Recovery of Nonlinearly Degraded Sparse Signals through Rational Optimization

Marc Castella
SAMOVAR, Télécom SudParis, CNRS,
Université Paris-Saclay,
9 rue Charles Fourier, 91011 Evry Cedex, France
Email: marc.castella@telecom-sudparis.eu

Jean-Christophe Pesquet
Center for Visual Computing, CentraleSupelec,
Université Paris-Saclay,
Grande Voie des Vignes, 92295, Châtenay-Malabry, France
Email: jean-christophe@pesquet.eu

***Abstract*—We show the benefit which can be drawn from recent global rational optimization methods for the minimization of a regularized criterion. The regularization term is a rational Geman-McClure like potential, approximating the ℓ_0 norm and the fit term is a least-squares criterion suitable for a wide class of nonlinear degradation models.**

I. Introduction

Over the last decade, much attention has been paid to inverse problems involving sparse signals. A popular approach consists in formulating such problems under a variational form where one minimizes the sum of a data fidelity term and a regularization term incorporating prior information. For sparse signals, the regularization term may involve the ℓ_0 norm, or an approximation of it [1]. This generally results in difficult optimization problems with many local minima and weak global convergence guarantees [2]–[6]. In this work, we consider rational optimization algorithms offering global optimality guarantees. In addition, our method allows us to address the challenging case of a nonlinear model [7]–[9].

II. Model and criterion

Consider a sparse vector with unknown nonnegative samples $\overline{\mathbf{x}} := (\overline{x}_1, \dots, \overline{x}_T)^\top$, only a few of which are nonzero. We aim at recovering it from measurements $\mathbf{y} := (y_1, \dots, y_T)^\top$ related to $\overline{\mathbf{x}}$ through a linear transformation (typically, a convolution) followed by some nonlinear effects:

$$\mathbf{y} = \boldsymbol{\phi}(\mathbf{H}\overline{\mathbf{x}}) + \mathbf{n}\ , \tag{1}$$

where $\mathbf{n} := (n_1, \dots, n_T)^\top$ is a realization of a random noise vector, and $\boldsymbol{\phi} : \mathbb{R}^T \to \mathbb{R}^T$ is a rational nonlinear function with components $[\boldsymbol{\phi}(\mathbf{u})]_k = \phi(u_k)$ depending on the k^{th} entry u_k only. $\mathbf{H} \in \mathbb{R}^{T\times T}$ is a given convolution matrix, which is assumed Toeplitz banded under suitable vanishing boundary conditions. To estimate $\overline{\mathbf{x}}$, we minimize a penalized criterion having the following form:

$$(\forall \mathbf{x} \in \mathbb{R}_+^T) \quad \mathcal{J}(\mathbf{x}) = \|\mathbf{y} - \boldsymbol{\phi}(\mathbf{H}\mathbf{x})\|^2 + \lambda \sum_{t=1}^{T} \frac{x_t}{\delta + x_t}\ , \tag{2}$$

where λ and δ are positive regularization and smoothing parameters. The last term is a Geman-McClure like potential as in [10]. We assume that an upper-bound B on the values $(\overline{x}_t)_{t=1}^T$ is available and the minimization is thus performed over a compact set defined and represented by $\mathbf{K} = \{\mathbf{x} \in \mathbb{R}^T \,|\, x_t(B - x_t) \geq 0, t = 1, \dots, T\}$. The optimization problem consists then in finding $\mathcal{J}^\star := \inf_{\mathbf{x}\in\mathbf{K}} \mathcal{J}(\mathbf{x})$.

III. Rational minimization

Given $\mathcal{J}$ in (2), the previous minimization is a rational problem. The methodology in [11, 12] builds for different orders k a hierarchical sequence of semi-definite programming (SDP) relaxations $\mathcal{P}_k^\star$ for which the following optimality result holds: $\mathcal{P}_k^\star \uparrow \mathcal{J}^\star$ as $k \to +\infty$.

By using SDP solvers to solve $\mathcal{P}_k^\star$, one can hence theoretically obtain the global optimum [10]. Due to the maximum tractable size of state of the art SDP solvers, this approach is however limited to small/medium size problems having small degree, even when restricting the hierarchy to a finite and small order k. To overcome this difficulty, we exploit the problem structure in the sum of rational terms in (2). Using the sparse Toeplitz banded shape of $\mathbf{H}$, it can be noticed that:

$$\mathcal{J}(\mathbf{x}) = \sum_{t=1}^{T} \underbrace{\left[y_t - \phi\left(\sum_{i=1}^{L} h_i x_{t-i+1}\right)\right]^2}_{\text{depends on } x_k \text{ for } k\,\in\, J_t} + \underbrace{\lambda \frac{x_t}{\delta + x_t}}_{\text{depends on } x_t \text{ only}}\ ,$$

where $J_t = \{\min\{1, t - L - 1\}, \dots, t\}$ and $J_{t+T} = \{t\}$ for any $t \in \{1, \dots, T\}$. These index subsets satisfy the so-called "Running Intersection Property" [13]. As a consequence, it is possible to introduce a much smaller SDP relaxation $\mathcal{P}_k^{\star\mathrm{s}}$ instead of $\mathcal{P}_k^\star$. The fundamental idea is that the SDP relaxations involve variables representing monomials in $(x_1, \dots, x_T)$. Using the above split form, many monomials can be discarded, the most striking case being when $\mathcal{J}$ is fully separable.

IV. Experiments

We have generated 100 Monte-Carlo realizations of vector $\overline{\mathbf{x}}$ containing $T = 200$ sparse samples, exactly 20 of which are nonzero. The nonzero sample values were randomly drawn in $[\frac{2}{3}; 1]$. We have generated $\mathbf{y}$ according to (1) with the nonlinearity $\phi(u_k) = \frac{u_k}{0.3+u_k}$ and with additive i.i.d. zero-mean Gaussian noise with standard deviation $\sigma = 0.15$. The banded Toeplitz matrix $\mathbf{H}$ has been set in accordance with two choices of FIR filters of length 3 (denoted $\mathbf{h}^{(a)}$ and $\mathbf{h}^{(b)}$). We have considered the estimate $\mathbf{x}_3^{\star\mathrm{s}}$ given by the optimal point of the SDP relaxation $\mathcal{P}_3^{\star\mathrm{s}}$ of order $k = 3$.

For comparison, we have implemented a classical gradient descent minimization of $\mathcal{J}$ and a proximal gradient algorithm based on Iterative Hard Thresholding (IHT) [3] extended to the the nonlinear model. Also, we have tested a convex relaxation based on a linearized reconstruction with ℓ_1 penalization. The local optimization algorithms have been started with different initializations and Table I indicates the existence of local minima.

On Figure 1, we have plotted the value $\mathcal{P}_3^{\star\mathrm{s}}$ reached by the SDP relaxation (which is a lower bound on $\mathcal{J}^\star$), the objective value $\mathcal{J}(\mathbf{x}_3^{\star\mathrm{s}})$ and the objective value reached using IHT using two different initializations. Clearly, our method provides a point close to a global minimizer and is very useful in providing a good initialization point for local optimization algorithms.

Finally, the estimation error has been quantified by $\|\hat{\mathbf{x}} - \overline{\mathbf{x}}\|$ for a given estimate $\hat{\mathbf{x}}$. The average error and objective values are summarized in Table II.

TABLE I

FINAL VALUES OF THE OBJECTIVE $\mathcal{J}(\mathbf{x})$ FOR THE CLASSICAL GRADIENT AND IHT LOCAL OPTIMIZATIONS (AVERAGE OVER 100 MONTE-CARLO REALIZATIONS). NOTE THAT OUR PROPOSED INITIALIZATION $\mathbf{x}_3^{\star s}$ LEADS TO THE LOWEST OBJECTIVE VALUE.

Gradient minimization					
Filter param.	Initialization				
	$\mathbf{x}_3^{\star s}$	ℓ_1	$\mathbf{y}$	zero	$\overline{\mathbf{x}}$
$\mathbf{h}^{(a)}$	6.9219	15.136	31.338	16.041	7.0894
$\mathbf{h}^{(b)}$	6.7078	13.245	30.222	18.060	7.0894

IHT minimization					
Filter param.	Initialization				
	$\mathbf{x}_3^{\star s}$	ℓ_1	$\mathbf{y}$	zero	$\overline{\mathbf{x}}$
$\mathbf{h}^{(a)}$	6.6943	8.4078	8.4129	16.041	6.7628
$\mathbf{h}^{(b)}$	6.6292	8.3442	8.2598	14.664	6.7372

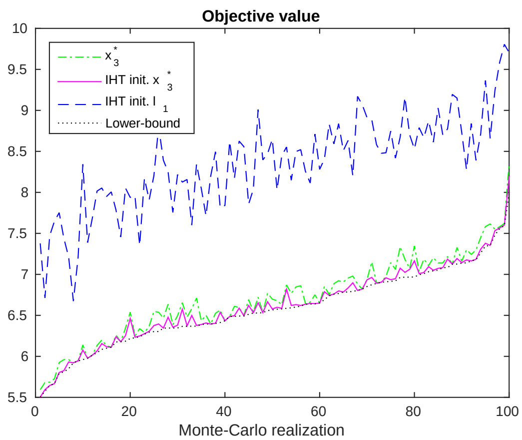

Fig. 1. Objective values provided by the different algorithms and lower-bound (using filter $\mathbf{h}^{(a)}$).

TABLE II

FINAL VALUES OF THE OBJECTIVE $\mathcal{J}(\mathbf{x})$ AND ESTIMATION ERROR GIVEN BY THE PROPOSED METHOD AND IHT WITH DIFFERENT INITIALIZATIONS (AVERAGE OVER 1000 MONTE-CARLO REALIZATIONS), SHOWING THAT A LINEARIZED RECONSTRUCTION IS NOT ADAPTED FOR THE CONSIDERED NONLINEAR MODEL.

	Objective		Error	
Filter param.	$\mathbf{h}^{(a)}$	$\mathbf{h}^{(b)}$	$\mathbf{h}^{(a)}$	$\mathbf{h}^{(b)}$
Proposed method	6.9219	6.7078	1.3278	1.5408
Proposed method + IHT	6.6943	6.6292	1.3374	1.5393
linear + ℓ_1 +IHT	8.4078	8.3442	1.5575	1.6833

REFERENCES

[1] 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, 2015.

[2] 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, 2013.

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

[4] 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, Jul. 2015.

[5] C. Soussen, J. Idier, J. Duan, and D. Brie, "Homotopy based algorithms for l0-regularized least-squares," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3301–3316, 2015.

[6] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 563–591, 2013.

[7] M. Shetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Wiley and sons, 1980.

[8] 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.

[9] Y. Deville and L. T. Duarte, "An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures," in *Proc. of the 12th Int. Conf. LVA/ICA*, ser. LNCS, vol. 9237. Liberec, Czech Republic: Springer, 2015, pp. 155–167.

[10] 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. 317–320.

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

[12] ——, *Moments, Positive Polynomials and Their Applications*, ser. Optimization Series. Imperial College Press, 2010, vol. 1.

[13] F. Bugarin, D. Henrion, and J.-B. Lasserre, "Minimizing the sum of many rational functions," *Mathematical Programming Computations*, vol. 8, no. 1, pp. 83–111, 2015.