



HAL
open science

Generalized Subspace Pursuit and an application to sparse Poisson denoising

François-Xavier Dupé, Sandrine Anthoine

► **To cite this version:**

François-Xavier Dupé, Sandrine Anthoine. Generalized Subspace Pursuit and an application to sparse Poisson denoising. International Conference on Image Processing (ICIP) 2014, IEEE conference on,, Oct 2014, Paris, France. hal-01071760

HAL Id: hal-01071760

<https://hal.science/hal-01071760v1>

Submitted on 6 Oct 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.

GENERALIZED SUBSPACE PURSUIT AND AN APPLICATION TO SPARSE POISSON DENOISING

François-Xavier Dupé

Aix Marseille Université, CNRS,
LIF, UMR 7279,
13288 Marseille, France

Sandrine Anthoine

Aix Marseille Université, CNRS, Centrale Marseille,
I2M, UMR 7373,
13453 Marseille, France

ABSTRACT

We present a generalization of Subspace Pursuit, which seeks the k -sparse vector that minimizes a generic cost function. We introduce the *Restricted Diagonal Property*, which much like RIP in the classical setting, enables to control the convergence of Generalized Subspace Pursuit (GSP). To tackle the problem of Poisson denoising, we propose to use GSP together with the Moreau-Yosida approximation of the Poisson likelihood. Experiments were conducted on synthetic, exact sparse and natural images corrupted by Poisson noise. We study the influence of the different parameters and show that our approach performs better than Subspace Pursuit or ℓ_1 -relaxed methods and compares favorably to state-of-art methods.

Index Terms— Sparse regularization, Subspace Pursuit, Poisson denoising, Greedy algorithm.

1. INTRODUCTION

Classical greedy algorithms such as MP [3] or CoSaMP [4] that minimize a quadratic cost under a l_0 sparsity constraint have recently been generalized to other costs. [1, 2] for example show that a k -sparse minimizer of a function can be found if its gradient satisfies specific properties. Furthermore these analyses lead to a better understanding of the behavior of these greedy methods. Such generalization open the doors to the use of greedy algorithms for general sparse optimization problems for which the theory is developed so far almost only when the l_0 -semi-norm is relaxed to a l_1 -norm.

In the context of Poisson denoising, we recently suggested a way to fill the theoretical gap by providing a CoSaMP-like greedy algorithm and its analysis [5]. Such effort is relevant because while some Poisson denoising algorithms have a solid theoretical ground, e.g. variance stabilization combined with Gaussian methods [6, 7] or convex minimization with l_1 -norm for sparsity [8], others are transposition from the Gaussian world and still need a thorough analysis, e.g. non-local mean [9] or greedy (hybrid) algorithms [10]. In this

paper we pursue these efforts by providing a new greedy algorithm for sparse Poisson denoising and exhibiting sufficient conditions for its convergence. Our contributions are twofold. 1) We propose the first (to our knowledge) generalization of the Subspace Pursuit (called GSP) to generic cost with a convergence theorem. 2) We devise a new algorithm for Poisson denoising using GSP and the same regularization procedure as in [5] and demonstrate its efficiency.

The structure of this paper is the following. In Section 2, we present Generalized Subspace Pursuit and introduce the *Restricted Diagonal Property* which ensures its convergence for a wide class of cost function (including non-convex ones). In Section 3, we use GSP with a Moreau-Yosida regularization to devise a new Poisson denoising algorithm. Experiments are conducted to see the behavior of this new method depending on the sparsity, the noise level and the regularization parameter. Finally, we compare it to state-of-art methods like SAFIR [9], MSVST [6], the classical l_1 -norm relaxation method and Subspace Pursuit (SP) [11].

2. GENERALIZED SUBSPACE PURSUIT

Classical greedy algorithms such as MP, CoSaMP or SP attempt to find the “best” k -sparse vector x that explains the observed measured vector y under a linear model: $y = \Phi x$, where “best” is understood in the l_2 sense which means that these algorithms solve the following problem:

$$\text{Find } x \in \underset{\hat{x}}{\text{argmin}} \|y - \Phi \hat{x}\|_2^2 \text{ s.t. } \|\hat{x}\|_0 \leq k. \quad (\text{P1})$$

($\|x\|_0 = \text{Card}\{i : x_i \neq 0\}$ is the number of non-zero entries of x). This can also be viewed as solving the linear inverse problem $y = \Phi x$ under the assumption of Gaussian noise and with a hard sparsity constraint as regularization.

In many inverse problems, the quadratic cost $\|y - \Phi \hat{x}\|_2$ is not appropriate and one rather wishes to solve:

$$\text{Find } x \in \underset{\hat{x}}{\text{argmin}} f(x) \text{ s.t. } \|x\|_0 \leq k, \quad (\text{P2})$$

for a more complex f . A vast literature gives solutions for a relaxed sparsity regularization (i.e. using the l_1 -norm instead

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

of the ℓ_0 -semi-norm) in the case of convex f for example. In contrast, only few attempts have successfully extended greedy algorithms to Problem (P2). Indeed the convergence proofs of MP, SP etc. heavily rely on linear algebra useful because the gradient of $\|y - \Phi x\|_2^2$ is affine (see [11, 4, 12]), which is not the case for a generic f .

2.1. Algorithm

We propose an extension of Subspace Pursuit [11] designed to solve Problem (P2) for a cost function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ differentiable. This algorithm named *Generalized Subspace Pursuit* (GSP) is detailed in Algorithm 1.

Here, we denote by $\text{supp}(x)$ the support of x : $\text{supp}(x) = \{i = 1, \dots, n \mid x_i \neq 0\}$ and $x|_k$ the restriction of x to its k largest coefficients in magnitude.

As SP, GSP seeks the support of the solution of size k (denoted by \mathcal{T}). It updates it at each iteration by adding a set of k candidate directions (step 3-5), obtaining an extended support set \mathcal{S} of size $2k$. The problem is solved on this extended support (step 6). \mathcal{T} is updated by the best support of size k extracted from this solution (step 7). Finally, one seeks the best solution of support \mathcal{T} (step 8).

GSP exactly mimics SP, except for steps 3, 6 and 8 where it accounts for f instead of $\|y - \Phi \hat{x}\|_2^2$. Steps 6 and 8 consist in finding an optimal solution to the original problem on a predefined support set, which can usually be done efficiently (e.g. with a projected gradient descent).

Algorithm 1 Generalized Subspace Pursuit

- 1: **initialization:** $\hat{a}^0 \leftarrow 0$.
 - 2: **for** $t = 0 \dots N - 1$ **do**
 - 3: $g \leftarrow \nabla f(\hat{a}^t)$, \triangleright Compute local directions.
 - 4: $\mathcal{G} \leftarrow g|_k$, \triangleright Select new directions.
 - 5: $\mathcal{S} \leftarrow \mathcal{G} \cup \text{supp}(\hat{a}^t)$, \triangleright Set extended support.
 - 6: $b \in \underset{\{a: \text{supp}(a) \subseteq \mathcal{S}\}}{\text{argmin}} f(a)$, \triangleright Solve on extended support.
 - 7: $\mathcal{T} \leftarrow \text{supp}(b|_k)$, \triangleright Set support.
 - 8: $\hat{a}^{t+1} \in \underset{\{a: \text{supp}(a) \subseteq \mathcal{T}\}}{\text{argmin}} f(a)$, \triangleright Solve on the support.
 - 9: **end for**
 - 10: **outputs:** \hat{a}^N .
-

2.2. Theoretical guarantees

In the classical case ($f = \|\Phi x - y\|^2$), a sufficient condition for MP, SP or CoSaMP to converge is that the matrix Φ satisfies the RIP condition with a small constant δ_k . This essentially ensures that Φ acts as almost isometry on the set of k -sparse vectors so that $\Phi^* \Phi$ which controls the increments $\nabla f(x) - \nabla f(x')$ is close to the identity. It turns out that in the generalized case, ensuring the increments of ∇f are close to the identity on the set of k -sparse vectors also guarantees the convergence of GSP. This condition can even be weakened by

replacing the identity by any diagonal operator bounded away from zero and having this property hold only locally.

To state the convergence of GSP, let us introduce \mathcal{D}_1 the set of diagonal operators bounded away by 1:

$$\mathcal{D}_1 = \{\mathbf{D} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \mathbf{D} \text{ diagonal}, \forall x \|\mathbf{D}x\| \geq \|x\|\}. \quad (1)$$

Definition 1 (Restricted Diagonal Property). \mathbf{T} is said to have the Restricted Diagonal Property (RDP) of order k if there exists $\delta_k > 0$ such that if $a, b \in \mathbb{R}^n$ then,

$$\begin{aligned} \text{Card}\{\text{supp}(a) \cup \text{supp}(b)\} \leq k &\Rightarrow \exists \mathbf{D}_{ab} \in \mathcal{D}_1, \\ \|\mathbf{T}(a) - \mathbf{T}(b) - \mathbf{D}_{ab}(a - b)\| &\leq \delta_k \|a - b\|. \end{aligned} \quad (2)$$

Remark 1. $\delta_{2k} < 1$ implies that \mathbf{T} is injective on the set of k -sparse vectors since $\|\mathbf{T}(a) - \mathbf{T}(b)\| \geq (1 - \delta_{2k}) \|x - y\|$.

Remark 2. If \mathbf{D}_{ab} is the Identity for all a and b , there is a connection with RIP since then $(1 - \delta_{2k}) \|x - y\| \leq \|\mathbf{T}(x) - \mathbf{T}(y)\| \leq (1 + \delta_{2k}) \|x - y\|$.

However the *Restricted Diagonal Property* is a more local property. Let us now state the convergence theorem.

Theorem 2. Assume $f : \mathbb{R}^n \rightarrow \mathbb{R}$, differentiable and (P2) has at least one solution. Denote by a^* any solution of (P2) and δ_0 the unique real root of $g(x) = x^3 + x^2 + 7x - 1$.

If there exists $\beta > 0$ such that $\frac{\nabla f}{\beta}$ has the Restricted Diagonal Property of order $3k$ with $\delta_{3k} \leq \delta_0$. Then GSP verifies

$$\|\hat{a}^N - a^*\| \leq \frac{1}{2^N} \|\hat{a}^0 - a^*\| + \frac{15}{\beta} \|\nabla f(a^*)|_{2k}\|. \quad (3)$$

The distance between the iterates of GSP and a solution of (P2) are thus controlled by two terms. The first one decreases exponentially fast to zero while the second expresses the error made, and is controlled by the gradient. Theorem 2 thus shows that if f admits global minimizer that is k -sparse, GSP recovers it exactly (since in that case $\nabla f(a^*) = 0$): this is the analog to the *exact recovery* case for SP, MP and the like. Otherwise, the quality of the iterates is controlled by the error term $\|\nabla f(a^*)|_{2k}\|$, which is analog the the uncompressible error term exhibited in the classical case (see [4]).

The proof of this theorem (not given here due to lack of space) is very similar to that of [5], where a generalization of CoSaMP is proposed and analyzed using a stronger property.

Let us pinpoint the pros and cons of the proposed analysis. First, this is the first generalization of SP to non-quadratic costs, and we provide a convergence theorem. An advantage of extending SP versus CoSaMP (as in GRASP [2, 5]) lies in GSP's last step (step 8), which consists in finding the best k -sparse vector with the updated support \mathcal{T} of size k . This is more costly than the simple pruning used in CoSaMP or GRASP, but has been empirically proved to speed up the convergence in the classical setting (SP vs CoSaMP). More importantly, this enables to consider cost functions f incorporating hard constraints since those will be naturally respected at all steps (not shown here).

Second, although the RDP is not easy to check in practice, it is clear that it does not require f to be convex (for example $f(x) = x^T(D + \gamma\mathbf{I})x$ with $D = \text{diag}(\pm 1, \dots, \pm 1)$, so that $D_{ab} = D$ and $\delta_k = \gamma$). So the proposed analysis extends the scope of greedy algorithms to non-convex costs, as opposed to previous analyses like [2].

In the next Section we shall apply GSP to the problem of Poisson denoising, which is a typical example where the quadratic cost function is not adapted.

3. APPLICATION TO POISSON DENOISING

3.1. Poisson denoising under a hard sparsity constraint

Let us assume that we observe $y \in \mathbb{R}^n$, a Poisson corrupted version of the true image $x \in \mathbb{R}^n$, both containing n pixels. In our model, we also assume that x has a k -sparse representation on the dictionary $\Phi = (\varphi_1, \dots, \varphi_m) \in \mathbb{R}^{n \times m}$:

$$x = \Phi\alpha = \sum \alpha_j \varphi_j \text{ with } \|\alpha\|_0 = k \ll n,$$

where the atoms are normalized ($\|\varphi_j\| = 1$).

Our goal is to reconstruct x given the data y , the sparsity k and the dictionary Φ , which may be done by solving:

$$\hat{x} = \Phi\hat{\alpha}, \quad \text{where } \hat{\alpha} = \underset{\alpha \in \mathbb{R}^m \text{ s.t. } \|\alpha\|_0 \leq k}{\text{argmin}} F_y(\Phi\alpha), \quad (\text{P3})$$

where $F_y(\hat{x})$ is a data fidelity term that quantifies how well an estimated image \hat{x} fits the observed data y . A natural fidelity term is the negative-log-likelihood $F_y(x) = -\log \mathbb{P}(y|x)$ which reads in the case of Poisson noise

$$F_y(x) = -\log \mathbb{P}(y|x) = \sum_{i=1}^n f(x_i, y_i), \quad \text{with}$$

$$f(\xi, \eta) = \begin{cases} -\eta \log(\xi) + \xi & \text{if } \eta > 0 \text{ and } \xi > 0, \\ \xi & \text{if } \eta = 0 \text{ and } \xi \geq 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (4)$$

Notice that $F_y(x)$ is finite only when x complies with the data, which implies $x \in \mathbb{R}_+^n$ and $x_i > 0$ if $y_i > 0$. Moreover, due to the logarithm, its gradient is not defined on its all domain. Hence GSP may not be applied on $F_y \circ \Phi$. These technical difficulties reflect a deeper conceptual problem in solving Problem (P3): in fact the hard constraints imposed by the Poisson negative-log-likelihood on $\Phi\alpha$ ($(\Phi\alpha)_i \geq 0$ for all i and $(\Phi\alpha)_i > 0$ if $y_i > 0$) may not be compatible with hard ℓ_0 constraint $\|\alpha\|_0 \leq k$.

To alleviate this problem, we substitute in (P3) $F_y \circ \Phi$ for its Moreau-Yosida regularization [13] of parameter λ :

$$\mathcal{M}_{\lambda, F_y \circ \Phi}(\alpha) = \inf_{s \in \mathbb{R}^m} \left[\frac{1}{2\lambda} \|s - \alpha\|^2 + F_y \circ \Phi(s) \right]. \quad (5)$$

The Moreau-Yosida envelope of a convex l.s.c. function has full domain, its gradient is Lipschitz. Moreover, its minimizer is also minimizer of the original function. It does provides a good candidate for relaxing $F_y \circ \Phi$ and using it with

GSP. To do so, we need the gradient of this function, which is computable when the dictionary Φ is a tight frame.

Proposition 3 (Gradient of the Moreau-Yosida regularization of the Poisson neg-log-likelihood). *If Φ is a tight frame of constant $\nu > 0$, then the gradient of $\mathcal{M}_{\lambda, F_y \circ \Phi}$ is [8]:*

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

$$\text{prox}_{\nu\lambda F_y}(x)_i = \frac{x_i - \nu\lambda + \sqrt{|x_i - \nu\lambda|^2 + 4\nu\lambda y_i}}{2}. \quad (6)$$

In the following, we thus denoise y using GSP to solve

$$\hat{x} = \Phi\hat{\alpha}, \quad \text{where } \hat{\alpha} = \underset{\alpha \text{ s.t. } \|\alpha\|_0 \leq k}{\text{argmin}} \mathcal{M}_{\lambda, F_y \circ \Phi}(\alpha). \quad (\text{P4})$$

It is not clear whether $\mathcal{M}_{\lambda, F_y \circ \Phi}$ has the RDP, however we empirically show in the next section that the proposed algorithm converges.

3.2. Experiments

The strengths and the weaknesses of our approach are assessed on three kinds of experiments. For each one, the quantitative values are averaged over 10 realizations of the noise. The dictionary Φ is fixed to cycle spinning wavelets with a redundancy of 9. In the first experiment, we study the evolution of the support recovery with the noise and the sparsity levels. Next, we study the influence of the parameter of the Moreau-Yosida regularization on the recovered support. Finally, we compare the performance of our method to that of state-of-the-art ones.

3.2.1. Support recovery

In this experiment, we generate random k -sparse images of size 64×64 with different sparsities and peak intensities. They are subsequently corrupted by Poisson noise. Each image is denoised using our method with the regularization parameter $\lambda = 1$. Fig 1(a) shows the phase diagram of support recovery (number of atoms of the true support recovered divided by k) as a function of sparsity level (k/n) and peak intensity (higher intensity means lower noise level). As expected when the image is sparser or less degraded (lower left), the support is easier to find.

3.2.2. Influence of the regularization

Let us assess the influence of the Moreau-Yosida parameter (λ) on support recovery. Fig. 1(b) pictures the evolution of support recovery with λ for a sparse version of the Cameraman ($k/n = 0.15$) with a peak intensity at 11 obtained by Subspace Pursuit. This illustrates the following facts. First, the lower the λ is, the better we reconstruct the support. This is expected as when λ tends to 0, the Moreau-Yosida regularized function tends to the *true* function. Second, the quality

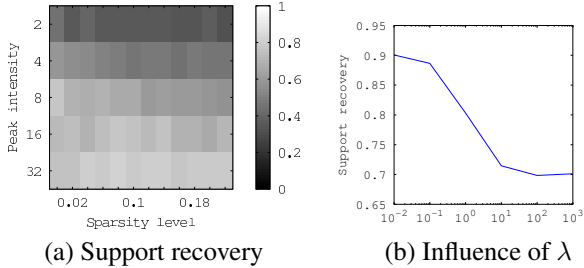


Fig. 1. Support recovery as a function of sparsity and noise level (a) and of the Moreau-Yosida regularization (b).

	Sparse Cameraman			Galaxy		
	PSNR	MAE	SSIM	PSNR	MAE	SSIM
Noisy	19.3	1.57	0.32	22.9	0.63	0.19
GSP	28.6	0.32	0.87	26.0	0.17	0.71
SP [11]	26.1	0.55	0.63	27.6	0.28	0.55
SAFIR [9]	27.6	0.36	0.86	29.9	0.15	0.84
MSVST [6]	29.5	0.31	0.84	30.5	0.12	0.83
ℓ_1 -relaxation	22.7	0.64	0.73	21.5	0.32	0.50

Table 1. Comparison of denoising methods on a sparse version of Cameraman ($k/n = 0.15$) and the NGC 2997 Galaxy.

degrades for high λ , eventually becoming worse than that of Subspace Pursuit (0.79), which shows that too much regularization degrades the estimation. Finally, the plateau reached for small λ indicates that λ does not need to be specifically tuned but only small enough. For the next experiments, we choose λ to obtain the best support recovery.

3.2.3. Performances

We compare our approach (GSP) with other state-of-the-art methods: Subspace Pursuit (SP) [11], SAFIR [9] (with the parameters from [7]), MSVST [6] and a convex ℓ_1 -relaxation of (P3) i.e. a procedure minimizing the Poisson likelihood on a ℓ_1 -ball (using a projection onto the ℓ_1 -ball [14, 15]).

We apply these methods on two images, a sparse version of the Cameraman (see 3.2.2) and the NGC 2997 galaxy (peak intensity at 5, see Fig. 2). We use the exact parameters (sparsity for GSP and SP and ℓ_1 -norm for the ℓ_1 -relaxation) for the sparse cameraman and tune them for the Galaxy. For each method we compute the PSNR, the mean absolute deviation (MAE) and the SSIM [16] and display them in Table 1.

The MSVST gives the best results in terms of PSNR and MAE and SAFIR does so for the SSIM. GSP is competitive with both these state-of-the-art methods in Poisson denoising. Moreover GSP is in both cases better than SP and the ℓ_1 -relaxation. While the performance over SP shows the benefits of using non-quadratic cost in the case of high Poisson noise level, the performance over the ℓ_1 -relaxation emphasizes that using the exact sparse ℓ_0 a-priori enables to recover a better dynamic than its convex relaxation counterpart. These remarks are visually confirmed for the galaxy (Fig. 2).

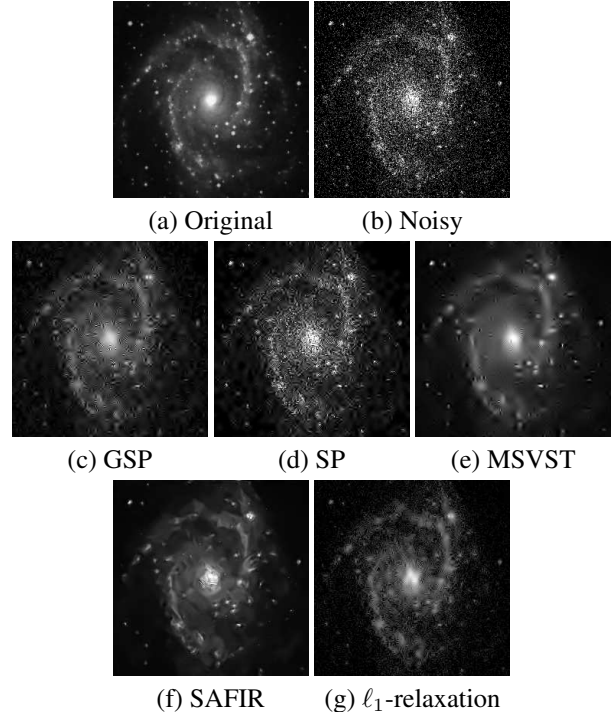


Fig. 2. NGC 2997 Galaxy image (a), a noisy version (b) and several denoising results (c-g).

4. CONCLUSION

In this paper we proposed a greedy algorithm, *Generalized Subspace Pursuit*, seeking a k -sparse minimizer for a generic cost function f . We exhibited a property on the gradient of f , the *Restricted Diagonal Property* that controls its convergence. Although hard to verify in practice, this property opens the way for using greedy algorithms on non-convex functions.

A method of Poisson denoising is proposed using GSP on the Moreau-Yosida regularization of the Poisson likelihood. The experiments show that it is competitive to state-of-the-art methods and improves over classical sparse approaches.

Future work on the theoretical side include improving the constants in the theorem, which seem quite conservative in view of the conducted experiments; studying possible links to notions used in the linear case, such as coherence; and adapting the analysis to the co-sparse [17] case. One may also want to reduce the complexity of GSP by removing one of the minimization steps, preserving the last one (step 8), as is done in OMP [12].

A further analysis of the regularization of the Poisson likelihood needs to be conducted to ensure the convergence of the proposed denoising method. Furthermore, it would be interesting to consider learning a dictionary to improve the quality of the results. However, such dictionaries are not likely to be tight frames or even frames and would make the previous analysis even more complex.

5. REFERENCES

- [1] T. Zhang, “Sparse recovery with orthogonal matching pursuit under RIP,” *Information Theory, IEEE Trans. on*, vol. 57, no. 9, pp. 6215–6221, 2011.
- [2] S. Bahmani, B. Raj, and P. Boufounos, “Greedy sparsity-constrained optimization,” *J. of Machine Learning Research*, vol. 14, no. 3, pp. 807–841, 2013.
- [3] 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.
- [4] 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.
- [5] F. X. Dupé and S. Anthoine, “A greedy approach to sparse poisson denoising,” in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2013.
- [6] 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.
- [7] 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.
- [8] P. L. Combettes and J.-. Pesquet, “A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,” *IEEE J. Selec. Top. Sig. Pro.*, vol. 1, no. 4, pp. 564–574, 2007.
- [9] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero, “Patch-based nonlocal functional for denoising fluorescence microscopy image sequences,” *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 442–454, 2010.
- [10] 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.
- [11] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [12] P. Jain, A. Tewari, and I. S. Dhillon, “Orthogonal matching pursuit with replacement,” in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, Eds., pp. 1215–1223. 2011.
- [13] 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.
- [14] 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*, vol. 20, no. 2, pp. 307–330, 2012.
- [15] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, “Epigraphical projection and proximal tools for solving constrained convex optimization problems: Part i,” *arXiv preprint arXiv:1210.5844*, 2012.
- [16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *Image Processing, IEEE Trans. on*, vol. 13, no. 4, 2004.
- [17] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, “The cosparsity analysis model and algorithms,” *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30 – 56, 2013.