

# High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation

Simon Gazagnes, Emmanuel Soubies, Laure Blanc-Féraud

► **To cite this version:**

Simon Gazagnes, Emmanuel Soubies, Laure Blanc-Féraud. High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation. ISBI 2017 - IEEE International Symposium on Biomedical Imaging, Apr 2017, Melbourne, Australia. pp.4, Proceedings of ISBI 2017. <hal-01443565>

**HAL Id: hal-01443565**

**<https://hal.inria.fr/hal-01443565>**

Submitted on 23 Jan 2017

**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.

# HIGH DENSITY MOLECULE LOCALIZATION FOR SUPER-RESOLUTION MICROSCOPY USING CEL0 BASED SPARSE APPROXIMATION

Simon Gazagnes<sup>(1)</sup>, Emmanuel Soubies<sup>(2)</sup>, Laure Blanc-Féraud<sup>(2)</sup>

(1) Université Côte d’Azur, INRIA, CNRS, I3S, France

(2) Université Côte d’Azur, CNRS, I3S, France

## ABSTRACT

Single molecule localization microscopy has made great improvements in spatial resolution achieving performance beyond the diffraction limit by sequentially activating and imaging small subsets of molecules. Here, we present an algorithm designed for high-density molecule localization which is of a major importance in order to improve the temporal resolution of such microscopy techniques. We formulate the localization problem as a sparse approximation problem which is then relaxed using the recently proposed CEL0 penalty, allowing an optimization through recent nonsmooth nonconvex algorithms. Finally, performances of the proposed method are compared with one of the best current method for high-density molecules localization on simulated and real data.

**Index Terms**— Super-resolution, Molecules localization, PALM, STORM,  $\ell_0$ -optimization, CEL0 relaxation.

## 1. INTRODUCTION

In conventional microscopy techniques, the spatial resolution of an image is limited by the diffraction phenomena. Recent methods like photo-activated localisation microscopy (PALM) [1, 2] and stochastic optical reconstruction microscopy (STORM) [3] allow high-precision molecule localization by sequentially *activating* and *imaging* a small random set of fluorescent molecules in the sample. This technique requires photoactivable or photoconvertible fluorescent probes with two states, let’s say for simplicity “Off” and “On”. Such molecules can be imaged only when they are in the “On” state and the activation step in PALM/STORM consists thus in changing the state to “On” of a small number of molecules in the sample which can then be imaged and located with high precision. The process is repeated until a maximum of molecules have been recovered. By assembling all the obtained molecule positions, we can construct a super-resolved image and go beyond the diffraction barrier. However, the quality of this super-resolved image is related to the density of emitters activated at each acquisition and the numerical method used to locate molecules.

Applications for these microscopy techniques are mainly restricted by the number of acquisitions needed to obtain the super-resolved image which cost time and memory. One way to overcome these limitations is to increase the density of emitters activated at each acquisition. Nevertheless, it will cause overlapping for a certain number of spots on the acquired image which makes the localization of the underlying molecules a harder task. For example, classical

localization methods using Gaussian fitting or centroid estimation to recover molecules positions [4], fail under such a high density configuration. Over the past few years, new methods have emerged to handle high density acquisitions using various techniques. Some of them are improved version of Gaussian fitting [5], while others use deconvolution under sparsity constraint [6]. A review of a large number of recent methods can be found in [7].

In the present paper, we are concerned in molecule localization for such high density settings. We propose to perform the localization by solving a  $\ell_0$ -penalized least squares criteria through the minimization of the recently proposed CEL0 relaxation [8]. After formalizing the optimization problem, presenting the CEL0 relaxation with its properties and describing the used algorithm for the minimization task, we use simulated and real dataset to compare the proposed method to another high density localization software, DAOSTORM [5].

## 2. A SPARSE APPROXIMATION PROBLEM

### 2.1. Acquisition system modelling

Let  $Y \in \mathbb{R}^{N \times N}$  be one PALM/STORM acquisition following a molecules activation. Then, our goal is to locate the underlying activated molecules with a high precision. To this end, we will estimate their positions on a  $L^2$  thinner grid  $X \in \mathbb{R}^{NL \times NL}$ , defined by discretizing each pixel of the acquisition grid  $Y$  in  $L \times L$  pixels. The linear acquisition process can then be formulated as,

$$Y = \mathcal{P}(M_L(H(X))), \quad (1)$$

where  $\mathcal{P}$  denotes the Poisson distribution,  $H: \mathbb{R}^{NL \times NL} \rightarrow \mathbb{R}^{NL \times NL}$  is a convolution operator with the Gaussian kernel  $H_{\text{ker}}$ :

$$H_{\text{ker}}(x, y) := \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left(-\frac{x^2 + y^2}{2\sigma_s^2}\right), \quad (2)$$

$\sigma_s$  denoting the spatial standard deviation, and  $M_L: \mathbb{R}^{NL \times NL} \rightarrow \mathbb{R}^{N \times N}$  is the downsizing operator which averages pixels by patches of size  $L \times L$  in order to map the high-resolved grid to the coarser one. Formally, we have,

$$M_L(X) := MXM^t \quad (3)$$

where the matrix  $M \in \mathbb{R}^{N \times NL}$  is given by:

$$M := \begin{bmatrix} \mathbf{1}_L & \mathbf{0}_L & \cdots & \mathbf{0}_L \\ \mathbf{0}_L & \mathbf{1}_L & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_L & \cdots & \cdots & \mathbf{1}_L \end{bmatrix} \quad (4)$$

Authors are belonging to MORPHEME team. This work was supported by the French Government (National Research Agency) through the Investments for the Future LABEX SIGNALIFE : program reference ANR-11-LABX-0028-01.

with  $\mathbf{1}_L \in \mathbb{R}^{1 \times L}$  (resp.  $\mathbf{0}_L \in \mathbb{R}^{1 \times L}$ ) a vector of ones (resp. zeros). Then we propose to estimate  $X$  from the acquired data  $Y$  by solving the following  $\ell_0$ -regularized least squares problem in order to impose sparse solutions:

$$\hat{X} \in \arg \min_{X \in \mathbb{R}^{NL \times NL}} \frac{1}{2} \|\mathbb{M}_L(H(X)) - Y\|_2^2 + \lambda \|X\|_0 + i_{\geq 0}(X), \quad (5)$$

where  $\|\cdot\|_0$  stands for the  $\ell_0$ -pseudo-norm defined by<sup>1</sup>

$$\|X\|_0 := \sum_{i=1}^{(NL)^2} |X_i|_0 \quad \text{with} \quad |u|_0 := \begin{cases} 0 & \text{if } u = 0, \\ 1 & \text{if } u \neq 0, \end{cases} \quad (6)$$

$\lambda > 0$  is a parameter allowing a trade-off between data fidelity and sparsity, and  $i_{\geq 0}$  is a positivity constraint (since we are reconstructing molecules intensities) given by:

$$i_{\geq 0}(X) := \begin{cases} 0 & \text{if } X \in \mathbb{R}_+^{NL \times NL}, \\ +\infty & \text{otherwise.} \end{cases} \quad (7)$$

We denotes the cost function in (5) by  $G_{\ell_0}$  and its minimization is known to be NP-Hard [9]. However, there exists suboptimal methods with local convergence guaranties. Among them, we find for instance greedy algorithms like *single best replacement* (SBR) [10], the well known *iterative hard thresholding* (IHT) algorithm [11] or *branch and bounds* algorithms combined with *cutting plane methods* [12] (global convergence). Except for IHT, the aforementioned methods are limited by the size of the considered problem. Another alternative consists in replacing the  $\ell_0$ -norm by a *continuous* nonconvex penalty allowing then the use of nonsmooth nonconvex optimization methods like the iteratively reweighted  $\ell_1$  algorithm (IRL1) [13]. In this paper, we will consider the recently proposed *continuous exact  $\ell_0$*  (CEL0) penalty [8], briefly presented in the next section, which combined with the IRL1 algorithm have been shown to better minimize  $G_{\ell_0}$  than IHT [8].

## 2.2. The CEL0 penalty

From [8], we consider the following continuous relaxation of  $G_{\ell_0}$ :

$$G_{\text{CEL0}}(X) := \frac{1}{2} \|\mathbb{M}_L(H(X)) - Y\|_2^2 + \Phi_{\text{CEL0}}(X) + i_{\geq 0}(X), \quad (8)$$

where  $\Phi_{\text{CEL0}}$  denotes the CEL0 penalty defined by

$$\Phi_{\text{CEL0}}(X) := \sum_{i=1}^{(NL)^2} \phi_{\text{CEL0}}(\|\mathbb{M}_L(H(E^i))\|, \lambda; |X_i|), \quad (9)$$

with, for  $a > 0$ ,  $\lambda > 0$  and  $u \in \mathbb{R}_+$ ,

$$\phi_{\text{CEL0}}(a, \lambda; u) := \lambda - \frac{a^2}{2} \left( u - \frac{\sqrt{2\lambda}}{a} \right)^2 \mathbb{1}_{\{u \leq \frac{\sqrt{2\lambda}}{a}\}}. \quad (10)$$

In (9),  $E^i \in \mathbb{R}^{NL \times NL}$  is the  $i^{\text{th}}$  element of the standard  $\mathbb{R}^{NL \times NL}$  basis that is  $(E^i)_j = 1$  if  $j = i$  and  $(E^i)_j = 0$  otherwise. Hence,  $\|\mathbb{M}_L(H(E^i))\|$  denotes the norm of the  $i^{\text{th}}$  ‘‘column’’ of the linear operator  $\mathbb{M}_L(H(\cdot))$ .

The interest in minimizing  $G_{\text{CEL0}}$  in place of  $G_{\ell_0}$  comes from its *continuity* and the following properties [8]:

- $\arg \min_{X \in \mathbb{R}^{NL \times NL}} G_{\ell_0}(X) \subseteq \arg \min_{X \in \mathbb{R}^{NL \times NL}} G_{\text{CEL0}}(X)$ ,

<sup>1</sup>Elements of  $X$  are refereed by a single index  $i \in \{1, \dots, (NL)^2\}$ .

- if  $\hat{X}$  is a (local) minimizer of  $G_{\text{CEL0}}$ , then  $\hat{X}^0$ , defined by  $\forall i$

$$\hat{X}_i^0 := \hat{X}_i \mathbb{1}_{\{|\hat{X}_i| \geq \sqrt{2\lambda} / \|\mathbb{M}_L(H(E^i))\|\}}, \quad (11)$$

is a (local) minimizer of  $G_{\ell_0}$ .

- some local (not global) minimizers of  $G_{\ell_0}$  are not critical points of  $G_{\text{CEL0}}$  ( $G_{\text{CEL0}}$  removes some local minimizers of  $G_{\ell_0}$ ).

## 3. IRL1 ALGORITHM FOR $G_{\text{CEL0}}$ MINIMIZATION

The minimization of the *continuous* functional  $G_{\text{CEL0}}$  is addressed through the *iterative reweighted  $\ell_1$*  (IRL1) algorithm [13] belonging to the class of majorization-minimization (MM) methods. The main idea relies on the minimization of a series of *convex* weighted- $\ell_1$  penalized functionals, majorizing  $G_{\text{CEL0}}$  and being equal to it at the current solution. At iteration  $n \in \mathbb{N}$  (with current solution  $X^n$ ), the weights  $\omega_i^{X^n}$ ,  $i \in \{1, \dots, (NL)^2\}$ , are nothing else than the derivative of  $\phi_{\text{CEL0}}$  (right derivative at 0) at point  $|X_i^n|$ . Finally, the sequence  $(X^n)_{n \in \mathbb{N}}$  is generated as summarized in Algorithm 1.

---

### Algorithm 1: Iterative Reweighted $\ell_1$ algorithm [13]

---

**Input:**  $X^0 \in \mathbb{R}^{NL \times NL}$

**1 repeat**

$$2 \omega_i^{X^n} = \left( \sqrt{2\lambda} a - a^2 |X_i^n| \right) \mathbb{1}_{\{|X_i^n| \leq \frac{\sqrt{2\lambda}}{a}\}} \\ \text{with } a = \|\mathbb{M}_L(H(E^i))\|;$$

$$3 X^{n+1} \in \arg \min_{X \in \mathbb{R}^{NL \times NL}} \frac{1}{2} \|\mathbb{M}_L(H(X)) - Y\|_2^2 + \sum_{i=0}^{(NL)^2} \omega_i^{X^n} |X_i| + i_{\geq 0}(X);$$

**4 until convergence;**

**Output:**  $X^n$

---

Convergence of IRL1 to a critical point of  $G_{\text{CEL0}}$  is proven in [13] (since  $G_{\text{CEL0}}$  verifies the Kurdyka-Lojasiewicz property [8]). Finally, the convex minimization problem at line 3 of Algorithm 1 is solved using the well known FISTA [14].

## 4. NUMERICAL RESULTS

In this section, we compare our method with DAOSTORM [5], a super-resolution localization algorithm which is currently among the best<sup>2</sup> state of the art methods for high-density molecules estimation according to the 2013 IEEE ISBI Single-Molecule Localisation Microscopy (SMLM) challenge [7].

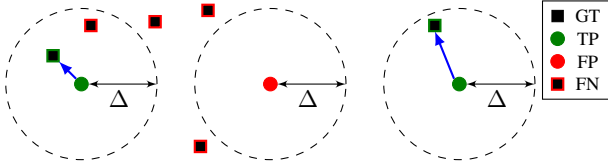
### 4.1. Performance evaluation

The performances are evaluated in terms of true positive (TP), false positives (FP) and false negatives (FN) molecules. To this end, a bipartite graph is constructed between ground-truth (GT) molecules and estimated ones allowing a matching between them which minimizes the sum of the distances between elements of each matched pair [7]. Moreover, an estimated molecule can be paired with a GT one only if the distance between them is lower than a tolerance  $\Delta$  as depicted on Figure 1. All paired estimated molecules are then categorized as TP, the remaining estimated molecules are defined as

<sup>2</sup>Note that in the ranking of [7] for high-density datasets, there is two better methods than DAOSTORM which are unfortunately not open access.

FP and the unpaired GT molecules are called FN. Finally, we will consider the Jaccard index defined by

$$\text{Jaccard index (\%)} := \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \times 100. \quad (12)$$



**Fig. 1.** Example of GT and estimated molecules pairing (blue arrows). Dashed circles represent tolerance discs of radius  $\Delta$  used in the matching. TP, FP and FN are also represented.

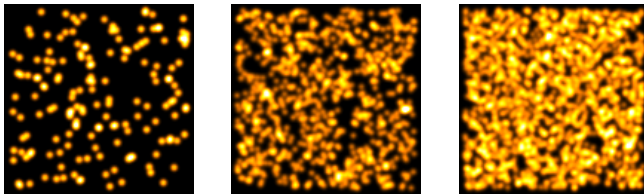
For both algorithm, we use different set of parameters (between 6 and 10 lambdas for IRL1, and approximately 500 combinations of two main parameters required by DAOSTORM) in order to obtain a large set of results. We initialize IRL1 with the “retro-projected” image  $X^0 = H^*(M_L^*(Y))$  where  $M_L^*$  and  $H^*$  are respectively the adjoint operators of  $M_L$  and  $H$ .

#### 4.2. Efficiency in function of density and SNR (simulation)

We first study the performance of the two methods in function of the density level and the signal-to-noise ratio (SNR) given by:

$$\text{SNR (dB)} := 10 \times \log_{10} \left( \frac{\|M_L(H(X^*))\|^2}{\|M_L(H(X^*)) - Y\|^2} \right), \quad (13)$$

where  $X^* \in \mathbb{R}^{NL \times NL}$  contains the true molecules locations. To this end, we create a  $512 \times 512$  image  $X^*$  with pixel size 25 nm. We maintain a 400 nm restricted area along the borders and we randomly distributed emitters on this image using five different densities (1, 2.5, 5, 7.5 and  $10 \mu\text{m}^{-2}$ ). We created a point spread function (PSF) based on the Gaussian model (2) with  $\sigma_s = 150$  nm. Then, simulated  $128 \times 128$  noisy acquisitions  $Y$  (pixel size of 100 nm) are obtained following model (1) with  $L = 4$ . For each density, a set of 20 random realisations of noise is generated for two different SNR (10 and 20 dB). Figure 2 shows some examples of generated data  $Y$ .

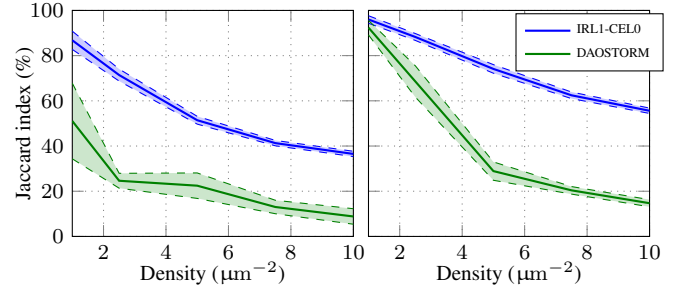


**Fig. 2.** Example of simulated acquisitions for three densities and SNR 20 dB. From left to right: density 1, 5 and  $10 \mu\text{m}^{-2}$ .

For each configuration (density, SNR, noise generation), both algorithms are run with different parameters and the Jaccard index is computed for the estimated molecules locations. Note that, since our method performs the localisation task on the same grid as the one used to simulate the molecules, the grid-free positions given by DAOSTORM are projected on this grid before to compute the Jaccard index in order to makes the results comparable. Here the size of the tolerance disc to compute the TP located molecules is set to 100 nm, that is a third of the full width at half maximum (FWHM) of the

used PSF ( $\approx 2.355\sigma_s = 350.25$  nm). Finally, only the result leading to the best Jaccard index among the tested algorithm parameters is retained for each simulated data.

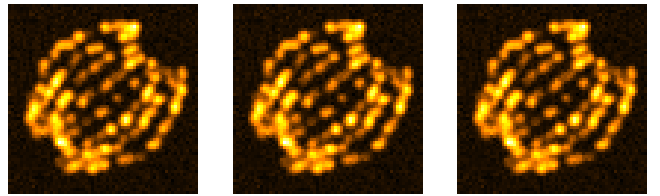
Figure 3 shows the evolution of this “best” Jaccard index (averaged over the 20 noise realizations) in function of the density of simulated molecules and for two different SNR levels. Clearly, the Jaccard index decreases as the density increases which is concordant with the increasing difficulty of the localization problem. Then, we can see that the proposed method performs better than DAOSTORM in terms of Jaccard index and that the gap between the two methods increases with the density of activated molecules on the acquisition.



**Fig. 3.** Jaccard index evolution (averaged over the 20 noise realizations) in function of the density of simulated molecules. Dashed curves represent the standard deviation. Left: SNR 10 dB. Right: SNR 20 dB.

#### 4.3. IEEE ISBI simulated dataset

The 2013 IEEE ISBI SMLM challenge is now a permanent challenge from which we can obtain realistic noisy simulated data with the corresponding ground truth localisation table<sup>3</sup>, allowing us to further compare our method with DAOSTORM. We downloaded a set of high density simulated acquisitions of a bundle of 8 simulated tubes of 30 nm diameter. For this simulation, the camera is composed by  $64 \times 64$  pixels of size 100 nm, the PSF is modelled by a Gaussian function whose FWHM equals to 258.21 nm, and the stack simulates 81049 emitters activated on 361 different frames. Figure 4 presents three different frames of the acquisition stack.



**Fig. 4.** Three frames of the downloaded simulated acquisition stack.

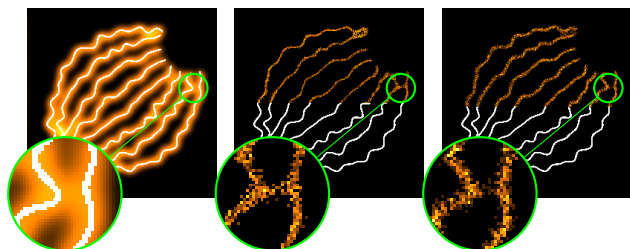
Performances are once again compared in terms of Jaccard index (12) computed on the whole reconstructed stack using the Java tool for evaluation available on the challenge website. Table 1 presents the best Jaccard index (obtained among the set of tested parameters for each algorithm) for several sizes of tolerance disc, varying from 50 nm to 250 nm. Note that these tolerances are all lower than the FWHM of the Gaussian PSF. We can see from Table 1 that  $G_{\text{CELO}}$  minimization with the IRL1 algorithm is more efficient than the DAOSTORM method. For IRL1-CELO, the reconstruction is performed on a  $256 \times 256$  pixels grid, corresponding thus to

<sup>3</sup><http://bigwww.epfl.ch/smlm/datasets/index.html>

$L = 4$  (pixel size of 25 nm) and the center of the detected pixels are retained as the estimated positions to compute the Jaccard index.

Method—Tolerance(nm)	50	100	150	200	250
IRL1-CEL0	<b>18.3</b>	<b>38.0</b>	<b>42.1</b>	<b>42.5</b>	<b>42.6</b>
DAOSTORM	3.0	13.4	22.4	27.5	30.5

**Table 1.** Best Jaccard index obtained (among the set of tested algorithm parameters) for the two methods and different tolerance discs for the reconstruction of ISBI simulated dataset.

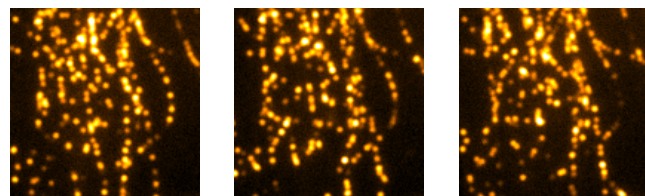


**Fig. 5.** Reconstruction results of the whole stack for the simulated dataset of Figure 4. From left to right: mean of the acquisition stack (*i.e.* wide field image) and ground truth in white, DAOSTORM reconstruction, and IRL1-CEL0 reconstruction for the best Jaccard index obtained with tolerance 250 nm. The bottom half part of the reconstructed images corresponds to the ground truth for comparison.

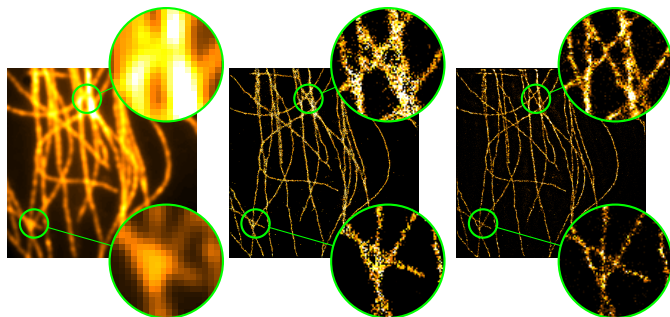
Figure 5 shows reconstruction results corresponding to the best Jaccard index for a 250 nm tolerance. For this representation, the DAOSTORM list of detected molecules positions is projected on the same grid as the one we use with the proposed method. As highlighted by the zooms, our method performs better to separate close structures and locate them with precision. However, one can notice that the reconstructed tubes by IRL1-CEL0 are a bit wider than those of the ground truth (also projected on the grid, in white on the figure) and those obtained with DAOSTORM. This can be explained by the fact that we are constrained to recover positions on the grid while they have been simulated using absolute coordinates.

#### 4.4. IEEE ISBI real dataset

Finally, we compare the two methods on a real high density dataset of tubulins also provided by the 2013 IEEE ISBI SMLM challenge. The acquisition stack contains 500 frames of  $128 \times 128$  pixels of size 100 nm (Figure 6). The FWHM of the system PSF was previously estimated at 351.8 nm [6] and we use it to construct our kernel (2). Then, the IRL1-CEL0 reconstruction is once again performed using a 4-downsizing operator (*i.e.*  $L = 4$ ) leading to a high resolution grid of  $512 \times 512$  pixels of size 25 nm. Figure 7 presents the reconstruction results where we can appreciate the better separation of the different tubulins obtained with the IRL1-CEL0 method.



**Fig. 6.** Three frames of the downloaded real acquisition stack of tubulins.



**Fig. 7.** Reconstruction results for the real dataset of Figure 6. From left to right: mean of the acquisition stack (*i.e.* wide field image), DAOSTORM reconstruction, and IRL1-CEL0 reconstruction.

## 5. CONCLUSION

In this paper, we presented a method designed for high density single molecules localization based on a sparse approximation formulation of the problem involving the  $\ell_0$  pseudo-norm. This NP-Hard combinatorial problem is then relaxed using the CEL0 penalty and the resulting continuous functional  $G_{\text{CEL0}}$  is minimized through the IRL1 algorithm. The performances of the proposed method are compared with the DAOSTORM algorithm, one among the best current methods for high-density localization. From both simulated and real dataset we showed that our method outperforms DAOSTORM in terms of Jaccard index and allows to reveal finer details of the acquired specimens. Moreover, we showed that the differences between the two tested methods increases with the density of molecules which makes the proposed method promising for very high density datasets.

## 6. REFERENCES

- [1] Eric Betzig et al., “Imaging intracellular fluorescent proteins at nanometer resolution,” *Science*, vol. 313, no. 5793, 2006.
- [2] Samuel T Hess et al., “Ultra-high resolution imaging by fluorescence photoactivation localization microscopy,” *Biophysical journal*, vol. 91, no. 11, 2006.
- [3] Michael J Rust et al., “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),” *Nature methods*, vol. 3, no. 10, 2006.
- [4] Ricardo Henriques et al., “Quickpalm: 3d real-time photoactivation nanoscopy image processing in imagej,” *Nature methods*, vol. 7, no. 5, 2010.
- [5] Seamus J Holden et al., “DAOSTORM: an algorithm for high-density super-resolution microscopy,” *Nature Methods*, vol. 8, no. 4, 2011.
- [6] Makhlad Chahid, *Echantillonnage compressif appliqué à la microscopie de fluorescence et à la microscopie de super résolution*, Ph.D. thesis, Bordeaux, 2014.
- [7] Daniel Sage et al., “Quantitative evaluation of software packages for single-molecule localization microscopy,” *Nature methods*, vol. 12, no. 8, 2015.
- [8] Emmanuel Soubies et al., “A Continuous Exact  $l_0$  Penalty (CEL0) for Least Squares Regularized Problem,” *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, 2015.
- [9] Balas Kausik Natarajan, “Sparse approximate solutions to linear systems,” *SIAM journal on computing*, vol. 24, no. 2, 1995.
- [10] Charles Soussen et al., “From Bernoulli–Gaussian deconvolution to sparse signal restoration,” *IEEE Transactions on Signal Processing*, vol. 59, no. 10, 2011.
- [11] Thomas Blumensath and Mike E Davies, “Iterative thresholding for sparse approximations,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, 2008.
- [12] Sébastien Bourguignon et al., “Exact Sparse Approximation Problems via Mixed-Integer Programming: Formulations and Computational Performance,” *IEEE Transactions on Signal Processing*, vol. 64, no. 6, 2016.
- [13] Peter Ochs et al., “On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision,” *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, 2015.
- [14] Amir Beck and Marc Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, 2009.