

A PRIMAL-DUAL PROXIMAL SPLITTING APPROACH FOR RESTORING DATA CORRUPTED WITH POISSON-GAUSSIAN NOISE

Anna Jezierska, Emilie Chouzenoux, Jean-Christophe Pesquet and Hugues Talbot

Université Paris-Est, Lab. Informatique Gaspard Monge, UMR CNRS 8049,
Champs-sur-Marne, 77454 Marne-la-Vallée, France
`{first.last}@univ-paris-est.fr`

ABSTRACT

A Poisson-Gaussian model accurately describes the noise present in many imaging systems such as CCD cameras or fluorescence microscopy. However most existing restoration strategies rely on approximations of the Poisson-Gaussian noise statistics. We propose a convex optimization algorithm for the reconstruction of signals degraded by a linear operator and corrupted with mixed Poisson-Gaussian noise. The originality of our approach consists of considering the exact continuous-discrete model corresponding to the data statistics. After establishing the Lipschitz differentiability of the Poisson-Gaussian log-likelihood, we derive a primal-dual iterative scheme for minimizing the associated penalized criterion. The proposed method is applicable to a large choice of penalty terms. The robustness of our scheme allows us to handle computational difficulties due to infinite sums arising from the computation of the gradient of the criterion. The proposed approach is validated on image restoration examples.

Index Terms— convex optimization, image restoration, denoising, deconvolution.

1. INTRODUCTION

Noise in many real imaging systems can be accurately described through a mixed Poisson-Gaussian model. For example, it is frequently encountered in astronomy [1, 2], medicine [3] and biology [4]. There has been a growing interest for denoising problems involving images corrupted in this fashion [5, 6, 7]. However, the literature involving this model together with deconvolution and reconstruction problems remains scarce. Among existing works, Benvenuto *et al.* [1] proposed a scaled gradient method and more recently Gil-Rodrigo *et al.* [8] developed an alternating minimization algorithm.

Generally, restoration strategies are grounded on some approximations of the noise statistics, which may be detrimental to the quality of the results. The use of approximations

THIS WORK WAS SUPPORTED BY THE AGENCE NATIONALE DE LA RECHERCHE UNDER GRANT ANR-09-EMER-004-03.

is motivated by the mathematical difficulties raised by the Poisson-Gaussian model. Indeed, the corresponding probability distribution has a discrete-continuous nature, and the expression of the associated log-likelihood function involves an infinite sum. For simplification, one usually neglects either the Poisson or the Gaussian component, or performs an approximation of the Poisson-Gaussian model based on variance stabilization techniques [9, 10].

In this paper, we investigate the properties of the Poisson-Gaussian negative log-likelihood, showing that it is a convex Lipschitz differentiable function. Since the gradient of the Poisson-Gaussian log-likelihood requires the computation of infinite series, we propose utilizing proximal optimization methods, since their convergence is guaranteed even in the presence of summable numerical errors. Among recent approaches, only a few primal-dual splitting algorithms [11, 12] can cope with the sum of a gradient Lipschitz term and a possibly non-smooth penalty term. Such a term can model a wide range of prior information, e.g. criteria promoting sparsity in a frame, total-variation and more generally hybrid regularization functions.

The paper is organized as follows: Section 2 investigates the Poisson-Gaussian model and introduces the notation used in this work. Section 3 describes the proposed optimization framework. Our approach is illustrated via experiments in Section 4. Finally, some conclusions are drawn in Section 5.

2. DEGRADATION MODEL

Let $\mathbf{y} \in \mathbb{R}^Q$ be a vector of observations related to an original signal $\mathbf{x} \in [0, +\infty)^N$ through the model $\mathbf{y} = \mathbf{z}(\mathbf{x}) + \mathbf{w}$, where $\mathbf{z}(\mathbf{x}) = (z_i(\mathbf{x}))_{1 \leq i \leq Q}$ and $\mathbf{w} = (w_i)_{1 \leq i \leq Q}$ are realizations of mutually independent random vectors $\mathbf{Z}(\mathbf{x}) = (Z_i(\mathbf{x}))_{1 \leq i \leq Q}$ and $\mathbf{W} = (W_i)_{1 \leq i \leq Q}$ having independent components. It is further assumed that, for every $i \in \{1, \dots, Q\}$,

$$Z_i(\mathbf{x}) \sim \mathcal{P}_\alpha([\mathbf{H}\mathbf{x}]_i) \quad (1)$$

$$W_i \sim \mathcal{N}(b, \sigma^2), \quad (2)$$

where \mathbf{H} is a matrix in $[0, +\infty)^{Q \times N}$ modeling the degradation process (e.g., a convolution operator), $\alpha > 0$ is the scaling parameter of Poisson noise and $(b, \sigma) \in \mathbb{R} \times (0, +\infty)$ are the mean and standard deviation, respectively, of the Gaussian noise component. Hence, \mathbf{y} is a realization of a random vector \mathbf{Y} with probability density function

$$p_{\mathbf{Y}}(\mathbf{y}; \mathbf{x}) = \prod_{i=1}^Q \left(\sum_{n=0}^{+\infty} \frac{e^{-\alpha[\mathbf{H}\mathbf{x}]_i} (\alpha[\mathbf{H}\mathbf{x}]_i)^n}{n!} e^{-\frac{1}{2\sigma^2}(y_i - b - n)^2} \right) \quad (3)$$

In the context of inverse problems, the original signal can be recovered by minimizing a penalized criterion:

$$\min_{\mathbf{x} \in \mathbb{R}^N} (f(\mathbf{x}) = g(\mathbf{x}) + r(\mathbf{x})), \quad (4)$$

where g is the so-called data fidelity term and r is a regularization function incorporating *a priori* information, so as to guarantee the stability of the solution w.r.t. the observation noise. In the Bayesian framework, this allows us to compute the *maximum a posteriori* (MAP) estimate [13] of the original image. In this context, the data fidelity term is defined as the negative logarithm of $p_{\mathbf{Y}}(\mathbf{y}; \mathbf{x})$:

$$\begin{aligned} g(\mathbf{x}) &= -\log(p_{\mathbf{Y}}(\mathbf{y}; \mathbf{x})) \\ &= -\sum_{i=1}^Q \log \left(\sum_{n=0}^{+\infty} \frac{e^{-\alpha[\mathbf{H}\mathbf{x}]_i} (\alpha[\mathbf{H}\mathbf{x}]_i)^n}{n!} e^{-\frac{1}{2\sigma^2}(y_i - b - n)^2} \right) \end{aligned} \quad (5)$$

and the regularization term r corresponds to the potential of the chosen prior probability distribution. The gradient of g on the positive orthant is given by

$$(\forall \mathbf{x} \in [0, +\infty)^N) \quad \nabla g(\mathbf{x}) = \alpha \mathbf{H}^\top (\mathbf{1} - \mathbf{u}(\alpha \mathbf{H}\mathbf{x})) \quad (6)$$

where, for every $\boldsymbol{\xi} = (\xi_i)_{1 \leq i \leq Q} \in [0, +\infty)^Q$, $\mathbf{u}(\boldsymbol{\xi}) = (s(\xi_i, y_i - b - 1)/s(\xi_i, y_i - b))_{1 \leq i \leq Q}$, and, for every $(\xi, v) \in \mathbb{R}^2$,

$$s(\xi, v) = \sum_{n=0}^{+\infty} \frac{\xi^n}{n!} e^{-\frac{1}{2\sigma^2}(v-n)^2}. \quad (7)$$

Based on the previous expressions, the following result can be proved.

Theorem 1. *The function g is convex and μ -Lipschitz differentiable on $[0, +\infty)^N$ with*

$$\mu = \alpha^2 \|\mathbf{H}\|^2 \left(1 - e^{-\frac{1}{\sigma^2}} \right) \max_{i \in \{1, \dots, Q\}} e^{\frac{2(y_i - b) - 1}{\sigma^2}}. \quad (8)$$

While a proof of the convexity property is provided in [1], the Lipschitz differentiability of g is a novel result.

For the optimization methods that are developed in the next section, it is also important to note that the definition of the negative log-likelihood can be extended to the whole space \mathbb{R}^N by setting

$$g(\mathbf{x}) = h(\mathbf{x}) + \iota_{[0, +\infty)^N}(\mathbf{x}), \quad (9)$$

where

$$h(\mathbf{x}) = \sum_{i=1}^Q \varphi_i(\alpha[\mathbf{H}\mathbf{x}]_i) \quad (10)$$

$$\iota_{[0, +\infty)^N}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in [0, +\infty)^N \\ +\infty & \text{otherwise.} \end{cases} \quad (11)$$

Here, for every $i \in \{1, \dots, Q\}$, $\varphi_i: \mathbb{R} \rightarrow \mathbb{R}$ is a convex, twice-differentiable function, whose expression is readily derived from (5) for non-negative values of its arguments, and which takes a quadratic form on $(-\infty, 0]$. By appropriately choosing the quadratic form, h is a convex function with a μ -Lipschitz gradient on \mathbb{R}^N .

3. PROPOSED OPTIMIZATION METHOD

3.1. Minimization problem

According to the analysis carried out in Section 2, the objective function takes the following form

$$f(\mathbf{x}) = h(\mathbf{x}) + r_0(\mathbf{x}) + \sum_{m=1}^M r_m(\mathbf{L}_m \mathbf{x}), \quad (12)$$

where the regularization term has been split in a sum of simpler functions. More precisely, it will be assumed that $r_0 \in \Gamma_0(\mathbb{R}^N)$ and, for every $m \in \{1, \dots, M\}$, $r_m \in \Gamma_0(\mathbb{R}^{P_m})$ and $\mathbf{L}_m \in \mathbb{R}^{P_m \times N}$.¹ Note that (12) covers a large range of penalization strategies. For instance, a sparsity prior in an analysis frame with frame operator \mathbf{L}_m is introduced by taking r_m equal to $\lambda_m \|\cdot\|_1$ with $\lambda_m > 0$. Block sparsity measures [14] can also be easily addressed in the proposed framework. Another popular example in image restoration is the total variation penalization [15]. In this case, $P_m = 2N$, $\mathbf{L}_m = [(\Delta^h)^\top \ (\Delta^v)^\top]^\top$, where $\Delta^h \in \mathbb{R}^{N \times N}$ (resp. $\Delta^v \in \mathbb{R}^{N \times N}$) corresponds to a horizontal (resp. vertical) gradient operator, and, for every $\mathbf{x} \in \mathbb{R}^N$, $r_m(\mathbf{L}_m \mathbf{x}) = \lambda_m \sum_{n=1}^N \left(([\Delta^h \mathbf{x}]_n)^2 + ([\Delta^v \mathbf{x}]_n)^2 \right)^{1/2}$ with $\lambda_m > 0$. The above penalties can be considered individually ($M = 1$) or combined in a hybrid manner ($M > 1$) [16]. Finally, following (9), r_0 should be equal to $\iota_{[0, +\infty)^N}$. However, to take into account the dynamic range of the expected output image, it can be more generally chosen equal to ι_C , where C is a closed convex subset of $[0, +\infty)^N$.

3.2. Primal-dual splitting algorithm

We are now ready to present our primal-dual splitting algorithm. We first require the notion of proximity operator.

¹ $\Gamma_0(\mathbb{R}^N)$ is the class of lower-semicontinuous, proper, convex functions from \mathbb{R}^N to $(-\infty, +\infty]$.

Definition 1. Let $\psi \in \Gamma_0(\mathbb{R}^N)$. For every $\mathbf{x} \in \mathbb{R}^N$, the minimization problem

$$\min_{\mathbf{y} \in \mathbb{R}^N} \psi(\mathbf{y}) + \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 \quad (13)$$

admits a unique solution, which is denoted by $\text{prox}_\psi(\mathbf{x})$. The so-defined operator $\text{prox}_\psi: \mathbb{R}^N \rightarrow \mathbb{R}^N$ is the proximity operator of ψ .

Numerous convex optimization algorithms are based on this notion (see [17] for a tutorial). Problem (4) where f takes the form (12) can be efficiently addressed using proximal splitting algorithms [18, 19, 11, 12]. The solution is obtained iteratively by evaluating the individual proximity operators of the functions $(r_m)_{0 \leq m \leq M}$, provided that they have an explicit expression. The main advantage of the primal-dual splitting algorithm that we employ is that it allows us to solve (4) for any Lipschitz differentiable function h while allowing for arbitrary linear operators $(\mathbf{L}_m)_{1 \leq m \leq M}$. This algorithm, proposed recently in [11], is summarized below.

Algorithm 1 Proposed algorithm.

Let $\gamma \in (0, +\infty)$. Let $(\mathbf{a}_k)_{k \in \mathbb{N}}$ and $(\mathbf{c}_k)_{k \in \mathbb{N}}$ be some sequences of elements of \mathbb{R}^N corresponding to possible errors in the computation of the gradient of h .

Initialization:

Set $\mathbf{x}_0 \in \mathbb{R}^N$, and $(\forall m \in \{1, \dots, M\}) \mathbf{v}_{m,0} \in \mathbb{R}^{P_m}$.

Iterations:

For $k = 0, \dots$

$$\begin{cases} \mathbf{y}_{1,k} = \mathbf{x}_k - \gamma \left(\nabla h(\mathbf{x}_k) + \sum_{m=1}^M \mathbf{L}_m^\top \mathbf{v}_{m,k} \right) + \mathbf{a}_k \\ \mathbf{p}_{1,k} = \text{prox}_{\gamma r_0}(\mathbf{y}_{1,k}) \\ \text{For } m = 1, \dots, M \\ \quad \mathbf{y}_{2,m,k} = \mathbf{v}_{m,k} + \gamma \mathbf{L}_m \mathbf{x}_k \\ \quad \mathbf{p}_{2,m,k} = \mathbf{y}_{2,m,k} - \gamma \text{prox}_{\gamma^{-1} r_m}(\gamma^{-1} \mathbf{y}_{2,m,k}) \\ \quad \mathbf{q}_{2,m,k} = \mathbf{p}_{2,m,k} + \gamma \mathbf{L}_m \mathbf{p}_{1,k} \\ \quad \mathbf{v}_{m,k+1} = \mathbf{v}_{m,k} - \mathbf{y}_{2,m,k} + \mathbf{q}_{2,m,k} \\ \quad \mathbf{q}_{1,k} = \mathbf{p}_{1,k} - \gamma \left(\nabla h(\mathbf{p}_{1,k}) + \sum_{m=1}^M \mathbf{L}_m^\top \mathbf{p}_{2,m,k} \right) + \mathbf{c}_k \\ \quad \mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{y}_{1,k} + \mathbf{q}_{1,k} \end{cases}$$

3.3. Convergence result

The convergence of the proposed primal-dual proximal splitting algorithm is guaranteed by the following result deduced from Theorem 1 and [11, Theorem 4.2]:

Theorem 2. Given the following assumptions:

- (i) f is coercive, i.e. $\lim_{\|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty$,
- (ii) for every $m \in \{1, \dots, M\}$, r_m is finite valued,
- (iii) $\gamma \in [\epsilon, (1 - \epsilon)/\beta]$ where $\epsilon \in (0, 1/(\beta + 1))$ and

$$\beta = \mu + \sqrt{\sum_{m=1}^M \|\mathbf{L}_m\|^2},$$

- (iv) $(\mathbf{a}_k)_{k \in \mathbb{N}}$ and $(\mathbf{c}_k)_{k \in \mathbb{N}}$ are absolutely summable sequences,

then there exists a minimizer $\bar{\mathbf{x}}$ of (12) such that the sequences $(\mathbf{x}_k)_{k \in \mathbb{N}}$ and $(\mathbf{p}_{1,k})_{k \in \mathbb{N}}$ converge to $\bar{\mathbf{x}}$.

3.4. Implementation issues

Note that Algorithm 1 is robust to numerical errors. This feature is essential in our problem, as the gradient of the Poisson-Gaussian negative log-likelihood given by (6) involves infinite sums and cannot be computed exactly. We propose to perform the sum of (7) between n_{\min} and n_{\max} , to include only the significant coefficient indices. Using Stirling's formula leads us to $n_{\min} \simeq \max(\min(\xi, v) - \delta, 0)$ and $n_{\max} \simeq \max(\xi, v) + \delta$ where $\delta = 3 \max(\sigma, \sqrt{\xi})$.

4. SIMULATION EXAMPLES

We now demonstrate the practical performance of our method on image denoising and restoration experiments. The restoration involves the minimization of

$$f = h + \iota_C + \lambda \text{tv}, \quad (14)$$

where ι_C is the indicator function of $C = [0, 255]^N$, tv denotes the total variation semi-norm as defined in Section 3.1 and $\lambda > 0$ is the regularization parameter. To reflect the Poisson-Gaussian noise statistics, we choose the data fidelity term h to be either derived from the Generalized Anscombe Transform (GAST) or to be given by the exact expression in (5). The quality of the results is evaluated in terms of Mean Absolute Error (MAE). In our experiments, we use the standard house image of size 256×256 from (sipi.usc.edu/database/). To generate the observed image \mathbf{y} , we degraded the original image with a convolution operator \mathbf{H} , which reduces to identity in the pure denoising case and corresponds to a truncated Gaussian point spread function of standard deviation 0.5 and kernel size 7×7 in the case of restoration. The image was further corrupted with a scaled Poisson noise and a zero-mean additive Gaussian noise. Table 1 presents the results of three experiments, which differ in the values of the scaling parameter α of the Poisson noise and the variance σ^2 of the Gaussian noise. Experiment 1 is defined by $\alpha = 0.05$ and $\sigma^2 = 9$, Experiment 2 by $\alpha = 0.1$ and $\sigma^2 = 16$ and Experiment 3 by $\alpha = 0.4$ and $\sigma^2 = 50$. For each model, parameter λ was adjusted to achieve the minimum MAE under all experimental conditions.

Since our approach can cope with a wide range of data fidelity terms, we were able to compare its performance for the two models. GAST was handled in a manner similar to [20]. One can observe by inspecting the MAE values in Table 1, that the exact model leads to the best result. The difference is more significant for low-count images.

	Denoising			Restoration		
	Exp.1	Exp.2	Exp.3	Exp.1	Exp.2	Exp.3
Initial image	63.61	43.58	20.51	63.81	43.83	21.25
GAST model	10.84	7.97	5.19	10.98	8.05	5.60
Exact model	8.96	7.28	5.11	8.95	7.44	5.55

Table 1. Obtained MAE for different models.

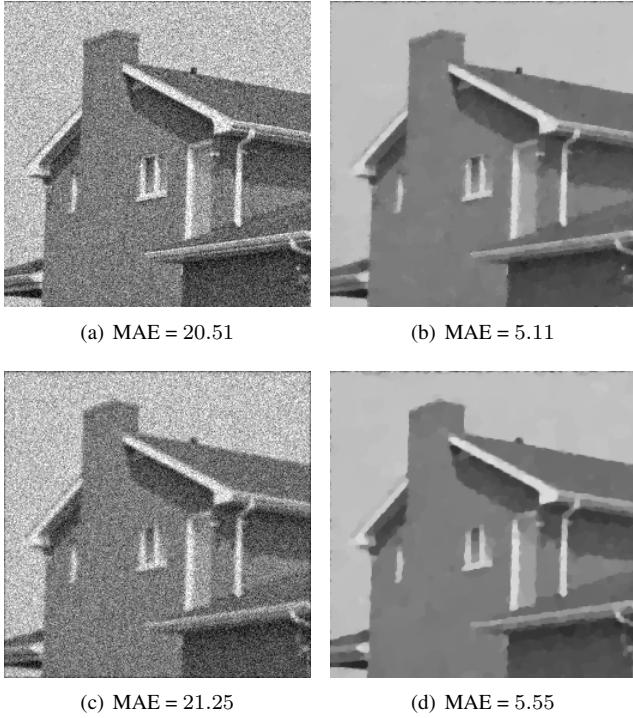


Fig. 1. (a,c) noisy image, and noisy blurred one, respectively. (b,d) resulting images using the exact Poisson-Gaussian model for denoising and restoration problems, respectively ($\alpha = 0.4$, $\sigma^2 = 50$).

5. CONCLUSION

We have proposed a new variational approach for solving data recovery problems in the presence of Poisson-Gaussian noise. Taking advantage of the convexity and Lipschitz-differentiability properties of the Poisson-Gaussian negative log-likelihood, an efficient primal-dual proximal algorithm has been developed. Although the proposed framework was applied to TV-based image restoration, its versatility allows us to address a wide range of applications by making use of various forms of convex penalty functions.

6. REFERENCES

- [1] F. Benvenuto, A. La Camera, C. Theys, A. Ferrari, H. Lantéri, and M. Bertero, “The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise,” *Inverse Problems*, vol. 24, no. 3, 2008.
- [2] D. L. Snyder, A. M. Hammoud, and R. L. White, “Image recovery from data acquired with a charge-coupled-device camera,” *Journal of the Optical Society of America A*, vol. 10, no. 5, pp. 1014–1023, May 1993.
- [3] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy, “Spatiotemporal reconstruction of list-mode PET data,” *IEEE Transactions on Medical Imaging*, vol. 21, no. 4, pp. 396–404, Apr. 2002.
- [4] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser, “Multiframe SURELET denoising of timelapse fluorescence microscopy images,” in *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, Paris, France, May 2008, pp. 149–152.
- [5] 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, Mar. 2011.
- [6] G. Boracchi and A. Foi, “Multiframe raw-data denoising based on block-matching and 3-D filtering for low-light imaging and stabilization,” in *The 2008 International Workshop on Local and Non-Local Approximation in Image Processing*, Lausanne, Switzerland, Aug. 2008.
- [7] B. Begovic, V. Stankovic, and L. Stankovic, “Contrast enhancement and denoising of Poisson and Gaussian mixture noise for solar images,” in *IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sep. 2011, pp. 185 – 188.
- [8] E. Gil-Rodrigo, J. Portilla, D. Miraut, and R. Suarez-Mesa, “Efficient joint Poisson-Gauss restoration using multi-frame ℓ_2 -relaxed- ℓ_0 analysis-based sparsity,” in *IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sep. 2011, pp. 1385 – 1388.
- [9] F. Murtagh, J.-L. Starck, and A. Bijaoui, “Image restoration with noise suppression using a multiresolution support,” *Astronomy and Astrophysics Supplement*, vol. 112, pp. 179–189, 1995.
- [10] A. Foi, “Clipped noisy images: Heteroskedastic modeling and practical denoising,” *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, Dec. 2009.
- [11] P. L. Combettes and J.-C. Pesquet, “Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators,” *Set-Valued and Variational Analysis*, pp. 1–24, 2011, 10.1007/s11228-011-0191-y.
- [12] H. Raguet, J. Fadili, and G. Peyré, “Generalized Forward-Backward splitting,” Tech. Rep., 2011, <http://hal.archives-ouvertes.fr/hal-00613637>.
- [13] G. Demont, “Image reconstruction and restoration: Overview of common estimation structure and problems,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 12, pp. 2024–2036, Dec. 1989.
- [14] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042 –3054, Jun. 2010.
- [15] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Journal Physica D*, vol. 60, pp. 259–268, 1992.
- [16] N. Pustelnik, C. Chaux, and J.-C. Pesquet, “Parallel proximal algorithm for image restoration using hybrid regularization,” *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2450 –2462, Sep. 2011.
- [17] P. L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185 – 212. Springer-Verlag, New York, 2010.
- [18] P. L. Combettes and J.-C. Pesquet, “A proximal decomposition method for solving convex variational inverse problems,” *Inverse Problems*, vol. 24, no. 6, Dec. 2008.
- [19] L. M. Briceños Arias and P. L. Combettes, “A monotone + skew splitting model for composite monotone inclusions in duality,” *SIAM Journal on Optimization*, vol. 21, no. 4, pp. 1230–1250, Oct. 2011.
- [20] F.-X. Dupe, J.M. Fadili, and J.-L. Starck, “A proximal iteration for deconvolving Poisson noisy images using sparse representations,” *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 310 –321, Feb. 2009.