left\{-\frac{1}{2}(f - Au - c)^T \Sigma^{-1}(f - Au - c)\right\} \quad (19)$$

where Σ is the covariance matrix. By the independence of noise at each pixel, we have

$$\Sigma = \operatorname{diag}(Au + c) \quad (20)$$

where $\text{diag}(Au + c)$ is the diagonal matrix defined by M -dimensional vector $Au + c$. Using maximum likelihood, we take the negative log of the normal distribution (19) to get the following fidelity term as

$$-\log \mathbf{P}(\epsilon|Au + c) \propto \frac{1}{2}(f - Au - c)^T \Sigma^{-1}(f - Au - c) \quad (21)$$

We denote weighted ℓ^2 norm of a vector $x \in \mathbb{R}^N$ as $\|\mathbf{x}\|_Q^2 = \mathbf{x}^T Q \mathbf{x}$, for Q being a symmetric positive definite matrix. Then (21) is reformulated as

$$F(u) = \|Au + c - f\|_{\Sigma^{-1}}^2 = \|Au + c - f\|_{\text{diag}(Au+c)^{-1}}^2 = \left\| \frac{Au + c - f}{\sqrt{Au + c}} \right\|_2^2 \quad (22)$$

Notice that the division and square root operator in the ℓ^2 norm on the right hand side of (22) are both element wise. Also by the assumption that $c > 0$ and $Au \geq 0$ for $u \geq 0$, we have the positive definitiveness of Σ^{-1} . Note that (22) was used as a discrepancy principle for choosing the regularization parameter λ in [3]. However, this fidelity term (22) has a weight involving unknown u , and in practice, we need to either approximate or directly solve the nonlinear square problem. For example, the unknown weight $Au + c$ can be approximated by the observed data f as

$$\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{f}} \right\|_2^2 \quad (23)$$

Under this simplification, efficient least square based method can be applied together with regularization.

Combining with framelets sparse regularization and nonnegativity constraint, we have the following restoration model

$$\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au + c}} \right\|_2^2 + \lambda \|\mathcal{W}u\|_1 \quad (24)$$

This formulation is considered in [32] and a scale gradient projection method is applied to directly solve the nonlinear least square term with sparse regularization, for which a smoothed version of the ℓ^1 regularization term was adopted for calculating the gradient.

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. In fact, suppose that there exists an iteration solving (24) and u_k is the sequence of solutions derived from it. If u_k converges to u^* , then we are more or less solving the following minimization problem with a fixed u^*

$$\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au^* + c}} \right\|_2^2 + \lambda \|\mathcal{W}u\|_1 \quad (25)$$

for k is big enough. As u_k becomes stable, we may approximate the weight $\frac{1}{Au^* + c}$ by $\frac{1}{Au_k + c}$. With such an idea in mind, we combine with the popular Split Bregman iteration and derives a new algorithm called the reweighted ℓ^2 algorithm with split Bregman for Poisson noise:

$$\begin{cases} u_{k+1} = \operatorname{argmin}_{u \geq 0} \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au_k + c}} \right\|_2^2 + \frac{\mu}{2} \|\mathcal{W}u - d_k + b_k\|_2^2 \\ d_{k+1} = \operatorname{argmin}_d \lambda \|d\|_1 + \frac{\mu}{2} \|d - \mathcal{W}u_{k+1} - b_k\|_2^2 \\ b_{k+1} = b_k + (\mathcal{W}u_{k+1} - d_{k+1}) \end{cases} \quad (26)$$

With an idea of being close to previous iteration for stableness of solutions, we can add a proximal term on the update of u and d . For given $\gamma_1, \gamma \geq 0$, the numerical scheme is described

as followed:

$$\begin{cases} u_{k+1} = \operatorname{argmin}_{u \geq 0} \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au_k + c}} \right\|_2^2 + \frac{\mu}{2} \|\mathcal{W}u - d_k + b_k\|_2^2 + \frac{\gamma_1}{2} \|u - u_k\|_2^2 \\ d_{k+1} = \operatorname{argmin}_d \lambda \|d\|_1 + \frac{\mu}{2} \|\mathcal{W}u_{k+1} - d + b_k\|_2^2 + \frac{\gamma_2}{2} \|d - d_k\|_2^2 \\ b_{k+1} = b_k + \mathcal{W}u_{k+1} - d_{k+1} \end{cases} \quad (27)$$

This proximal iteration was largely studied by Rockefeller [25] in the context of augmented lagrangian. It was used for least square minimization with ℓ^0 regularization in [13] to stabilize the iterations. With $\gamma_1 = \gamma_2 = 0$, it reduces to the previous algorithm.

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.

Algorithm 1 Reweighted ℓ^2 with Spit Bregman for Poisson noise restoration Algorithm

Step0. Set the initial value, $u_0 = 0; d_0 = \mathcal{W}f; b_0 = 0; k = 1$; Σ_0 as the initial covariance matrix.

Step1.

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

$$u_k = \operatorname{argmin}_{u \geq 0} \frac{1}{2} \|Au + c - f\|_{\Sigma_{k-1}}^2 + \frac{\mu}{2} \|\mathcal{W}u - d_{k-1} + b_{k-1}\|_2^2 + \frac{\gamma_1}{2} \|u - u_{k-1}\|_2^2$$

$$d_k = \operatorname{sign}\left(\frac{\mu(\mathcal{W}u_k + b_{k-1}) + \gamma_2 d_{k-1}}{\mu + \gamma_2}\right) \cdot \max(|\mu(\mathcal{W}u_k + b_{k-1}) + \gamma_2 d_{k-1}| - \lambda / (\mu + \gamma_2), 0)$$

$$b_k = b_{k-1} + \mathcal{W}u_k - d_k$$

$$\Sigma_k = \operatorname{diag}(Au_k + c)$$

$$k = k + 1$$

end while

3.3 Analysis

In this section, we justify the proposed algorithm by presenting some mathematical analysis. Our main theorem shows that the sequence of solutions u_k generated by (27), if converge, it converges to a minimizer of KL-divergence fidelity model (14). We first present some lemmas to reveal the relation between the solutions of different models. In the following, we will refer to the KL-divergence fidelity model (14) replacing \mathcal{D} by \mathcal{W} .

Lemma 3.1 *Under the assumptions 1, we have that for $u \geq 0$, $Au = 0$ if and only if $u = 0$.*

Proof. By the assumption, $\sum_j A_{ij} > 0, \forall i$, we have $(Au)_i > 0$ for any $u \geq 0$ but $u \neq 0$. Hence $Au = 0$ has trivial non-negative solution. ■

Lemma 3.2 *The KL-divergence model (14) exist a unique minimizer.*

Proof. See [7] Theorem 4.8, 4.9. ■

Lemma 3.3 *The model (25) for a given u^* has a unique minimizer.*

Proof. We can see that the minimizers of the problem (25) exists due to the usual coercivity of the objective function. Furthermore, the minimizer is unique due to the assumption $Au = 0$ if and only if $u = 0$ in nonnegative orthant. In fact, since A has trivial null space in nonnegative

orthant, and $Au + c > 0$ for any $u \geq 0$, thus the objective function is strict convex. Therefore we obtain the uniqueness of the solution. ■

In the following, we will denote the unique minimizer of (14) as \hat{u} and show that the sequence generated by (27), if converge, converges to \hat{u} .

Theorem 3.4 *Let (u_k, d_k, b_k) be the sequence generated by the algorithm (27). If (u_k, d_k, b_k) converges to $(\tilde{u}, \tilde{d}, \tilde{b})$, then $\tilde{u} = \hat{u}$.*

Proof. If (u_k, d_k, b_k) generated by (27) converges to $(\tilde{u}, \tilde{d}, \tilde{b})$, we immediately get

$$\begin{aligned} A^T \tilde{\Sigma}^{-1}(A\tilde{u} + c - f) + \mu W^T \tilde{b} - \tilde{q} &= 0 \\ \lambda \tilde{p} - \mu \tilde{b} &= 0 \quad \text{with } \tilde{p} \in \partial \|\tilde{d}\|_1 \\ W\tilde{u} &= \tilde{d} \end{aligned} \tag{28}$$

where $\tilde{q} \geq 0$ satisfies $\langle \tilde{q}, \tilde{u} \rangle = 0$, and $\tilde{\Sigma} = \text{diag}(A\tilde{u} + c)$. Therefore

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

Comparing it with the first order condition of the model (14)

$$0 = A^T \left(\frac{A\hat{u} + c - f}{A\hat{u} + c} \right) + \lambda W^T \hat{p} - \hat{q} \tag{29}$$

where $\hat{p} \in \partial \|\hat{d}\|_1$ with $\hat{d} = W\hat{u}$ and $\hat{q} \geq 0$ is a lagrangian multiplier such that $\langle \hat{q}, \hat{u} \rangle = 0$. We can see that $\tilde{u} = \hat{u}$. ■

3.4 Acceleration

As recently proposed in [19], we show that Algorithm 1 can be accelerated using optimal first order methods, namely Nesterov method [24]. We refer to Algorithm 2 as the accelerated version of Algorithm 1. Note that the convergence of the accelerated algorithm for alternating direction multiplier method (ADMM) is demonstrated under some assumptions. In our reweighed case, it is not hard to see that Algorithm 2, if converges, converges to the same solution as Algorithm 1, based on the similar arguments in Theorem 3.4.

Algorithm 2 Accelerated Weighted ℓ^2 Fidelity Denoising-Deblurring Algorithm

Step0. Set the initial value, $u_0 = \mathbf{0}$; $d_0 = \tilde{d}_1 = \mathcal{W}f$; $b_0 = \tilde{b}_1 = \mathbf{0}$; $\alpha_0 = 1$; Σ_0

Step1.

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

$$u_k = \underset{u \geq 0}{\text{argmin}} \frac{1}{2} \|Au + c - f\|_{(\Sigma_{k-1})^{-1}}^2 + \frac{\mu}{2} \|\mathcal{W}u - \tilde{d}_k + \tilde{b}_k\|_2^2 + \frac{\gamma_1}{2} \|u - u_{k-1}\|_2^2$$

$$d_k = \text{sign}(\mu(\mathcal{W}u_k + \tilde{b}_k) + \gamma_2 d_{k-1}) / (\mu + \gamma_2) \cdot \max(|\mu(\mathcal{W}u_k + \tilde{b}_k) + \gamma_2 d_{k-1}| - \lambda / (\mu + \gamma_2), 0)$$

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

$$b_k = \tilde{b}_k + \mathcal{W}u_k - d_k$$

$$\alpha_{k+1} = \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2}$$

$$\tilde{d}_{k+1} = d_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (d_k - d_{k-1})$$

$$\tilde{b}_{k+1} = b_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (b_k - b_{k-1})$$

$$k = k + 1$$

end while

3.5 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-born noise that can be approximated by AWGN, as considered in [31]. The literature on Poisson-Gaussian mixed noise are limited despite of its importance. We now explore the extended application of the reweighted ℓ^2 fidelity to mixed Gaussian-Poisson noise case. With a small modification of the Poisson noise version, weighted ℓ^2 fidelity can be adapted to the mixed noise case.

Let σ^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 PDF of f again by normal distribution (19) with covariance matrix $\Sigma = \text{diag}(Au + c) + \sigma^2 I$, we obtain the new fidelity term as followed for mixed noise:

$$F(u) = \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au + c + \sigma^2 I}} \right\|_2^2 \quad (30)$$

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

$$\min_{u \geq 0} \frac{1}{2} \left\| \frac{Au + c - f}{\sqrt{Au + c + \sigma^2 I}} \right\|_2^2 + \lambda \|\mathcal{W}u\|_1 \quad (31)$$

The algorithm solving this model is the same as Algorithm 1 except adding σ^2 in estimation and updating of covariance matrix Σ .

4 Numerical results

In this section, numerical experiments on Poisson-Gaussian mixed noise models (31) will be shown for synthesized and real digital photos for demonstrating the performance of the proposed Algorithm (1) and (2). In all implementations, we choose the piecewise linear B-spline wavelets with 2 to 3 levels of decomposition. The corresponding filter masks in discrete version are $h_0 = [\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$, $h_1 = [-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}]$, $h_2 = [\frac{\sqrt{2}}{4}, 0, -\frac{\sqrt{2}}{4}]$.

For the initial guess of the variance matrix Σ_0 , although the observed data f is an acceptable estimation, we use a preprocessed data as a closer approximation to $Au + c$. In particular, we apply a bilateral filter [33] to preprocess the noised data f . This processing can be used for the first few steps of iteration till the sequence of solutions is stabilized. The bilateral filter \mathcal{B} of an 2D digital image x is defined as

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

where $[i, j], [p, q]$ are indices of pixels,

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

and $G_{\sigma_s}, G_{\sigma_r}$ are Gaussian functions with variance σ_s and σ_r respectively.

Given a $M \times N$ reference image u , the peak signal-to-noise ratios (PSNR) value of image \tilde{u} to u is defined by

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

where u_{\max} and u_{\min} are the maximum and minimum respectively of the reference image u , and M, N are the size of 2D image. It is a quantitative measure of image quality as \tilde{u} being the denoised result and u the ground truth.

4.1 Poisson denoise

For image denoising, the linear operator A reduces to the identity operator I . We compare first the three fidelities through their corresponding denoising method: the reweighted ℓ^2 method (27) with Algorithm 1, the KL-divergence model (14) with the algorithm (15) and the Anscombe transform model (17). 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 'poissrnd'. The parameters in each algorithm are tuned to get the best visual outcome for one simulated noisy image; they are then fixed when applied on the rest images. For the denoised results, we also add a post procedure of bilateral filter to remove some obvious artifacts.

The comparisons are performed in Figure 1 on several test images. The denoising results of the reweighted ℓ^2 model and that of the KL-divergence model are more or less the same both qualitatively and in terms of PSNR. The Anscombe transform model gives a slightly poorer result than the other two models. There are some artifacts in the high intensity part of the denoised image which may be caused by the post inverse Anscombe transform.

4.2 Deblurring from Poisson data

The deblurring case is in general more difficult than denoising. We compare our reweighted ℓ^2 method and the KL-divergence model (14) since Ancombe transform can not be easily applied in this case. For this application, we compare both Algorithm 1 and Algorithm 2 as the direct version and the accelerated version respectively. The EM+ ℓ^1 algorithm is applied for the KL-divergence model (14) as the denoising case. 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 2, the result of the KL-divergence fidelity algorithm is more blurry in contrast to the result of weighted ℓ^2 fidelity, although they have similar PSNR, see Table 1 for more comparisons.

	reweighted ℓ^2	KL-divergence
goldhill	26.48	26.29
cameraman	26.28	25.94
boat	25.37	25.16

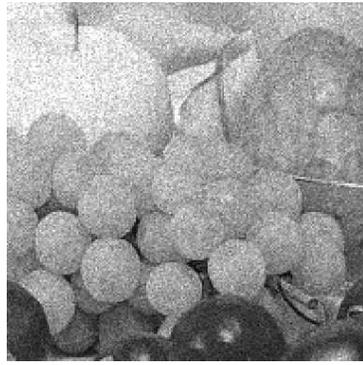
Table 1: PSNR value of the deblurring result.

In Figure 3, we compare the energy evolution of different algorithm. As we shown previously, Algorithm 1 and 2, if converge, converge to the solution of the model (14). We thus compare the KL-divergence model $E(u) = \mathbf{1}^T(Au + c) - f^T \log(Au + c) + \lambda \|Wu\|_1$ of the three algorithms (Algorithm 1, the accelerated Algorithm 2, and EM- ℓ^1). We can see that the energy for both Algorithm 1 and 2 decrease with the iterations, and Algorithm 2 has faster speed of convergence numerically. The evolution of weighted ℓ^2 -norm energy $E(u) = \frac{1}{2} \left\| \frac{Au+c-f}{\sqrt{Au^*+c}} \right\|_2^2 + \lambda \|Wu\|_1$ is also shown in the right figure of Fig. 3, where u^* is set as the ground truth image. Both reweighted based algorithms have decreasing energy, while EM- ℓ^1 has a divergence energy for this fidelity. To reach the same PSNR level of the result given by the direct algorithm after 100 iterations, the accelerated Algorithm 2 needs less iterations (around 30 steps), See Fig. 4.

In terms of computation time, the three denoising algorithms run generally for 20 to 30s. In the deblurring case, the reweighted algorithms (1) needs around 100s while the EM- ℓ^1 around 240s for the same relative error stopping criteria.



cameraman noisy



fruits noisy



goldhill noisy



PSNR: 31.23dB



PSNR: 31.17dB



PSNR: 29.49dB



PSNR: 30.72dB



PSNR: 30.83dB



PSNR: 29.07dB



PSNR: 31.00dB



PSNR: 31.08dB



PSNR: 29.35dB

Figure 1: Denoising results of simulated noisy images with peak intensity 120. From up to down: noisy image, reweighted ℓ^2 method by Algorithm 1, Anscombe transform (17) and KL-divergence model with EM- ℓ^1 algorithm (15).



Figure 2: Deblurring results of simulated noisy-blurred images with peak intensity 1200. from left to right: noisy-blurred image, direct reweighted ℓ^2 algorithm 1 and KL-divergence model (14). From top to bottom: full images and zoomed out regions.

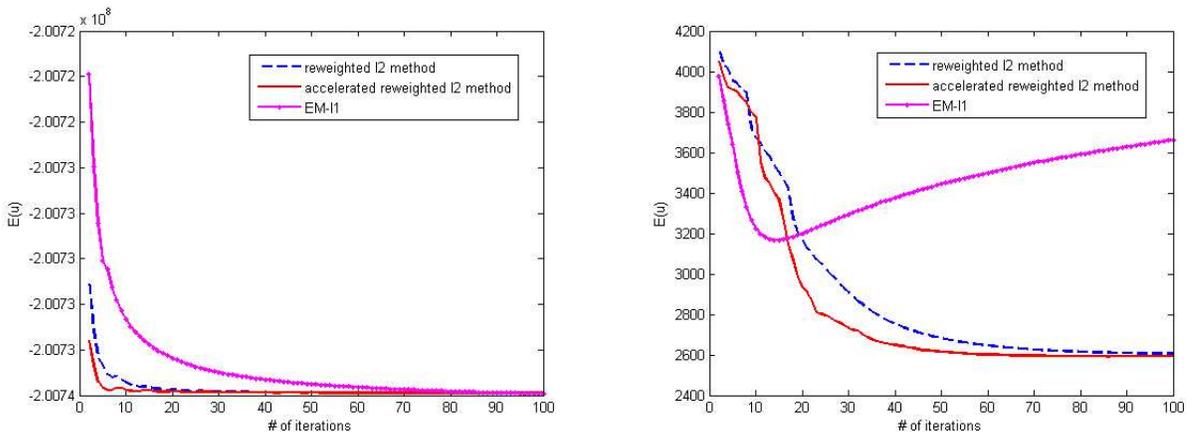


Figure 3: Left: KL-divergence ($E(u) = \mathbf{1}^T(Au + c) - f^T \log(Au + c) + \lambda \|Wu\|_1$) energy evolution of Algorithm 1, 2 and (15). Right: reweighted ℓ^2 -norm energy ($E(u) = \frac{1}{2} \left\| \frac{Au+c-f}{\sqrt{Au^*+c}} \right\|_2^2 + \lambda \|Wu\|_1$) evolution of Algorithm 1, 2 and (15).

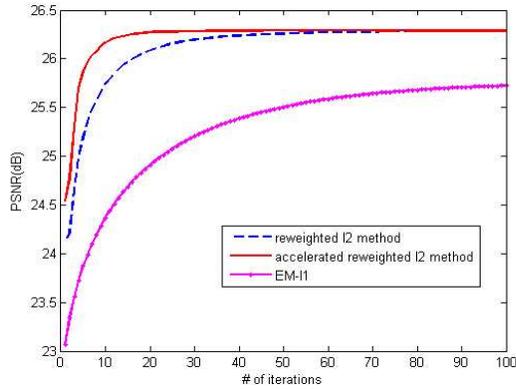


Figure 4: PSNR evolution of three algorithms: Algorithm 1, Algorithm 2 and EM- ℓ^1 (15).

Finally, the weighted ℓ^2 and the KL-divergence denoising model have superior results than the Anscombe transform model; while in deblurring case, the reweighted ℓ^2 outperforms the KL-divergence one. From these observations, our reweighted ℓ^2 fidelity term is a competitive potential choice when working with complex noisy models.

4.3 Mixed Poisson-Gaussian noise restoration

In this section, we test the extension of reweighted ℓ^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. 5 shows the denoising result of model (31) with A being identity operator.



Figure 5: Left: input noisy image with mixed Poisson-Gaussian noise, peak intensity: 120, PSNR: 18.59dB; right: denoised by model (31), PSNR: 27.42dB.

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 6, 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. 7 for the deblurring result of model (31).



Figure 6: Top: real photo; middle: output of camera using the noise reduction option. bottom: denoised by model (31).



disk,3 ; PSNR:22.61dB



PSNR:26.00dB



disk,5; PSNR:20.83dB



PSNR:24.51dB

Figure 7: left: input blurred image with mixed Poisson-Gaussian noise, $\sigma = 12$, peak intensity: 1200; right: deblurred by model (31).

5 Conclusion

In this paper, we studied weighted ℓ^2 fidelity term derived from Gaussian approximation of Poisson statistics. In solving denoising and deblurring models, we proposed a reweighted algorithm based on Split Bregman iteration and its Nesterov accelerated one. The algorithms have the same limit solution as the minimizer of KL-divergence fidelity model (14) if they converge. The proposed algorithms give competitive result with respect to classical fidelity terms in denoising and deblurring simulations. In addition, the reweighted ℓ^2 model can be easily extended to the mixed Poisson-Gaussian noise case. As shown by numerical results, the reweighted ℓ^2 with split Bregman framework is promising in a wide perspective.

Acknowledgments

Xiaoqun Zhang was partially supported by the NNSFC (Grant nos. 11101277, 11161130004) and the Shanghai Pujiang Talent Program (Grant no. 11PJ1405900). Zuwei Shen was partially supported by National University of Singapore (Grant no. MOE2011-T2-1-116).

References

- [1] F. J. Anscombe. *The Transform of Poisson, Binomial and Negative-Binomial Data*. Biometrika, **35** (1948).
- [2] M. Bertero, P. Boccacci, G. Desidera and G. Vicidomini. *Image deblurring with Poisson data: from cells to galaxies*. Inverse Problems, **12(25)**(2009).
- [3] M. Bertero, P. Boccacci, G. Talenti, R. Zanella and L. Zanni. *A discrepancy principle for Poisson data*. Inverse Problems, **26(10)** (2010).
- [4] P. J. Bickel and E. Levina. *Regularized estimation of large covariance matrices*. Ann. Statistics, **36** (2008),199–227.
- [5] C. Brune, A. Sawatzky and M. Burger. *Bregman-EM-TV Methods with Application to Optical Nanoscopy*. In Proc. SSVM 2009, LNCS, Springer, 5567 (2009).
- [6] C. Brune, A. Sawatzky and M. Burger. *Primal and Dual Bregman Methods with Application to Optical Nanoscopy*. Int J Comput Vis, **72** (2011), 211–229.
- [7] C. Brune, M. Burger, A. Sawatzky, T. Kösters and Frank Wübblerling. *Forward-Backward EM-TV methods for inverse problems with Poisson noise*. Preprint, August 2009.
- [8] J.-F. Cai, S. Osher and Z. Shen. *Split Bregman Methods and Frame Based Image Restoration*. Multiscale Modeling & Simulation, **8(2)** (2009) , 337-369.
- [9] J.-F. Cai, B. Dong, S. Osher and Z. Shen *Image Restoration: Total Variation; Wavelet Frames; and beyond* J. Amer. Math. Soc., **25(4)** (2012), pp. 1033-1089.
- [10] Imre Csiszár. *Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems*. Ann. Statist. **19(4)** (1991) , 2032-2066.
- [11] I. Daubechies, B. Han, A. Ron and Z. Shen. *Framelets: MRA-based constructions of wavelet frames*. Applied and Computational Harmonic Analysis, **14**(2003), 1-46.

- [12] B. Dong and Z. Shen. *MRA-based wavelet frames and applications, IAS Lecture Notes Series*. Park City Mathematics Institute, expected to appear in 2012 (manuscript revised on Sep 28, 2011).
- [13] B. Dong and Y. Zhang. *An efficient algorithm for l^0 minimization in wavelet frame based image restoration*. Journal of Scientific Computing, to appear.
- [14] E. Esser. *Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman*. UCLA CAM report 09-31,2009.
- [15] E. Esser, X. Zhang X and T.-F. Chan. *A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science*. SIAM J. Imag. Sci. **3(4)** (2010) 1015-1046.
- [16] M. Figueiredo and J. Bioucas-Dias. *Restoration of Poissonian images using alternating direction optimization*. IEEE Transactions on Image Processing, to appear
- [17] F. Luisier, T. Blu and M. Unser. *Image denoising in mixed Poisson-Gaussian noise*. Image restoration, IEEE Transactions, **20** (2011), 696-708.
- [18] J. Friedman, T. Hastie and R. Tibshirani. *Sparse inverse covariance estimation with the graphical lasso*. Biostatistics,**9**(2008), 432-441.
- [19] T. Goldstein, Brendan O'Donoghue, and Simon Setzer, *Fast Alternating Direction Optimization Methods*. UCLA CAM report 12-35. May 2012.
- [20] T. Goldstein and S. Osher. *The split bregman method for l^1 regularized problems*. SIAM Journal on Imaging Sciences, **2(2)** (2009), 323-343 .
- [21] Z. Gong, Z. Shen and K.-C. Toh. *Image Restoration with Mixed or Unknown Noises*. preprint, private communication.
- [22] K. Lange and R. Carson. *EM reconstruction algorithms for emission and transmission tomography*. J Comput Assist Tomogr. 8(2), pp.306-16 (1984).
- [23] F. Luisier, T. Blu and M. Unser *Image denoising in mixed Poisson-Gaussian noise* IEEE Transactions on Image Processing, vol. 20, no. 3, pp. 696-708, March 2011
- [24] Y. Nesterov. *A method of solving a convex programming problem with convergence rate $o(1/k^2)$* . Soviet Math. Dokl., **27** (1983) :372C376.
- [25] R. Rockafellar, *Augmented lagrangians and applications of the proximal point algorithm in convex programming*. Mathematics of Operations Research. **1(2)** (1976). 97-116.
- [26] A. Ron, Z. Shen. *Affine Systems in $L_2(\mathbb{R}^d)$: The Analysis of the Analysis Operator*. Journal of Functional Analysis, **148(2)**(1997), 408-447.
- [27] L.I. Rudin, S. Osher and E. Fatemi *Nonlinear total variation based noise removal algorithms* Physics D, **60** (1992), 259-268.
- [28] S. Setzer *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, in SSVM '09 Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision, 2008.
- [29] S. Setzer, G. Steidl, and T. Teuber. *Deblurring Poissonian images by split Bregman techniques*. Journal of Visual Communication and Image Representation, **21** (2010), pp. 193 - 199.

- [30] L. A. Shepp and Y. Vardi. *Maximum Likelihood Reconstruction in Positron Emission Tomography*. IEEE Transactions on Medical Imaging, **1** (1982), 113-122.
- [31] D. L. Snyder, A. M. Hammoud, and R. L. White. *Image recovery from data acquired with a charge-coupled-device camera*. J. Opt. Soc. Am. A **10**(1983), 1014-1023.
- [32] A. Staglianò, P. Boccacci and M. Bertero. *Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle*. Inverse Problems **27** (2011). doi:10.1088/0266-5611/27/12/125003.
- [33] C. Tomasi and R. Manduchi *Bilateral filtering for gray and color images* Sixth International Conference on Computer Vision (1998).
- [34] X. Zhang, M. Burger and S. Osher. *A Unified Primal-Dual Algorithm Framework Based on Bregman Iteration*. Journal of Scientific Computing, **46 (1)** (2010), 20-46.