



**HAL**  
open science

## A greedy approach to sparse poisson denoising

François-Xavier Dupé, Sandrine Anthoine

► **To cite this version:**

François-Xavier Dupé, Sandrine Anthoine. A greedy approach to sparse poisson denoising. Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on, Sep 2013, Southampton, United Kingdom. pp.1-6, 10.1109/MLSP.2013.6661993 . hal-00998189

**HAL Id: hal-00998189**

**<https://hal.science/hal-00998189v1>**

Submitted on 30 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A greedy approach to sparse Poisson denoising

François-Xavier Dupé\*

Aix-Marseille Université, CNRS, LIF, UMR 7279, 13453 Marseille France

Sandrine Anthoine†

Aix Marseille Université, CNRS, Centrale Marseille, LATP, UMR 7353, 13453 Marseille France

## Abstract

In this paper we propose a greedy method combined with the Moreau-Yosida regularization of the Poisson likelihood in order to restore images corrupted by Poisson noise. The regularization provides us with a data fidelity term with nice properties which we minimize under sparsity constraints. To do so, we use a greedy method based on a generalization of the well-known CoSaMP algorithm. We introduce a new convergence analysis of the algorithm which extends its use outside of the usual scope of convex functions. We provide numerical experiments which show the soundness of the method compared to the convex  $\ell_1$ -norm relaxation of the problem.

**Keywords** Sparsity, greedy methods, Poisson noise, Moreau-Yosida regularization, proximal calculus.

## 1 Introduction

In this paper, we tackle the problem of denoising an image  $y$  corrupted by Poisson noise, under the assumption that the underlying object  $x$  can be well modeled by a sparse decomposition on a dictionary  $\Phi$ , i.e.  $x = \Phi\alpha$  with  $\alpha$  containing only a few non-zero coefficients.

Poisson denoising without sparsity constraints has been tackled quite extensively [1, 2, 3]. The main difficulty is to be accurate at low photon counts (i.e. for very low signal to noise ratios), which is an important issue in e.g. astrophysical and biological imaging. Indeed in these cases, the classical methods (such as Variance Stabilizing Transforms) that attempt to transform the Poisson noise to a Gaussian noise fail. Adding the sparsity assumption alleviates this problem by treating directly the Poisson noise case and thus allows for denoising images with very low photon counts [4, 5].

The literature on sparsity, from theory to applications and including algorithms, is vast and abundant. However it mostly concerns the case when the observed data is corrupted by a

Gaussian noise (i.e. a quadratic data-fidelity term). Our aim is to tackle the Poisson noise case in this context.

One can distinguish mainly two types of algorithms seeking for sparsity : 1) those directly minimizing the  $\ell_0$ -pseudo-norm that counts the non-zero coefficients and 2) the ones considering instead a relaxation such as the  $\ell_1$ -norm. The former consider the original NP-hard problem while the latter lead to convex optimization problems [6] which are more tractable. In this paper, we consider a greedy method such as *Matching Pursuit* [7] to solve the original  $\ell_0$  minimization.

While initially designed to tackle a quadratic data fidelity term, *Matching Pursuit* and its variants were recently extended [8, 9, 4, 5]. Among these, [4, 5] tackle the Poisson noise problem, but their algorithms are not theoretically grounded, while the reverse holds for [8, 9]. Let us note that handling the Poisson case is hard because it leads to a data fidelity term that is both not differentiable and strongly constrained (as we will explain in Section 2).

Here, we propose a greedy procedure that is theoretically grounded together with a solution to handle the Poisson noise case. Our contributions are thus twofold: i) we propose and analyze a generic greedy approach, that works for generic non-convex data-fidelity terms; ii) we propose and experiment a regularization of the the Poisson denoising problem in this context.

The paper is organized as follows. Section 2 is devoted to Poisson corrupted data, their compliance with sparsity and the proposed regularization. Then in Section 3, we proposed a novel analysis, using new conditions, of the greedy algorithm first proposed in [9]. In Section 4, we discuss the application of the algorithm to the regularized Poisson data and the bias introduced by it. Our numerical experiments (Section 3) confirm that i) the regularization does not add to much bias and ii) we recover the good properties of the  $\ell_0$  pseudo-norm like the preservation of the photometry.

\*This work is partially supported by the French GIP ANR under contract ANR GRETA 12-BS02-004-01 Greediness: theory and algorithms.

†Preprint of the article published in 2013 IEEE International Workshop on Machine Learning for Signal Processing.

## 2 Denoising a Poisson noise corrupted image under sparsity constraint

### 2.1 Sparse image representation

The true image  $x$  and observed image  $y$  containing  $n$  pixels are considered as vectors in the real Hilbert space  $\mathcal{H} = \mathbb{R}^n$ . We denote by  $\|\cdot\|$  the norm associated with the inner product in  $\mathcal{H}$  and by  $\mathbf{I}$  the identity operator on  $\mathcal{H}$ .

Our a priori is that the input image  $x$  as a sparse representation on a dictionary  $\Phi \in \mathbb{R}^{n \times m}$ . This means that

$$x = \Phi\alpha = \sum_{i=1, \dots, m} \alpha_i \varphi_i \text{ with } \|\alpha\|_0 = L \ll n.$$

Here, the  $\ell_0$ -norm  $\|\alpha\|_0 = \text{Card}\{i : \alpha_i \neq 0\}$  denotes the number of non-zero coefficients of  $\alpha$  aka the cardinal of its support  $\text{supp } \alpha = \{i : \alpha_i \neq 0\}$ . The  $\varphi_i$  are the columns of  $\Phi$ . They are called the atoms and are normalized ( $\|\varphi_i\| = 1$ ).

### 2.2 Poisson noise model

We assume we directly observe  $y$ , a version of  $x$  corrupted by Poisson noise. I.e. for each pixel  $i = 1, \dots, n$ , pixel  $y[i]$  follows a Poisson law distribution of parameter  $x[i]$ , which we write in short as:

$$y \sim \mathcal{P}(x). \quad (1)$$

Using the Poisson probability density function, the likelihood of observing an image  $y$  knowing that the input data is  $x$  is:

$$p(y|x) = \prod_i \frac{(x[i])^{y[i]} \exp(-x[i])}{y[i]!}. \quad (2)$$

Assuming the observed data is  $y$ , we derive the negative log-likelihood  $F_y(x) = -\log p(y|x)$ :

$$F_y : \eta \in \mathbb{R}^n \mapsto \sum_{i=1}^n f_y^i(\eta[i]), \quad (3)$$

where

$$f_y^i(\xi) = \begin{cases} -y[i] \log(\xi) + \xi & \text{if } y[i] > 0 \text{ and } \xi > 0, \\ \xi & \text{if } y[i] = 0 \text{ and } \xi \geq 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (4)$$

The negative log-likelihood measures how likely is the observation  $y$  when the input is  $x$ . It thus provides us with a data fidelity term  $F_y(\hat{x})$  that quantifies how well an estimated image  $\hat{x}$  fits the observed data  $y$ .

### 2.3 Poisson-sparsity

Here we seek an image  $x$  that is both sparsely coded on  $\Phi$  and that explains well the data  $y$ . A natural way to do so is to solve the following problem:

$$\hat{x} = \Phi\hat{\alpha}, \quad \text{where } \hat{\alpha} = \underset{\alpha \in \mathbb{R}^m \text{ s.t. } \|\alpha\|_0 \leq K}{\text{argmin}} F_y(\Phi\alpha). \quad (5)$$

Here,  $K$  denotes the expected sparsity of  $x$ . Note that this problem is complicated for two reasons: i) the  $\|\cdot\|_0$  constraint leads to a combinatorial problem, ii) the data fidelity term is not smooth and even not defined everywhere. Greedy algorithms such as *Matching Pursuit* give sub-optimal but good solutions of the  $\|\cdot\|_0$  problem for quadratic data fidelity terms. We will present and use here a greedy method that handles the non-smooth case, which solves i) (see Section 3).

Let us now analyze ii).  $F_y$  is a lower semi-continuous (l.s.c.) convex function. Notice that  $F_y(x) = +\infty$  not only for images  $x$  with at least a pixel of negative value ( $x[i] < 0$ ) but also for images  $x$  with have a pixel of value 0 while it is not the case in the observed data (i.e. if there an  $i$  such that  $x[i] = 0$  and  $y[i] > 0$ ). This property stems from the Poisson law and thus reflects the process generating the data, however it yields a data fidelity term which restricts a lot the estimates  $\hat{x}$  one can look for.

Imagine we look for a sparse signal in the pixel domain (i.e.  $\Phi = \mathbf{I}$ ), then whatever the level of sparsity  $K$  demanded, a solution  $\hat{x}$  of Problem (5) must verify  $\hat{x}[i] > 0$  if  $y[i] > 0$ . It is thus impossible to find a solution of this problem for  $K < \text{Card}\{i : y[i] > 0\}$ . In other words, the Poisson neg-log-likelihood does not allow for any compromise towards nullity. It thus seems odd to look for a sparse estimate in this setting.

Of course, using another dictionary  $\Phi$  than the identity allows to transport these hard support constraints on  $\Phi\alpha$  and thus decouple the problems. However, the previous reasoning shows that the type of restrictions imposed by  $F_y$  are quite different from a classical close convex constraints (typically the non-negativity constraint  $x \in (\mathbb{R}^+)^n$ ), which can be dealt with using projections.

A way around this problem could be to use a variance stabilizing function [1, 2]. But such methods can still be delicate to use because they involve a square-root and are not efficient for low intensity regimes. Here, we rather propose a different view that sticks more to the original Poisson data fidelity term. We propose to use a regularized version of it, namely its Moreau-Yosida regularization.

### 2.4 Moreau-Yosida regularization

**Definition 1** (Moreau-Yosida regularization [10]).  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semi-continuous (l.s.c.) convex function. Its Moreau-Yosida regularization for  $\lambda > 0$  is:

$$\begin{aligned} \mathcal{M}_{\lambda, f} : \mathbb{R}^d &\rightarrow \mathbb{R}, \\ s &\mapsto \inf_{x \in \mathbb{R}^d} \left[ \frac{1}{2\lambda} \|s - x\|^2 + f(x) \right]. \end{aligned} \quad (6)$$

The Moreau-Yosida regularization is a  $C^1$  function with a Lipschitz continuous gradient. It converges pointwise to  $f$  as  $\lambda$  tends to 0. Its gradient is related to the proximity mapping:

**Definition 2** (proximity mapping).  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper l.s.c. convex function. The proximity mapping of  $f$  is:

$$\text{prox}_f : \mathbb{R}^d \rightarrow \mathbb{R}^d, \\ s \mapsto \underset{x \in \mathbb{R}^d}{\text{argmin}} \left[ \frac{1}{2} \|s - x\|^2 + f(x) \right]. \quad (7)$$

The proximity mapping is a well-known convex analysis tool. Closed-form solutions exist for numerous *simple* functions (see [11]).

The gradient of the Moreau-Yosida regularization is [10]:

$$\nabla \mathcal{M}_{\lambda, f}(x) = \frac{1}{\lambda} (\mathbf{I} - \text{prox}_{\lambda f})(x). \quad (8)$$

Let us return to Problem (5), we shall replace the minimization of the neg-log-likelihood by its Moreau-Yosida regularization and thus solve

$$\hat{x} = \Phi \hat{\alpha}, \text{ where } \hat{\alpha} = \underset{\alpha \in \mathbb{R}^m \text{ s.t. } \|\alpha\|_0 \leq K}{\text{argmin}} \mathcal{M}_{\lambda, F_y \circ \Phi}(\alpha), \quad (9)$$

with  $\circ$  the composition. (The influence of  $\lambda$  will be discussed in Section 4.) To do so, the greedy algorithm we present in the next section requires the gradient of the regularized function.

**Proposition 3** (Gradient of the Moreau-Yosida regularization of the Poisson neg-log-likelihood [6]). *If  $\Phi$  is a tight frame (i.e.  $\exists \nu > 0$ , s. t.  $\Phi \circ \Phi^* = \nu \mathbf{I}$ ), then the gradient of the Moreau-Yosida envelope of  $F_y \circ \Phi$  is:*

$$\nabla \mathcal{M}_{\lambda, F_y \circ \Phi}(x) = \frac{1}{\nu \lambda} \Phi^* \circ (\mathbf{I} - \text{prox}_{\nu \lambda F_y}) \circ \Phi \quad (10)$$

with

$$\text{prox}_{\nu \lambda F_y}(x)[i] = \frac{x[i] - \nu \lambda + \sqrt{|x[i] - \nu \lambda|^2 + 4\nu \lambda y[i]}}{2}. \quad (11)$$

Let us now propose a greedy approach to solve Problem (9).

### 3 Greedy Minimization Algorithms

In this Section, we consider the general problem of minimizing a function under sparsity constraints. We wish to solve,

$$\min_{a \in \mathcal{H}} f(a) \quad \text{s. t. } \|a\|_0 \leq K, \quad (12)$$

where  $K$  is the maximal sparsity of the estimate  $a$  and  $f : \mathcal{H} \rightarrow \mathbb{R}$  is a smooth but not necessarily convex function. We denote by  $a^*$  a solution of (12).

Greedy procedures look for a sub-optimal solution since such a problem is combinatorial and NP-complete even in

simple cases. Here, we use (the greedy) Algorithm 1 (also known as GRASP [9]), which generalizes CoSaMP [12]. Just like the latter, this algorithm seeks the support of the optimal solution, iteratively updating it through five steps. The first three aim at updating the extended guess support of size  $3K$  ( $\mathcal{S}$ ) using both gradient information and the support of the current estimate. ( $w_k$  denotes the version of  $w$  truncated to its  $k$  largest coefficients in magnitude.) One then minimizes the function  $f$  over the new support (step four) and restricts the solution its  $K$  largest coefficient (step five). Notice that the forth step may require to solve an optimization problem of small size (if  $f$  is differentiable, a gradient descent can be sufficient).

---

**Algorithm 1:** Find a greedy solution to the non-convex optimization Problem (12).

---

**Initialization:**  $\hat{a}^0 = 0$ .

**Main iteration:**

**For**  $t = 0$  **to**  $N - 1$ ,

- 1) *Compute the local directions:*  $g = \nabla f(\hat{a}^t)$ ,
- 2) *Identify the next directions:*  $\mathcal{G} = \text{supp}(g_{2K})$ ,
- 3) *Merge the supports:*  $\mathcal{S} = \mathcal{G} \cup \text{supp}(\hat{a}^t)$ ,
- 4) *Minimize over the support:*  
 $z \in \underset{a: \text{supp}(a) \subset \mathcal{S}}{\text{argmin}} f(a)$
- 5) *Prune estimate:*  $\hat{a}^{t+1} = z_K$ .

**End main iteration**

**Output:** An approximated solution of (12):  $\hat{a}^N$ .

---

To show the convergence of Algorithm 1, we introduce the two following definitions.

**Definition 4** (Restricted Lipschitz gradient). *A function  $f$  is said to have a Lipschitz gradient of order  $K$  if there exists  $\beta_K > 0$  such that,*

$$\forall a, b \in \mathcal{H}, \quad \text{s. t. } \text{Card}\{\text{supp}(a) \cup \text{supp}(b)\} \leq K, \\ \|\nabla f(a) - \nabla f(b)\| \leq \beta_K \|a - b\|. \quad (13)$$

Of course  $\beta_K$  exists if  $f$  has a Lipschitz gradient. This variable is used to scale the gradient in the next definition.

**Definition 5** (Restricted Lipschitz dual gradient). *If  $f$  has a  $\beta_K$  restricted Lipschitz gradient, we define its dual gradient upper local bound of order  $K$  as:*

$$\delta_K = \sup_{a, b \in \mathcal{H}} \left\{ \frac{1}{\|a - b\|} \left\| (\mathbf{I} - \frac{1}{\beta_K} \nabla f)(a) - (\mathbf{I} - \frac{1}{\beta_K} \nabla f)(b) \right\| \mid \right. \\ \left. a \neq b, \text{Card}\{\text{supp}(a) \cup \text{supp}(b)\} \leq K \right\}. \quad (14)$$

Notice that by definition  $\delta_K \leq 2$ . Furthermore, if we replace  $\nabla f$  by a matrix  $\mathbf{A}$  (the Compressed sensing case with

Gaussian noise [12]), Eq. (14) leads directly to the famous *Restricted Isometry Property*. In fact,  $\delta_K$  is linked with the generalization of the power spectra for non-linear operators.

Using this bound allows to prove the following theorem<sup>1</sup>:

**Theorem 6.** *Suppose that  $\nabla f$  satisfies the restricted Lipschitz dual gradient condition (Definition 5) with  $\delta_{4K} \leq (2\sqrt{3} - 3)/3$ , then the distance between  $(\hat{a}^i)$  the estimate at the  $i$ -th iteration of Algorithm 1 and an optimal solution  $a^*$  of Problem (12) verifies,*

$$\|\hat{a}^{i+1} - a^*\| \leq \frac{\|\hat{a}^i - a^*\|}{2^i} + \frac{8\|\nabla f(a^*)\|_{3K}}{(1 - \delta_{4K})^2}. \quad (15)$$

The right part of Equation (15) can be split in two, on the left we have the exponentially decreasing convergence term, on the right an approximation term. This last term depends on the optimal solution and the gradient. It is easy to prove that in the Gaussian noise case, this term only depends of the variance of the noise (see [12]). More generally, the quality of the solution is function of the noise and the properties of the gradient (for example its smoothness).

Let us also note that the algorithm was first proposed and analyzed in [9]. Their analysis relies strongly on the convexity of  $f$  (via either the Hessian or the Bregman divergence) at least on sparse directions. By contrast, the conditions required for Theorem 6 can be easily generalized to operators and not just gradients, and do not require convexity. We believe that such results open the field of application of Algorithm 1.

## 4 Discussion

Let us now discuss the application of Algorithm 1 to the problem of Poisson denoising via the Moreau-Yosida regularization of the Poisson neg-log-likelihood.

The gradient  $\mathcal{M}_{\lambda, F_y \circ \Phi}$  is Lipschitz with a Lipschitz constant of  $\frac{1}{\lambda}$ . Therefore it has a *Restricted Lipschitz gradient* as defined in Eq.(13). Since  $\mathcal{M}_{\lambda, F_y \circ \Phi}$  is convex its *dual gradient upper local bounds of order  $K$*  verifies  $\delta_K \leq 1$  (using the Baillon-Haddad theorem [13]). We have not been able to prove that in this case  $\delta_{4K} \leq (2\sqrt{3} - 3)/3$  as required by Theorem 6, however the smoothness properties of  $\mathcal{M}_{\lambda, F_y \circ \Phi}$  and its derivative let us think that this is the case for a large range of  $\lambda$ . Notice that these values are located away from 0 where  $\mathcal{M}_{\lambda, F_y \circ \Phi}$  acts as the Poisson neg-log-likelihood, and also away from  $+\infty$  where  $\mathcal{M}_{\lambda, F_y \circ \Phi}$  tends to the constant value  $\inf_{\alpha} \{F_y \circ \Phi(\alpha)\}$ . This is supported by the experiments we display in the following section.

Let us now analyze how far is the solution we find from the solution of the original problem. Indeed our original problem was to find the best  $K$  - *sparse* signal minimizing the Poisson neg-log-likelihood, let us denote by  $x^o = \Phi \alpha^o$  this

signal. Instead we consider the output of Algorithm 1 after  $i$  iterations  $x^i = \Phi \alpha^i$ , when minimizing the Moreau-Yosida regularization  $\mathcal{M}_{\lambda, F_y \circ \Phi}$ . To evaluate the distance from  $x^o$  to  $x^i$ , we need to consider  $x^\lambda = \Phi \alpha^\lambda$  the best  $K$  - *sparse* signal minimizing the Moreau-Yosida regularization  $\mathcal{M}_{\lambda, F_y \circ \Phi}$ .

Since  $\Phi$  is a tight frame of constant  $\nu$ , we have:

$$\begin{aligned} \|x^i - x^o\| &\leq \|x^i - x^\lambda\| + \|x^\lambda - x^o\| \\ &\leq \sqrt{\nu} \|\alpha^i - \alpha^\lambda\| + \sqrt{\nu} \|\alpha^\lambda - \alpha^o\|. \end{aligned}$$

Thus using Eq.(15), we obtain:

$$\begin{aligned} \|x^i - x^o\| &\leq \sqrt{\nu} \frac{\|\alpha^{i-1} - \alpha^\lambda\|}{2^{i-1}} + 8\sqrt{\nu} \frac{\|\nabla \mathcal{M}_{\lambda, F_y \circ \Phi}(\alpha^\lambda)\|_{3K}}{(1 - \delta_{4K})^2} \\ &\quad + \sqrt{\nu} \|\alpha^\lambda - \alpha^o\| \end{aligned} \quad (16)$$

The error we commit is thus separable into three terms. The first one quantifies the optimization error done by stopping Algorithm 1 at a finite time, it decreases exponentially fast. The second source of error (second term) is an approximation error that is due to the noise in the data. This term is fully analyzed in the compressed sensing case in CoSaMP [12], and is proportional to the variance of the noise in the Gaussian case. In the Poisson case, it is more difficult to analyze it. The last term, called bias, quantifies the error we make by using the Moreau-Yosida regularization instead of the Poisson neg-log-likelihood. Assuming we could bound beforehand the maximum intensities of the  $x^i$ ,  $x^o$  and  $x^\lambda$ , we could use a result in [14] showing that  $\|\alpha^\lambda - \alpha^o\| = O(\sqrt{\lambda})$  which quantifies this error (remember that the Moreau-Yosida regularization converges to the original function when  $\lambda$  goes to 0).

## 5 Experiments

In this section, we evaluate the performance of our method (aka  $\ell_0$  method) and compare it to the classical convex relaxation way (aka  $\ell_1$  method). The latter was implemented using the Poisson model as described in [11] using a convex optimization algorithm. The experiments will shed light on the effects of using the  $\ell_0$ -norm instead the  $\ell_1$  and also on the consequences of the use of Moreau-Yosida regularization.

Two experiments are proposed using two classical images (*Camerman* and *Barbara*) with two different dictionaries, the undecimated wavelet transform (with the symlet 6) and the curvelet transform. Notice that both transforms are redundant and thus well fit to the denoising task.

For all the experiments, we set the Moreau-Yosida regularization parameter to 10. This value may lead to a non-negligible bias but allows for a better convergence rate. The regularization parameter of both  $\ell_0$  and  $\ell_1$  methods have been fixed to give comparable sparsity levels. Remark that finding *good* (or optimal) parameters is an open problem in both cases (see [15] for an example for the Gaussian noise case).

<sup>1</sup>The proof is provided in [http://www.latp.univ-mrs.fr/~anthoine/TechRep\\_072013.pdf](http://www.latp.univ-mrs.fr/~anthoine/TechRep_072013.pdf)

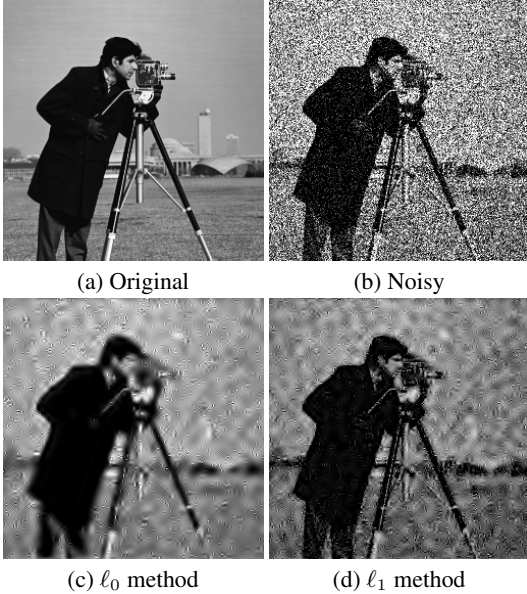


Figure 1: Denoising *Cameraman* with a maximal intensity of 5 with the undecimated wavelet transform.

Figure 1 shows the results for the *Cameraman* with a maximal intensity of 5. To show the differences of photometry, the images in a same figure are always displayed using the same grayscale colormap. Assuming that the image is sparse in the undecimated wavelet domain (which is mostly true), we apply the  $\ell_0$  method (Fig. 1(c)) and compare it to the  $\ell_1$  method (Fig. 1(d)). Notice that the  $\ell_0$ -norm leads a smoother image, while most of the details are preserved with the  $\ell_1$ -norm (but with more noise). Because the *Cameraman* is not truly sparse in the chosen domain, enforcing the sparsity for the reconstruction is not relevant. However, the  $\ell_0$  preserves the photometry better: for example the coat of the *Camera-man* is darker in Fig. 1(c) than in Fig. 1(d).

We repeat the experiment with a maximal intensity of 30 (thus with a higher SNR). As the noise is weaker, more details should be recovered. Figure 2 shows the results with both methods. The  $\ell_0$  method (Fig. 2(c)) preserves the details as well as the  $\ell_1$  method (Fig. 2(d)). As with the previous experiment, the most important difference between Fig. 2(c) and Fig. 2(d) is the photometry. For example, the camera is brighter in Fig. 2(c) (like in the original) than in Fig. 2(d).

Maximal Intensity	$\ell_0$ method			$\ell_1$ method		
	PSNR	MAE	SSIM	PSNR	MAE	SSIM
5	25.2	0.36	0.53	20.2	0.47	0.57
30	26.9	1.22	0.74	22.0	1.61	0.80

Table 1: Denoising performance for our method and the relaxed method for the *Cameraman* at two levels of noise.

To quantify our assertions, we repeat both experiments 10 times each and for each result, we compute the peak signal to

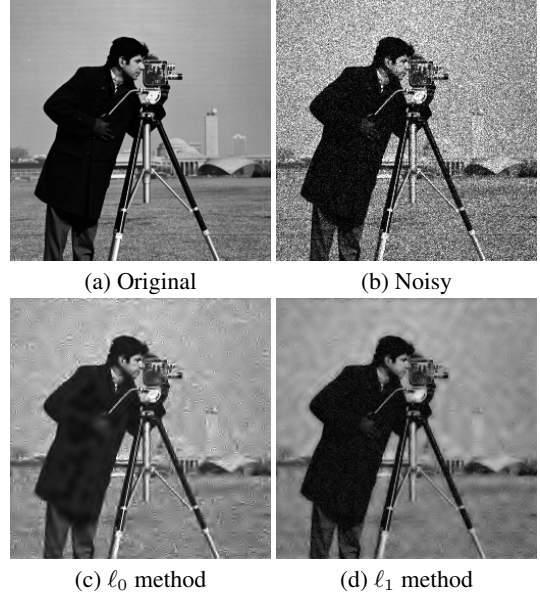


Figure 2: Denoising *Cameraman* with a maximal intensity of 30 with the undecimated wavelet transform.

noise ratio (PSNR), the mean absolute error (MAE) and the structural similarity index (SSIM) [16]. Table 1 shows the averaged numerical results for the two levels of intensity (i.e. of noise) we tested, using the very same parameters. Notice that for both intensity levels, the  $\ell_0$  method gives a better PSNR and MAE than the  $\ell_1$  method, but not for the SSIM. These results confirm that the  $\ell_0$  method is better as preserving the photometry (to which PSNR and MAE are sensitive), but  $\ell_1$  is better at restoring the details (as reflected by the SSIM).

To show the effect of the choice of the dictionary, we repeat the experiment at a maximal intensity of 30 (medium level of noise). For this experiment, we use *Barbara* (Fig. 3(a)) because of the curve-like textures on the pants. Figure 3 shows the results for each method using two different dictionaries, Fig. 3(c) and Fig. 3(d) for the curvelet transform and Fig. 3(e) and Fig. 3(f) for undecimated wavelet transform. As expected  $\ell_0$  method shows also a better photometry. Moreover, for both methods the curvelet transform is better at restoring the textures. With the undecimated wavelet transform, part of the textures is lost and the  $\ell_0$  method is less efficient than the  $\ell_1$  method (see specifically the shawl). This shows the importance of the selection of the dictionary while using sparse method. Using the wrong one may lead to artifacts and loss of some structures (like textures).

## 6 Conclusion

In this paper, we presented a new analysis of a generalization of the CoSaMP algorithm and proposed to apply it to remove Poisson noise under a sparsity constraint. In order to avoid the shortcomings of a direct use of the Poisson log-likelihood, we

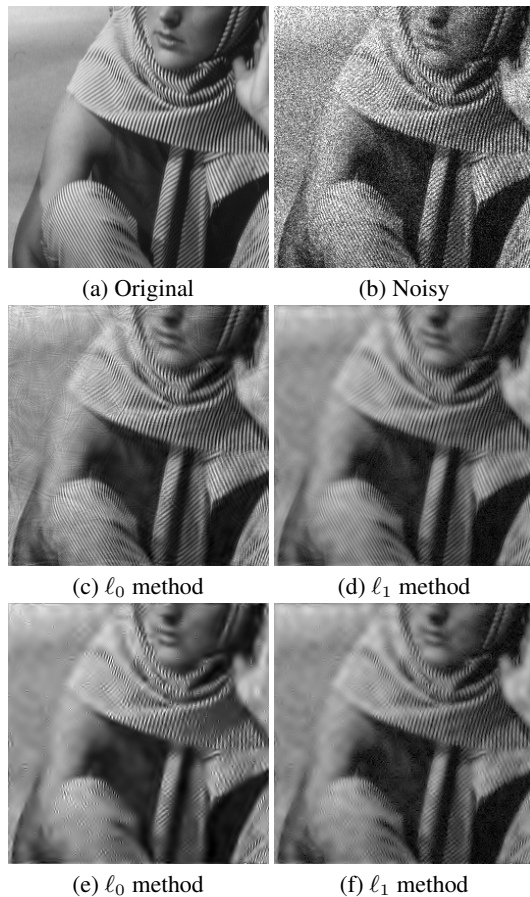


Figure 3: Denoising *Barbara* with a maximal intensity of 30. (c) and (d) using the curvelet transform. (e) and (f) using the undecimated wavelet transform.

exploited instead its Moreau-Yosida regularization. Experiments showed that the obtained method is competitive with a convex relaxation approach. The proposed algorithm is provably convergent. Moreover, the proposed convergence conditions allow to apply it to non-convex functions, which was not the case of the original analysis. However a deeper analysis of these conditions of convergence is still to be done. Future works also include considering dictionary learning to adjust the dictionary to the sparsity prior.

## References

- [1] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal," *Image Processing, IEEE Trans. on*, vol. 17, no. 7, pp. 1093–1108, 2008.
- [2] M. Makitalo and A. Foi, "Optimal inversion of the anscombe transformation in low-count poisson image denoising," *Image Processing, IEEE Trans. on*, vol. 20, no. 1, pp. 99–109, 2011.
- [3] J. Salmon, C. A. Deledalle, R. Willett, and Z. Harmany, "Poisson noise reduction with non-local PCA," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE Int. Conf. on*. IEEE, 2012, pp. 1109–1112.
- [4] X. Zhang, Y. Lu, and T. Chan, "A novel sparsity reconstruction method from poisson data for 3D bioluminescence tomography," *J. of Scientific Computing*, vol. 50, no. 3, pp. 519–535, 2012.
- [5] R. Giryes and M. Elad, "Sparsity based Poisson denoising," in *Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of*, 2012, pp. 1–5.
- [6] P. L. Combettes and J.-. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Top. Sig. Pro.*, vol. 1, no. 4, pp. 564–574, 2007.
- [7] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Trans. on*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [8] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," *Information Theory, IEEE Trans. on*, vol. 57, no. 9, pp. 6215–6221, 2011.
- [9] S. Bahmani, B. Raj, and P. Boufounos, "Greedy sparsity-constrained optimization," *J. of Machine Learning Research*, vol. 14, no. 3, pp. 807–841, 2013.
- [10] C. Lemaréchal and C. Sagastizábal, "Practical Aspects of the Moreau–Yosida Regularization: Theoretical Preliminaries," *SIAM J. on Optimization*, vol. 7, no. 2, pp. 367–385, 1997.
- [11] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inv. Prob.*, vol. 24, no. 6, 2008.
- [12] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [13] H. H. Bauschke and P. L. Combettes, "The baillon-haddad theorem revisited," *J. of Convex Analysis*, vol. 17, no. 4, pp. 781–787, 2010.
- [14] P. Mahey and P. D. Tao, "Partial regularization of the sum of two maximal monotone operators," *Modélisation mathématique et analyse numérique*, vol. 27, no. 3, pp. 375–392, 1993.
- [15] S. Vaiter, C.-A. Deledalle, G. Peyré, C. Dossal, and J. Fadili, "Local behavior of sparse analysis regularization: Applications to risk estimation," *Applied and Computational Harmonic Analysis*, 2012.

- [16] Z. Wang, A. Bovik, H. R. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.