



HAL
open science

Signal reconstruction from sub-sampled and nonlinearly distorted observations

Arthur Marmin, Marc Castella, Jean-Christophe Pesquet, Laurent Duval

► **To cite this version:**

Arthur Marmin, Marc Castella, Jean-Christophe Pesquet, Laurent Duval. Signal reconstruction from sub-sampled and nonlinearly distorted observations. EUSIPCO 2018: 26th European Signal Processing Conference, Sep 2018, Roma, Italy. pp.1970-1974, 10.23919/EUSIPCO.2018.8553174. hal-01957568

HAL Id: hal-01957568

<https://hal.science/hal-01957568>

Submitted on 17 Dec 2018

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.

Signal Reconstruction from Sub-sampled and Nonlinearly Distorted Observations

Arthur Marmin*, Marc Castella†, Jean-Christophe Pesquet*, Laurent Duval‡

* *Center for Visual Computing, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France*
 {arthur.marmin,jean-christophe.pesquet}@centralesupelec.fr

† *SAMOVAR, Télécom SudParis, CNRS Université Paris-Saclay, Evry, France*
 marc.castella@telecom-sudparis.eu

‡ *IFP Energies nouvelles Rueil-Malmaison, France*
 laurent.duval@ifpen.fr

Abstract—Faithful short-time acquisition of a sparse signal is still a challenging issue. Instead of an idealized sampling, one has only access to an altered version of it through a measurement system. This paper proposes a reconstruction method for the original sparse signal when the measurement degradation is composed of a nonlinearity, an additive noise, and a sub-sampling scheme. A rational criterion based on a least-squares fitting penalized with a suitable approximation of ℓ_0 is minimized using a recent approach guaranteeing global optimality for rational optimization. We provide a complexity analysis and show that the sub-sampling offers a significant gain in terms of computational time. This allows us to tackle practical problems such as chromatography. Finally, experimental results illustrate that our method compares very favorably to existing methods in terms of accuracy in the signal reconstruction.

I. INTRODUCTION

Accurate data acquisition of sparse signals from real-world measurements remains an open challenge. Our present motivation can be illustrated by an actual situation, encountered in analytical chemistry, especially in chromatography. Let us consider a mixture of chemical compounds in different concentrations. An optimal measurement would produce a set of isolated peaks whose locations and amplitudes characterize each compound and its relative concentration. A discrete version is illustrated with the three-sample parsimonious black-dot signal modeled in Figure 1. Usually, such a sparse information is subject to undesirable fluctuations and degradations, requiring to employ restoration methods from observed responses. Here, the linear part of the system assumes the convolution by a limited-support kernel. In Figure 1, the kernel is a five-sample discrete approximation of a binomial filter. Such models are commonly used in physico-chemical data processing [1]–[3]. The resulting peak signal is further affected by a nonlinear distortion, representing a saturation, that flattens higher amplitudes. This is depicted on the blue-cross signal. Those smoothed data are further degraded by an additive Gaussian noise as shown in solid red lines.

This paper extends results in [4]–[6] and provides a novel recovery method for sparse signals from sub-sampled observations obtained through a nonlinear and noisy model. This stands in contrast with the vast majority of available methods,

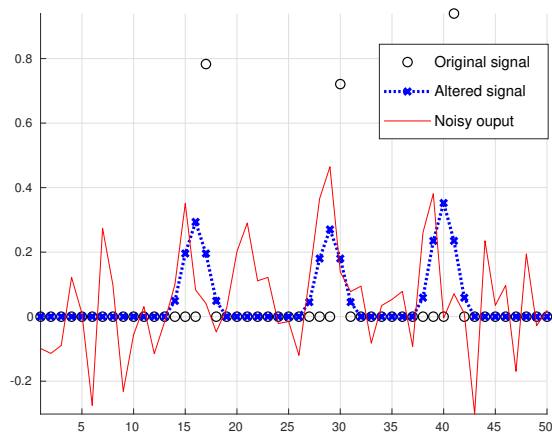


Fig. 1. Example of chromatography signals

which only apply in a linear context [7]–[9]. Among the few methods that have tried to weaken the linearity assumption, one can mention [10]–[12].

Our method builds on the minimization of a sum of rational functions, for which recent methodologies can be applied. Here, we specifically focus on the possibility to acquire the data at a reduced rate. This feature is important in applicative areas, since it permits faster acquisitions for high-throughput experiments and analysis. In addition, the structure induced by the sub-sampled acquisition reduces the overall complexity, which was a major difficulty in the previous work in [6]. We provide an analysis of the benefit of the sub-sampling on the complexity of the problem.

Our paper is organized as follows: Section II introduces our model for the observed signal. Section III explains how our method recovers the target signal. It also provides a study of the computational difficulties encountered. Simulation results can be found in Section IV and Section V concludes our work.

We introduce the following notation: for any nonnegative integers n and p , \mathbb{S}_+^n is the cone of $n \times n$ real positive semi-definite matrices, $\binom{n}{p}$ is the binomial coefficient “among n choose p ” and $\lfloor \cdot \rfloor$ (resp. $\lceil \cdot \rceil$) is the greatest (resp. smallest)

integer lower (resp. bigger) than its argument.

II. PROBLEM STATEMENT

A. Observation model

We consider the reconstruction from sub-sampled measurements of an unknown sparse discrete-time signal $\bar{\mathbf{x}}$ of length T , comprising few peaks. The measurement process deteriorates $\bar{\mathbf{x}}$ in the following way: peak enlargement, nonlinear distortion due to sensor saturation, and noise. Finally, the measured signal is sub-sampled during the acquisition. This is modelled by a convolution with a finite impulse response filter given by vector \mathbf{h} followed by a nonlinear function Φ , the addition of a noise \mathbf{w} with samples drawn from an i.i.d. zero-mean Gaussian distribution, and a decimation operator D . Defining the observation vector \mathbf{y} of size U after sub-sampling, the corresponding modelling equation reads

$$\mathbf{y} = D(\Phi(\mathbf{h} * \bar{\mathbf{x}}) + \mathbf{w}).$$

We will be mainly interested in regular decimations D_α where all the elements indexed with a multiple of α are deleted, namely $D_\alpha((v_t)_{1 \leq t \leq T}) = (v_{\Delta(u,\alpha)})_{1 \leq u \leq U}$ where Δ is defined as $\Delta(u, \alpha) = u + \lfloor \frac{u-1}{\alpha} \rfloor$. We denote by D_∞ the identity operator that preserves the entire signal.

B. Signal recovery criterion

To estimate the original signal $\bar{\mathbf{x}}$, we minimize a criterion \mathcal{J} composed of two terms. The first one is a fit measure between the recorded measurements \mathbf{y} and the output of the noiseless model for a given estimate \mathbf{x} of the original signal $\bar{\mathbf{x}}$. The second term is a sparsity-promoting penalization approximating the ℓ_0 pseudo-norm weighted by a positive parameter λ , i.e.

$$\mathcal{J}(\mathbf{x}) = \underbrace{\|\mathbf{y} - D_\alpha(\Phi(\mathbf{h} * \mathbf{x}))\|_2^2}_{\text{data fidelity}} + \lambda \underbrace{\sum_{t=1}^T \Psi(x_t)}_{\text{penalization}}.$$

In the following, we choose Φ and Ψ as rational functions and Φ acts component-wise. To simplify our notation, we set the components of \mathbf{x} with nonpositive index to be identically zero. Then \mathcal{J} can be written as a sum of rational functions

$$\begin{aligned} \mathcal{J}(\mathbf{x}) &= \sum_{u=1}^U \left(y_u - \Phi \left(\sum_{l=1}^L h_l x_{\Delta(u,\alpha)-l+1} \right) \right)^2 + \lambda \sum_{t=1}^T \Psi(x_t) \\ &= \sum_{u=1}^U \frac{p_u(x_{\Delta(u,\alpha)-L+1}, \dots, x_{\Delta(u,\alpha)})}{q_u(x_{\Delta(u,\alpha)-L+1}, \dots, x_{\Delta(u,\alpha)})} + \sum_{t=1}^T \frac{r(x_t)}{s(x_t)}. \end{aligned}$$

The polynomials $(p_u)_{1 \leq u \leq U}$ and $(q_u)_{1 \leq u \leq U}$ are functions of L variables whereas the polynomials r and s depend on one variable. We define also

$$d_q = \left\lceil \frac{\text{degree}(q_u)}{2} \right\rceil \quad \text{and} \quad d_s = \left\lceil \frac{\text{degree}(s)}{2} \right\rceil.$$

III. METHOD

A. Lasserre's relaxation

We here extend the methodology developed in [6] to the case of decimated observations. Using a recent approach introduced by Lasserre [13] and related to the sum of squares relaxation, the solution to the rational optimization problem

$$\mathcal{J}^* = \min_{\mathbf{x} \in \mathbb{R}^T} \mathcal{J}(\mathbf{x}), \quad (1)$$

can be found by solving a hierarchy of convex semi-definite programming (SDP) problems in their standard dual form. Defining a vector \mathbf{z} whose components are related to the monomials up to degree $2k$ in the variables of Problem (1), the relaxation of order k takes the following form

$$\mathcal{P}_k^* = \min_{\mathbf{z} \in \mathbb{R}^m} \mathbf{b}^\top \mathbf{z} \quad (2)$$

$$\text{subject to} \quad \mathbf{C} - \sum_{i=1}^m z_i \mathbf{A}_i \in \mathbb{S}_+^n \quad (3)$$

$$\mathbf{f} - \mathbf{G}^\top \mathbf{z} = 0, \quad (4)$$

where \mathbf{C} and $(\mathbf{A}_i)_{1 \leq i \leq m}$ are symmetric matrices, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{f} \in \mathbb{R}^{n_i}$, and \mathbf{G} is a matrix of $\mathbb{R}^{m \times n_i}$.

The size of the SDP, expressed by the integers n , m and n_i , is related to Problem (1) and to the order k of the relaxation. This is detailed in Section III-B, where it will appear that the size of the SDP grows tremendously as k increases. In addition, it has been proved [13], [14] that the SDP relaxations \mathcal{P}_k^* converge to the optimum \mathcal{J}^* of the rational optimization problem when the relaxation order goes to infinity, that is $\lim_{k \rightarrow \infty} \mathcal{P}_k^* = \mathcal{J}^*$. Fortunately, low relaxation orders suffice in practice to provide correct solutions. The order of relaxation yet needs to be at least equal to half of the maximal degree of the polynomials involved in the criterion to allow all the polynomials in the original problem to be represented. Finally, the convergence results require the original variables to belong to a compact subset of \mathbb{R}^T . We thereby restrict the domain of our problem to a closed bounded box chosen such that a solution to Problem (1) belongs to it.

B. Structure and complexity of the relaxation

This section studies the structure and the size of the SDP relaxation. For simplicity, we do not take into account here that some variables in the expression of \mathcal{J} are set identically to zero. The SDP size presented here will hence be slightly overestimated.

In the SDP relaxation, each rational fraction in \mathcal{J} is associated with a measure and the components of the variable \mathbf{z} are moments (truncated to the order $2k$) of these measures. To ensure that each sub-sequence corresponding to a given rational term is a valid moment sequence for a set of say p variables, $1 + 2p$ positive semi-definite constraints must be enforced (one for the whole set of variables, and $2p$ for the two bound constraints coming from the lower and upper bounds imposed to each variable). Since our criterion is a sum of U rational functions over L variables and T rational functions

over 1 variable, the total number of such semi-definite positive constraints is equal to

$$n_s = U(1 + 2L) + 3T.$$

These constraints can be expressed more concisely under the form (3) provided that matrices $(\mathbf{A}_i)_{1 \leq i \leq m}$ and \mathbf{C} have a block diagonal structure involving n_s blocks.

In addition, since any given variable x_t may appear in multiple rational functions, many moments must be identical in \mathbf{z} . These moment equality constraints correspond to the linear equality constraints in (4). Let us denote by θ_u the overlap parameter, defined as the number of variables in common between $\frac{p_{u+1}}{q_{u+1}}$ and $\frac{p_u}{q_u}$ when $u \in \{1, \dots, U-1\}$. θ_u depends on u , but also on L and α .

According to the general methodology [15], we have to consider the equality of the moments of all monomials of these θ_u overlapping variables up to degree $2(k - d_q)$.

As a consequence, this gives $\binom{\theta_u + 2(k - d_q)}{2(k - d_q)}$ equality constraints for every $u \in \{1, \dots, U-1\}$. Similar arguments hold for the T rational functions $\frac{r}{s}$ appearing in \mathcal{J} , but since it depends on one variable only, the above binomial coefficient reduces to $2(k - d_s)$ equality constraints on moments. Hence,

$$n_l = \sum_{u=1}^{U-1} \binom{\theta_u + 2(k - d_q)}{2(k - d_q)} + 2(k - d_s)T.$$

Concerning the size m of the vector \mathbf{z} , it is given by the sum of the number of monomials up to degree $2k$ for all rational terms in \mathcal{J} . A total of U of these terms depend on L variables, and T terms depend on one variable. Noting that $\binom{1+2k}{2k} = 1 + 2k$, it follows that

$$m = U \binom{L + 2k}{2k} + T(1 + 2k).$$

Furthermore, n is the sum of the block sizes in $(\mathbf{A}_i)_{1 \leq i \leq m}$ and \mathbf{C} . The blocks can be viewed as matrices whose rows and columns are indexed by monomials. Each of their entries is a sum of moments whose maximal degree is the product of the row index, the column index and, if any, a polynomial defining a bound constraint. For a maximal degree $2k$, and for the first U terms in \mathcal{J} the block sizes are thus $\binom{L+k}{k}$ for the whole set of variables and $\binom{L+k-1}{k-1}$ for the upper and lower bound constraints. The last T terms in \mathcal{J} can be considered similarly. Finally, we obtain

$$n = U \left(\binom{L+k}{k} + 2L \binom{L+k-1}{k-1} \right) + T(1 + 3k).$$

The order of relaxation k appears as a binomial coefficient in both SDP size variables n and m , so that increasing k quickly blows up the size of the SDP. The previous remark can be extended to L . Actually, for a given order of relaxation k , the size of the SDP asymptotically becomes of the order:

$$m = \mathcal{O}(UL^{2k}), \quad n = \mathcal{O}(UL^k). \quad (5)$$

In our context, decimation yet reduces efficiently the size of the SDP relaxation since it decreases U and $(\theta_u)_{1 \leq u \leq U-1}$.

Both play a prominent role in the values of n , m , n_s , and n_l which determine the complexity of the optimization problem. Table I compares the size of SDP relaxations without decimation (D_∞) and with the two-fold D_2 decimation. As discussed above, the size variables n and m increase quickly with the order of relaxation k and the length of the filter L . In the decimated case, the variables n and m are decreased about one half while n_s and n_l are respectively reduced about one third and two thirds.

TABLE I
SIZE OF SDP RELAXATION FOR DIFFERENT DECIMATIONS —
CONSIDERING VARIABLES WITH NEGATIVE INDEX AS IDENTICALLY ZERO

T	L	k	m		n		n_s		n_l	
			D_∞	D_2	D_∞	D_2	D_∞	D_2	D_∞	D_2
50	3	3	4417	2373	4384	2430	494	321	976	366
100	3	3	8497	4823	8884	4930	994	646	1976	741
50	4	3	10339	5271	7909	4155	588	367	1916	596
100	4	3	21189	10871	16159	8530	1188	742	3916	1221
50	3	4	8424	4419	8158	4383	494	321	1702	512
100	3	4	17124	8994	16558	8908	994	646	3452	1037
100	4	4	49134	24834	35473	18268	1188	742	8884	2045
100	5	3	45381	22967	26815	13858	1380	838	7276	2181

C. Exact relaxation of ℓ_0

In this section, we show how to theoretically address an exact ℓ_0 penalty with our relaxation. We explain why this relaxation is intractable in practice by relying on our previous complexity estimation. This justifies our choice of a rational approximation of ℓ_0 as a regularization.

An equivalent form of the minimization of \mathcal{J} , where Ψ is replaced by the exact ℓ_0 penalization is as follows

$$\begin{aligned} & \underset{(\mathbf{x}, \boldsymbol{\xi}) \in (\mathbb{R}^T)^2}{\text{minimize}} && \|\mathbf{y} - D_\alpha(\Phi(\mathbf{h} * (\mathbf{x} \odot \boldsymbol{\xi}))\|^2 + \lambda \sum_{i=1}^T \xi_i \\ & \text{subject to} && (\forall i \in \{1, \dots, T\}) \quad \xi_i = \xi_i^2, \end{aligned}$$

where the operator \odot denotes the element-wise Hadamard product and λ is a nonnegative regularization parameter. Unfortunately, in the above formulation, the introduction of the ξ_i necessary to formulate the ℓ_0 penalization in a polynomial form results in twice more variables than in the original problem. As a consequence, the variables n and m defining the size of the SDP relaxation become much higher. A rough and underestimated idea of the SDP size is obtained by replacing L with $2L$. Our method may then become intractable.

The substantial increase in the computation time has been confirmed in our experiments. Even for a signal of length 50 with a filter of length 2, we needed about three hours of computations to reconstruct the solution compared to a few seconds using our rational approximation of the ℓ_0 penalty.

IV. SIMULATIONS AND RESULTS

A. Simulation setup

We have run our simulations for nonnegative signals $\bar{\mathbf{x}}$ with 50, 100 and 200 samples and for filters of length 3, 4 and 5 whose coefficients are binomial coefficients on the

corresponding line of Pascal’s triangle. These specific filter coefficients as well as the nonnegativity of input signals are of practical importance for chemistry problems, especially in chromatography. Indeed the concentrations of compounds are always positive and the coefficients of the filters correspond to peak enlargement.

We have set the approximation Ψ of ℓ_0 to the following Geman-McClure-like potential [4]

$$(\forall t \in \mathbb{R}) \quad \Psi(t) = \frac{|t|}{\delta + |t|},$$

where the parameter δ controls the quality of the ℓ_0 approximation. Smaller δ leads to better approximation, but also to a harder optimization problem. We have set δ to 0.01. Note that this approximation of ℓ_0 is well-chosen because it is rational with low degree, which allows a low order of relaxation k . The saturation function Φ acts component-wise and is defined as $(\forall t \in \mathbb{R}) \quad \Phi(t) = \frac{t}{\chi + |t|}$ where the parameter χ has been set to 0.3 and controls the level of the saturation.

For each test, we have run 100 simulations and we show here the average results. We have set the order of relaxation to 3 and have used GloptiPoly [16] coupled with the SDP solver SDPT3-4 [17] to solve the optimization problem. We have run the simulations on a standard computer with an Intel I7 CPU running at 3.60 GHz and 32 GB of RAM.

We have compared our method to an improved LASSO approach (named iLASSO). It consists of applying first the least absolute shrinkage and selection operator (LASSO) [7] using a linearization of Φ . More precisely

$$\mathbf{x}_{\text{lasso}} = \underset{\mathbf{x} \in [0, +\infty[^T}{\text{argmin}} \|\mathbf{y} - D_\alpha(\chi^{-1}(\mathbf{h} * \mathbf{x}))\|^2 + \lambda \|\mathbf{x}\|_1.$$

The solution $\mathbf{x}_{\text{lasso}}$ is then used to initialize a proximal gradient algorithm based on iterative hard thresholding (IHT) [8] extended to the nonlinear model.

The value of the relaxation parameter λ has been optimized empirically and set to 0.10 for our method and 0.05 for LASSO and IHT. Decimation D_∞ represents the case when the signal is not sub-sampled and the corresponding results are provided for comparison with results discussed in [6].

B. Results

In this section, we study the effect of the different decimation patterns first on the reconstructed signal quality and then on the computational time.

1) *Reconstruction quality*: Figure 2 illustrates, from top to bottom, a sub-sampled observed signal \mathbf{y} , the corresponding original signal $\bar{\mathbf{x}}$, the signal reconstructed with our method and the signal reconstructed using iLASSO. Within the framework of our nonlinear model, iLASSO performs poorly whereas our estimation of $\bar{\mathbf{x}}$ is close to the original signal.

To assess peak detection, we introduce a varying threshold on the estimated amplitude of $\bar{\mathbf{x}}$. Figure 3 shows the Receiver Operating Characteristic (ROC) for D_∞ , D_4 and D_2 decimations respectively. ROC curves are drawn in solid red for our method and in dashed blue for iLASSO. iLASSO detects more

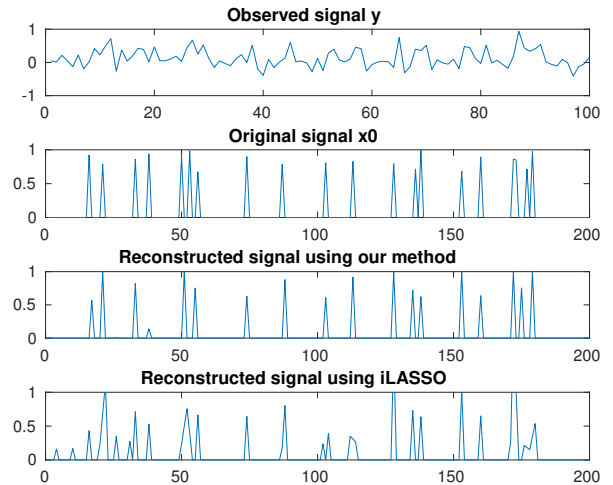


Fig. 2. Comparison between iLASSO and our method using D_2 decimation

peaks but is also more prone to false detection. Conversely, our method detects a high number of true peaks for the different thresholds and make few mistakes. In accordance with our observations in Figure 2, our method has a very low false positive rate which maintains its ROC curve close to the ordinate axis. Moreover, the amplitudes of the peaks are sensibly closer to the real ones with our method than with iLASSO. Amplitudes are highly important in chromatography since they generally correspond to the concentration of the different compounds that we are analyzing.

Comparing the ROC of our method for the different decimation patterns in Figure 3, we notice that the higher the decimation is, the fewer peaks detected in the true position. We observe that our model keeps a very low false positive rate which is related to the fact that very few peaks are detected at a shifted position. Decimation mainly causes the system to miss some peaks and weakly affect their positions.

2) *Computational load*: Table II shows the computational time corresponding to the different test cases for the three decimation operators D_∞ , D_4 and D_2 .

The computational time explodes when the length of the filter increases. This is due to the presence of L in the binomial coefficients of the expressions of n and especially of m in (5). Indeed, state-of-the-art SDP solvers exhibit high computational time when the dimension m is large.

However, the higher the decimation, the lower the computational cost. The effect is especially beneficial for high values of L and it allows us to deal with filter lengths that are intractable without decimation. This is of high practical interest, because filter lengths from 5 to 7 can be sufficient for dedicated chemistry needs.

V. CONCLUSION

We have considered the challenging problem of reconstructing sparse signals degraded by a decimated non-linear model. We have defined a least-squares fitting criterion penalized with an ℓ_0 approximation that can be written as a sum of rational

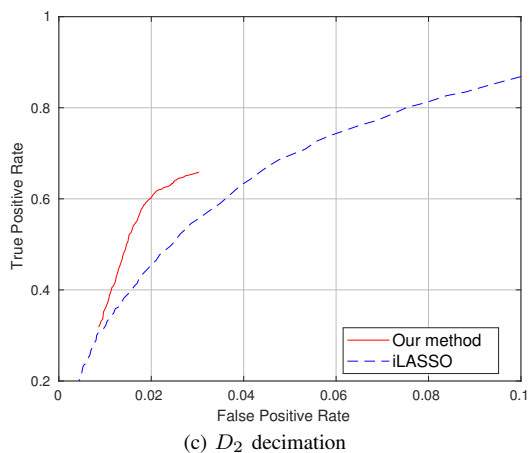
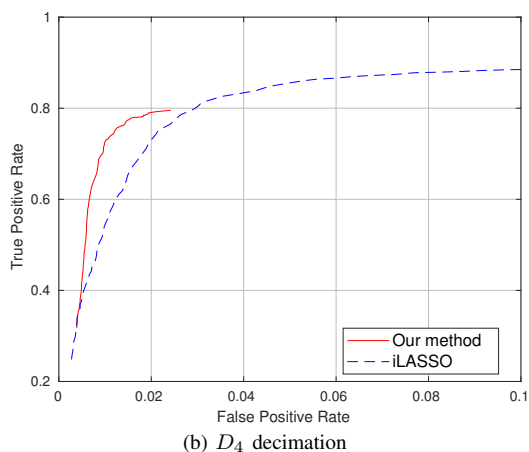
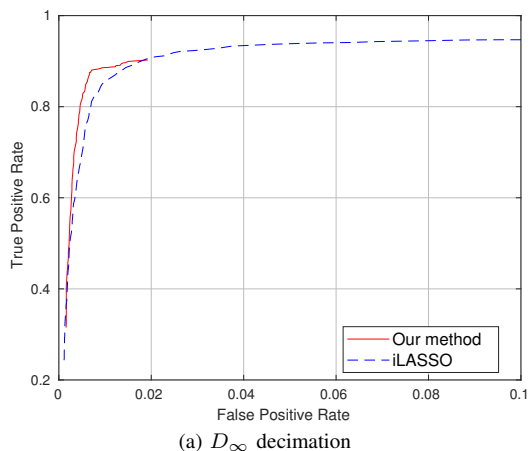


Fig. 3. Comparison of ROC between our method and iLASSO

TABLE II
COMPUTATIONAL TIME DEPENDING ON DECIMATION OPERATOR

T	L	Time (s)		
		D_∞	D_4	D_2
50	3	18	12	8
100	3	45	30	18
200	3	88	65	42
50	4	1285	522	219
100	4	9871	4468	1356
50	5	5559	2269	620
100	5	Overload	22354	5105

functions. Thereby we have applied a recent methodology for rational minimization and relaxed the original problem into a hierarchical sequence of convex SDP. One of the main advantages of the proposed approach is that the SDP size is significantly reduced when the decimation factor increases, which may be of high interest in applications such as chromatography. The reduction of the computation time also allows us to consider filters with a larger length as encountered in signal processing applications. Finally, we have shown that our method compares favorably to LASSO followed by IHT.

REFERENCES

- [1] C. Vendevre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaud, and M.-C. Hennion, "Characterisation of middle-distillates by comprehensive two-dimensional gas chromatography (GC \times GC): A powerful alternative for performing various standard analysis of middle-distillates," *J. Chrom. A*, vol. 1086, no. 1-2, pp. 21–28, 2005.
- [2] C. Vendevre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, and D. Thiébaud, "Comprehensive two-dimensional gas chromatography for detailed characterisation of petroleum products," *Oil Gas Sci. Tech.*, vol. 62, no. 1, pp. 43–55, 2007.
- [3] X. Ning, I. W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometr. Intell. Lab. Syst.*, vol. 139, pp. 156–167, Dec. 2014.
- [4] M. Castella and J.-C. Pesquet, "Optimization of a Geman-McClure like criterion for sparse signal deconvolution," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Mexico, Dec. 2015, pp. 309–312.
- [5] —, "Recovery of nonlinearly degraded sparse signals through rational optimization," in *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop*, Lisbon, Portugal, Jun. 2017.
- [6] —, "A global optimization approach for rational sparsity promoting criteria," in *Proc. Eur. Sig. Image Proc. Conf.*, Kos, Greece, Aug. 2017, pp. 156–160.
- [7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
- [8] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 629–654, Sep. 2008.
- [9] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A continuous exact ℓ_0 penalty (CELO) for least squares regularized problem," *SIAM J. Imaging Sci.*, vol. 8, no. 3, pp. 1607–1639, Jan. 2015.
- [10] M. Schetzen, "Nonlinear system modelling and analysis from the Volterra and Wiener perspective," in *Lecture Notes in Control and Information Sciences*. Springer London, 2010, pp. 13–24.
- [11] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.
- [12] Y. Deville and L. T. Duarte, "An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures," in *Latent Variable Analysis and Signal Separation*. Springer International Publishing, 2015, pp. 155–167.
- [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] D. Jibeteau and E. de Klerk, "Global optimization of rational functions: a semidefinite programming approach," *Math. Program.*, vol. 106, no. 1, pp. 93–109, Apr. 2005.
- [16] 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.
- [17] R. H. Tütüncü, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Math. Program.*, vol. 95, no. 2, pp. 189–217, Feb. 2003.