ROBUST SPARSE ANALYSIS REGULARIZATION

Samuel Vaiter

(joint work with Gabriel Peyré, Charles Dossal and Jalal Fadili)

CEREMADE, Paris, France
Université Paris-Dauphine
vaiter@ceremade.dauphine.fr

ABSTRACT

This work studies some properties of ℓ^1-analysis regularization for the resolution of linear inverse problems. Analysis regularization minimizes the ℓ^1 norm of the correlations between the signal and the atoms in the dictionary. The corresponding variational problem includes several well-known regularizations such as the discrete total variation and the fused lasso. We give sufficient conditions such that analysis regularization is robust to noise.

1. ANALYSIS VERSUS SYNTHESIS

Regularization through variational analysis is a popular way to compute an approximation of $x_0 \in \mathbb{R}^N$ from the measurements $y \in \mathbb{R}^Q$ as defined by an inverse problem $y = \Phi x_0 + w$ where w is some additive noise and Φ is a linear operator, for instance a super-resolution or an inpainting operator.

A dictionary $D = (d_i)_{i=1}^P$ is a (possibly redundant) collection of P atoms $d_i \in \mathbb{R}^N$ which is used to synthesize a signal

$$x = D\alpha = \sum_{i=1}^{P} \alpha_i d_i.$$

Common examples in signal processing of dictionary include the wavelet transform or a finite-difference operator.

Synthesis regularization corresponds to the following minimization problem

$$\min_{\alpha \in \mathbb{R}^P} \frac{1}{2}\|y - \Psi\alpha\|_2^2 + \lambda\|\alpha\|_1, \tag{1}$$

where $\Psi = \Phi D$, and $x = D\alpha$. Properties of synthesis prior had been studied intensively, see for instance [1, 3].

Analysis regularization corresponds to the following minimization problem

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda\|D^* x\|_1. \tag{$\mathcal{P}_\lambda(y)$}$$

In the noiseless case, $w = 0$, one uses the constrained optimization which reads

$$\min_{x \in \mathbb{R}^N} \|D^* x\|_1 \quad \text{subject to} \quad \Phi x = y. \tag{$\mathcal{P}_0(y)$}$$

This prior had been less studied than the synthesis prior, see for instance [2]. In a synthesis prior, the generative vector α is sparse in the dictionary D whereas in analysis prior, the correlation between the signal x and the dictionary D is sparse. Synthesis and analysis regularizations differ significantly when D is redundant.

2. UNION OF SUBSPACES MODEL

It is natural to keep track of the support of this correlation vector, as done in the following definition.

Definition 1. *The* D-support I *of a vector* $x \in \mathbb{R}^N$ *is defined as* $I = \operatorname{supp}(D^* x)$. *Its* D-cosupport J *is defined as* $J = I^c$.

A signal x such that $D^* x$ is sparse lives in a cospace $\mathcal{G}_J$ of small dimension where $\mathcal{G}_J$ is defined as follow.

Definition 2. *Given a dictionary* D*, and* J *a subset of* $\{1 \cdots P\}$*, the* cospace $\mathcal{G}_J$ *is defined as*

$$\mathcal{G}_J = \operatorname{Ker} D_J^*,$$

where D_J *is the subdictionary whose columns are indexed by* J.

The signal space can thus be decomposed as a union of subspaces of increasing dimensions

$$\Theta_k = \{\mathcal{G}_J \setminus \dim \mathcal{G}_J = k\}.$$

For the 1-D total variation prior, Θ_k is the set of piecewise constant signals with $k-1$ steps.

3. ROBUSTNESS AND IDENTIFIABILITY

We give [4] a sufficient condition on x_0 ensuring that the solution of $\mathcal{P}_\lambda(y)$ is close to x_0 when w is small enough.

Definition 3. *Let $s \in \{-1, 0, +1\}^P$, I its D-support and J its D-cosupport. We suppose*

$$\mathcal{G}_J \cap \operatorname{Ker} \Phi = \{0\}, \qquad (H_J)$$

holds. The analysis Identifiabiltiy Criterion **IC** *of s is defined as*

$$\mathbf{IC}(s) = \min_{u \in \operatorname{Ker} D_J} \|\Omega s_I - u\|_\infty,$$

where $\Omega = D_J^+(\Phi^*\Phi A^{[J]} - \operatorname{Id})D_I$ and $A^{[J]}$ denotes the matrix $U(U^*\Phi^*\Phi U)^{-1}U^*$ where U is a matrix which columns form a basis of $\mathcal{G}_J$.

Theorem 1. *Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-cosupport J, and of D-support $I = J^c$. Suppose (H_J) holds and $\mathbf{IC}(\operatorname{sign}(D^*x_0)) < 1$. There exist two constants $c_J > 0$ and $\tilde{c}_J > 0$, such that if $y = \Phi x_0 + w$, where*

$$\frac{\|w\|_2}{T} < \frac{\tilde{c}_J}{c_J} \quad \text{and} \quad T = \min_{i \in \{1,\cdots,|I|\}} |D_I^* x_0|_i,$$

and if λ satisfies

$$c_J \|w\|_2 < \lambda < T\tilde{c}_J,$$

the vector defined by

$$\hat{x}^\star = x_0 + A^{[J]}\Phi^* w - \lambda A^{[J]} D_I s_I, \qquad (2)$$

is the unique solution of $\mathcal{P}_\lambda(y)$. Moreover,

$$\hat{x}^\star \in \mathcal{G}_J \quad \text{and} \quad \operatorname{sign}(D^*x_0) = \operatorname{sign}(D^*\hat{x}^\star).$$

Note that it is possible to choose λ proportional to the noise level $\|w\|_2$. Hence, for $\|w\|_2$ small enough, equation (2) gives

$$\|\hat{x}^\star - x_0\| = O(\|w\|_2).$$

The same condition ensures that x_0 is the unique solution of $\mathcal{P}_0(y)$ when $w = 0$.

Theorem 2. *Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-cosupport J. Suppose that condition (H_J) holds and $\mathbf{IC}(\operatorname{sign}(D^*x_0)) < 1$. Then x_0 is identifiable (i.e x_0 is the unique solution of $\mathcal{P}_0(\Phi x_0)$).*

Our last contribution defines a stronger criterion that ensures robustness to an arbitrary bounded noise.

Definition 4. *The analysis* Recovery Criterion (RC) *of $I \subset \{1 \dots P\}$ is defined as*

$$\mathbf{RC}(I) = \max_{x \in \mathcal{G}_J} \mathbf{IC}\left(\operatorname{sign}(D^*x)\right).$$

Theorem 3. *Let I be a fixed D-support and J its associated D-cosupport $J = I^c$. Suppose that (H_J) holds. If $\mathbf{RC}(I) < 1$ and*

$$\lambda = \rho \|w\|_2 \frac{c_J}{1 - \mathbf{RC}(I)} \quad \text{with} \quad \rho > 1,$$

where c_J is a positive constant, then for every x_0 of D-support I, there exists a unique solution $x^\star$ of D-support included in I, verifying

$$\|x^\star - x_0\|_2 = O(\|w\|_2).$$

This theorem shows that if the parameter λ is big enough, then $\mathcal{P}_\lambda(y)$ recovers a unique vector which is close enough in the ℓ^2 sense and lives in the same $\mathcal{G}_J$ as the unknown signal x_0.

4. EXAMPLE : TOTAL VARIATION

The most popular analysis sparse regularization is the total variation which corresponds to using a derivative operator, i.e a convolution by the vector $(1, -1)$.

A signal is said to contain a staircase subsignal if there exists $i \in \{1 \dots |I| - 1\}$ such that

$$\operatorname{sign}(D_I^* x)_i = \operatorname{sign}(D_I^* x)_{i+1} = \pm 1.$$

Proposition 1. *We consider the denoising case, $\Phi = \operatorname{Id}$. If x does not contain a staircase subsignal, then $\mathbf{IC}(\operatorname{sign}(D^*x)) < 1$. Otherwise, $\mathbf{IC}(\operatorname{sign}(D^*x)) = 1$.*

This proposition together with Theorem 1 shows that if a signal does not have a staircase subsignal, TV denoising is robust to a small noise.

References

[1] J.J. Fuchs, On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.

[2] S. Nam, M.E., Davies, M. Elad and R. Gribonval, The Cosparse Analysis Model and Algorithms, Preprint arxiv-1106.4987v1, 2011.

[3] J.A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.

[4] S. Vaiter, G. Peyré, C. Dossal and J. Fadili, Robust Sparse Analysis Regularization, Preprint arxiv-1109.6222v1, 2011.