

# Global Optimization for Recovery of Clipped Signals Corrupted with Poisson-Gaussian Noise

A. Marmin, A. Jezierska, M. Castella, *Member IEEE*, and J.-C. Pesquet, *Fellow IEEE*

**Abstract**—We study a variational formulation for reconstructing nonlinearly distorted signals corrupted with a Poisson-Gaussian noise. In this situation, the data fidelity term consists of a sum of a weighted least squares term and a logarithmic one. Both of them are precomposed by a nonlinearity, modelling a clipping effect, which is assumed to be rational. A regularization term, being a piecewise rational approximation of the  $\ell_0$  function provides a suitable sparsity measure with respect to a preset linear operator. We propose a global optimization approach for such a problem. More specifically, it is first transformed into a generalized moment problem by introducing some auxiliary variables. Then, a hierarchy of semidefinite programming relaxations is built. Numerical examples show the good performance of the proposed approach.

**Index Terms**—signal reconstruction, polynomial optimization, Poisson-Gaussian noise,  $\ell_0$  penalization, nonconvex models, Padé approximation

## I. INTRODUCTION

Over the last years, there has been a growing interest for signal reconstruction from measurements corrupted by Poisson-Gaussian (PG) noise. Examples of application areas are fluorescence microscopy [1], low dose computer tomography [2]–[4], and visible light communication (VLC) [5]. Most reconstruction methods rely on some approximations to PG statistics. Among them, the weighted least squares approximation is one of the most popular [6], [7]. In existing works, the log-likelihood is approximated by the log-likelihood of a Gaussian variable whose variance depends on the data model. The data fit term is then a sum of a weighted least squares plus a logarithm term, the latter being often omitted in order to ensure the convexity of the problem, and to simplify it. Recently in [2], the authors proposed to keep this term and to handle the resulting nonconvex optimization problem. The latter problem was addressed by employing the alternating direction method of multipliers which, in this nonconvex setting, provides only a guarantee to return a local minimizer.

In this letter, we propose a novel approach with global convergence properties. More precisely, our contributions consist in (i) proposing a new rational approximation to the PG data fidelity term, (ii) showing that rational approximations to the  $\ell_0$  sparsity measure introduced in [8] can be combined efficiently with a linear operator such as a discrete gradient, (iii) solving

A. Marmin and J.-C. Pesquet are with Université Paris-Saclay, Centrale-Supélec, Center for Visual Computing, Inria, Gif-sur-Yvette, France (e-mail: {arthur.marmin,jean-christophe.pesquet}@centralesupelec.fr).

A. Jezierska is with the Systems Research Institute Polish Academy of Sciences, 01-447 Warsaw, Poland and with Gdansk University of Technology, ETI, 80-233 Gdansk, Poland (e-mail: Anna.Jezierska@ibspan.waw.pl).

M. Castella is with SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France (e-mail: marc.castella@telecom-sudparis.eu).

the resulting nonconvex maximum a posteriori (MAP) problem with a method that guarantees the global optimality of the solution by using recent advances in polynomial optimization and moment problems, (iv) illustrating the good performance of the method on several practical examples.

Our paper is organized as follows: Section II introduces the considered data model as well as the general form of the addressed optimization problem. Section III reformulates the original optimization problem into a rational optimization one and describes how to solve it globally. Numerical simulations and results are presented in Section IV to validate our approach. Finally, Section V concludes the paper.

## II. PG NOISE MODEL AND OPTIMIZATION PROBLEM

### A. Considered model and problem

We consider the reconstruction of a discrete positive signal  $\bar{\mathbf{x}}$  of length  $T$  from an observation vector  $\mathbf{y}$ . The signal  $\bar{\mathbf{x}}$  is first degraded by a linear operator represented by a matrix with positive coefficients  $\mathbf{H}$  and then by a scalar nonlinear function  $\phi$  which takes positive values. Finally, the output signal is corrupted with a PG noise. For instance, matrix  $\mathbf{H}$  can model the convolution with a filter impulse response, while  $\phi$  can represent saturation effects such as clipping [9], [10]. Our model hence reads

$$(\forall t \in \llbracket 1, T \rrbracket) \quad y_t \sim \mathcal{N}(c + \mathcal{P}(\phi((\mathbf{H}\bar{\mathbf{x}})_t)), \sigma^2), \quad (1)$$

where  $(\mathbf{H}\bar{\mathbf{x}})_t$  denotes the  $t$ -th component of the vector  $\mathbf{H}\bar{\mathbf{x}}$ ,  $\mathcal{P}$  and  $\mathcal{N}$  denote respectively the PG distributions,  $c$  is a nonnegative constant modelling the average background noise, and  $\sigma^2$  is the variance of the Gaussian noise.

To reconstruct the original signal  $\bar{\mathbf{x}}$ , a MAP estimator  $\hat{\mathbf{x}}$  is computed, which amounts to minimizing the sum of the negative log-likelihood  $f_{\mathbf{y}}$  plus a regularization  $\mathcal{R}_\lambda$  balanced by a parameter  $\lambda > 0$

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^T}{\operatorname{argmin}} f_{\mathbf{y}}(\mathbf{x}) + \mathcal{R}_\lambda(\mathbf{x}). \quad (2)$$

The likelihood of PG Model (1) is given by

$$\prod_{t=1}^T \left( \sum_{n=0}^{+\infty} \frac{e^{-\phi((\mathbf{H}\mathbf{x})_t)} (\phi((\mathbf{H}\mathbf{x})_t))^n e^{-\left(\frac{y_t - n - c}{\sqrt{2}\sigma}\right)^2}}{n! \sqrt{2\pi}\sigma} \right). \quad (3)$$

The corresponding log-likelihood is intricate and surrogates are often used for  $f_{\mathbf{y}}$ .

When  $\phi$  is the identity function, classical approaches often approximate the negative logarithm of (3) with a more tractable function of the form  $\sum_{t=1}^T g((\mathbf{H}\mathbf{x})_t, y_t)$  [1], [3], [7],

[11]. Such a good approximation is the Weighted least squares with a logarithm term (WLOG):

$$g_{\text{wlog}}(x, y) = \frac{1}{2} \frac{(y - x)^2}{x + \sigma^2} + \frac{1}{2} \log(x + \sigma^2).$$

### B. Sparse regularization

The regularization  $\mathcal{R}_\lambda$  is chosen according to the prior information available on  $\bar{\mathbf{x}}$ . We suppose here that, after a suitable linear transformation  $\mathbf{D}$ ,  $\bar{\mathbf{x}}$  is a sparse signal, i.e. only a few of its components are nonzero. This extends our previous work where the signal itself was composed of only a few spikes and the operator  $\mathbf{D}$  was the identity [8]. Alternatively,  $\mathbf{D}$  can be a gradient operator for a signal with sharp discontinuities. The  $\ell_0$  regularization function is an effective way to enforce the sparsity of the solution since it penalizes equally the nonzero components of the vector  $\mathbf{D}\bar{\mathbf{x}}$ . However, this choice yields NP-hard optimization problems.

A standard approximation to the  $\ell_0$  function is the  $\ell_1$  norm whose main benefit is to be convex. However, it also introduces a bias in the estimate. In this work, we therefore choose to replace  $\ell_0$  with a tight continuous approximation such as the ones described in [8]. Those approximations, although nonconvex, have been shown to possess nice properties like unbiasedness, continuity, and stability [12]. The pre-composition with the operator  $\mathbf{D}$  which is introduced in this paper adds flexibility to the criterion.

Finally, in the following, to build our MAP estimator, we consider the optimization Problem (2) with

$$f_{\mathbf{y}}(\mathbf{x}) = \sum_{t=1}^T g_{\text{wlog}}(\phi((\mathbf{H}\mathbf{x})_t) + c, y_t) \quad (4)$$

$$\mathcal{R}_\lambda(\mathbf{x}) = \sum_{t=1}^T \Psi_\lambda((\mathbf{D}\mathbf{x})_t), \quad (5)$$

where  $\Psi_\lambda$  is a continuous approximation to the  $\ell_0$  function such as the capped  $\ell_p$  function [13]–[15], the smoothly clipped absolute deviation (SCAD) [12], or the minimax concave penalty (MCP) [16].

## III. RATIONAL OPTIMIZATION FORMULATION

### A. Rational formulation

Due to our choice in (4) and (5), Problem (2) is highly nonconvex and is difficult to solve globally with standard optimization methods. However, it can be noticed that the approximations to the  $\ell_0$  function considered in Section II-B are piecewise rational functions, i.e. on different intervals, they are ratios of two polynomial functions. We assume here that the nonlinearity  $\phi$  is rational, which is not restrictive as any function can be tightly approximated with a suitable rational function. Furthermore, we substitute a rational approximation  $\widehat{\log}$  for the log function in  $g_{\text{wlog}}$ . An example of a good rational approximation for the log function is given by the following Padé approximant [17]

$$(\forall x \in \mathbb{R}_+^*) \quad \widehat{\log}(x) = (x - 1) \frac{x + 5}{2(2x + 1)}, \quad (6)$$

such that, for all  $x$  in  $\mathbb{R}_+^*$ ,  $\log(x) \leq \widehat{\log}(x)$ . This approximation is satisfactory on a broad interval and especially accurate on  $[0.2, 3.9]$  where the relative error is less than 8%.

Finally, substituting the log function with its rational approximation  $\widehat{\log}$ , Equation (4) becomes a sum of rational functions while (5) is a sum of piecewise rational functions. Problem (2) hence becomes the unconstrained minimization of a sum of rational and piecewise rational functions. This can be reformulated as the minimization of a sum of solely rational functions by replacing the characteristic functions of each subdomain of the piecewise functions with a binary variable as described in [8], [18]. As an example, we express Problem (2) as a rational optimization problem in the case of interest for many practical applications where  $\mathbf{D}$  is the discrete gradient operator and  $\Psi_\lambda$  is the capped  $\ell_1$  regularization:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}, \mathbf{r}} \quad & \hat{f}_{\mathbf{y}}(\mathbf{x}) + \sum_{t=1}^{T-1} (r_t(1 - z_t) + \lambda z_t) \\ \text{s.t.} \quad & (\forall t \in \llbracket 1, T - 1 \rrbracket) \begin{cases} z_t = z_t^2 \\ (z_t - 1/2)(r_t - \lambda) \geq 0 \\ r_t \geq 0 \\ r_t^2 = (x_{t+1} - x_t)^2, \end{cases} \end{aligned} \quad (7)$$

where  $\hat{f}_{\mathbf{y}}$  is the function in (4) with the log term replaced by  $\widehat{\log}$ ,  $(z_t)_{t \in \llbracket 1, T - 1 \rrbracket}$  and  $(r_t)_{t \in \llbracket 1, T - 1 \rrbracket}$  are extra real-valued variables used to handle respectively the indicator function and the absolute value of the capped  $\ell_1$  function.

Note that the operator  $\mathbf{D}$  imposes extra complexity to (7). Instead of having a polynomial in the three variables  $(x_t, z_t, r_t)$  as in [8], we now have a polynomial in four variables  $(x_{t+1}, x_t, z_t, r_t)$ . The choice of the operator  $\mathbf{D}$  is thereby important for the complexity of the model: using a higher-order discrete difference operator indeed leads to an increase of the number of variables per polynomial.

### B. Moment problem formulation

In (7), the objective is rational and the feasible set is compact. It can hence be written  $\min_{\mathbf{x} \in \mathcal{K}} \frac{p(\mathbf{x})}{q(\mathbf{x})}$  where  $p$  and  $q$  are polynomials and  $\mathcal{K}$  is the constraint set. Any such problem can be formulated equivalently as an infinite-dimensional general moment problem [19], where the variable  $\mu$  is a positive measure on  $\mathcal{K}$ :

$$\min_{\mu \geq 0} \int_{\mathcal{K}} p(\mathbf{x}) \mu(d\mathbf{x}) \quad \text{s.t.} \quad \int_{\mathcal{K}} q(\mathbf{x}) \mu(d\mathbf{x}) = 1. \quad (8)$$

The main bottleneck for numerically solving (8) is the number of variables on which the measure  $\mu$  is applied. This number has a critical influence on the size of the semidefinite programming (SDP) relaxation obtained in Section III-C. To get around this problem, we use the structure of the criterion. Indeed, under mild assumptions, Formulation (8) can be generalized to a criterion expressed as a sum of rational functions [20]. We hence define measures  $\mu_t^{(1)}$  and  $\mu_t^{(2)}$  for respectively each least square and  $\widehat{\log}$  term of  $g_{\text{wlog}}$  as well as a measure  $\nu_t$  for each term of the regularization. Under the assumption that the matrix  $\mathbf{H}$  has a Toeplitz-band structure, the number of variables per measure is now much fewer than  $T$ . For example

in Problem (7), each measure  $\nu_t$  is defined on the 4 variables  $(x_{t+1}, x_t, z_t, r_t)$  and the following criterion is minimized with respect to the variables  $\mu^{(1)}$ ,  $\mu^{(2)}$ , and  $\nu$ :

$$\sum_{t=1}^T \int p_t^{(1)} d\mu_t^{(1)} + \sum_{t=1}^T \int p_t^{(2)} d\mu_t^{(2)} + \sum_{t=1}^{T-1} \int p_t^{(3)} d\nu_t, \quad (9)$$

where  $p_t^{(1)}$  and  $p_t^{(2)}$  are the polynomial numerators of respectively  $\frac{1}{2} \frac{(y_t - \phi((\mathbf{H}\mathbf{x})_t))^2}{\phi((\mathbf{H}\mathbf{x})_t) + \sigma^2}$  and  $\widehat{\log}(\phi((\mathbf{H}\mathbf{x})_t))$  while  $p_t^{(3)}$  is the polynomial  $r_t(1 - z_t) + \lambda z_t$ .

However, this formulation using several measures comes at the cost of extra linear constraints linking the marginal moments of different measures that are equal. For example the moments of  $x_1$  in  $\mu_1^{(1)}$  in  $\mu_1^{(2)}$ , and in  $\nu_1$  should be all equal. Although there might be numerous additional linear constraints, they have a negligible effect on the complexity of the problem (see [18] for a more detailed discussion).

### C. SDP relaxation

As the feasible set is defined by polynomial constraints, the algorithm to solve minimization problems on measures like (8) and (9) consists in (i) replacing the measures by their moments with the additional semidefinite constraints on the corresponding moment and localizing matrices, (ii) truncating the moments up to a given maximal degree to obtain an SDP relaxation, (iii) solving these SDP problems, for instance with an interior point method, to obtain the moments of the sought measures, (iv) extracting the global minimizers of Problem (7) from the latter moments by an algebraic method [21]. Solving the successive SDP problems for different maximal degree results in an increasing sequence of lower bounds converging to the optimal objective value of the original optimization problem [19]. The main advantage of this method is its guarantee on the global optimality of the returned solutions.

Using Criterion (9) instead of (8) results in block SDP problems in Lasserre's hierarchy. Instead of having a single high-dimensional moment matrix in the semidefinite constraints, we have then lower-dimensional moment matrices, one for each introduced measure. Therefore, by using the structure of (2), we can drastically decrease the dimensions of the SDP problem to solve, which is essential for the computability.

## IV. NUMERICAL RESULTS

Throughout this part,  $\mathbf{H}$  is a Toeplitz matrix corresponding to the convolution with the lowpass filter with impulse response  $\mathbf{h} = [0.25, 0.5, 0.25]$  and with convolution boundaries using zero padding. The variance  $\sigma^2$  of the Gaussian noise is set to 0.15 (except in Table II) and  $c$  to 0. We solve Problem (2) using (4) and (6) for the data fit term together with a SCAD regularization (with parameter equal to 2.1) in (5). We use the software GloptiPoly [22] to perform the relaxation of the rational problem into SDP problems which are then solved using SDPT3 [23]. We set the relaxation order in the SDP hierarchy to 3, i.e. we consider moments up to degree 6.

We compare our method to proximal methods applied on the following convex problem:

$$\underset{\mathbf{x} \in \mathbb{R}^T}{\text{minimize}} \sum_{t=1}^T g((\mathbf{H}\mathbf{x})_t, y_t) + \lambda \ell_1(\mathbf{D}\mathbf{x}), \quad (10)$$

where  $g$  is one of the convex approximations to the negative log-likelihood mentioned in Section II. We use the peak signal to noise ratio (PSNR) between the original signal  $\bar{\mathbf{x}}$  and the estimator  $\hat{\mathbf{x}}$  to assess the quality of the reconstruction.

### A. Sparse signal reconstruction

The linear operator  $\mathbf{D}$  is first set to identity. We performed simulations on 100 randomly generated sparse signals of length  $T = 200$  with 20 nonzero elements. The positions of the latter are drawn uniformly between 1 and  $T$ .

1) *Linear case*: We first consider the case when  $\phi = \text{Id}$ . We compare our method to the classic Forward-Backward algorithm (FB) applied to Problem (10) as the considered function  $g$  is differentiable. More precisely, we compare our method with the results obtained with FB applied for the four approximations GAST (Generalized Anscombe transform), WLS (Weighted least squares), WLOG, and SPOI (Shifted Poisson) [1], [11] of the likelihood (3). The value of the regularization parameter  $\lambda$  has been tuned empirically to 5.5 for our method and 0.5, 2.5, 0.9, and 1 for FB respectively on GAST, WLS, WLOG, and SPOI.

Table I shows statistics on the PSNR of the estimated signal over 100 runs for each tested methods. We set the maximum number of iterations to 10000 for FB. Furthermore, we show in Table II the impact of the variance  $\sigma^2$  on the quality of the signal reconstructed with our method.

TABLE I  
STATISTICS ON THE PSNR (IN DB) BETWEEN THE ORIGINAL SPARSE SIGNAL AND THE ESTIMATED SIGNAL (100 REALIZATIONS)

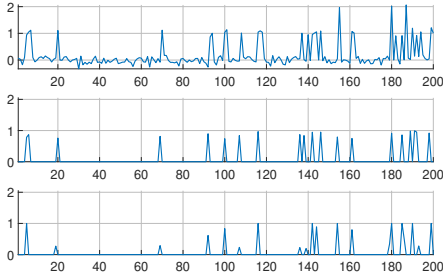
	Average	Median	Minimum	Maximum
FB (GAST)	12.0	12.0	10.5	13.8
FB (WLS)	7.3	7.1	4.1	10.9
FB (WLOG)	8.8	8.7	7.1	11.8
FB (SPOI)	10.7	11.2	6.9	13.5
Our method	12.5	12.5	11.3	14.1

TABLE II  
IMPACT OF THE VARIANCE  $\sigma^2$  ON THE RECONSTRUCTION QUALITY

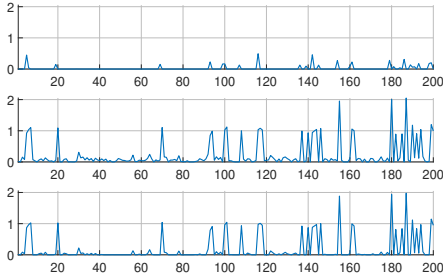
	$\sigma^2$	0.15	0.25	0.5	0.85
Average PSNR on 100 realizations (in dB)		12.5	11.3	10.5	8.1

2) *Nonlinear case*: We now choose  $\phi(x) = \frac{x}{\delta + |x|}$  with  $\delta = 0.3$ . Since  $\phi$  is nonconvex, we linearize it around 0 in (10) in order to apply FB. Figure 1 shows an example of the reconstruction in the nonlinear case where our method performs better than FB. Moreover, we observe that in both linear and nonlinear cases, the convergence in the SDP hierarchy occurs and that our method returns a global solution to (2). Indeed, the relative gap between the lower bound returned by solving the SDP problem of order 3 and the value of the criterion in (2) is in the order of magnitude of  $10^{-6}$ .

3) *Impact of logarithmic term*: In this section, we study the impact of the log term by dropping it in (4) and compare it with the previous criterion. We use the same methodology developed in Section III and the same experimental settings of Section IV-A2.



(a) From top to bottom: the observed signal  $\mathbf{y}$ , the original signal  $\bar{\mathbf{x}}$ , and the signal reconstructed with our method.



(b) From top to bottom, the estimator returned by FB for respectively GAST, WLOG, and SPOI approximations.

Fig. 1. Example of sparse signal reconstruction with PG noise and a nonlinear model.

In Figure 2, we zoom in the first 60 samples of the signals to compare them. We observe that without the log term, the weighted least squares (WLS) approximation always overestimate the amplitude of the peaks. This is consistent with [3] where the authors evidence that WLS approximation, as well as SPOI and GAST, introduce bias since the observations  $\mathbf{y}$  must be nonnegative. Conversely, we also note that after adding the log term, the amplitude of the peaks are slightly underestimated. However, many small artifacts are removed. Table III shows that adding the log term results in an 1.4 dB PSNR improvement in average.

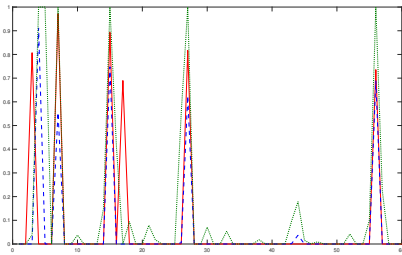


Fig. 2. Comparison of the reconstructed signal using our method. In red, the original signal  $\bar{\mathbf{x}}$ , in dashed blue, the signal reconstructed using  $g_{wlog}$ , and in dotted green, the result by using  $g_{wls}$

TABLE III  
COMPARISON ON THE PSNR (IN dB) FOR WLS AND WLOG APPROXIMATIONS USING OUR METHOD (100 RANDOM REALIZATIONS)

	Average	Median	Minimum	Maximum
WLS	11.0	10.8	9.2	13.4
WLOG	12.0	11.9	10.5	15.1

## B. Visible light communication (VLC) signals

We now observe  $\mathbf{y}$  according to Model (1) when  $\phi(x) = \frac{x}{\delta + |x|}$  and  $\mathbf{D}$  is the discrete gradient operator. We consider 100 binary signals  $\bar{\mathbf{x}}$  such that there are only 20 transitions between the 0 and 1 states. The location of the transitions are drawn uniformly between 1 and  $T$ . This model is inspired from VLC where a digital signal is transmitted by a LED that is alternatively switching between high and low intensity [5].

As the proximal operator of  $\ell_1 \circ \mathbf{D}$  does not have a closed form, our method is now compared to the forward-backward primal-dual algorithm (FBPD) [24] instead of FB. Moreover, we use the GAST approximation for the log-likelihood in FBPD as it gives the best reconstruction results. The value of the regularization parameters  $\lambda$  is tuned empirically to 2 for our method and to 1.1 for FBPD.

Figure 3 illustrates the obtained signals for a single test. We observe that our method provides a better estimator of the original signal. This is confirmed in Table IV that shows different statistics on the PSNR between the original signal  $\bar{\mathbf{x}}$  and the estimated one  $\hat{\mathbf{x}}$ .

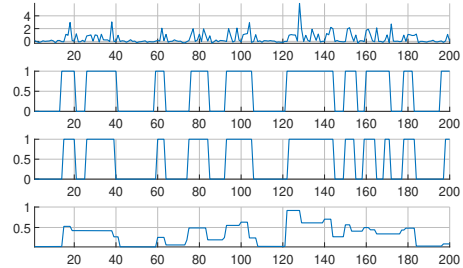


Fig. 3. Examples of VLC signals reconstruction. From top to bottom: the observed signal  $\mathbf{y}$ , the original signal  $\bar{\mathbf{x}}$ , the estimator obtained with our method, and the estimator returned by FBPD.

TABLE IV  
STATISTICS ON PSNR (IN dB) BETWEEN THE ORIGINAL VLC SIGNAL AND THE ESTIMATED SIGNAL (100 REALIZATIONS)

	Average	Median	Minimum	Maximum
FBPD	8.16	8.18	5.87	9.94
Our method	9.20	9.12	6.38	12.22

## V. CONCLUSION

We have considered the challenging problem of reconstructing a sparse signal in the presence of Poisson-Gaussian noise. We have adopted a quite general degradation model involving a convolutive effect and a nonlinear degradation. We have defined a rational approximation to the MAP optimization problem where we have proposed a nonconvex approximation to the log-likelihood and a nonconvex approximation to the  $\ell_0$  regularization that promotes sparsity after a given linear transform. We have applied a recent methodology based on a general moment problem to compute the MAP estimate as a global minimizer of the so-defined objective function. Finally, we have shown that our approximation to the MAP problem together with our computational approach provide better reconstruction results than standard ones for two different types of application.

## REFERENCES

- [1] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, “A convex approach for image restoration with exact Poisson–Gaussian likelihood,” *SIAM J. Imaging Sci.*, vol. 8, no. 4, pp. 2662–2682, Jan 2015.
- [2] Q. Ding, Y. Long, X. Zhang, and J. A. Fessler, “Statistical image reconstruction using mixed Poisson-Gaussian noise model for X-ray CT,” ArXiv Preprint, arXiv:1801.09533, Jan. 2018.
- [3] S. Lee, M. S. Lee, and M. G. Kang, “Poisson-Gaussian noise analysis and estimation for low-dose X-ray images in the NSCT domain,” *Sensors*, vol. 18, no. 4, pp. 1019, Apr. 2018.
- [4] S. Zhang, Y. Xia, and C. Zou, “An adaptive regularization method for low-dose CT reconstruction from CT transmission data in Poisson-Gaussian noise,” *Optik*, vol. 188, pp. 172–186, July 2019.
- [5] Chao Wang, Hong-Yi Yu, Yi-Jun Zhu, and Tao Wang, “Blind detection for SPAD-based underwater VLC system under P–G mixed noise model,” *IEEE Communications Letters*, vol. 21, no. 12, pp. 2602–2605, Dec. 2017.
- [6] P. J. Green, “Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 46, no. 2, pp. 149–192, 1984.
- [7] J. Li, Z. Shen, R. Yin, and X. Zhang, “A reweighted  $\ell_2$  method for image restoration with Poisson and mixed Poisson-Gaussian noise,” *Inverse Problems and Imaging*, vol. 9, no. 3, pp. 875–894, July 2015.
- [8] A. Marmin, M. Castella, and J.-C. Pesquet, “How to globally solve non-convex optimization problems involving an approximate  $\ell_0$  penalization,” in *Proc. Int. Conf. Acoust. Speech Signal Process.* May 2019, IEEE.
- [9] N. Taherkhani and K. Kiasaleh, “Statistical modelling of the clipping noise in OFDM-based visible light communication system,” arXiv:1909.03587 [eess.SP], 2019.
- [10] Z. Jiang, C. Gong, and Z. Xu, “Clipping noise and power allocation for OFDM-based optical wireless communication using photon detection,” *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 237–240, Feb. 2019.
- [11] Y. Marnissi, Y. Zheng, E. Chouzenoux, and J.-C. Pesquet, “A variational bayesian approach for image restoration—application to image deblurring with Poisson–Gaussian noise,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 722–737, Dec 2017.
- [12] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,” *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.
- [13] T. Zhang, “Analysis of multi-stage convex relaxation for sparse regularization,” *J. Mach. Learn. Res.*, vol. 11, pp. 1081–1107, Mar. 2010.
- [14] M. Artina, M. Fornasier, and F. Solombrino, “Linearly constrained nonsmooth and nonconvex minimization,” *SIAM J. Optim.*, vol. 23, no. 3, pp. 1904–1937, Jan. 2013.
- [15] A. Jezierska, H. Talbot, O. Veksler, and D. Wesierski, “A fast solver for truncated-convex priors: Quantized-convex split moves,” in *Lecture Notes in Computer Science*, pp. 45–58. Springer Berlin Heidelberg, 2011.
- [16] C.-H. Zhang, “Nearly unbiased variable selection under minimax concave penalty,” *Ann. Appl. Stat.*, vol. 38, no. 2, pp. 894–942, Apr. 2010.
- [17] G. Dahlquist and Å. Björck, *Numerical Methods in Scientific Computing, Volume 1*, Society for Industrial and Applied Mathematics, Jan. 2008.
- [18] M. Castella, J.-C. Pesquet, and A. Marmin, “Rational optimization for nonlinear reconstruction with approximate  $\ell_0$  penalization,” *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1407–1417, Mar. 2019.
- [19] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, U.K., 2009.
- [20] F. Bugarin, D. Henrion, and J. B. Lasserre, “Minimizing the sum of many rational functions,” *Math. Program. Comput.*, vol. 8, no. 1, pp. 83–111, Aug. 2015.
- [21] D. Henrion and J.-B. Lasserre, “Detecting global optimality and extracting solutions in GloptiPoly,” in *Positive Polynomials in Control*, vol. 312, pp. 293–310. Springer Berlin Heidelberg, Sept. 2005.
- [22] D. Henrion, J.-B. Lasserre, and J. Löfberg, “GloptiPoly 3: moments, optimization and semidefinite programming,” *Optim. Methods Softw.*, vol. 24, no. 4-5, pp. 761–779, Oct. 2009.
- [23] K. C. Toh, M. J. Todd, and R. H. Tütüncü, “SDPT3 — a Matlab software package for semidefinite programming, version 1.3,” *Optim. Methods Softw.*, vol. 11, no. 1-4, pp. 545–581, Jan. 1999.
- [24] N. Komodakis and J.-C. Pesquet, “Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems,” *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, Nov. 2015.