Exact biconvex reformulation of the $\ell_2 - \ell_0$ minimization problem

Arne BECHENSTEEN[1], Laure BLANC-FÉRAUD[1], Gilles AUBERT[2]

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

[2]Université Côte d'Azur, UNS, Laboratoire J. A. Dieudonné UMR 7351,
06100 Nice , France

{abechens,blancf,gaubert}@unice.fr

Résumé – Nous étudions dans ce travail la minimisation du critère des moindres carrés sous une contrainte de parcimonie. Inspirés par des résultats récents, nous reformulons le pseudo-norme ℓ_0 comme une minimisation convexe par rapport à une variable auxiliaire. Nous proposons une reformulation biconvexe et exacte du problème de départ. Nous montrons qu'il y a une correspondance entre les minimiseurs du problème original et ceux du problème reformulé. La biconvexité suggère d'appliquer un algorithme de minimisation alternée. Ces résultats sont appliqués sur un problème d'imagerie super-résolue de molécules individuelles et nous comparons les résultats avec ceux donnés par l'algorithme IHT.

Abstract – We focus on the minimization of the least square loss function under a k-sparse constraint. Based on recent results, we reformulate the ℓ_0 pseudo-norm as a convex minimization problem by introducing an auxiliary variable. We then propose an exact biconvex reformulation of the $\ell_2 - \ell_0$ constrained problem. We give correspondence results between minimizers of the initial function and the reformulated one. The reformulation is biconvex which allows efficient alternating minimization methods to be used. The reformulation is tested numerically on Single Molecule Localization Microscopy and compared to IHT.

1 Introduction

In this paper, we are interested in the $\ell_2 - \ell_0$ constrained problem with a positivity constraint. We search for $\hat{x} \in \mathbb{R}^N$ that minimizes the cost function G_k:

$$G_k(x) := \frac{1}{2}\|Ax - d\|_2^2 + \iota_{\cdot \geq 0}(x) \text{ s.t. } \|x\|_0 \leq k \qquad (1)$$

where the observation $d \in \mathbb{R}^M$, the matrix $A \in \mathbb{R}^{M\times N}$, with $N > M$. $\|\cdot\|_0$ is the counting function, and is, by abuse of terminology, referred to as the ℓ_0-norm. The indicator function ι_X is defined for $X \subset \mathbb{R}^N$ as $+\infty$ if $x \notin X$ and 0 if $x \in X$, and we assume that the observation, d, is affected by some additive white Gaussian noise. The non-negativity constraint is commonly used as a priori in imaging problems.

The above problem is not continuous, nor convex and the problem is known to be NP-hard due to the combinatorial nature of the ℓ_0-norm. However, it has been greatly studied due to its countless applications such as sparse reconstruction of signals, variable selection, and single-molecule localization microscopy to cite a few. Among the approaches to solve the problem, we find greedy algorithms (see [10] and references therein), relaxations on the penalized problem (see [8] and references therein) and Mixed Integer Programming (MIP) on relatively small dimension problem. A more recent approach is Mathematical Program with Equilibrium Constraint (MPEC) (see [2, 11, 5, 9]).

The aim of this paper is to present and study a MPEC method for optimizing the constrained $\ell_2 - \ell_0$ problem (1). We start in section 2 by introducing a reformulation of the ℓ_0-norm by a variational characterization. The norm is rewritten as a convex minimization problem by introducing an auxiliary variable, and we can reformulate (1) as a MPEC problem. Then we define a Lagrangian cost function $G_\rho : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ which is biconvex, and thus easier to minimize than (1). Theorem 1 shows that minimizing G_ρ is equivalent, in the sense of minimizers, as to find a solution to the initial constrained problem.

Notations: Unless otherwise stated, for a matrix $A \in \mathbb{R}^{M\times N}$, we denote $\|A\|$ the spectral norm of A defined as $\|A\| = \sigma(A)$, where $\sigma(A)$ is the largest singular value of A. The subgradient of the convex function f at point x is the set of vectors v such that $\forall z \in dom(f)$, $f(z) \geq f(x) + v^T(z - x)$. For a vector $x \in \mathbb{R}^N$ we denote $|x| \in \mathbb{R}^N$, a vector containing the absolute value of each component of the vector x. The notation $-\mathbf{1} \leq u \leq \mathbf{1}$ is a component-wise notation, i.e, $\forall\ i$, $-1 \leq u_i \leq 1$.

2 Exact reformulation

In this section we focus on a reformulation of the ℓ_0-norm given in [9, 2]. The ℓ_0-norm can be rewritten as a convex minimization problem by introducing an auxiliary variable.

Lemma 1. *[9, Lemma 1] For any $x \in \mathbb{R}^N$*

$$\|x\|_0 = \min_{-\mathbf{1}\leq u \leq \mathbf{1}} \|u\|_1 \ s.t\ \|x\|_1 =< u, x > \qquad (2)$$

The introduction of the auxiliary variable u increases the dimension of the problem, but the non-convex and non-continuous ℓ_0-norm is now written as a *convex* minimization problem. In this paper, we study the $\ell_2 - \ell_0$ constrained problem (1) using the reformulation of the ℓ_0-norm, then written as:

$$\min_{x,u} \frac{1}{2}\|Ax-d\|^2 + I(u) + \iota_{\cdot \geq 0}(x) \text{ s.t. } \|x\|_1 =< x, u > \quad (3)$$

where $I(u)$ is:

$$I(u) = \begin{cases} 0 \text{ if } \|u\|_1 \leq k \text{ and } \forall\ i,\ -1 \leq u_i \leq 1 \\ \infty \text{ otherwise} \end{cases} \quad (4)$$

which is equivalent to problem (1). We note $\mathcal{S} = \{(x,u); \|x\|_1 =< x, u >\}$, and define the functional G as

$$G(x,u) = \frac{1}{2}\|Ax-d\|^2 + I(u) + \iota_{\cdot \geq 0}(x) + \iota_{\cdot \in \mathcal{S}}(x,u) \quad (5)$$

The functional (5) is continuous but non-convex due to the equality constraint. However it is biconvex: the minimization of (5) with respect to x while u is fixed is convex, and conversely. Based on Lagrange Multipliers method, the equality constraint in (5) is relaxed by introducing a penalty term, $\rho(\|x\|_1 - < x, u >)$. Note that it is not necessary to add the absolute value to this penalty term since $\forall\ i, |u_i| \leq 1$ and therefore the penalty term is never strictly negative.

The Lagrangian cost function $G_\rho(x,u) : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is defined as

$$G_\rho(x,u) = \frac{1}{2}\|Ax-d\|^2 + I(u) + \iota_{\cdot \geq 0}(x) + \rho(\|x\|_1 - < x, u >) \quad (6)$$

In this paper we are focusing on exact penalty methods, meaning that any minimizer of (6) is also a minimizer of (5). The following theorem ensures this.

Theorem 1. *Assume that $\rho > \sigma(A)\|d\|_2$, and A is of full rank. Let G_ρ and G be defined respectively in (6) and (5). We have:*

1. *If (x_ρ, u_ρ) is a local (respectively global) minimizer of G_ρ, then (x_ρ, u_ρ) is a local (respectively global) minimizer of G.*
2. *If $(\hat{x}, \hat{u})$ is a global minimizer of G, then $(\hat{x}, \hat{u})$ is a global minimizer of G_ρ.*

This theorem differs from [5, Corollary 3.2] as their ρ may be arbitrarily large in the case of the ℓ_2 data fitting term, and we can calculate ρ exactly. Furthermore, they work with a slightly different reformulation of the ℓ_0-norm and not explicitly with problem (1) since they assume their loss-function to be Lipschitzian.

Multiple lemmas are needed to prove Theorem 1. The proofs of Lemma 2, 3, 5 are omitted since they are obvious. For the proofs, see [12, Appendix].

Lemma 2. *Let $A \in \mathbb{R}^{M\times N}$, let a_i denote the ith column of A. Defining ω to be a set of indices, $\omega \subseteq \{1,\dots,N\}$. Let the restriction of A to the columns indexed by the elements of ω be denoted as $A_\omega = (a_{\omega[1]}, \dots, a_{\omega[\#\omega]}) \in \mathbb{R}^{M\times\#\omega}$. Then $\|A_\omega\| \leq \|A\|$.*

Lemma 3. *Given the problem*

$$\arg\min_x \frac{1}{2}\|Ax-d\|^2 + < w, |x| > + \iota_{\cdot \geq 0}(x) \quad (7)$$

where $A \in \mathbb{R}^{M\times N}$ is a full rank matrix and w a non-negative vector. Let $\hat{x}$ be a solution of problem (7). Then $\|A\hat{x}-d\|_2$ is bounded independently of w:

$$\|A\hat{x}-d\| \leq \|d\| \quad (8)$$

Lemma 4. *Let $f(x) = \frac{1}{2}\|Ax-d\|_2^2 + < w, |x| > + \iota_{\cdot \geq 0}(x)$, A be a full rank matrix and w is a non-negative vector. If $w_i > \sigma(A)\|d\|_2$ then the optimal solution of the following optimization problem:*

$$\hat{x} = \arg\min_x f(x) \quad (9)$$

is achieved with $\hat{x}_i = 0$.

Proof. It is clear from Lemma 3 that $\sigma(A)\|d\|_2 \geq \left|\left(A^T(A\hat{x}-d)\right)_i\right|$. A necessary and sufficient condition for $\hat{x}$ to be a minimizer of f on $\mathbb{R}^N_+$ is that

$$\mathbf{0} \in A^T(A\hat{x}-d) + \partial < w, |\hat{x}| > + N_{\mathbb{R}^d_+}(\hat{x})$$

where

$$(\partial < w, |\hat{x}| >)_i \begin{cases} = w_i \text{ if } \hat{x}_i > 0 \\ = -w_i \text{ if } \hat{x}_i < 0 \\ \in [-w_i, w_i] \text{ if } \hat{x}_i = 0 \end{cases}$$

and

$$(N_{\mathbb{R}^d_+}(\hat{x}))_i \begin{cases} = 0 \text{ if } \hat{x}_i > 0 \\ \in]-\infty, 0] \text{ if } \hat{x}_i = 0 \end{cases}$$

Since $\hat{x}_i \geq 0$ we get

$$-A^T(A\hat{x}-d)_i \begin{cases} = w_i \text{ if } \hat{x}_i > 0 \\ \in [-w_i, w_i] +]-\infty, 0] \text{ if } \hat{x}_i = 0 \end{cases}$$

If $w_i > \sigma(A)\|d\|_2$, then $w_i > |A^T(A\hat{x}-d)_i|$ and $\hat{x}_i$ cannot be strictly positive, furthermore $\hat{x}_i$ cannot be strictly negative since we work in the non-negative space. Therefore $\hat{x}_i = 0$. $\square$

Lemma 5. *Let (x_ρ, u_ρ) be a local minimizer of G_ρ defined in (6). Let $G_{x_\rho}(u) = \frac{1}{2}\|Ax_\rho - d\|^2 + I(u) + \rho(\|x_\rho\|_1 - < x_\rho, u >)$. We denote O as the indices of the k largest values of $\{|(x_\rho)_i|, i = 1...N\}$. $Q \triangleq \{i | (x_\rho)_i > 0\}$, and $S \triangleq \{j | (x_\rho)_j < 0\}$. Moreover, we define $D \triangleq O \cap Q$, $L \triangleq O \cap S$ and $W \triangleq \{1, 2..., N\} \backslash \{D \cup L\}$. If $\#(D \cup L) = k$, that is, $\|x_\rho\|_0 \geq k$, then the minimum of $G_{x_\rho}(u)$ will be reached with u_ρ such that*

$$(u_\rho)_i \begin{cases} = 1 \text{ if } i \in D \\ = -1 \text{ if } i \in L \\ = 0 \text{ if } i \in W \end{cases} \quad (10)$$

If $\#(D \cup L) < k$, that is, $\|x_\rho\|_0 < k$, then

$$(u_\rho)_i \begin{cases} = 1 \text{ if } i \in D \\ = -1 \text{ if } i \in L \\ \in [-1, 1] \text{ if } i \in W \end{cases} \quad (11)$$

with $\sum_{i \in W} |u_i| \leq k - \#(D \cup L)$.

Lemma 6. *Let $\rho > \sigma(A)\|d\|_2$. Let (x_ρ, u_ρ) be a local or global minimizer of $G_\rho(x,u) := \frac{1}{2}\|Ax-d\|^2 + I(u) + \rho(\|x\|_1 - < x, u >)$ with $I(u)$ defined as in (4). Let $\omega = \{i \in \{1,\dots,N\}; (u_\rho)_i = 0\}$. Then $(x_\rho)_i = 0\, \forall i \in \omega$.*

Proof. Let J denote the set of indices: $J = \{1,\dots,N\}\backslash\omega$. If (x_ρ, u_ρ) is a local or global minimizer of G_ρ then $\forall (x,u) \in \mathcal{N}((x_\rho,u_\rho),\gamma)$, where $\mathcal{N}((x_\rho,u_\rho),\gamma)$ denotes a neighborhood of (x_ρ,u_ρ) of size γ, we have $G_\rho(x_\rho,u_\rho) \le G_\rho(x,u)$. By choosing $u = u_\rho$ and $x = \tilde{x}$ with $\tilde{x}_J = (x_\rho)_J$ and $\tilde{x}_\omega = x_\omega$, with $(x_\omega,(u_\rho)_\omega) \in \mathcal{N}(((x_\rho)_\omega,(u_\rho)_\omega),\gamma)$, we get

$$\begin{aligned}\frac{1}{2}\|Ax_\rho - d\|^2 + &\iota_{\cdot\ge 0}(x_\rho) + \rho\|(x_\rho)_\omega\|_1 \\ &\le \frac{1}{2}\|A\tilde{x} - d\|^2 + \iota_{\cdot\ge 0}(\tilde{x}) + \rho\|x_\omega\|_1\end{aligned} \tag{12}$$

We want to show that $(x_\rho)_\omega$ is zero. We have

$$\begin{aligned}\|Ax-d\|^2 = \sum_i &\left[(\sum_{j\in J} A_{ij}x_j)^2 + (\sum_{j\in\omega} A_{ij}x_j)^2 \right] + \|d\|^2 \\ &- 2\left[\sum_{i\in J} x_i(A^Td)_i + \sum_{i\in\omega} x_i(A^Td)_i \right]\end{aligned}$$

Using the above decomposition simplifies (12), and $\forall\ x_\omega$:

$$\begin{aligned}&\frac{1}{2}\sum_i(\sum_{j\in\omega} A_{ij}(x_\rho)_J)^2 - \sum_{i\in\omega}(x_\rho)_i(A^Td)_i \\ &\quad + \rho\|(x_\rho)_\omega\|_1 + \iota_{\cdot\ge 0}(x_\rho) \\ &\le \frac{1}{2}\sum_i(\sum_{j\in\omega} A_{ij}x_j)^2 - \sum_{i\in\omega} x_i(A^Td)_i + \rho\|x_\omega\|_1 + \iota_{\cdot\ge 0}(x_\omega)\end{aligned}$$

Thus $(x_\rho)_\omega$ minimizes

$$\frac{1}{2}\sum_i(\sum_{j\in\omega} A_{ij}x_j)^2 - \sum_{i\in\omega} x_i(A^Td)_i + \rho\|x_\omega\|_1 + \iota_{\cdot\ge 0}(x_\omega)$$

or, equivalently, is a solution of

$$\arg\min_{x_\omega} \frac{1}{2}\|A_\omega x_\omega - d\|^2 + \rho\|x_\omega\|_1 + \iota_{\cdot\ge 0}(x_\omega) \tag{13}$$

where A_ω is the $M \times \#\omega$ submatrix of A composed by the columns indexed by ω of A. With Lemma 2, we have that $\sigma(A) \ge \sigma(A_\omega)$. If $\rho > \sigma(A)\|d\|_2$ we apply Lemma 4 with $w = [\rho \dots \rho]^T$. We conclude that $(x_\rho)_\omega = 0$. □

Lemma 7. *If $\rho > \sigma(A)\|d\|_2$, let (x_ρ, u_ρ) be a local or global minimizer of*

$$\arg\min_{x,u} \frac{1}{2}\|Ax-d\|^2 + \iota_{\cdot\ge 0}(x) + \rho(\|x\|_1 - < x, u >) + I(u)$$

with $I(u)$ defined as in (4). Then $\|x_\rho\|_1 - < x_\rho, u_\rho >= 0$.

Proof. From Lemma 5, we have that $(u_\rho)_i(x_\rho)_i = |(x_\rho)_i| \forall\ i \in J$, and $(u_\rho)_i = 0\, \forall i \in \omega$. It suffices to prove $(x_\rho)_i = 0\, \forall i \in \omega$. For that we use Lemma 6, and conclude that $(x_\rho)_\omega = 0$. □

With the above lemmas we can prove Theorem 1

Proof. We start by proving the first part of the theorem. Let (x_ρ, u_ρ) be a local minimizer of G_ρ. Let $\mathcal{S} = \{(x,u); \|x\|_1 = < x, u >\}$. If $\rho > \sigma(A)\|d\|_2$ then, from Lemma 7, (x_ρ, u_ρ) verifies

$$\|x_\rho\|_1 =< x_\rho, u_\rho > .$$

Furthermore, from the definition of a minimizer, we have

$$G_\rho(x_\rho,u_\rho) \le G_\rho(x,u)\ \forall (x,u) \in \mathcal{N}((x_\rho,u_\rho),\gamma)$$

and so we have

$$G_\rho(x_\rho,u_\rho) \le G_\rho(x,u)\ \forall (x,u) \in \mathcal{N}((x_\rho,u_\rho),\gamma) \cap \mathcal{S}$$

Since $\forall (x,u) \in \mathcal{S}, G_\rho(x,u) = G(x_\rho,u_\rho)$, we have

$$G(x_\rho,u_\rho) \le G(x,u)\ \forall (x,u) \in \mathcal{N}((x_\rho,u_\rho),\gamma) \cap \mathcal{S} \tag{14}$$

By the definition, (x_ρ, u_ρ) is also a local minimizer of G.

Now we prove part 2 of Theorem 1.

Let $(\hat{x},\hat{u})$ be a global minimizer of G. We necessarily have $\|\hat{x}\|_1 =< \hat{x}, \hat{u} >$. First, we show that

$$G_\rho(\hat{x},\hat{u}) \le \min G_\rho(x,u).$$

This can be shown by contradiction. Assume the opposite, and denote (x_ρ, u_ρ) a global minimizer of G_ρ. We then have

$$G_\rho(\hat{x},\hat{u}) > \min G_\rho(x,u) = G_\rho(x_\rho,u_\rho) \tag{15}$$

Lemma 7 shows that $\|x_\rho\|_1 =< x_\rho, u_\rho >$, so $G_\rho(x_\rho,u_\rho) = G(x_\rho,u_\rho)$ and we have

$$\begin{aligned}G(\hat{x},\hat{u}) = G_\rho(\hat{x},\hat{u}) &> \min G_\rho(x,u) \\ &= G_\rho(x_\rho,u_\rho) = G(x_\rho,u_\rho)\end{aligned}$$

and more precisely, $G(\hat{x},\hat{u}) > G(x_\rho,u_\rho)$ which is not possible, since $(\hat{x},\hat{u})$ is a global minimizer of G.

We have shown that $G_\rho(\hat{x},\hat{u}) \le \min G_\rho(x,u)$, and we have

$$G_\rho(\hat{x},\hat{u}) \le \min G_\rho(x,u) \le G_\rho(x,u) \quad \forall (x,u)$$

$(\hat{x},\hat{u})$ is thus a global minimizer of G_ρ. □

Theorem 1 shows that, for ρ large enough, minimizing (6) is equivalent in terms of minimizers as minimizing (5).

Although $G_\rho(x,u)$ in (6) is non-convex, the formulation is biconvex. An algorithm to minimize G_ρ can be easily implemented using Proximal Alternating Minimization algorithm [1]. Good results have been obtained starting with a small ρ ($\rho^0 = 1$) and then resolve G_ρ with an increasing ρ, using the results from the previous iteration as initialization. We stop the minimization when $\rho = \sigma(A)\|d\|_2$.

3 Numerical results

We compare the minimization of the biconvex reformulation to the algorithm Iterative Hard Thresholding [3] with an added non-negativity constraint to x. They are applied to the problem of 2D Single-Molecule Localization Microscopy (SMLM).

SMLM is a microscopy method which is used to obtain images with a higher resolution than what is possible with normal

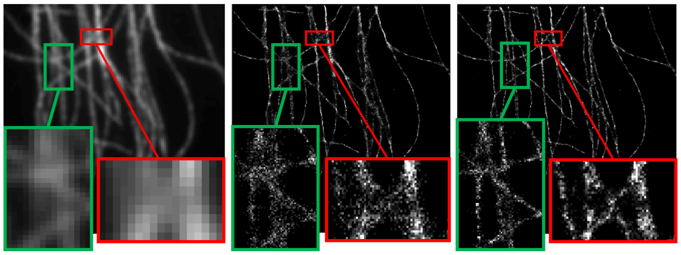

Figure 1: From left to right: Sum of the 500 acquisitions, IHT reconstruction and Biconvex constrained reconstruction.

optical microscopes (see [6]). SMLM exploits photoactivable fluorescent molecules, and for each acquisition activates only a sparse set of the fluorescent molecules in the sample. The localization of each molecule with a high precision is possible since the probability of two or more molecules to be in the same diffraction disk is small. The localization becomes harder if the density of emitting molecules is higher. Once each molecule has been precisely localized, they are switched off and the process is repeated until all the molecules have been activated. The final superresolved image (as in Fig.1) is the sum of all the sparse restored images (500 in the given example).

The localization problem of SMLM can be described as an optimization problem such as (1), with A as the operator that performs a convolution with the Point Spread function and a reduction of dimensions. Therefore, the biconvex formulation can be applied to the SMLM problem. The molecules are reconstructed on an $ML \times ML$ grid which is finer than the observed image $d \in \mathbb{R}^{M\times M}$, with $L > 1$. For a complete lecture on the mathematical model, see [4].

We compare the algorithms on a high-density dataset of tubulins which are provided from the 2013 ISBI SMLM challenge [7], where there are 500 acquisitions. Each acquisition is of size 128×128 pixels and each pixel is of size 100×100 nm^2. The FWHM is estimated to be 351.8 nm. We localize the molecules on a 512×512 pixel image, where each pixel is of size 25×25 nm^2.

We set the sparsity constraint $k = 140$ for the two algorithms. Figure 1 presents the reconstructions. We observe that the proposed algorithm distinguishes each tubelin correctly. For a more complete comparison with other state-of-the-art algorithms see [13].

4 Conclusion

In this paper, we have presented a reformulation of the $\ell_2 - \ell_0$ constrained problem. We have proved in Theorem 1 the exactness of the reformulation, that is, we can from a minimizer of the reformulation obtain a minimizer of the initial problem. Furthermore, the reformulation is biconvex. Numerically it performs well on the SMLM problem. The reformulation can also be applied to the penalized $\ell_2 - \ell_0$.

References

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[2] S. Bi, X. Liu, and S. Pan. Exact penalty decomposition method for zero-norm minimization based on MPEC formulation. *SIAM Journal on Scientific Computing*, 36(4):A1451–A1477, 2014.

[3] T. Blumensath and M E. Davies. Iterative thresholding for sparse approximations *Journal of Fourier Analysis and Applications, 14 (2008), pp. 629–654.*

[4] S. Gazagnes, E. Soubies, and L. 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*, pages 28–31. IEEE, 2017.

[5] Y. Liu, S. Bi, and S.Pan. Equivalent lipschitz surrogates for zero-norm and rank optimization problems. *Journal of Global Optimization*, pages 1–26, 2018.

[6] M. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature methods*, 3(10):793, 2006.

[7] 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(8):717, 2015.

[8] E. Soubies, L. Blanc-Féraud, and G. Aubert. A unified view of exact continuous penalties for \ell_2-\ell_0 minimization. *SIAM Journal on Optimization*, 27(3):2034–2060, 2017.

[9] G.Yuan and B. Ghanem. Sparsity Constrained Minimization via Mathematical Programming with Equilibrium Constraints. *arXiv:1608.04430*, August 2016.

[10] C. Soussen, J. Idier, D. Brie and D. Junbo. From Bernoulliian deconvolution to sparse signal restoration. *IEEE Transactions on Signal Processing*, 59(10) 4572–2484, 2011.

[11] A. d'Aspremont. A semidefinite representation for some minimum cardinality problems. *42nd IEEE International Conference on Decision and Control*, 5 4985–4990, 2003.

[12] A. Bechensteen, L. Blanc-Féraud, and G. Aubert. Research Report: Exact biconvex reformulation of the $\ell_2 - \ell_0$ minimization problem. *arXiv:1903.01162*, March 2019.

[13] A. Bechensteen, L. Blanc-Féraud, and G. Aubert. New methods for L_2-L_0 minimization and their applications to 2D Single-Molecule Localization Microscopy. *IEEE International Symposium on Biomedical Imaging 2019*, Apr 2019, Venice, Italy.