Neurocomputing 173 (2016) 278–289

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A greedy algorithm for the analysis transform domain

Raja Giryes

The Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form
7 February 2015
Accepted 20 February 2015
Available online 6 September 2015

Keywords:
Sparse representations
Compressed sensing
Sparsity for big data
Analysis
Iterative hard thresholding
Greedy algorithms

ABSTRACT

Many image processing applications benefited remarkably from the theory of sparsity. One model of sparsity is the cosparse analysis one. It was shown that using ℓ_1-minimization one might stably recover a cosparse signal from a small set of random linear measurements if the operator is a frame. Another effort has provided guarantee for dictionaries that have a near optimal projection procedure using greedy-like algorithms. However, no claims have been given for frames. A common drawback of all these existing techniques is their high computational cost for large dimensional problems.

In this work we propose a new greedy-like technique with theoretical recovery guarantees for frames as the analysis operator, closing the gap between greedy and relaxation techniques. Our results cover both the case of bounded adversarial noise, where we show that the algorithm provides us with a stable reconstruction, and the one of random Gaussian noise, for which we prove that it has a denoising effect, closing another gap in the analysis framework. Our proposed program, unlike the previous greedy-like ones that solely act in the signal domain, operates mainly in the analysis operator's transform domain. Besides the theoretical benefit, the main advantage of this strategy is its computational efficiency that makes it easily applicable to visually big data. We demonstrate its performance on several high dimensional images.

1. Introduction

For more than a decade the idea that signals may be represented sparsely has a great impact on the field of signal and image processing. New sampling theory has been developed [1] together with new tools for handling signals in different types of applications, such as image denoising [2], image deblurring [3], super-resolution [4], radar [5], medical imaging [6] and astronomy [7], to name a few [8]. Remarkably, in most of these fields the sparsity based techniques achieve state-of-the-art results.

The classical sparse model is the synthesis one. In this model the signal $\mathbf{x} \in \mathbb{R}^d$ is assumed to have a k-sparse representation $\boldsymbol{\alpha} \in \mathbb{R}^n$ under a given dictionary $\mathbf{D} \in \mathbb{R}^{d \times n}$. Formally,

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}, \quad \|\boldsymbol{\alpha}\|_0 \le k, \tag{1}$$

where $\|\cdot\|_0$ is the ℓ_0-pseudo norm that counts the number of non-zero entries in a vector. Notice, that the non-zero elements in α corresponds to a set of columns that creates a low-dimensional subspace in which $\mathbf{x}$ resides.

Recently, a new sparsity based model has been introduced: the analysis one [9,10]. In this framework, we look at the coefficients of $\boldsymbol{\Omega}\mathbf{x}$, the coefficients of the signal after applying the transform $\boldsymbol{\Omega} \in \mathbb{R}^{p \times d}$ on it. The sparsity of the signal is measured by the number of zeros in $\boldsymbol{\Omega}\mathbf{x}$. We say that a signal is ℓ-cosparse if $\boldsymbol{\Omega}\mathbf{x}$ has ℓ zero elements. Formally,

$$\|\boldsymbol{\Omega}\mathbf{x}\|_0 \le p - \ell. \tag{2}$$

Remark that each zero element in $\boldsymbol{\Omega}\mathbf{x}$ corresponds to a row in $\boldsymbol{\Omega}$ to which the signal is orthogonal and all these rows define a subspace to which the signal is orthogonal. Similar to synthesis, when the number of zeros is large the signal's subspace is low dimensional. Though the zeros are those that define the subspace, in some cases it is more convenient to use the number of non-zeros $k = p - \ell$ as done in [11,12].

The main setup in which the above models have been used is

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}, \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^m$ is a given set of measurements, $\mathbf{M} \in \mathbb{R}^{m \times d}$ is the measurement matrix and $\mathbf{e} \in \mathbb{R}^m$ is an additive noise which is assumed to be either adversarial bounded noise [1,8,13,14] or with a certain given distribution such as Gaussian [15–17]. The goal is to recover $\mathbf{x}$ from $\mathbf{y}$ and this is the focus of our work. For details about other setups, the curious reader may refer to [18,19,20–24].

Clearly, without a prior knowledge it is impossible to recover $\mathbf{x}$ from $\mathbf{y}$ in the case $m < d$, or have a significant denoising effect when $\mathbf{e}$ is random with a known distribution. Hence, having a prior, such as the sparsity one, is vital for these tasks. Both the

E-mail address: raja.giryes@duke.edu

http://dx.doi.org/10.1016/j.neucom.2015.02.100
0925-2312/

synthesis and the analysis models lead to (different) minimization problems that provide estimates for the original signal $\mathbf{x}$.

In synthesis, the signal is recovered by its representation, using

$$\hat{\boldsymbol{\alpha}}_{S-\ell_0} = \arg\min_{\tilde{\boldsymbol{\alpha}} \in \mathbb{R}^n} \|\tilde{\boldsymbol{\alpha}}\|_0 \quad s.t \quad \|\mathbf{y} - \mathbf{M}\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_2 \le \lambda_{\mathbf{e}}, \tag{4}$$

where $\lambda_{\mathbf{e}}$ is an upper bound for $\|\mathbf{e}\|_2$ if the noise is bounded and adversarial ($S-\ell_0$ refers to synthesis-ℓ_0). Otherwise, it is a scalar dependent on the noise distribution [15,16,25]. The recovered signal is simply $\hat{\mathbf{x}}_{S-\ell_0} = \mathbf{D}\hat{\boldsymbol{\alpha}}_{S-\ell_0}$. In analysis, we have the following minimization problem:

$$\hat{\mathbf{x}}_{A-\ell_0} = \mathrm{argmin}_{\tilde{\mathbf{x}} \in \mathbb{R}^d} \|\boldsymbol{\Omega}\tilde{\mathbf{x}}\|_0 \quad s.t \quad \|\mathbf{y} - \mathbf{M}\tilde{\mathbf{x}}_2 \le \lambda_{\mathbf{e}}. \tag{5}$$

The values of $\lambda_{\mathbf{e}}$ are selected as before depending on the noise properties ($A-\ell_0$ refers to analysis-ℓ_0).

Both (4) and (5) are NP-hard problems [10,26]. Hence, approximation techniques are required. These are divided mainly into two categories: relaxation methods and greedy algorithms. In the first category we have the ℓ_1-relaxation [9] and the Dantzig selector [15], where the latter has been proposed only for synthesis. The ℓ_1-relaxation leads to the following minimization problems for synthesis and analysis respectively[1]

$$\hat{\boldsymbol{\alpha}}_{S-\ell_1} = \arg\min_{\tilde{\boldsymbol{\alpha}} \in \mathbb{R}^n} \|\tilde{\boldsymbol{\alpha}}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{M}\mathbf{D}\tilde{\boldsymbol{\alpha}}_2 \le \lambda_{\mathbf{e}}, \tag{6}$$

$$\hat{\mathbf{x}}_{A-\ell_1} = \underset{\tilde{\mathbf{x}}}{\mathrm{argmin}} \in \mathbb{R}^d \|\boldsymbol{\Omega}\tilde{\mathbf{x}}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{M}\tilde{\mathbf{x}}\|_2 \le \lambda_{\mathbf{e}}. \tag{7}$$

Among the synthesis greedy strategies we mention the thresholding method, orthogonal matching pursuit (OMP) [29–31], CoSaMP [32], subspace pursuit (SP) [33], iterative hard thresholding [34] and hard thresholding pursuit (HTP) [35]. Their counterparts in analysis are thresholding [36], GAP [10], analysis CoSaMP (ACoSaMP), analysis SP (ASP), analysis IHT (AIHT) and analysis HTP (AHTP) [37].

An important question to ask is what are the recovery guarantees that exist for these methods. One main tool that was used for answering this question in the synthesis context is the restricted isometry property [13]. It has been shown that under some conditions on the RIP of $\mathbf{MD}$, we have in the adversarial bounded noise case that

$$\|\hat{\boldsymbol{\alpha}}_{alg} - \boldsymbol{\alpha}\|_2^2 \le C_{alg}\|\mathbf{e}\|_2^2, \tag{8}$$

where $\hat{\boldsymbol{\alpha}}_{alg}$ is the recovered representation by one of the approximation algorithms and $C_{alg} > 2$ is a constant that depends on the RIP of $\mathbf{MD}$ and differs for each of the methods [1,13,31–34,37–39]. This result implies that these programs achieve a stable recovery.

Similar results were provided for the case where the noise is random white Gaussian with variance σ^2 [15–17,40,41]. In this case the reconstruction error is guaranteed to be $O(k\log(n)\sigma^2)$ [15–17]. Unlike the adversarial noise case, here we may have a denoising effect, as the recovery error can be smaller than the initial noise power $d\sigma^2$. Remark that the above results can be extended also to the case where we have a model mismatch and the signal is not exactly k-sparse.

In the analysis framework we have similar guarantees for the adversarial noise case. However, since the analysis model treats the signal directly, the guarantees are in terms of the signal and not the representation like in 8. Two extensions for the RIP have been proposed providing guarantees for analysis algorithms. The first is the D-RIP [11]:

Definition 1.1 (*D-RIP (Candès et al. [11])*). A matrix $\mathbf{M}$ has the D-RIP with a dictionary $\mathbf{D}$ and a constant $\delta_k = \delta_{\mathbf{D},k}$, if δ_k is the smallest constant that satisfies

$$(1-\delta_k)\|\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_2^2 \le \|\mathbf{M}\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_2^2 \le (1+\delta_k)\|\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_2^2, \tag{9}$$

whenever $\tilde{\boldsymbol{\alpha}}$ is k-sparse.

The D-RIP has been used for studying the performance of the analysis ℓ_1-minimization [11,42,43]. It has been shown that if $\boldsymbol{\Omega}$ is a frame with frame constants A and B, $\mathbf{D} = \boldsymbol{\Omega}^\dagger$ and $\mathbf{M}$ has the D-RIP with $\delta_{ak} \le \delta_{A-\ell_1}(a, A, B)$ then

$$\|\hat{\mathbf{x}}_{A-\ell_1} - \mathbf{x}\|_2^2 \le C_{A-\ell_1}\left(\|\mathbf{e}\|_2^2 + \frac{\|\boldsymbol{\Omega}\mathbf{x} - [\boldsymbol{\Omega}\mathbf{x}]_k\|_1^2}{k}\right), \tag{10}$$

where the operator $[\cdot]_k$ is a hard thresholding operator that keeps the largest k elements in a vector, $\delta_{A-\ell_1}(a, A, B)$ is a function of a, A and B, and $a \ge 1$ and $C_{A-\ell_1}$ are some constants. A similar result has been proposed for analysis ℓ_1-minimization with the finite difference operator [28,44].

The second is the O-RIP [37], which was used for the study of the greedy-like algorithms ACoSaMP, ASP, AIHT and AHTP.

Definition 1.2 (*O-RIP (Giryes et al. [37])*). A matrix $\mathbf{M}$ has the O-RIP with an operator $\boldsymbol{\Omega}$ and a constant $\delta_{\boldsymbol{\Omega},\ell}$, if $\delta_{\boldsymbol{\Omega},\ell}$ is the smallest constant that satisfies

$$(1-\delta_{\boldsymbol{\Omega},\ell})\|\mathbf{v}_2^2 \le \|\mathbf{M}\mathbf{v}_2^2 \le (1+\delta_{\boldsymbol{\Omega},\ell})\|\mathbf{v}\|_2^2, \tag{11}$$

whenever $\boldsymbol{\Omega}\mathbf{v}$ has at least ℓzeroes.

With the assumption that there exists a cosupport selection procedure $\hat{\mathcal{S}}_\ell$ that implies a near optimal projection for $\boldsymbol{\Omega}$ with a constant C_ℓ (see Definition 3.1 in Section 3). It has been proven for such operators that if $\delta_{\boldsymbol{\Omega},a\ell} \le \delta_{alg}(C_\ell, C_{2\ell-p}, \sigma_{\mathbf{M}}^2)$ then

$$\|\hat{\mathbf{x}}_{A-\ell_1} - \mathbf{x}\|_2^2 \le C_{alg}\left(\|\mathbf{e}\|_2^2 + \|\mathbf{x} - \mathbf{x}^\ell\|s_2^2\right), \tag{12}$$

where $\sigma_{\mathbf{M}}^2$ is the largest singular value of $\mathbf{M}$, $\mathbf{x}^\ell$ is the best ℓ-cosparse approximation for $\mathbf{x}$, $\delta_{alg}(C_\ell, C_{2\ell-p}, \sigma_{\mathbf{M}}^2)$ is a function of C_ℓ, $C_{2\ell-p}$ and $\sigma_{\mathbf{M}}^2$, and $a \ge 3$ and C_{alg} are some constants that differ for each technique.

Notice that the conditions in synthesis imply that no linear dependencies can be allowed within small number of columns in the dictionary as the representation is the focus. The existence of such dependencies may cause ambiguity in its recovery. Since the analysis model performs in the signal domain, i.e. focus on the signal and not its representation, dependencies may be allowed within the dictionary. A recent series of contributions have shown that high correlations can be allowed in the dictionary also in the synthesis framework if the signal is the target and not the representation [45–51].

1.1. Our contribution

The conditions for greedy-like techniques require the constant C_ℓ to be close to 1. Having a general projection scheme with $C_\ell = 1$ is NP-hard [52]. The existence of a program with a constant close to one for a general operator is still an open problem. In particular, it is not known whether there exists a procedure that gives a small constant for frames. Thus, there is a gap between the results for the greedy techniques and the ones for the ℓ_1-minimization.

Another drawback of the existing analysis greedy strategies is their high complexity. All of them require applying a projection to an analysis cosparse subspace, which implies a high computational cost. Therefore, unlike in the synthesis case, they do not provide a “cheap” counterpart to the ℓ_1-minimization.

In this work we propose a new efficient greedy program, the transform domain IHT (TDIHT), which is an extension of IHT that operates in the analysis transform domain. Unlike AIHT, TDIHT has a low complexity, as it does not require applying computationally demanding projections like AIHT, and it inherits guarantees

[1] Note that setting $\boldsymbol{\Omega}$ to be the finite difference operator in 7 leads to the anisotropic version of the well-known total variation (TV) [27]. See [28,44] for more details.