New methods for $\ell\_2 - \ell\_0$ minimization and their applications to 2D Single-Molecule Localization Microscopy

Arne Bechensteen, Laure Blanc-Féraud, Gilles Aubert

NEW METHODS FOR $\ell_2 - \ell_0$ MINIMIZATION AND THEIR APPLICATIONS TO 2D SINGLE-MOLECULE LOCALIZATION MICROSCOPY

Arne Bechensteen, Laure Blanc-Féraud

Université Côte d'Azur
CNRS, INRIA, Laboratoire I3S UMR 7271
06903 Sophia Antipolis, France

Gilles Aubert

Universite Côte d'Azur, UNS
Laboratoire J. A. Dieudonné
UMR 7351, 06100 Nice, France

ABSTRACT

We present in this paper a biconvex reformulation of a $\ell_2 - \ell_0$ problem composed of a least-square data term plus a sparsity term introduced as a constraint or a penalization. Minimization algorithms are derived and compared with the state of the art in $\ell_2 - \ell_0$ minimization by relaxation or deep learning. Application results are shown on Single-Molecule Localization Microscopy.

Index Terms— Super-resolution, PALM, ℓ_0-optimization, CEL0 relaxation, Deep-STORM, IHT.

1. INTRODUCTION

In this paper we study the two following problems:

$$\arg\min_{x\in\mathbb{R}^{ML\times ML}} \frac{1}{2}\|Ax-d\|_2^2 + \iota_{\|\cdot\|_0\leq k}(x) + \iota_{.\geq 0}(x) \quad \text{(C)}$$

and

$$\arg\min_{x\in\mathbb{R}^{ML\times ML}} \frac{1}{2}\|Ax-d\|_2^2 + \lambda\|x\|_0 + \iota_{.\geq 0}(x) \quad \text{(P)}$$

where $A \in \mathbb{R}^{M\times M} \times \mathbb{R}^{ML\times ML}$, $L > 0$, $d \in \mathbb{R}^{M\times M}$. ι is the indicator function defined as

$$\iota_C(x) = \begin{cases} 0 \text{ if } x \in C \\ +\infty \text{ otherwise} \end{cases}$$

and the $\|\cdot\|_0$ designs the pseudo-norm which counts the number of non-zero components in a vector. Problem (C) is known as the constrained $\ell_2 - \ell_0$ problem, as the solution is constrained to have at maximum k non-zeros entries. Problem (P) is the penalized $\ell_2 - \ell_0$ problem, and λ is a positive trade-off parameter between the data fitting term and the sparsity term. Both problems are non-continuous, non-convex and NP-hard due to the nature of the ℓ_0 pseudo-norm. They have been greatly studied as they are important in applications such as the vast area of compressed sensing, variable selection, machine learning and, as we present in this paper, Single-Molecule Localization Microscopy (SMLM).

Among the approaches to resolve problems (C) and (P) we mainly find relaxations of the ℓ_0 pseudo-norm [1] and greedy algorithms [2].

In this paper we present a reformulation of problems (C) and (P), already introduced in [3], as well as a minimization algorithm. Numerical results are shown on SMLM ISBI challenge datasets and compared to state of the art algorithms which are based on relaxation and deep learning.

2. AN EXACT REFORMULATION OF THE ℓ_0 PSEUDO-NORM

In this section, following the work of [3], we introduce an auxiliary variable in order to rewrite the ℓ_0 pseudo-norm as a convex minimization problem.

Definition 1 *[3, Lemma 1] For any $x \in \mathbb{R}^N$*

$$\|x\|_0 = \min_{-\mathbf{1}\leq u\leq \mathbf{1}} \|u\|_1 \; s.t \; \|x\|_1 = <u,x> \quad \text{(1)}$$

By introducing the auxiliary variable u, the non-convex and non-continuous ℓ_0 pseudo-norm can now be written as a *convex* minimization problem. We can introduce the reformulation of the ℓ_0 pseudo-norm in the two problems (C) and (P), and obtain

$$\arg\min_{x,u} G(x,u) := \frac{1}{2}\|Ax-d\|_2^2 + \iota_{.\geq 0}(x) + I(u) \quad \text{(G)}$$

$$\text{s.t. } \|x\|_1 = <x,u>$$

where $I(u)$ is in the case of the constrained problem (C):

$$I_c(u) = \begin{cases} 0 \text{ if } \|u\|_1 \leq k \text{ and } -1 \leq u \leq 1 \\ \infty \text{ otherwise} \end{cases} \quad \text{(a)}$$

and in the penalized problem (P):

$$I_p(u) = \begin{cases} \lambda\|u\|_1 \text{ if } -1 \leq u \leq 1 \\ \infty \text{ otherwise} \end{cases} \quad \text{(b)}$$

Problem (G) is not convex with respect to (x, u) due to the constraint $\|x\|_1 = <x, u>$. However, the problem is biconvex, that is, the minimization of (G) with respect to each

variable, with the other fixed, is convex. We relax the constraint $\|x\|_1 = <x, u>$ by introducing the cost function $G_\rho(x, u) : \mathbb{R}^{ML \times ML} \times \mathbb{R}^{ML \times ML} \to \mathbb{R}$ defined as

$$G_\rho(x, u) = \frac{1}{2}\|Ax - d\|_2^2 + I(u) + \iota_{\cdot \geq 0}(x) + \rho(\|x\|_1 - <x, u>) \tag{2}$$

Assuming A is of full rank, if $\rho > \sigma(A)\|d\|_2$ where $\sigma(A)$ is the largest singular value of A, then (2) and (G) have the same local and global minimizers. This is an extension to the work in [3], where the data term is assumed Lipschitz continuous and the minimal bound is much larger. As the squared norm is not Lipschitz continuous, their work cannot be applied to (C) and (P).

In both cases, constrained (C) and penalized (P), we use the same main algorithm, which is described below.

Algorithm 1: Biconvex minimization

Input:
$\quad \rho^0 > 0$;

Initialisation:
$\quad x^0 = \mathbf{0} \in \mathbb{R}^{ML \times ML}; u^0 = \mathbf{0} \in \mathbb{R}^{ML \times ML};$

Repeat:
$\quad$ Solve problem G_{ρ^n}

$$\{x^{n+1}, u^{n+1}\} \in \arg\min G_{\rho^n}(x^n, u^n) \tag{3}$$

$\quad$ Update the penalty term

$$\rho^{n+1} = \min(\sigma(A)\|d\|_2, 2\rho^n) \tag{4}$$

Until: $\rho^{n+1} = \sigma(A)\|d\|_2$

Output: x^{n+1}

Note that we use the Proximal Alternating Minimization algorithm [4] to minimize (3).

3. APPLICATIONS & NUMERICAL RESULTS

Single-Molecule Localization Microscopy exploits photoactivable fluorescent molecules to obtain a resolution which surpasses the classical limit of optical microscopes. Photo-Activated Localization Microscopy (PALM) [5, 6] and Stochastic Optical Reconstruction Microscopy (STORM) [7] are both SMLM methods. They acquire a sequence of images where for each image only a sparse set of fluorescent molecules are activated. We can precisely locate the molecules for each acquired image since the probability of two molecules to be in the same diffraction disk is small. To reduce acquisition time, the density of activated molecules is increased on each acquired image, making the localization more difficult.

A discrete acquisition system model is presented in [8], and shows that 2D-SMLM can be modeled as a $\ell_2 - \ell_0$ minimization problem such as (C) or (P). The newly presented biconvex algorithm is therefore a good alternative for SMLM. Note that the acquisitions model presented in [8] is not the only model for SMLM, there exist algorithms which reconstruct the molecules in the continuous space (see [9, 10]). Other recent approaches are based on deep learning [11]. For a complete lecture on the SMLM and the different localization algorithms, see the ISBI-SMLM challenge [12].

In this section we compare different localization algorithms on two datasets, both accessible from the ISBI 2013 challenge [12]. The algorithms are compared to a simulated dataset where we can perform a numerical comparison to the ground truth and a real dataset where we can only compare visually the results. The biconvex algorithms are compared with two recent algorithms, the IRL1-CEL0 [8] and Deep-STORM [13], as well as the standard Iterative Hard Thresholding (IHT) [14]. The IRL1-CEL0 minimizes an exact relaxation [15] of the penalized formulation (P) and Deep-STORM is an algorithm that uses deep-learning and convolutional network to localize the molecules. We use the codes made public by the authors of Deep-STORM [16]. The IHT algorithm can be used to minimize the constrained formulation (C) and the penalized formulation (P). In this paper we limit the use of the IHT algorithm to the constrained formulation.
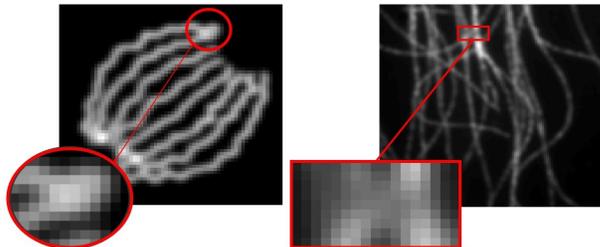

Fig. 1. The simulated and real dataset, all acquisitions summed

3.1. Parameters

The algorithms, except Deep-STORM, are based either on the formulation (P) or (C). The penalized formulation is mostly used when we do not have any information on the noise and the sparsity of the solution, and therefore we choose a trade-off parameter, λ, between the data fitting term and the sparsity term. The larger λ, the sparser the solution will be. There is no possible way to know the optimal λ to choose before running some simulations. An optimal λ is chosen by trial and error. The IRL1-CEL0 and the biconvex penalized algorithm are based on this model.

The constrained problem (C) is mostly used when the user has an idea of the number of non-zero elements to reconstruct.

The parameter to choose is the maximum number, k, of reconstructed molecules for each image. The acquisitions may give an idea of this number. In contrast to the parameter λ for the penalized case, we do not need to run any simulations to know the sparsity obtained. The IHT algorithm reconstructs exactly k molecules, while the biconvex constrained algorithm reconstructs possibly less.

Deep-STORM is a deep learning algorithm and the training involve a ℓ_1 penalization term with the trade-off parameter fixed to one. Thus Deep-STORM is parameter free. The algorithm needs, however, to be trained with an appropriate dataset with the right density and right level of noise. In contrast to the other algorithms, we cannot run one simulation on one acquisition to verify the choices, as Deep-STORM needs to train itself first which takes around two hours. However, note that the algorithm is faster to reconstruct the molecules than the other algorithms (see [13]) once the algorithm is trained with the right parameters.

3.2. Performance evaluation

The performance of the reconstruction can be evaluated with different criteria. In this paper we use the Jaccard index which evaluates the localization of the reconstructed molecules (see [12]). The Jaccard index is the ratio between the correctly reconstructed (CR) molecules and the sum of correctly reconstructed-, false negatives (FN)- and false positives (FP) molecules. The Jaccard index yields values from 0 to 1 and is defined as

$$Jac = \frac{CR}{CR + FP + FN}.$$

We also use the classical image comparison tool Signal to Noise Ratio (SNR) as this method compare both the localization and the intensity. The signal to noise ratio between the ground truth x and the reconstructed signal y is defined as

$$SNR(x,y) = 10 \log \frac{\|x\|^2}{\|x - y\|^2}.$$

3.3. ISBI simulated dataset

The ISBI simulated dataset represents 8 tubes of 30 nm diameter. The acquisition is simulated with a size of 64×64 pixels where each pixel is of size 100×100 nm^2. The PSF is modeled by a Gaussian function where the Full Width at Half Maximum (FWHM) is 258.21 nm. In total there are 81 049 molecules which are simulated on a total of 361 images.

We localize the molecules on a 256×256 pixel image, where the size of each pixel is 25×25 nm^2. This is equivalent of reconstructing an $x \in \mathbb{R}^{ML \times ML}$ for an acquisition $d \in \mathbb{R}^{M \times M}$, where $M = 64$ and $L = 4$. In the discrete reconstruction, the center of the pixel is used to estimate the position of the molecule in nm and each pixel which intensity is strictly larger than 0 is counted as a molecule.
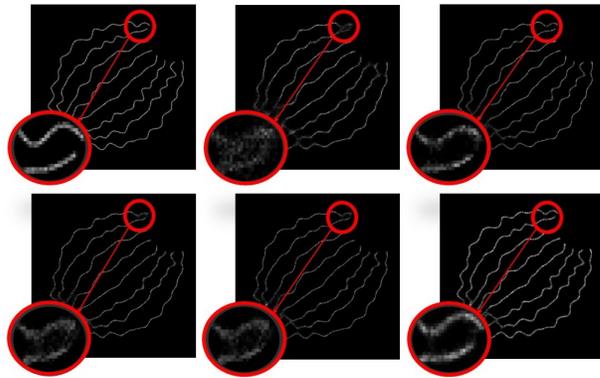

Fig. 2. Reconstructed of the simulated ISBI dataset. Top: From left to right: Ground truth, IHT, ILR1-CEL0. Bottom: From left to right: Biconvex constrained, Biconvex penalized and Deep-STORM
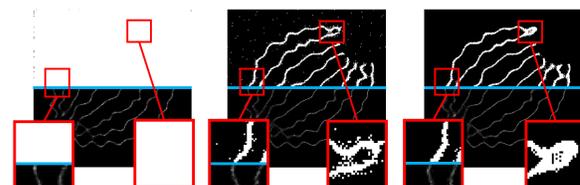

Fig. 3. Reconstructed of the simulated ISBI dataset. Deep-STORM, IRL1-CEL0 and biconvex constrained algorithm. Above blue line: Each molecule reconstructed is represented in white. Below blue line, normalized image.

We perform a reconstruction with the biconvex constrained algorithm first. We set k, the maximum number of molecules the algorithm reconstructs, equal to 220. This number is the average number of molecules for each acquisition, which we know from the ground truth. The biconvex constrained algorithm reconstructs on average 99 molecules per acquisition. We choose to reconstruct the same number of molecules for the other algorithms in order to properly compare the Jaccard index. Through trial and error we find $\lambda = 0.185$ for the IRL1-CEl0, $k = 100$ for IHT and for the biconvex penalized algorithm we obtain $\lambda = 100$. Deep-STORM is trained using the training set available at [16]. Note that we cannot control the sparsity of the solution using Deep-STORM and the algorithm is constructed to create an image. The reconstructions of the simulated ISBI dataset is shown in Figure 2. In order to observe the reconstruction, we normalize the image, that is, we let the smallest value in the image to be 0, and the largest to be 1. Each pixel has an intensity between 0 and 1, and the brighter the pixel the stronger the intensity. Figure 2 shows that IHT reconstructs less precisely than the other algorithms. Figure 3 illustrates the same reconstructions using Deep-STORM, IRL1-CELO and biconvex constrained algorithm with, above the blue line, each molecule reconstructed is displayed in white color (bi-

	Jaccard index (%)			
Method - Tolerance (nm)	50	100	150	200
IHT	20.1	35.9	40.4	41.3
IRL1-CEL0	**29.3**	**41.3**	42.4	42.6
Biconvex - Constrained	25.2	40.0	**43.2**	**43.9**
Biconvex - Penalized	25.0	39.3	42.2	42.8
Deep-STORM	×	×	×	×

Table 1. The Jaccard index with respect to the tolerance and the different algorithms. In bold is the best reconstruction.

Method	IHT	I-CEL0	B-C	B-P	D-STORM
SNR (dB)	1.76	**1.93**	1.39	1.45	0.44

Table 2. The Signal to Noise Ratio with respect to the reconstruction

nary image). Below the blue line is the normalized image. There is a clear difference of behavior between the biconvex constrained algorithm and IRL1-CEL0. The IRL1-CEL0 reconstructs false positive molecules that may be far from the tubelins, while the biconvex constrained algorithm reconstructs false positive molecules closer to the tubulins, but it makes the tubulins less resolved. Deep-STORM reconstructs a great number of molecules everywhere in the image with a concentration on the tubelins. That is why the image in Figure 3 is white over the blue line and why the Jaccard index cannot be computed (see Table 1). The intensity of the molecules which are not on the tubelins are very low so they are not visible in the normalized image in Figure 2 where Deep-STORM visually appears as the best result.

The IRL1-CEL0 algorithm has a high Jaccard index for a low tolerance compared to the other algorithms, which can be observed in Table 1. The biconvex constrained algorithm reconstructs better with respect to the Jaccard index with a tolerance larger than 150 nm. As Table 2 shows, the IRL1-CEL0 algorithm reconstructs the best among the five algorithms with respect to the SNR criterion. Deep-STORM has the lowest SNR value, which is the opposite of the results given in [13] where they use an equivalent numerical criterion. Again, this is due to the high number of reconstructed molecules. The Table 2 shows also the relatively high SNR number for the IHT algorithm, which is surprising, as this algorithm performs visibly worse than the other algorithms as seen in Figure 2. This may be explained by the fact that the localization of molecules is more important than their true intensity value.

3.4. ISBI real dataset

We compare the algorithms on a high-density dataset of tubulins which are provided from the 2013 ISBI SMLM challenge. There are 500 acquisitions, where each acquisition is of size 128×128 pixels and each pixel is of size 100×100 nm^2. The
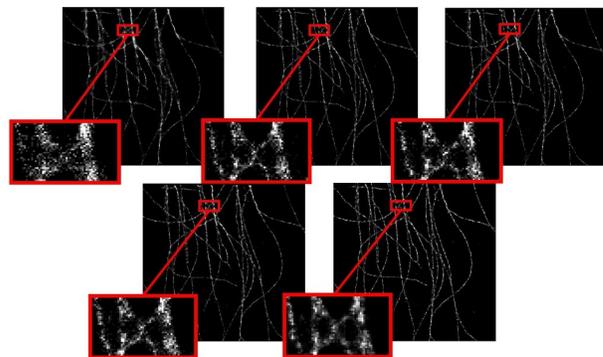

Fig. 4. Reconstructed of the real ISBI dataset. Fop: From left to right: IHT, ILR1-CEL0, Biconvex constrained. Bottom: From left to right: Biconvex penalized and Deep-STORM

FWHM has been estimated to be 351.8 nm [17]. We localize the molecules on a 512×512 pixel image, where each pixel is of size 25×25 nm^2.

In this section, we do not have any beforehand knowledge of the solution, and we set $k = 140$. In the case of IRL1-CEL0, we have chosen $\lambda = 0.5$, for the biconvex penalized algorithm we set $\lambda = 1200$ and for the algorithm IHT, we set the constraint $k = 100$. Deep-STORM is trained with a density of 1 molecule per μm^2. In Figure 4 we observe that IHT does not reconstruct as well as the other algorithms. Furthermore, each algorithm, apart from IHT, reconstructs the image with high precision, and for Deep-STORM we get a possibly slightly smoother reconstruction.

4. CONCLUSION

In this paper we have addressed the problem of minimizing an $\ell_2 - \ell_0$ problem. We have introduced two biconvex formulations for respectively the constrained and penalized form. We propose also an algorithm to numerically solve them. We test the algorithms and compare them with other algorithms on SMLM. We have shown that the parameter for the biconvex constrained algorithm is easy to regulate and that the algorithm localizes well with respect to the Jaccard index and visually. The IRL1-CEL0 algorithm reconstructs precisely with respect to the Jaccard index and with respect to SNR. The Deep-STORM algorithm reconstructs images precisely and it is fast once it has been trained with the right parameters.

5. REFERENCES

[1] Emmanuel Soubies, Laure Blanc-Féraud, and Gilles Aubert, "A unified view of exact continuous penalties for $\ell_2 - \ell_0$ minimization," *SIAM Journal on Optimization*, vol. 27, no. 3, pp. 2034–2060, 2017.

[2] Charles Soussen, Jérôme Idier, David Brie, and Junbo

Duan, "From Bernoulli–Gaussian deconvolution to sparse signal restoration," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4572–4584, 2011.

[3] Ganzhao Yuan and Bernard Ghanem, "Sparsity Constrained Minimization via Mathematical Programming with Equilibrium Constraints," *arXiv:1608.04430*, Aug. 2016.

[4] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.

[5] Samuel T Hess, Thanu PK Girirajan, and Michael D Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophysical journal*, vol. 91, no. 11, pp. 4258–4272, 2006.

[6] Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[7] Michael J Rust, Mark Bates, and Xiaowei Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nature methods*, vol. 3, no. 10, pp. 793, 2006.

[8] Simon Gazagnes, Emmanuel Soubies, and Laure Blanc-Féraud, "High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 28–31.

[9] Junhong Min, Cédric Vonesch, Hagai Kirshner, Lina Carlini, Nicolas Olivier, Seamus Holden, Suliana Manley, Jong Chul Ye, and Michael Unser, "FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data," *Scientific reports*, vol. 4, pp. 4577, 2014.

[10] Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[11] Nicholas Boyd, Eric Jonas, Hazen P Babcock, and Benjamin Recht, "Deeploco: Fast 3d localization microscopy using neural networks," *BioRxiv*, p. 267096, 2018.

[12] Daniel Sage, Hagai Kirshner, Thomas Pengo, Nico Stuurman, Junhong Min, Suliana Manley, and Michael Unser, "Quantitative evaluation of software packages for single-molecule localization microscopy," *Nature methods*, vol. 12, no. 8, pp. 717, 2015.

[13] Elias Nehme, Lucien E Weiss, Tomer Michaeli, and Yoav Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, 2018.

[14] Patrick L Combettes and Valérie R Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[15] E. Soubies, L. Blanc-Fraud, and G. Aubert, "A Continuous Exact ℓ_0 Penalty (CEL0) for Least Squares Regularized Problem," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1607–1639, Jan. 2015.

[16] "Software nano-bio-optics lab yoav shechtman, https://nanobiooptics.net.technion.ac.il/software/," .

[17] Makhlad Chahid, *Echantillonnage compressif appliqué à la microscopie de fluorescence et à la microscopie de super résolution*, Ph.D. thesis, Bordeaux, 2014.