How to globally solve non-convex optimization problems involving an approximate ℓ 0 penalization

Arthur Marmin, Marc Castella, Jean-Christophe Pesquet

HOW TO GLOBALLY SOLVE NON-CONVEX OPTIMIZATION PROBLEMS INVOLVING AN APPROXIMATE ℓ_0 PENALIZATION

Arthur Marmin[†] *Marc Castella*[⋆] *Jean-Christophe Pesquet*[†]

[†]Center for Visual Computing, CentraleSupélec, INRIA, Université Paris-Saclay, Gif-sur-Yvette, France
[⋆]SAMOVAR, Télécom SudParis, CNRS, Université Paris-Saclay, Evry, France

ABSTRACT

For dealing with sparse models, a large number of continuous approximations of the ℓ_0 penalization have been proposed. However, the most accurate ones lead to non-convex optimization problems. In this paper, by observing that many such approximations are piecewise rational functions, we show that the original optimization problem can be recast as a multivariate polynomial problem. The latter is then globally solved by using recent optimization methods which consist of building a hierarchy of convex problems. Finally, experimental results illustrate that our method always provides a global optimum of the initial problem for standard ℓ_0 approximations. This is in contrast with existing local algorithms whose results depend on the initialization.

***Index Terms*—** polynomial and rational optimization, global optimization, ℓ_0 penalization, sparse modeling

1. INTRODUCTION

When rich statistical or variational models are employed, it is often necessary to select the most economical one (i.e. involving the fewest parameters) so as to avoid overfitting. In this context, compressive sensing has become a key solution to represent data in a sparse form in order to compress them efficiently and extract the salient information.

To promote sparse solutions and estimators, a common approach consists of adding an ℓ_0 penalization to a data-fit cost function [1]. Nevertheless, this is known to lead to NP hard optimization problems. Consequently, several surrogates to the ℓ_0 penalization have been suggested, the simplest one being the ℓ_1 norm. The latter has the enjoyable property of being convex, which simplifies the optimization task [2, 3], but it also strongly penalizes high values of the variables and thus introduces a bias in the solutions. Therefore further relaxations of ℓ_0 function have been investigated (see [4]). A major drawback is that those relaxations are non-convex and result in optimization problems which are difficult to solve globally. Most of the current available algorithms are limited in the sense that they only converge to local solutions due to the non-convexity of the cost function and they are therefore highly dependent on their initialization. The iterative hard thresholding [5] is a well-known example of such an algorithm that provides a local minimum for an ℓ_0 penalized cost function. The work in [6] suggests stronger optimality conditions to provide local minima close to the global one. Non-convex approximations to the ℓ_0 function that maintain the convexity of the overall cost function have also been proposed [7] to tackle this issue but they require specific assumptions to be met.

In this work, we propose a method to find the global minimum of a wide class of criteria involving non-convex approximations to ℓ_0 function [4, 8–12]. We show that such criteria are leading to optimization problems which are actually piecewise rational. We use the framework of Lasserre's hierarchy [13] which allows polynomial optimization problems to be globally solved. This framework has been extended to rational optimization problems in [14, 15]. Here, we further extend the method to find the global extrema of a piecewise rational optimization problem. We compare the obtained solutions with the ones provided by classical algorithms in the literature [4, 16–18]. In contrast with existing methods, we are able to find and certify the global optimum.

Our paper is organized as follows: Section 2 introduces the problem and gives examples of widely used approximations to ℓ_0 function. Section 3 explains how to recast a piecewise rational problem into a rational problem and how to apply it on our optimization problem. Simulation results are presented in Section 4. Finally, some concluding remarks are drawn in Section 5. In the following, the characteristic function of a given set $\mathcal{X}$ is denoted by $\mathbb{1}_{\{\cdot\in\mathcal{X}\}}$ with $\mathbb{1}_{\{x\in\mathcal{X}\}} = 1$ if $x \in \mathcal{X}$ and 0 otherwise.

2. PROBLEM STATEMENT

We tackle the minimization problem of the following composite criterion $\mathcal{J}$:

$$(\forall \mathbf{x} \in \mathbb{R}^T) \quad \mathcal{J}(\mathbf{x}) = f_{\mathbf{y}}(\mathbf{x}) + \mathcal{R}_\lambda(\mathbf{x})\,.$$

The optimization variable is thus a real-valued vector of dimension T, $f_{\mathbf{y}}$ is a fitting function depending on a vector of observations $\mathbf{y}$ and assumed to be rational. $\mathcal{R}_\lambda$ is a regularization that promotes sparsity and depends on a parameter

$\lambda \in]0, +\infty[$. Ideally, we would like $\mathcal{R}_\lambda$ to be the sparsity measure ℓ_0 but, in order to derive computationally efficient optimization techniques, a suitable separable approximation is substituted for it:

$$\mathcal{R}_\lambda(\mathbf{x}) = \sum_{t=1}^{T} \Psi_\lambda(x_t)\,.$$

The function $\Psi_\lambda : \mathbb{R} \to \mathbb{R}$ then requires the following three properties [4]: unbiasedness for large values, sparsity to reduce the complexity of the model, and continuity to ensure the stability of the model. Those conditions lead to non-convex functions. Indeed, the unbiasedness condition implies that the penalization is constant for large values of the variable and the sparsity and continuity conditions lead to a null value at zero. Below and displayed in Figure 1, we give examples of some of the most famous penalizations that satisfy the above three properties.

- Capped ℓ_p [8, 10, 11]:

  $$(\forall p \in \mathbb{N}\backslash\{0\}) \quad \Psi_\lambda(x) = |x|^p \mathbb{1}_{\{|x|\leq\lambda\}} + \lambda^p \mathbb{1}_{\{|x|>\lambda\}}\,.$$

- Smoothly clipped absolute deviation (SCAD) [4]:

  $$\begin{aligned}\Psi_\lambda(x) = \lambda\,|x|\,\mathbb{1}_{\{|x|\leq\lambda\}} \\ - \frac{\lambda^2 - 2\gamma\lambda\,|x| + x^2}{2(\gamma-1)}\mathbb{1}_{\{\lambda<|x|\leq\gamma\lambda\}} \\ + \frac{(\gamma+1)\lambda^2}{2}\mathbb{1}_{\{|x|>\gamma\lambda\}}\,,\end{aligned}$$

  where γ is a parameter taking values in $]2, +\infty[$.

- Minimax concave penalty (MCP) [9]:

  $$\Psi_\lambda(x) = \left(\lambda\,|x| - \frac{x^2}{2\gamma}\right)\mathbb{1}_{\{|x|\leq\gamma\lambda\}} + \frac{\gamma\lambda^2}{2}\mathbb{1}_{\{|x|>\gamma\lambda\}}\,,$$

  where $\gamma \in]0, +\infty[$.

- Continuous exact ℓ_0 (CEL0) [12]:

  $$\Psi_\lambda(x) = \lambda - \frac{\gamma^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\gamma}\right)^2 \mathbb{1}_{\left\{|x|\leq\frac{\sqrt{2\lambda}}{\gamma}\right\}}\,,$$

  where $\gamma \in]0, +\infty[$.

Note that the lower the parameter γ, the closer the approximation to ℓ_0 function but the stronger also the non-convexity. Since those functions are non-convex, the existing first or second-order optimization algorithms are only guaranteed to deliver a local minimizer of the criterion $\mathcal{J}$ [4, 9]. However, these methods do not exploit the fact that the previous penalizations are all piecewise polynomial functions. We show in Section 3 that the minimization of a piecewise rational function like $\mathcal{J}$ can reformulated as the minimization of a polynomial function under polynomial constraints. Then, the latter is relaxed into a hierarchy of semi-definite programming (SDP) problems that yields the global minimizers of the original function.

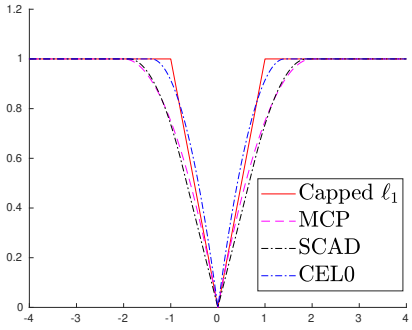

Fig. 1. Examples of continuous relaxation of ℓ_0 penalization ($\lambda = 1$, $\gamma_{\text{SCAD}} = 2.5$, $\gamma_{\text{MCP}} = 2$, $\gamma_{\text{CEL0}} = 1$).

3. RATIONAL FORMULATION

3.1. Reformulation of the considered class of penalties

Let us first show how to recast into a constrained rational optimization problem, the minimization of any piecewise rational function. Assume that this function reads

$$(\forall x \in \mathbb{R}) \quad \Psi(x) = \sum_{i=1}^{I} g_i(x)\mathbb{1}_{\{\sigma_{i-1}\leq x<\sigma_i\}}\,, \tag{1}$$

where $(g_i)_{1\leq i\leq I}$ are rational functions, I is a nonzero integer and $(\sigma_i)_{0\leq i\leq I}$ is an increasing sequence of real values. The function Ψ is our basic block for building the class of piecewise rational approximation to ℓ_0 function. All the functions discussed in the previous section can be rewritten under this form.

We introduce the binary variables $z^{(i)}$ such that

$$(\forall i \in [\![0, I]\!]) \quad z^{(i)} = \mathbb{1}_{\{\sigma_i\leq x\}}\,.$$

Note that we can possibly set $\sigma_0 = -\infty$, $z^{(0)} = 1$ and $\sigma_I = +\infty$, $z^{(I)} = 0$ to define Ψ on the whole real line $\mathbb{R}$. From the definition of $z^{(i)}$, we deduce that

$$\mathbb{1}_{\{\sigma_{i-1}\leq x<\sigma_i\}} = z^{(i-1)}(1 - z^{(i)})\,. \tag{2}$$

Finally, the constraint $z^{(i)} = \mathbb{1}_{\{\sigma_i\leq x\}}$ is equivalent to two polynomial constraints

$$z^{(i)} = \mathbb{1}_{\{\sigma_i\leq x\}} \Leftrightarrow \begin{cases} \left(z^{(i)}\right)^2 - z^{(i)} = 0 \\ \left(z^{(i)} - 0.5\right)(x - \sigma_i) \geq 0\,. \end{cases} \tag{3}$$

Indeed, the polynomial equality constraint enforces $z^{(i)}$ to be a binary variable while the polynomial inequality constraint ensures that it takes the same values as $\mathbb{1}_{\{\sigma_i\leq x\}}$. Therefore, by combining (1)-(3), we obtain the minimization of a rational function depending on x and $\left(z^{(i)}\right)_{0\leq i\leq I}$ under polynomial constraints. In turn, it has been shown in [14, 15] how rational problems can be reduced to polynomial ones.

3.2. Reformulation of our optimization problem

By applying the previous reasoning to penalization $\mathcal{R}_\lambda$, we reformulate the original minimization of $\mathcal{J}$. By introducing the binary variables $\mathbf{z} = (z_t^{(i)})_{0\leq i\leq I, 0\leq t\leq T}$, the optimization problem becomes

$$\begin{aligned} \underset{\mathbf{x},\mathbf{z}}{\text{minimize}} \quad & f_{\mathbf{y}}(\mathbf{x}) + \sum_{t=1}^{T}\sum_{i=1}^{I} g_i(x_t) z_t^{(i-1)}(1-z_t^{(i)}) \\ \text{s.t.} \quad & (\forall (i,t) \in [\![0,I]\!]\times[\![1,T]\!]) \\ & \left(z_t^{(i)}\right)^2 - z_t^{(i)} = 0 \\ & \left(z_t^{(i)} - 0.5\right)(x_t - \sigma_{i+1}) \geq 0\,. \end{aligned} \tag{4}$$

Problem (4) is a rational optimization problem since $f_{\mathbf{y}}$ and $(g_i)_{1\leq i\leq I}$ are rational by assumption and the constraints are polynomial. To find the global minimizers, we can thus exploit the problem structure and use the technique detailed in [15] which is based on Lasserre's method. It consists in relaxing a polynomial problem into a hierarchy of convex SDP problems indexed by an integer k corresponding to the relaxation order. Solving each SDP problem yields a lower bound $\mathcal{J}_k^*$ on the optimal value $\mathcal{J}^*$ of the criterion $\mathcal{J}$. Furthermore, the higher the order k, the tighter the bound $\mathcal{J}_k^*$ but the higher also the dimensions of the SDP problem. It has been proved [13] that $(\mathcal{J}_k^*)_{k\in\mathbb{N}}$ is an increasing convergent sequence whose limit is $\mathcal{J}^*$. Finally, the solution $\hat{\mathbf{x}}$ of the polynomial problem is extracted from the solution of the SDP problem [19] and we can theoretically certify that $\hat{\mathbf{x}}$ is a global minimizer by comparing $\mathcal{J}(\hat{\mathbf{x}})$ to $\mathcal{J}_k^*$.

3.3. Examples

To clarify the previous reformulation, we demonstrate it on the regularizations given in Section 2. Note that since the considered penalizations are even, we can use the resulting symmetry to decrease the number of variables and hence reduce the computations. Absolute values can be handled with an additional variable constrained to be nonnegative and whose square is equal to the square of the absolute value of the original variable [15]. Since SCAD has three pieces, it requires to introduce variables $z_t^{(1)}$ and $z_t^{(2)}$ so leading to

$$\begin{aligned} \underset{\mathbf{x},\mathbf{z}}{\text{minimize}} \quad & f_{\mathbf{y}}(\mathbf{x}) + \sum_{t=1}^{T}(1-z_t^{(1)})\lambda\,|x_t| \\ & - z_t^{(1)}(1-z_t^{(2)})\frac{\lambda^2 - 2\gamma\lambda\,|x_t| + x_t^2}{2(\gamma-1)} \\ & + z_t^{(2)}\frac{(\gamma+1)\lambda^2}{2} \\ \text{s.t.} \quad & (\forall (i,t) \in \{1,2\}\times[\![1,T]\!]) \quad \left(z_t^{(i)}\right)^2 - z_t^{(i)} = 0 \\ & (\forall t \in [\![1,T]\!]) \quad \left(z_t^{(1)} - 0.5\right)(|x_t| - \lambda) \geq 0 \\ & (\forall t \in [\![1,T]\!]) \quad \left(z_t^{(2)} - 0.5\right)(|x_t| - \gamma\lambda) \geq 0\,. \end{aligned}$$

A similar approach applies to Capped ℓ_p, MCP, and CEL0; the details are omitted for conciseness.

4. NUMERICAL SIMULATIONS

4.1. Simulation scenario

We consider the following degradation model:

$$\mathbf{y} = \mathbf{H}\overline{\mathbf{x}} + \mathbf{w}\,,$$

where $\mathbf{w}$ is a zero-mean white Gaussian noise and $\mathbf{H}$ is a Toeplitz band matrix corresponding to a Gaussian convolution filter. For each test, coefficients of the vector $\overline{\mathbf{x}}$ were randomly drawn according to a uniform distribution on $[0.6, 1]$. We set the size T of the vectors to 200. We defined the ideal criterion $\mathcal{J}_{\ell_0}$ as

$$\mathcal{J}_{\ell_0}(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{y} - \mathbf{H}\mathbf{x}\right\|^2 + \lambda\ell_0(\mathbf{x})\,,$$

where λ is set to 0.1.

Our goal is to approximate this criterion using Capped ℓ_1, SCAD, MCP, and CEL0 instead of ℓ_0 penalization. For each regularization, we compared out method to Forward-Backward (FB), Iteratively Reweighted ℓ_1 (IRL1) [16,17] and Coordinate Descent (CD) [18] algorithms. IRL1 minimizes iteratively the fitting function penalized with a weighted ℓ_1 norm whose weights are computed by linearizing the non-convex penalization function. CD solves a sequence of single-dimensional problems by fixing all the variables except one in the original problem.

We used GloptiPoly [20] to relax rational problems into SDP problems which are then solved with the solver SDPT3 [21]. The relaxation was performed at order k equal to 2 or 3. The parameter γ was set to 0.5 for MCP and 2.1 for SCAD. Following [12], we used the norm of the t-th column of $\mathbf{H}$ as parameter γ for the input variable x_t in CEL0.

4.2. Local methods and initial points

We first show that FB, IRL1 and CD can get trapped in local minima whereas our method is guaranteed to provide a global minimizer. We ran the algorithms using different initial points $\mathbf{x}_{\text{init}}$: a random point, zero, and the true value $\overline{\mathbf{x}}$ (which is not available in practice). Notice that our method does not need any initial point. Table 1 compares the criterion value at the solutions returned by the different algorithms.

We observe that all the local algorithms have different estimated optimal criterion values and thus return different minimizers depending on the starting point $\mathbf{x}_{\text{init}}$. Furthermore, it can be observed that local optimization algorithms often are not reliable. Hence for Capped ℓ_1, CD yields a better minimizer than IRL1 when using a random starting point, whilst the converse holds when 0 is the starting point. Without a global method, we cannot assess the quality of the obtained solution. Interestingly, our method provides the lowest

criterion value; the next section confirms the validity of the method.

Table 1. Optimal criterion value depending on initial point.

Alg. \ Ψ_λ	Capped ℓ_1	SCAD	MCP	CEL0
Proposed ($k=3$)	**3.725**	**3.506**	**2.749**	**4.425**
Random $\mathbf{x}_{\text{init}}$				
FB	4.372	3.759	3.194	4.638
IRL1	7.052	4.632	3.381	6.926
CD	4.099	3.927	3.520	4.638
$\mathbf{x}_{\text{init}} = 0$				
FB	4.208	3.763	3.149	4.638
IRL1	4.839	4.566	3.013	6.879
CD	5.455	4.371	3.717	5.316
$\mathbf{x}_{\text{init}} = \overline{\mathbf{x}}$				
FB	3.867	3.639	2.871	4.637
IRL1	4.766	4.567	2.914	6.874
CD	4.093	3.639	3.015	5.365

4.3. Global vs local algorithms

We now compare the different values of the criterion at the solutions returned by each algorithm as well as the lower bound $\mathcal{J}_k^*$. Figure 2 shows the values of the criterion for the SCAD regularization. We ran 200 tests on randomly generated data but, for the sake of clarity, only 100 are plotted. The results are ordered according to the value of the criterion at the minimizer $\hat{\mathbf{x}}$ obtained with our approach. We observe that our method always yields a better minimizer of the criterion while the other methods are sensitive to local minimizers.

Moreover, we compare the estimated optimal value of the criterion $\mathcal{J}(\hat{\mathbf{x}})$ given by our method with the lower bounds $\mathcal{J}_2^*$ and $\mathcal{J}_3^*$. Table 2 shows statistics over the 200 tests about the difference between these values. Up to a numerical precision of 10^{-4}, $\hat{\mathcal{J}}^*$ reaches the lower bound with a relaxation order of 2 for Capped ℓ_1 and 3 for the three other regularizations. The global optimality is therefore theoretically certified in these cases. It is remarkable that Lasserre's hierarchy converges for a low value of k.

Table 2. Statistics on $\mathcal{J}(\hat{\mathbf{x}}) - \mathcal{J}_k^*$ for positive signals.

(200 runs)		Capped ℓ_1	SCAD	MCP	CEL0
$\mathcal{J}(\hat{\mathbf{x}}) - \mathcal{J}_2^*$	Avg	**2.0e−4**	3.4e−2	5.0e−2	2.3e−3
	Med	**2.0e−4**	3.3e−2	5.0e−2	9.0e−4
$\mathcal{J}(\hat{\mathbf{x}}) - \mathcal{J}_3^*$	Avg	2.0e−4	**4.0e−5**	**1.0e−4**	**2.0e−4**
	Med	2.0e−4	**<1e−5**	**1.0e−4**	**<1e−5**

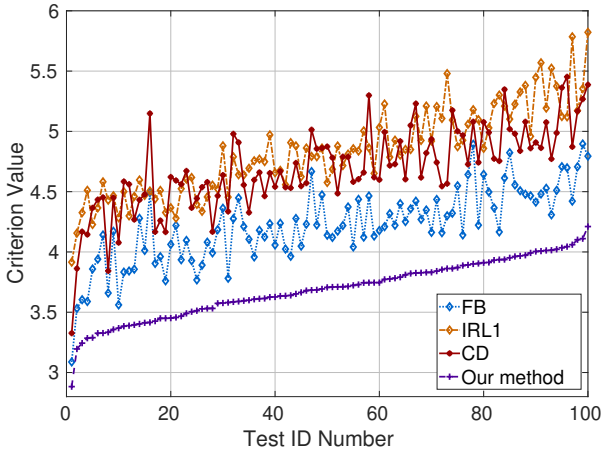

Fig. 2. Optimal criterion values with SCAD regularization.

4.4. Case of real-valued signals

We provide additional simulation results concerning real-valued signals $\overline{\mathbf{x}}$. In the case of positive signals, we could naturally drop the absolute values since $|\mathbf{x}| = \mathbf{x}$. In contrast, the case of real-valued signals is more intricate due to the introduction of additional variables to handle this absolute value. As a consequence, the convergence of Lasserre's hierarchy does not occur as fast as in the positive case. However, we can use the solution obtained with an order of relaxation $k = 3$ as an initial point of a local method and improve both solutions as pointed out by Table 3. The latter shows the statistics on $\mathcal{J}(\hat{\mathbf{x}})$ for FB and FB initialized with Lasserre's solution (FBwL) using the SCAD and MCP regularizations.

Table 3. Statistics on $\mathcal{J}(\hat{\mathbf{x}})$ for real-valued signals.

(200 runs)		FB	FBwL
SCAD	Average	4.317	**3.977**
	Median	4.301	**3.969**
MCP	Average	3.936	**3.426**
	Median	3.920	**3.419**

5. CONCLUSION

We study the global optimization of cost functions penalized by classic non-convex continuous approximations to ℓ_0 regularization. By showing that the original optimization problem can be reformulated as a polynomial problem, we obtain sparse solutions through convex relaxation techniques. Finally our simulation results illustrate that the main benefit of our approach is to secure a global optimum of the criterion in contrast with state-of-the-art local methods.

6. REFERENCES

[1] M. Nikolova, "Description of the minimizers of least squares regularized with ℓ_0 norm. Uniqueness of the global minimizer," *SIAM J. Imaging Sci.*, vol. 6, no. 2, pp. 904–937, Jan. 2013.

[2] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM J. Optim.*, vol. 18, no. 4, pp. 1351–1376, jan 2008.

[3] P. L. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer, 2011.

[4] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.

[5] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 629–654, Sept. 2008.

[6] A. Patrascu and I. Necoara, "Random coordinate descent methods for ℓ_0 regularized convex optimization," *IEEE Trans. Automat. Contr.*, vol. 60, no. 7, pp. 1811–1824, July 2015.

[7] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sept. 2017.

[8] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *J. Mach. Learn. Res.*, vol. 11, pp. 1081–1107, Mar. 2010.

[9] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Appl. Stat.*, vol. 38, no. 2, pp. 894–942, Apr. 2010.

[10] M. Artina, M. Fornasier, and F. Solombrino, "Linearly constrained nonsmooth and nonconvex minimization," *SIAM J. Optim.*, vol. 23, no. 3, pp. 1904–1937, Jan. 2013.

[11] A. Jezierska, H. Talbot, O. Veksler, and D. Wesierski, "A fast solver for truncated-convex priors: Quantized-convex split moves," in *Lecture Notes in Computer Science*, pp. 45–58. Springer Berlin Heidelberg, 2011.

[12] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem," *SIAM J. Imaging Sci.*, vol. 8, no. 3, pp. 1607–1639, jan 2015.

[13] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optim.*, vol. 11, no. 3, pp. 796–817, Jan. 2001.

[14] F. Bugarin, D. Henrion, and J. B. Lasserre, "Minimizing the sum of many rational functions," *Math. Program. Comput.*, vol. 8, no. 1, pp. 83–111, Aug. 2015.

[15] M. Castella, J.-C. Pesquet, and A. Marmin, "Rational optimization for nonlinear reconstruction with approximate ℓ_0 penalization," *IEEE Trans. Signal Process.*, vol. 67, pp. 1–1, 2018.

[16] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision," *SIAM J. Imaging Sci.*, vol. 8, no. 1, pp. 331–372, Jan. 2015.

[17] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, Oct. 2008.

[18] P. Breheny and J. Huang, "Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection," *Ann. Appl. Stat.*, vol. 5, no. 1, pp. 232–253, Mar. 2011.

[19] D. Henrion and J.-B. Lasserre, "Detecting global optimality and extracting solutions in GloptiPoly," in *Positive Polynomials in Control*, vol. 312, pp. 293–310. Springer Berlin Heidelberg, Sept. 2005.

[20] D. Henrion, J.-B. Lasserre, and J. Löfberg, "GloptiPoly 3: moments, optimization and semidefinite programming," *Optim. Methods Softw.*, vol. 24, no. 4-5, pp. 761–779, Oct. 2009.

[21] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3 — a Matlab software package for semidefinite programming, version 1.3," *Optim. Methods Softw.*, vol. 11, no. 1-4, pp. 545–581, Jan. 1999.