Weighted-CEL0 sparse regularisation for molecule localisation in super-resolution microscopy with Poisson data

Marta Lazzaretti, Luca Calatroni, Claudio Estatico

WEIGHTED-CEL0 SPARSE REGULARISATION FOR MOLECULE LOCALISATION IN SUPER-RESOLUTION MICROSCOPY WITH POISSON DATA

Marta Lazzaretti[1], *Luca Calatroni*[2] *and Claudio Estatico*[1]

[1] Università degli Studi di Genova, DIMA, Italy
[2] Université Côte d'Azur, CNRS, INRIA, I3S, France

ABSTRACT

We consider a variational model for Single Molecule Localisation Microscopy (SMLM) super-resolution. More specifically, we study a generalization of the Continuous Exact ℓ_0 (CEL0) penalty, recently introduced to relax the $\ell_2 - \ell_0$ problem, where a weighted-ℓ_2 data fidelity now models signal-dependent Poisson noise. For the numerical solution of the associated non-convex minimisation problem, we propose an iterative reweighted ℓ_1 (IRL1) algorithm, for which efficient parameter computation strategies are detailed. Both qualitative and quantitative molecule localisation results are reported, showing that the proposed weighted-CEL0 (wCEL0) model for Poisson noisy data improves the results obtained by CEL0 and state-of-the art deep-learning approaches for the high-density SMLM ISBI 2013 dataset.

Index Terms— Super-resolution, SMLM, ℓ_0-optimisation, Poisson noise, weighted-CEL0 relaxation.

1. INTRODUCTION

In fluorescence microscopy, Single Molecule Localisation Microscopy (SMLM) approaches (among which we mention PALM and STORM) allow to overcome the intrinsic limitations in optical resolution imposed by the light diffraction. SMLM techniques takes advantage of the absorption/emission properties of fluorescent molecules, which are sequentially activated and deactivated at random so as to limit the density of visible molecules in the sample. As a result, SMLM data consist of a stack of noisy and blurred images, whose individual frames represent sparse molecule samples, easier to analyse, which can be re-combined at a final stage to obtain the desired super-resolved image. In terms of localisation precision, the quality of the result strongly depends on the density of the molecules activated at each frame and most of the existing models fail whenever such value is too high (see [1] for a review).

In [2], the authors considered a $\ell_2 - \ell_0$-type continuous non-convex and sparsity-promoting variational model for super-resolution of SMLM high-density data. Such model had been previously studied in [3], where *exact* relaxation properties were shown to hold w.r.t. to the original, NP-hard, $\ell_2 - \ell_0$ model. The ℓ_2 data fidelity term considered in [2, 3] describes the presence of additive white Gaussian noise, although in [2] was shown to perform rather well also with Poisson distributed data, a more realistic scenario in biological imaging.

Inspired by [3], in this work we study a sparsity-promoting weighted $\ell_2 - \ell_0$-type model, accounting for signal-dependent Poisson noise in SMLM data. Our model approximates the Kullback-Leibler data fidelity functional corresponding to the Poisson negative log-likelihood as a weighted ℓ_2 data fidelity with local data intensity weights. Correspondingly, the CEL0-type associated penalty promotes sparsity depending both on the degradation model and local intensity information. To solve the corresponding composite non-convex optimisation problem, we consider an iterative-reweighted ℓ_1 algorithm and provide some algorithmic details regarding its (challenging) implementation. We validate our model on the high-density SMLM ISBI 2013 dataset and compare the results with CEL0 [2] and Deep-STORM [4] solutions.

2. WEIGHTED $\ell_2 - \ell_0$ OPTIMISATION

2.1. Inverse problem formulation

Let $\mathbf{y} \in \mathbb{R}^{M^2}_{>0}$ a vectorised $M \times M$ image acquired by means of a PALM/STORM technique and $\mathbf{x} \in \mathbb{R}^{N^2}_{\geq 0}$, with $N = LM$, the desired $N \times N$ image containing precise molecule localisations, defined on a L-times finer grid with $L \in \mathbb{N}$. The acquisition process can be described as:

$$\mathbf{y} = \mathcal{P}(\mathbf{R_L H x}),$$

where, for $\mathbf{z} \geq \mathbf{0}$, $\mathcal{P}(\mathbf{z})$ denotes the vector of realisations of Poisson random variables with parameters $z_i \geq 0$, $\mathbf{H} \in \mathbb{R}^{N^2 \times N^2}$ is the Block Circulant with Circulant Blocks (BCCB) matrix corresponding to the the two-dimensional periodic convolution with a specific Gaussian Point Spread Function (PSF) $\mathbf{h} \in \mathbb{R}^{N^2}$ and $\mathbf{R_L} \in \mathbb{R}^{M^2 \times N^2}$ is the down-sampling operator mapping the desired image from the fine grid to the coarser one. For shorthand notation, we further set $\mathbf{A} := \mathbf{R_L H} \in \mathbb{R}^{M^2 \times N^2}$.

For $\lambda > 0$, we consider the following non-convex sparsity-promoting model, for computing a sparse approximation of $\mathbf{x}$ given that the data $\mathbf{y}$ is Poisson-distributed:

$$\hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^{N^2}}{\arg\min}\ D_{KL}(\mathbf{A}\mathbf{x};\mathbf{y}) + \lambda\|\mathbf{x}\|_0 + i_{\geq 0}(\mathbf{x}), \quad (1)$$

where D_{KL} denotes the Kullback-Leibler fidelity term, which is derived via standard MAP estimation (see, e.g., [5]) and is defined for $\mathbf{v} \in \mathbb{R}^{M^2}_{>0}$ as $D_{KL}(\mathbf{v};\mathbf{y}) := \sum_{i=1}^{M^2} d_{KL}(v_i; y_i)$ with $d_{KL}(v_i; y_i) := v_i - y_i \log(v_i)$. Note that $\mathbf{y}$ is guaranteed to be strictly positive by adding a positive constant background term $\mathbf{b} = b\mathbf{1}_{M^2}$ where $0 < b \ll 1$ with $\mathbf{1}_{M^2}$ being the vector of all ones. The indicator function $i_{\geq 0}(\cdot)$ forces the desired solution $\mathbf{x}$ to be non-negative (since it represents molecule intensities which are indeed non-negative), while the regularisation term $\|\cdot\|_0$ denotes the N^2-dimensional ℓ_0 pseudo-norm defined by:

$$\|\mathbf{x}\|_0 = \sum_{i=1}^{N^2} |x_i|_0 \quad \text{with} \quad |x_i|_0 := \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{if } x_i = 0. \end{cases}$$

Dealing directly with the Kullback-Leibler functional D_{KL} above makes the problem quite challenging. To overcome such difficulties, we follow [5] and consider here a second-order Taylor approximation of $D_{KL}(\cdot;\mathbf{y})$ around $\mathbf{y}$ which leads to the following, symmetric weighted-ℓ_2 data term:

$$\begin{aligned} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2_{\mathbf{W}} &:= \frac{1}{2}\langle \mathbf{A}\mathbf{x}-\mathbf{y}, \mathbf{W}(\mathbf{A}\mathbf{x}-\mathbf{y})\rangle \\ &= \frac{1}{2}\sum_{i=1}^{M^2} \frac{((\mathbf{A}\mathbf{x})_i - y_i)^2}{y_i}, \end{aligned} \quad (2)$$

where the weighted norm is defined in terms of the diagonal, positive definite matrix $\mathbf{W} = \mathrm{diag}(\mathbf{1_{M^2}} ./\ \mathbf{y}) \in \mathbb{R}^{M^2\times M^2}$ and $\mathbf{1_{M^2}} ./\ \mathbf{y}$ denotes the Hadamard element-wise division between $\mathbf{1_{M^2}}$, defined above, and $\mathbf{y}$. This fidelity term can now be used in (1) as an approximation of D_{KL}. It weights locally the least-square discrepancy by the inverse intensity of the given low-resolution data. This choice thus enforces a large/low fidelity whenever low/high signal (corresponding to locally low/high noise) is measured, respectively.

Hence, instead of (1), we consider the following simplified weighted $\ell_2 - \ell_0$ problem:

$$\hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^{N^2}}{\arg\min}\ G_{w\ell_0} := \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2_{\mathbf{W}} + \lambda\|\mathbf{x}\|_0 + i_{\geq 0}(\mathbf{x}). \quad (3)$$

We remark that due to the presence of the ℓ_0 pseudo-norm, problems in the form (3) are known to be NP-hard. Several locally convergent methods can be alternatively used to solve these problems, such as, for instance, the Iterative Hard Thresholding (IHT) or branch and bounds algorithms, which, however, are highly dependent on the initialisation or hard to apply to large-scale data, respectively. To overcome this issue, in recent years a new class of continuous non-convex penalties has been studied (see, e.g., [3]), based also on the analytical properties of their local/global minimisers studied in [6]. The general idea for this type of penalties is to consider *continuous* non-convex relaxations of the ℓ_0 pseudo-norm obtained by repeated application of Fenchel conjugation. The continuity of the relaxed functional allows for the use of standard optimisation algorithms, such as, for instance, the iterative reweighted ℓ_1 (IRL1) algorithm [7]. We proceed similarly and consider a variation of the *continuous exact* ℓ_0 (CEL0) penalty introduced in [3] for the $\ell_2 - \ell_0$ problem, which is better suited to deal with the data term (2).

2.2. A weighted-CEL0 (wCEL0) penalty

To derive a continuous approximation of the non-convex functional $G_{w\ell_0}$ in (3), we follow the computations carried out in [3] to obtain its biconjugate functional by applying twice Fenchel conjugation (see [8] for the details).

We attain the following continuous relaxation of $G_{w\ell_0}$:

$$G_{\mathtt{wCEL0}}(x) := \frac{1}{2}||\mathbf{A}\mathbf{x}-\mathbf{y}||^2_{\mathbf{W}} + \Phi_{\mathtt{wCEL0}}(\mathbf{x};\lambda) + i_{\geq 0}(\mathbf{x}), \quad (4)$$

where, for $\lambda > 0$, $\Phi_{\mathtt{wCEL0}}(\cdot;\lambda)$ denotes the non-convex non-smooth continuous penalty defined by:

$$\Phi_{\mathtt{wCEL0}}(\mathbf{x};\lambda) := \sum_{i=1}^{N^2} \lambda - \frac{\|\mathbf{a}_i\|^2_{\mathbf{W}}}{2}\left(|x_i| - \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|_{\mathbf{W}}}\right)^2 \mathbb{1}_{\{|x_i| < \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|_{\mathbf{W}}}\}},$$

and $\mathbf{a}_i = (a_{j,i})_j \in \mathbb{R}^{M^2}$ denotes the i-th column of the operator $\mathbf{A}$. Here, the computation of the weighted norm $\|\mathbf{a}_i\|^2_{\mathbf{W}}$ contained in the expression of the penalty term $\Phi_{\mathtt{wCEL0}}$ encodes the dependence on the data $\mathbf{y}$ since, by definition:

$$\|\mathbf{a}_i\|^2_{\mathbf{W}} = \sum_{j=1}^{M^2} \frac{a^2_{j,i}}{y_j}. \quad (5)$$

Note that by setting $\mathbf{D} := \sqrt{\mathbf{W}}$ entry-wise, the fidelity term can be re-written as

$$\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2_{\mathbf{W}} = \|\mathbf{D}\mathbf{A}\mathbf{x}-\mathbf{D}\mathbf{y}\|^2_2 = \|\tilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{y}}\|^2_2, \quad (6)$$

where $\tilde{\mathbf{A}} = \mathbf{D}\mathbf{A}$ and $\tilde{\mathbf{y}} = \mathbf{D}\mathbf{y}$. This shows that we can obtain (4) by considering the CEL0 model associated with $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{y}}$. Thus, exact continuous relaxation properties of wCEL0 model follow from [3]. In particular, the global minima of $G_{w\ell_0}$ are also global minima of $G_{\mathtt{wCEL0}}$ and $G_{\mathtt{wCEL0}}$ eliminates some local minimisers of $G_{w\ell_0}$.

Remark (Comparison with CEL0)**.** *Compared to the Φ_{CEL0} penalty considered in [2, 3] for the standard $\ell_2 - \ell_0$ problem, the new penalty Φ_{wCEL0} presents an explicit dependence on both the model (i.e. the columns of the operator $\mathbf{A}$, as for CEL0) and the data $\mathbf{y}$. This reflects the intrinsic signal-dependence encoded into the considered Poisson modelling and, numerically, leads to the introduction of a threshold which is different for each component x_i of the solution (as it involves the i-th column of $\mathbf{A}$) and adapts to any data $\mathbf{y}$.*

3. ALGORITHMIC IMPLEMENTATION

We describe here the implementation of the model-dependent algorithmic parameters required to minimise the functional $G_{\texttt{wCEL0}}$ in (4). We follow [2] and consider the IRL1 algorithm whose pseudocode is reported in Algorithm 1. The presence of the weighted norm induced by $\mathbf{W}$ makes the numerical solution of (4) quite challenging.

Weighted column norms computation. The computation of the N^2 weighted norms (5) $\|\mathbf{a}_i\|_{\mathbf{W}}$ of the M^2-dimensional columns of the operator $\mathbf{A} = \mathbf{R_L H}$ is required for the computation of the penalty term $\Phi_{\texttt{wCEL0}}$. To do so, we proceed as follows. Since the operator $\mathbf{R_L}$computes down-sampling via the sum of $L \times L$ neighbourhood pixel values, it can be viewed as a restriction of the two-dimensional periodic convolution operator with kernel $\mathbf{k_L} \in \mathbb{R}^{N\times N}$ defined as

$$(\mathbf{k_L})_{i,j} = \begin{cases} 1, & i,j \in \{\frac{ML}{2} - \frac{L}{2} + 1, \cdots, \frac{ML}{2} + \frac{L}{2}\} \\ 0, & \text{otherwise,} \end{cases}$$

where L here is assumed even for simplicity. Indeed, by denoting by $\mathbf{K_L} \in \mathbb{R}^{N^2\times N^2}$ the BCCB matrix corresponding to the kernel $\mathbf{k_L}$, we can compute any matrix-vector product with $\mathbf{A} \in \mathbb{R}^{M^2\times N^2}$ by means of horizontal and vertical L-equispaced selections of the result of the matrix-vector product with the symmetric BCCB matrix $\mathbf{A_E} = \mathbf{K_L H} \in \mathbb{R}^{N^2\times N^2}$. In this way, although $\mathbf{A_E}$ is larger than the original $\mathbf{A} \in \mathbb{R}^{M^2\times N^2}$, it is fully BCCB, so that its usage only involves the two-dimensional FFTs of the kernels $\mathbf{k_L}$ and $\mathbf{h}$ and a $O(N^2 \log N)$ numerical complexity. The enlarged structured matrix $\mathbf{A_E}$ allows us to compute the weighted norms $\|\mathbf{a}_i\|_{\mathbf{W}}$. Let us first insert the acquired $M \times M$ matrix image $\mathbf{Y}$ into the $N \times N$ matrix $\tilde{\mathbf{Y}}$ as follows

$$(\tilde{\mathbf{Y}})_{i,j} = \begin{cases} (\mathbf{Y})_{p+1,q+1} & \text{if } i = 1 + Lp,\ j = 1 + Lq, \\ & \text{for } p,q = 0, \cdots, M-1 \\ 0 & \text{otherwise} \end{cases}$$

and let $\tilde{\mathbf{y}} \in \mathbb{R}^{N^2}$ be the vectorisation of $\tilde{\mathbf{Y}}$. The matrix-vector product $\mathbf{v} = \mathbf{A_E}^{\cdot 2}(\mathbf{1_{M^2}} ./ \tilde{\mathbf{y}})$, where $^{\cdot 2}$ denotes element-wise matrix square, gives the vector $\mathbf{v} \in \mathbb{R}^{N^2}$, whose i-th component is just the weighted sum of the square of the elements of the i-th column of $\mathbf{A}$, since the corresponding weights are the values $\mathbf{1_{M^2}} ./ \tilde{\mathbf{y}}$. These N^2 quantities just correspond to $\|\mathbf{a}_i\|^2_{\mathbf{W}}$, for $i = 1, \cdots, N^2$. We stress that the penalty term in [2] involves the computation of the euclidean norms $\|\mathbf{a}_i\|^2_2$, which can be computed by exploiting the structure of $\mathbf{A}$ in a straightforward way. When considering the weighted norms $\|\mathbf{a}_i\|^2_{\mathbf{W}} = \|\tilde{\mathbf{a}}_i\|^2_2$ the representation of the down-sampling operation as restricted convolution and the described procedure are crucial here for the actual (fast) computation of (5).

Backtracking. As efficient solver for the inner weighted-ℓ_1 problems of the IRL1 algorithm, we use a FISTA algorithm endowed with an adaptive backtracking strategy of the step-size $\tau \in (1, 1/\mathcal{L}]$ where $\mathcal{L} = \|\mathbf{A^T W A}\| \leq \|\mathbf{A}\|^2\|\mathbf{W}\|$ is the Lipschitz constant of the gradient of the data term (2), see [9]. An accurate estimation of $\mathcal{L}$ is indeed required to ensure good convergence properties. However, due to the sub-multiplicative property of $\|\cdot\|$, an estimate of the type $\|\mathbf{A}\|^2\|\mathbf{W}\| = L^2 \max(\mathcal{F}(\mathbf{h}))^2 \min(\mathbf{Y})^{-1}$ with $\mathcal{F}(\cdot)$ being the 2D FFT, tends, typically, to overestimate $\mathcal{L}$ due to the possible small values, very close to zero, assumed by $\mathbf{y}$. This thus corresponds to consider extremely small step-sizes, which may badly affect convergence speed. The use of an adaptive backtracking strategy providing at each iteration of the IRL1 inner loop a good estimate τ_k avoids this drawback.

Algorithm 1 Weighted CEL0 (wCEL0) via IRL1

Require: $\mathbf{y} \in \mathbb{R}^{M^2}, \mathbf{x}^0 \in \mathbb{R}^{N^2}, \lambda > 0$
  repeat
    update weights $\omega_i^{\mathbf{x^k}} \in \partial\Phi_{\texttt{wCEL0}}(\mathbf{x^k}; \lambda)$
    $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2_{\mathbf{W}} + \lambda \sum_{i=1}^{N^2} \omega_i^{\mathbf{x^k}} |x_i|$
  until convergence
  return $\mathbf{x}$

Parameters. We initialise the IRL1 algorithm for both CEL0 and wCEL0 models by choosing $\mathbf{x^0} = \mathbf{A^T y}$ and assess convergence by a joint criterion based on the relative difference between consecutive iterates, the difference of their corresponding function values and a maximum number of iterations for both the inner and the outer loop. Finally, we consider an heuristic tuning of the regularisation parameter $\lambda > 0$ for both methods by averaging the parameters optimising the results for 8 randomly chosen temporal frames.

4. NUMERICAL RESULTS

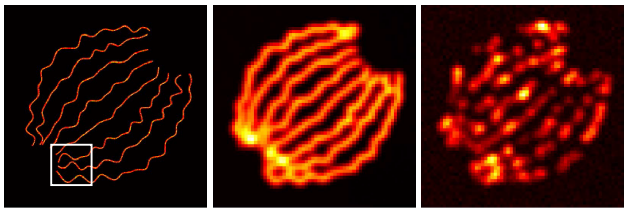
We test the proposed wCEL0 model on the high-density ISBI SMLM 2013 dataset, composed of 361 images representing 8 tubes of 30nm diameter. The size of each acquisition is 64×64 pixels with 100×100nm^2 pixels. We localise the molecules on a 256×256 pixel image corresponding to a factor $L = 4$ and a pixel size of 25×25nm^2. The total number of molecules is 81049 and 217 fluorophores are activated on average at each time acquisition. The Gaussian PSF has FWHM $= 258.2$nm. We report in Fig. 1a the ground-truth image, in Fig. 1b the sum of all acquisitions, and in Fig. 1c an example of a single frame from the dataset. We report in Figure 2 the solution computed by wCEL0 Algorithm 1 in comparison with the ones obtained by CEL0 [3] and by Deep-STORM [4], a deep-learning based model for super-resolution microscopy whose COLAB notebook[1] was used to generate an ad-hoc training data using the parameters above. For a quantitative assessment of localisation precision, up to some tolerance radius $\delta > 0$ we consider the Jaccard index $J_\delta := \frac{\text{CD}}{\text{CD+FN+FP}} \in [0, 1]$, where the number of correct de-

[1] https://github.com/EliasNehme/Deep-STORM

	J_0	J_2	J_4	CD	FN	FP
wCEL0	**0.057**	**0.552**	**0.659**	151	67	14
CEL0	0.042	0.467	0.552	121	96	**3**
Deep-STORM	0.025	0.037	0.038	**217**	**1**	8157

Table 1: Jaccard index for $\delta \in \{0, 2, 4\}$ and CD, FN, FP for $\delta = 4$ computed as mean over the frames.

tections is denoted by CD, of false negatives by FN and of false positives by FP, and then we take the average over the frames. We compute J_δ for each reconstruction and then we take the average over the frames. We test three different values of $\delta \in \{0, 2, 4\}$ corresponding to a tolerance of 0, 50 and 100 nm, respectively. Our results are reported in Table 1, which contains also the values of CD, FN and FP computed for each method for the case $\delta = 4$. We observe that the Jaccard values for wCEL0 are significantly better than the ones computed for both CEL0 and Deep-STORM. We observe that, while in terms of CD Deep-STORM outperforms the other methods, its reconstruction shows a large number of FP, as it can be observed in the close-ups in Figure 2. To solve this drawback (which would of course improve also the performance of CEL0 and wCEL0), post-processing techniques can be used.



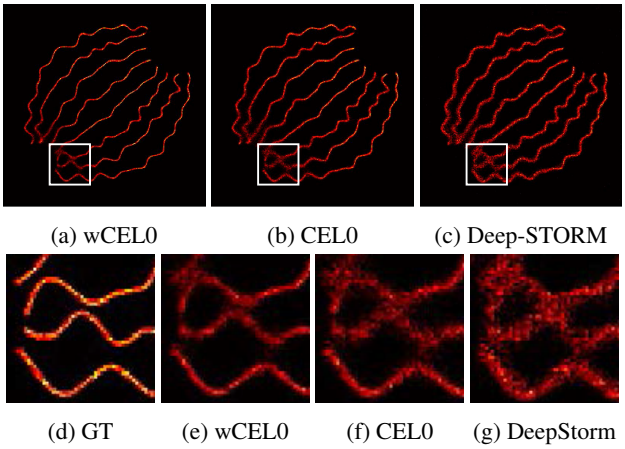
(a) GT (b) $\overline{\mathbf{y}}$ (c) Single frame

Fig. 1: ISBI SMLM 2013 dataset: (a) Ground Truth data, (b) sum of all acquisitions (x4), (c) single acquisition (x4).

5. CONCLUSIONS

We proposed a weighted $\ell_2 - \ell_0$ model for sparse super-resolution of high-density SMLM data, suited to model the presence of signal-dependent noise. We followed [3] and considered its continuous exact relaxation, defined in terms of a weighted-CEL0 penalty depending both on model parameters and observed data. The numerical solution of the weighted problem is challenging due to non-trivial to compute algorithmic parameters. To overcome these issues, we detail suitable estimation strategies allowing to solve the problem efficiently via IRL1 algorithm. Our numerical results show improvements in molecule localisation in comparison with standard CEL0 and Deep-STORM approaches.

Future research should address the case of general data fidelities, in order to deal directly with the case of non-symmetric terms, such as the Kullback-Leibler fidelity.



(a) wCEL0 (b) CEL0 (c) Deep-STORM

(d) GT (e) wCEL0 (f) CEL0 (g) DeepStorm

Fig. 2: **First-row**: (a) wCEL0 result, (b) CEL0 result, (c) Deep-STORM result. **Second row**: close-up on a detail.

6. COMPLIANCE WITH ETHICAL STANDARDS

This work was conducted using biological data from the SMLM ISBI 2013 dataset. Ethical approval was not required as confirmed by the license attached with the open access data.

7. ACKNOWLEDGEMENTS

ML and LC acknowledge the support of UCA IDEX JEDI. LC acknowledges the support received by UCA IDEX JEDI and by the NoMADS RISE H2020 project 777826. The authors thank L. Blanc-Féraud for useful discussions and suggestions. The authors have no relevant financial/non-financial interests to disclose.

8. REFERENCES

[1] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser, "Quantitative evaluation of software packages for single-molecule localization microscopy," *Nature methods, 12*, 2015.

[2] S. Gazagnes, E. Soubies, and L. Blanc-Féraud, "High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation," in *IEEE ISBI 2017*, 2017.

[3] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A continuous exact ℓ^0 penalty (CEL0) for least squares regularized problem," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, 2015.

[4] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, Apr 2018.

[5] J. Li, Z. Shen, R. Yin, and X. Zhang, "A reweighted l^2 method for image restoration with Poisson and mixed Poisson-Gaussian noise," *Inverse Probl. Imag.*, vol. 9, 2015.

[6] M. Nikolova, "Description of the minimizers of least squares regularized with ℓ_0 -norm. uniqueness of the global minimizer," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, 2013.

[7] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, 2015.

[8] M. Lazzaretti, "Continuous relaxation of sparse ℓ_0 optimisation problems in fluorescence microscopy with Poisson data," M.S. thesis, 2020, Università degli Studi di Genova.

[9] L. Calatroni and A. Chambolle, "Backtracking strategies for accelerated descent methods with smooth composite objectives," *SIAM Journal on Optimization*, vol. 29, no. 3, pp. 1772–1798, 2019.