Sampling in the Analysis Transform Domain

Raja Giryes\textsuperscript{a}

\textsuperscript{a}The Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA

Abstract

Many signal and image processing applications have benefited remarkably from the fact that the underlying signals reside in a low dimensional subspace. One of the main models for such a low dimensionality is the sparsity one. Within this framework there are two main options for the sparse modeling: the synthesis and the analysis ones, where the first is considered the standard paradigm for which much more research has been dedicated. In it the signals are assumed to have a sparse representation under a given dictionary. On the other hand, in the analysis approach the sparsity is measured in the coefficients of the signal after applying a certain transformation, the analysis dictionary, on it. Though several algorithms with some theory have been developed for this framework, they are outnumbered by the ones proposed for the synthesis methodology.

Given that the analysis dictionary is either a frame or the two dimensional finite difference operator, we propose a new sampling scheme for signals from the analysis model that allows to recover them from their samples using any existing algorithm from the synthesis model. The advantage of this new sampling strategy is that it makes the existing synthesis methods with their theory also available for signals from the analysis framework.

Keywords: Sparse representations, Compressed sensing, Synthesis, Analysis, Transform Domain.

2010 MSC: 94A20, 94A12, 62H12

1. Introduction

The idea that signals reside in a union of low dimensional subspaces has been used extensively in the recent decade in many fields and applications [1]. One of the main problems that has benefited remarkably from this theory is the one of compressed sensing. In this problem we want to recover an unknown signal $x \in \mathbb{R}^d$ from a small number of noisy linear measurements:

$$y = Mx + e,$$

where $M \in \mathbb{R}^{m \times d}$ is the measurements matrix, $e \in \mathbb{R}^d$ is an additive noise and $y \in \mathbb{R}^m$ is the noisy measurement.

If the signal $x$ can be any signal then we are in a hopeless situation in the task of recovering it from $y$. However, if we restrict it to a low-dimensional manifold that does not intersect with the null space of $M$ at any point except the origin then we are more likely to be able to recover $x$ from $y$ by looking for the signal at this manifold, which is closest to $y$ after multiplying it by $M$.

An example for such a low dimensional manifold is the one of $k$-sparse signals under a given dictionary $D \in \mathbb{R}^{d \times n}$. In this case our signal $x$ satisfies

$$x = D\alpha, \|\alpha\|_0 \leq k,$$

where $\|\alpha\|_0$ is the $\ell_0$-pseudo norm that counts the number of non-zero entries in a vector. In this case we may recover $x$ from $y$ by minimizing the following problem,

$$\hat{\alpha}_{S-k} = \arg\min_{\alpha} \|\alpha\|_0 \quad s.t. \quad \|y - MD\alpha\|_2 \leq \lambda_e.$$

References

[1] Raja Giryes, et al., "Sparse representations, Compressed sensing, Synthesis, Analysis, Transform Domain.

2010 MSC: 94A20, 94A12, 62H12

1. Introduction

The idea that signals reside in a union of low dimensional subspaces has been used extensively in the recent decade in many fields and applications [1]. One of the main problems that has benefited remarkably from this theory is the one of compressed sensing. In this problem we want to recover an unknown signal $x \in \mathbb{R}^d$ from a small number of noisy linear measurements:

$$y = Mx + e,$$

where $M \in \mathbb{R}^{m \times d}$ is the measurements matrix, $e \in \mathbb{R}^d$ is an additive noise and $y \in \mathbb{R}^m$ is the noisy measurement.

If the signal $x$ can be any signal then we are in a hopeless situation in the task of recovering it from $y$. However, if we restrict it to a low-dimensional manifold that does not intersect with the null space of $M$ at any point except the origin then we are more likely to be able to recover $x$ from $y$ by looking for the signal at this manifold, which is closest to $y$ after multiplying it by $M$.

An example for such a low dimensional manifold is the one of $k$-sparse signals under a given dictionary $D \in \mathbb{R}^{d \times n}$. In this case our signal $x$ satisfies

$$x = D\alpha, \|\alpha\|_0 \leq k,$$

where $\|\alpha\|_0$ is the $\ell_0$-pseudo norm that counts the number of non-zero entries in a vector. In this case we may recover $x$ from $y$ by minimizing the following problem,

$$\hat{\alpha}_{S-k} = \arg\min_{\alpha} \|\alpha\|_0 \quad s.t. \quad \|y - MD\alpha\|_2 \leq \lambda_e.$$
where \( \lambda_e \) is an upper bound for \( \|e\|_2 \) if the noise is bounded and adversarial, or a scalar dependent on the noise distribution [2]. As this problem is NP-hard [3] many approximation methods have been proposed for it [4, 5], such as orthogonal matching pursuit (OMP) [6] and the \( \ell_1 \)-relaxation strategy that replaces the \( \ell_0 \)-pseudo norm with the \( \ell_1 \)-norm in (3) [7].

One of the main theoretical questions being asked with regard to these algorithms is what are the requirements on \( \mathbf{M}, \mathbf{D}, m, k \) and \( \Omega \) such that the representation, \( \alpha \), of \( \mathbf{x} \) may be stably recovered from \( \mathbf{y} \) using these techniques, i.e., their recovery \( \hat{\alpha} \) will satisfy

\[
\|\hat{\alpha} - \alpha\|_2 \leq C\|e\|_2, \tag{4}
\]

where \( C \) is a certain constant (different for each algorithm).

Two main tools have been used to answer this question. The first is the coherence of \( \mathbf{MD} \) [8], which is the maximal (normalized) inner product between the columns of \( \mathbf{MD} \). It has been shown that if the matrix \( \mathbf{MD} \) is incoherent (has a small coherence) then it is possible to get a stable recovery using OMP and the \( \ell_1 \)-relaxation. The problem with the coherence based recovery conditions is that they limit the number of measurements \( m \) to be of the order of \( k^2 \), while \( m = 2k \) is enough to guarantees uniqueness for (1) in the noiseless case and \( m = O(k \log(n)) \) is enough for stability in the noisy one.

The second property of \( \mathbf{MD} \) used to derive reconstruction performance guarantees is the restricted isometry property (RIP). This property provides us with a bound on the minimal and maximal eigenvalues of every sub-matrix consisting of any \( k \)-columns from a given matrix. Formally,

**Definition 1.1 (RIP [9]).** A matrix \( \mathbf{A} \in \mathbb{R}^{m \times n} \) has the RIP with a constant \( \delta_k \), if \( \delta_k \) is the smallest constant that satisfies

\[
(1 - \delta_k) \|\mathbf{A}\|_2^2 \leq \|\mathbf{A} \hat{\mathbf{\alpha}}\|_2^2 \leq (1 + \delta_k) \|\mathbf{A}\|_2^2, \tag{5}
\]

whenever \( \hat{\alpha} \in \mathbb{R}^n \) is \( k \)-sparse.

It has been shown for many approximation algorithms that they get stable recovery in the form of (4), if \( \mathbf{MD} \) has the RIP with a constant \( \delta_{ak} < \delta_{ref} \), where \( a \) and \( \delta_{ref} \) are two constants dependent on the algorithm in question [9, 10, 11, 12, 13, 14]. The true force behind these RIP conditions is that it has been shown that many matrices (typically random subgaussian matrices) satisfy this bound given that \( m = O(k \log(n)) \) [9, 15, 16]. Notice that the main significance of this result is that it shows that it is possible to recover a signal from a number of measurements proportional to its manifold dimension \( k \).

An alternative model for low dimensional signals that relies on sparsity is the analysis framework [17, 18]. In this paradigm, we look at the behavior of the signal after applying a certain operator \( \mathbf{\Omega} \in \mathbb{R}^{md} \) on it, assuming that \( \mathbf{\Omega} \mathbf{x} \) has \( \ell \) zeros. With this prior at hand, we may recover \( \mathbf{x} \) from (1) by solving

\[
\hat{\mathbf{x}}_{\Lambda - \ell \text{-zero}} = \arg\min_{\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_0 \quad s.t. \quad \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2 \leq \lambda_e, \tag{6}
\]

where here also \( \lambda_e \) depends on the noise properties.

Note that as we minimize the number of non-zeros in \( \mathbf{\Omega}\mathbf{x} \) in (6), the number of zeros is the one that defines the manifold dimension in which \( \mathbf{x} \) resides. Each zero in \( \mathbf{\Omega}\mathbf{x} \) corresponds to a row in \( \mathbf{\Omega} \) to which \( \mathbf{x} \) is orthogonal. Denoting by \( T \) the support of \( \mathbf{\Omega}\mathbf{x} \) and \( T^c \) it complimentary, we may say that \( \mathbf{x} \) resides in a subspace of dimension \( d - \text{rank}(\mathbf{\Omega}_{T^c}) \). Therefore if \( \mathbf{\Omega}\mathbf{x} \) has \( \ell \) rows, \( k \) is the number of non-zeros in it, and \( \mathbf{\Omega} \) is in general position, i.e., every \( d \) rows in it are independent, then the manifold dimension is \( d - \ell \).

In the noiseless case (\( e = 0 \)), the requirement \( m = 2(d - \text{rank}(\mathbf{\Omega}_{T^c})) \) is enough to guarantee uniqueness in the solution of (6) (and therefore the recovery of \( \mathbf{x} \)) under very mild assumptions on the relation between \( \mathbf{\Omega} \) and \( \mathbf{M} \) [18]. However, in the noisy case having a number of samples at the order of the manifold dimension, i.e., \( m = O(d - \text{rank}(\mathbf{\Omega}_{T^c})) \) is not enough to guarantee stability even by solving (6) [19]. Therefore, it is not surprising that the recovery conditions for algorithms that approximate (6) require \( m = O(k \log(n)) \) [20, 21, 22, 23, 24, 25], where \( \mathbf{\Omega} \) is assumed to be either a frame [20, 21, 24, 25], the 2D-DIF operator [23] or an operator that generates a manifold with a tractable projection onto it [22].

Though the number of measurements in synthesis and analysis are similar there are two major differences between the two: (i) In synthesis the number of measurements are proportional to the manifold dimension, while in analysis
this is not necessarily the case as \( k = n - \ell \) might be remarkably larger than \( d - \text{rank}(\Omega_\ell^T) \) (See [22] for more details); (ii) In synthesis the dictionary \( D \) must be incoherent as otherwise the RIP condition will no longer hold [26], while in the analysis case there is no such restriction on the analysis dictionary \( \Omega \) but only on \( M \).

An interesting relation between analysis and synthesis, which is depicted in [17], is that if \( \Omega \) is a frame and \( \Omega x \) is \( k \)-sparse then \( x \) has a \( k \)-sparse representation under \( D = \Omega^\dagger \) (the pseudo-inverse of \( \Omega \)), i.e., \( x = D\Omega x \). Therefore, if \( k \) is small enough then relying on the uniqueness of the sparse representation [27], we can recover \( x \) by minimizing (3). The problem we encounter in this case is that unless \( \Omega \) is an incoherent matrix (its rows are incoherent) and therefore \( D \) is incoherent, non of the existing synthesis approximation algorithms is guaranteed to provide us with a good estimate\(^1\).

1.1. Our Contribution

In this work we provide a new sampling strategy that allows recovering signals from the analysis model using any existing synthesis algorithm, given that the analysis dictionary is either a frame or the 2D-DIF operator. Our scheme is general and can be easily extended to other types of analysis dictionaries. Instead of sampling the signal itself, we sample the signal in the analysis transform domain and then perform the recovery in this domain. From the proxy in the transform domain we get a reconstruction of our original signal. The idea to recover an analysis signal in the transform domain is not a new idea and was used before [25, 34, 35, 36]. However, the uniqueness in our approach compared to previous works is that (i) we sample with one matrix and then use another one for recovery; and (ii) we make use of existing synthesis algorithms as a black box without changing them for recovering the transform domain coefficients of the signal. Our sampling and recovery strategy is presented in Section 2 for the case that the analysis dictionary is a general frame or the 2D-DIF operator. In Section 3 we provide a simple demonstration of the usage of our scheme and in Section 4 we conclude the paper.

2. Sampling in the Transform Domain

Before we turn to present our scheme let us recall the problem we aim at solving in the analysis case:

**Definition 2.1 (Problem \( \mathcal{P} \)).** Consider a measurement vector \( y \in \mathbb{R}^m \) such that \( y = Mx + e \) where \( \Omega x \in \mathbb{R}^d \) is either \( k \)-sparse for a given and fixed analysis operator \( \Omega \in \mathbb{R}^{m \times d} \) or almost \( k \)-sparse, i.e. \( \Omega x \) has \( k = n - \ell \) leading elements. The non-zero locations of the \( k \) leading elements is denoted by \( T \). \( M \in \mathbb{R}^{m \times d} \) is a degradation operator and \( e \in \mathbb{R}^m \) is an additive noise. Our task is to recover \( x \) from \( y \). The recovery result is denoted by \( \hat{x} \).

2.1. Guarantees for Frames

Let \( A \in \mathbb{R}^{m \times n} \) be a given matrix and \( \mathcal{P}(y) = \mathcal{P}(y|A, k) \) be an algorithm that receives a signal \( y \) such that \( y = A\alpha + e \), where \( \alpha \in \mathbb{R}^n \) is either \( k \)-sparse or almost \( k \)-sparse, such that either one of the following (or the two of them) holds: (i) for the case that \( e \) is an adversarial noise with a bounded energy it is guaranteed that

\[
||\alpha - \mathcal{P}(y)||_2 \leq C_1 ||e||_2 + C_2 \left( ||\alpha||_2^2 + \frac{1}{\sqrt{k}} ||\alpha||_1^2 \right),
\]

where \( ||\alpha||_1 \) is the best \( k \)-term approximation of \( \alpha \), and \( C_1 \) and \( C_2 \) are two constants depending on \( A \) and the algorithms\(^2\) (See [9, 10, 11, 12, 13, 14]); or (ii) for the case that \( e \) is a zero-mean white Gaussian noise with variance \( \sigma^2 \) it is guaranteed that

\[
||\alpha - \mathcal{P}(y)||_2^2 \leq C_3 \sigma^2 \log(n) + C_4 \left( ||\alpha||_2^2 + \frac{1}{\sqrt{k}} ||\alpha||_1^2 \right),
\]

where \( C_3 \) and \( C_4 \) are two constants depending on \( A \) and the algorithms (See [37, 38, 39, 40]).

Assuming that \( \Omega \) in Problem \( \mathcal{P} \) is a frame, we propose the following sampling and reconstruction strategy:

\(^1\)Some recent works have addressed the case of coherent dictionaries in the synthesis case [28, 29, 30, 31, 32, 33]. However, they are very limited to specific cases and do not apply to general types of dictionaries such as frames.

\(^2\)Note that (7) is a generalization of the bound in (4) for the case that \( \alpha \) is a non-exact \( k \)-sparse vector.
• Set the sensing matrix to be \( M = A \Omega \). In this case we have \( y = Mx + e = A \Omega x + e \) and therefore we can apply algorithm \( \mathcal{P} \) to recover \( \Omega x \) as it is a \( k \)-sparse (or approximately so) vector.

• Compute an estimate for \( \Omega x: \hat{\alpha} = \mathcal{P}(y) \).

• Use the frame’s Moore-Penrose pseudo-inverse to recover \( x: \hat{x} = \Omega^\dagger \hat{\alpha} \).

This algorithm is summarized also in Algorithm 1. Remark that we sample in the transform domain of \( \Omega \), as we sample with \( M = A \Omega \), and then recover only with \( A \) the transform coefficients of \( x \), i.e. \( \Omega x \). Note also that in the final step, where we calculate \( \hat{x} = \Omega^\dagger \hat{\alpha} \), we may replace \( \Omega^\dagger \) with any dictionary that satisfies \( D \Omega = I \).

\[ \text{Algorithm 1 Signal Recovery from Samples of Frames in the Transform Domain} \]

\begin{itemize}
  \item \textbf{Require:} \( k, A \in \mathbb{R}^{m \times p}, \Omega \in \mathbb{R}^{p \times d}, y, \mathcal{P} \), where \( y = A \Omega x + e \), \( \Omega x \) is a \( k \)-sparse vector or approximately so, \( e \) is an additive noise, and \( \mathcal{P}(\cdot) = \mathcal{P}(\cdot | A, k) \) is a synthesis recovery program for \( k \)-sparse signals under the matrix \( A \).
  \item \textbf{Ensure:} \( x \): Approximation of \( \Omega x \; \hat{w} = \mathcal{P}(y | A, k) \)
  \item Signal recovery: \( \hat{x} = \Omega^\dagger \hat{w} \), generating a signal estimate using the transform domain proxy.
\end{itemize}

The following theorem provides guarantees for signal recovery using the above scheme given that the synthesis reconstruction program used in it \( \mathcal{P} \) satisfies either (7) or (8), or both of them.

\[ \text{Theorem 2.2 (Signal recovery from samples of frames in the transform domain).} \]

Consider the problem \( \mathcal{P} \) such that \( M = A \Omega \) and \( \Omega \) is a frame with a lower frame bound \( A \). Let \( \hat{x} \) be the output of Algorithm 1 with the synthesis program \( \mathcal{P}(\cdot | A, k) \). If \( e \) is a bounded additive adversarial noise and (7) holds for \( \mathcal{P}(\cdot | A, k) \) then

\[ \| x - \hat{x} \|_2^2 \leq \frac{C_1}{A} \| e \|_2^2 + \frac{C_2}{A} \left( \| \Omega \mathcal{P}(x) \|_2^2 + \frac{1}{k} \| \Omega \mathcal{P}(x) \|_1^2 \right), \]

implying a stable recovery. If \( e \) is a zero-mean white Gaussian noise with variance \( \sigma^2 \) and (8) holds for \( \mathcal{P}(\cdot | A, k) \) then

\[ E \| x - \hat{x} \|_2^2 \leq \frac{C_1}{A^2} k \log n \sigma^2 + \frac{C_4}{A^2} \left( \| \Omega \mathcal{P}(x) \|_2^2 + \frac{1}{k^2} \| \Omega \mathcal{P}(x) \|_1^2 \right), \]

implying a denoising effect. The constants \( C_1, C_2, C_3, C_4 \) are the same as in (7) and (8).

\[ \text{Proof:} \]

We prove only the bound in (9). The proof for (10) is very similar and omitted. Assume that (7) holds. Then since \( y = A \Omega x + e \), we have that

\[ \| \Omega x - \hat{w} \|_2 = \| \Omega x - \mathcal{P}(y) \|_2 \leq C_1 \| e \|_2^2 + C_2 \left( \| \Omega x - [\Omega x]_k \|_2^2 + \frac{1}{\sqrt{k}} \| \Omega x - [\Omega x]_k \|_1^2 \right). \]

We get (9) by using the facts that (i) \( [\Omega x]_k = \Omega \mathcal{P} x \) and therefore \( \Omega x - [\Omega x]_k = \Omega \mathcal{P} x \); (ii) \( \Omega \) is a frame with a lower frame bound \( A \) and therefore \( \| \Omega \|_2^2 \leq \frac{1}{A} \); and (iii) \( x = \Omega \Omega x \) and thus \( \| x - \hat{x} \|_2 = \| \Omega \Omega (\Omega x - \hat{w}) \|_2 \).

\[ \square \]

2.2. Guarantees for 2D-DIF Operator

Having a guarantee for frames we turn to provide a guarantee for 2D-DIF, the two dimensional finite difference operator. For convenience we assume that \( x \) is an image (column stacked) of size \( N \times N = d \) \( (N = \sqrt{d}) \). Notice that unlike frames, for the 2D-DIF operator a small distance in the transform domain does not imply a small distance in the signal domain. For example, the distance between two constant images is zero in the transform domain of the 2D-DIF operator. However, it can be arbitrarily as large as we want depending on the constant value we assign to each image. Therefore, it is impossible to recover a signal by just using the scheme we have in Algorithm 1. Note that the problem lies in the last stage of the algorithm as we do not have enough information to get back stably from the transform domain to the signal domain. Note that also if we will add rows to \( \Omega_{2D-DIF} \) and then apply a pseudo
in the transform domain, we will not have a stable recovery in the signal domain given the recovery in the transform domain (See [23] for more details).

Therefore we utilize the tools used in [23] that studies the performance of the 2D-DIF operator with the analysis $\ell_1$-minimization, which is known also as the anisotropic total variation (TV). Two key steps are used in that work for developing the result for TV:

- The construction of the measurements:

$$y = \begin{pmatrix} M_1 x_{nfr} \\ M_1 x_{nlr} \\ M_2 x_{nfc} \\ M_2 x_{nlc} \\ M_3 x \end{pmatrix} + e,$$

(12)

where $x_{nfr}$, $x_{nlr}$, $x_{nfc}$ and $x_{nlc}$ are versions of $x$ with no first row, last row, first column or last column respectively. In addition, $M_1 \in \mathbb{R}^{m_1 \times N(N-1)}$ and $M_2 \in \mathbb{R}^{m_2 \times N(N-1)}$ are assumed to satisfy the RIP with $\delta_{5k} < \frac{1}{4}$ and $M_3 H^{-1}$ is assumed to satisfy the RIP with $\delta_{2k} < 1$, where $H$ is the bivariate Haar transform and $M_3 \in \mathbb{R}^{m_3 \times d}$.

- The usage of the relationship between $\Omega_{2D-DIF}$ and $H$: For any vector $v$, if $\|\Omega_{2D-DIF}v\|_0 \leq k$ then $\|Hv\|_0 \leq k \log(d)$.

The first two measurement matrices $M_1$ and $M_2$ provide information about the derivatives of $x$ and lead to a stable recovery of $\Omega_{2D-DIF} x$, the discrete gradient vector of $x$. As we have mentioned before $\Omega_{2D-DIF}$ is non-invertible. Therefore, the reconstruction of the derivatives is not enough for recovering the signal. For this purpose the third matrix $M_3$ is used to guarantee stable recovery also in the signal domain. This is achieved using the following theorem:

**Theorem 2.3 (Strong Sobolev inequality. Theorem 8 in [23]).** Let $N$ be a power of 2 and $M_1$ be a linear map which, composed with the inverse bivariate Haar transform $M_1 H^{-1} \in \mathbb{R}^{m_3 \times d}$, has the RIP with a constant $\delta_{2k} < 1$. Suppose that for $z \in \mathbb{R}^d$ we have $\|M_1 z\|_2 \leq \epsilon$. Then

$$\|z\|_2 \leq \frac{2C_H}{1 - \delta_{2k}} \frac{1}{\sqrt{k}} \log(d/k) \|\Omega_{2D-DIF} z\|_1 + \frac{1}{1 - \delta_{2k}} \epsilon,$$

(13)

where $C_H = 36(480 \sqrt{5} + 168 \sqrt{3})$.

We utilize the above theorem for extending our sampling technique for the 2D-DIF operator. By observing again the samples generated by $M_1$ and $M_2$, and denoting by $\Omega_{nfr}$ and $\Omega_{nlr}$ the vertical and horizontal difference of $\Omega_{2D-DIF}$ respectively, we can write $M_1(x_{nfr} - M_1 x_{nlr}) = M_1(\partial_z x)$ and $M_1(x_{nfc} - M_1 x_{nlc}) = M_2 \partial_h x$. Alternatively, we can rewrite it as

$$\begin{pmatrix} M_1 \\ 0 \\ M_2 \end{pmatrix} \Omega_{2D-DIF} x,$$

(14)

and we end up with having samples from the derivatives domain.

Notice that we do not have to restrict ourselves to a block diagonal matrix composed of two linear maps for sampling each derivative direction. We can use any sampling operator that has recovery guarantees in the synthesis framework for reconstructing the coefficients in the transform domain. We denote this reconstruction by $\hat{w}$. In order to recover the signal from its proxy $\hat{w}$, we take more measurements of the original signal $x$. These are taken using a matrix $B$ for which $BH^{-1}$ (its composition with the inverse bivariate Haar transform) has the RIP with a constant $\delta_{2k} < 1$. Given these measurements, $y_2 = Bx + e_2$, we get a recovery of the signal by solving

$$\hat{x}_{2D-DIF} = \arg\min_x \|\Omega x - \hat{w}_{2D-DIF}\|_1 \quad s.t. \quad \|B x - y_2\| \leq \|e_2\|_2.$$

(15)

To sum it up, our sampling strategy for the 2D-DIF operator consists of taking two sets of measurements. The first in the transform domain, $y_1 = A \Omega_{2D-DIF} x + e_1$, leads to reconstruction of the gradient components. The second is
Therefore, setting the bounds in \((7)\) and \((8)\) we assume that the following holds:

\[
\|x\|_2 \leq C_5 \sqrt{k} \|e\|_2 + C_6 \|x\|_1.
\]

Such a bound holds for the synthesis recovery program for \(k\)-sparse representation under the matrix \(A\).

**Proof:** Since \(\alpha\) is a minimizer of \((2)\) we have that

\[
\|B\hat{x} - y_2\|_2 \leq \|e_2\|_2. 
\]

Since \(y_2 = Bx + e_2\) we have from the triangle inequality that

\[
\|B(\hat{x} - x)\|_2 \leq 2 \|e_2\|_2. 
\]

Therefore, setting \(z = \hat{x} - x\) in Theorem 2.3 we have

\[
\|z\|_2 \leq \frac{2C_5}{1 - \delta_{2k}} \frac{1}{\sqrt{k}} \log(d/k) \|\Omega_{2D-DIF}(\hat{x} - x)\|_1 + \frac{2}{1 - \delta_{2k}} \|e_2\|_2.
\]

From the triangle inequality we have

\[
\|\Omega_{2D-DIF}(\hat{x} - x)\|_1 \leq \|\Omega_{2D-DIF}(\hat{x} - \hat{w})\|_1 + \|\Omega_{2D-DIF}(\hat{w} - x)\|_1.
\]

**Theorem 2.4 (Stable signal recovery from samples of 2D-DIF in the transform domain).** Consider the problem \(P\) such that \(M = \begin{bmatrix} A\Omega_{2D-DIF} & B \end{bmatrix}\), where \(\Omega_{2D-DIF}\) is the 2D-DIF operator and \(BH^{-1}\) has the RIP with a constant \(\delta_{2k} < 1\). Let \(\hat{x}\) be the output of Algorithm 1 with the synthesis program \(P(\cdot|A, k)\). If \(e\) is a bounded additive adversarial noise and \((16)\) holds for \(P(\cdot|A, k)\) then

\[
\|\hat{x} - x\|_2 \leq \log(d/k) \left(C_7 \|e\|_2 + \frac{C_8}{\sqrt{k}} \|\Omega_{2D-DIF}x\|_1 \right),
\]

implying a stable recovery, where \(C_7\) and \(C_8\) are functions of \(C_5\) and \(\delta_{2k}\).

**Proof:** Since \(\hat{x}\) is a minimizer of \((2)\) we have that

\[
\|B\hat{x} - y_2\|_2 \leq \|e_2\|_2.
\]

Algorithm 2 Signal Recovery from Samples of 2D-DIF in the Transform Domain

**Require:** \(k, A \in \mathbb{R}^{m \times p}, B \in \mathbb{R}^{n \times d}, \Omega_{2D-DIF} \in \mathbb{R}^{2d \times d}, P, \) where \(y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\) such that \(y_1 = A\Omega_{2D-DIF}x + e_1\) and \(y_2 = Bx + e_2\). \(\Omega_{2D-DIF}\) is \(k\) sparse or approximately so, \(e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}\) is an additive noise, and \(P(\cdot) = P(\cdot|A, k)\) is a signal recovery program for \(k\)-sparse representation under the matrix \(A\).

**Ensure:** \(\hat{x}\): Approximation of \(x\).

Get a transform domain proxy for \(\Omega x\): \(\hat{w} = P(y_1, A, k)\)

Signal recovery: Calculate \(\hat{x}\) using (2) with \(y_2\) and \(\hat{w}\).
Since \( \mathbf{x} \) is a feasible solution to (2) and \( \hat{\mathbf{x}} \) is its minimizer we have

\[
\| \mathbf{\Omega}_{2D-DIF}(\hat{\mathbf{x}} - \hat{\mathbf{w}}) \|_1 \leq \| \mathbf{\Omega}_{2D-DIF}(\mathbf{x} - \hat{\mathbf{w}}) \|_1.
\]  

(22)

Plugging (22) in (21) we have

\[
\| \mathbf{\Omega}_{2D-DIF}(\hat{\mathbf{x}} - \mathbf{x}) \|_1 \leq 2 \| \mathbf{\Omega}_{2D-DIF}(\mathbf{x} - \hat{\mathbf{w}}) \|_1.
\]  

(23)

Notice that we can bound the right hand side (rhs) of (23) with (16), where \( \alpha = \mathbf{\Omega} \mathbf{x} \) and \( \hat{\alpha} = \hat{\mathbf{w}} \). Therefore, by combining (23) and (16) with (20) we have

\[
\| \mathbf{x} - \mathbf{\hat{x}} \|_2 \leq \frac{2C_\|}{1 - \sigma_{2k}} \log(d/k) \left( C_5 \| \mathbf{\epsilon} \|_2 + \frac{C_6}{\sqrt{k}} \| \mathbf{\Omega}_{2D-DIF}\mathbf{x} \|_1 \right) + \frac{2}{1 - \sigma_{2k}} \| \mathbf{\epsilon}_2 \|_2.
\]  

(24)

Extending this idea further we do not restrict the sampling strategy in Algorithm 2 only to \( \mathbf{\Omega}_{2D-DIF} \). It can be applied for any operator for which a stable recovery in the coefficients domain implies a stable recovery in the signal domain by some additional measurements of the signal.

3. Epilogue - Do We Still Need Analysis Algorithms?

Following the fact that our proposed recovery guarantees are similar to the ones achieved for the existing analysis algorithms and that sampling in the manifold dimension of analysis signals lead to unstable recovery [19], one may ask whether there is a need at all for reconstruction strategies that rely on the analysis model. For this reason we perform several experiments to compare the empirical recovery performance of our new sampling scheme, with synthesis \( \ell_1 \)-minimization, and the standard sampling scheme, with analysis \( \ell_1 \)-minimization, for signals from the analysis framework.

We start with the case of signals that are sparse after applying randomly generated tight-frames. We set \( \mathbf{\Omega} \in \mathbb{R}^{144 \times 120} \), where the signal dimension is \( d = 200 \), and \( k = 144 - 110 \) (setting the signal intrinsic dimension to be 10, see [29] for more details). In the standard sampling setup, the entries of the sensing matrix \( \mathbf{M} \in \mathbb{R}^{d \times d} \), where \( \gamma \in [0.05, 0.1, 0.15, \ldots, 1] \), are randomly generated from an i.i.d random Gaussian distribution, followed by a normalization of each column to have a unit \( \ell_2 \)-norm. For the new scheme we set \( \mathbf{M} = \mathbf{A} \mathbf{\Omega} \) with \( \mathbf{A} \in \mathbb{R}^{d \times 1.2d} \) a random Gaussian matrix selected in the same way that \( \mathbf{M} \) is selected in the standard sampling scheme. For each value of \( \gamma \) we generate 1000 different sensing matrices and signals \( \mathbf{x} \) that have sparsity \( k \) under \( \mathbf{\Omega} \). The signals are generated by projecting a randomly selected Gaussian vector to the subspace orthogonal to randomly selected \( n - k \) rows from \( \mathbf{\Omega} \), followed by normalization of the vector.

In Fig. 1 we preset the recovery rate of the two algorithms in the noiseless and noisy cases. The noise is set to be i.i.d white Gaussian with \( \sigma = 0.01 \). It can be seen that it is possible to recover signals from the analysis model using Algorithm 1. However, this comes at the cost of using more samples in order to achieve the same recovery rate and error. This shows us that though the theoretical guarantees of the analysis algorithms take into account only \( k \) and not the intrinsic dimension of the signals, losing the information about the latter, which happens when we sample in the transform domain, may harm the recovery. On the other hand, if we can afford having more measurements, then we have the privilege of using existing synthesis algorithms. This is an important advantage as synthesis algorithms are generally more efficient than the analysis ones. Using cvx [41, 42], the running time of synthesis \( \ell_1 \)-minimization is shorter than the one of analysis \( \ell_1 \) minimization. This advantage in efficiency is not unique to the \( \ell_1 \)-relaxation alone. For more examples, we refer the reader to compare OMP with GAP [18] or the synthesis greedy-like algorithms with their analysis counterparts [29].

We repeat the experiment with the 2D-DIF operator and compare analysis \( \ell_1 \)-minimization with the scheme in Algorithm 2 that uses synthesis \( \ell_1 \)-minimization for recovery. The signals we generate are random \( 14 \times 14 \) images with four connected components. We start with a constant image and then add to it three additional connected components using a random walk on the image using the same technique in [19]. The sensing matrices are selected as in the previous experiment, where in the new sampling scheme we assign 2 measurements (from the total number of...
measurements we use) in the noiseless case for the signal recovery from the transform domain proxy and \( m/10 \) in the noisy case.

Figure 2 presents the reconstruction rate in the noiseless case and the recovery error in the noisy case, where the noise is the same as in the previous experiments. We see the same phenomenon that we saw in the previous experiment but stronger. As the redundancy the in analysis operator is bigger in this experiment, the number of measurements we need for the new scheme is relatively larger and the recovery error in the noisy case is higher. Another reason, other than the bigger redundancy, for the inferior performance in this case is that we separate the measurements we have into two parts, where in the standard scheme the analysis \( \ell_1 \)-minimization uses all the measurements at once for the recovery of the signal. Note that this causes that even in the case that \( m = d \) we do not get 100% recovery. Clearly in this case we will just invert the measurement matrix instead of using neither of the two schemes.
4. Discussion and Conclusion

In this work we have presented a new sampling and recovery strategy for signals that are sparse under frames or the 2D-DIF operator in the analysis model. Our scheme utilize existing algorithms from the synthesis sparsity model to recover signals that belong to the analysis framework. The advantage of this technique is that it enables the usage of existing tools for recovering signals from another model. Though in theory there is no additional cost for the usage of this scheme, it seems that in practice its advantage comes at the cost of the usage of more measurements in the sampling stage. This gap between the theory and practical performance gives us a hint that the existing guarantees are not tight and that there is a need for further investigation of the field.

Acknowledgment

The authors would like to thank Michael Elad and Yaniv Plan for fruitful discussions. Raja Giryes is partially supported by AFOSR.

References

[1] A. M. Bruckstein, D. L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review 51 (1) (2009) 34–81.

[2] E. Candès, Modern statistical estimation via oracle inequalities, Acta Numerica 15 (2006) 257–325.

[3] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, Journal of Constructive Approximation 50 (1997) 57–98.

[4] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, 1st Edition, Springer Publishing Company, Incorporated, 2010.

[5] H. R. Simon Foucart, A Mathematical Introduction to Compressive Sensing, 1st Edition, Springer Publishing Company, Incorporated, 2013.

[6] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3397–3415.

[7] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203 – 4215.

[8] D. L. Donoho, M. Elad, V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inf. Theory 52 (1) (2006) 6–18.

[9] E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. Inf. Theory 52 (12) (2006) 5406 –5425.

[10] D. Needell, J. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. A. 26 (3) (2009) 301 – 321.

[11] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inf. Theory 55 (5) (2009) 2230 –2249.

[12] T. Blumensath, M. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal 27 (3) (2009) 265 – 274.

[13] S. Foucart, Hard thresholding pursuit: an algorithm for compressive sensing, SIAM J. Numer. Anal. 49 (6) (2011) 2543–2563.

[14] T. Zhang, Sparse recovery with orthogonal matching pursuit under RIP, IEEE Trans. Inf. Theory 57 (9) (2011) 6215 –6221.

[15] M. Rudelson, R. Vershynin, Sparse reconstruction by convex relaxation: Fourier and gaussian measurements, in: Information Sciences and Systems, 2006-40th Annual Conference on, 2006, pp. 207 –212.

[16] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, Constructive Approximation 28 (3) (2008) 253–263.

[17] M. Elad, P. Milanfar, R. Rubinstein, Analysis versus synthesis in signal priors, Inverse Problems 23 (3) (2007) 947–968.

[18] S. Nam, M. Davies, M. Elad, R. Gribonval, The cosparse analysis model and algorithms, Appl. Comput. Harmon. Anal. 34 (1) (2013) 30 – 56.

[19] S. Giryes, Y. Plan, R. Vershynin, On the effective measure of dimension in analysis cosparse models (2014).

[20] E. J. Candès, Y. C. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmon. Anal 31 (1) (2011) 59 – 73.

[21] Y. Liu, T. Mi, S. Li, Compressed sensing with general frames via optimal-dual-based l1-analysis, IEEE Trans. Inf. Theory 58 (7) (2012) 4201–4214. doi:10.1109/TIT.2012.2191612.

[22] R. Giryes, S. Nam, M. Elad, R. Gribonval, M. Davies, Greedy-like algorithms for the cosparse analysis model, Linear Algebra and its Applications 441 (0) (2014) 22 – 60, special Issue on Sparse Approximate Solution of Linear Systems.

[23] D. Needell, R. Ward, Stable image reconstruction using total variation minimization, SIAM Journal on Imaging Sciences 6 (2) (2013) 1035–1058.

[24] M. Kabanava, H. Rauhut, Analysis l1-recovery with frames and Gaussian measurements, Preprint.

[25] R. Giryes, Greedy algorithm for the analysis transform domain (2014).

[26] H. Rauhut, K. Schnass, P. Vandergheynst, Compressed sensing and redundant dictionaries, IEEE Trans. Inf. Theory 54 (5) (2008) 2210 –2219.

[27] D. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization, Proc. Nat. Aca. Sci. 100 (5) (2003) 2197–2202.

[28] M. Davenport, D. Needell, M. Wakin, Signal space CoSaMP for sparse recovery with redundant dictionaries, IEEE Trans. Inf. Theory. 59 (10) (2013) 6820–6829.

[29] R. Giryes, D. Needell, Greedy signal space methods for incoherence and beyond, Appl. Comput. Harmon. Anal. To appear.
[30] R. Giryes, M. Elad, Can we allow linear dependencies in the dictionary in the synthesis framework?, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.
[31] R. Giryes, M. Elad, Iterative hard thresholding for signal recovery using near optimal projections?, in: 10th Int. Conf. on Sampling Theory Appl. (SAMPTA), 2013.
[32] R. Giryes, M. Elad, OMP with highly coherent dictionaries, in: 10th Int. Conf. on Sampling Theory Appl. (SAMPTA), 2013.
[33] C. Hegde, P. Indyk, L. Schmidt, Approximation-tolerant model-based compressive sensing, in: ACM Symposium on Discrete Algorithms (SODA), 2014.
[34] B. Ophir, M. Lustig, M. Elad, Multi-scale dictionary learning using wavelets, IEEE Journal of Selected Topics in Signal Processing 5 (5) (2011) 1014–1024.
[35] S. Ravishankar, Y. Bresler, MR image reconstruction from highly undersampled k-space data by dictionary learning, IEEE Trans. Medical Imaging 30 (5) (2011) 1028–1041.
[36] S. Ravishankar, Y. Bresler, Sparseifying transform learning for compressed sensing MRI, in: 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI), 2013, pp. 17–20.
[37] E. Candès, T. Tao, The Dantzig selector: Statistical estimation when p is much larger than n, Annals Of Statistics 35 (2007) 2313.
[38] P. Bickel, Y. Ritov, A. Tsybakov, Simultaneous analysis of lasso and dantzig selector, Annals of Statistics 37 (4) (2009) 1705–1732.
[39] Z. Ben-Haim, Y. Eldar, M. Elad, Coherence-based performance guarantees for estimating a sparse vector under random noise, IEEE Trans. Signal Process. 58 (10) (2010) 5030–5043.
[40] R. Giryes, M. Elad, RIP-based near-oracle performance guarantees for SP, CoSaMP, and IHT, IEEE Trans. Signal Process. 60 (3) (2012) 1465–1468.
[41] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1 (Mar. 2014).
[42] M. Grant, S. Boyd, Graph implementations for nonsmooth convex programs, in: V. Blondel, S. Boyd, H. Kimura (Eds.), Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, pp. 95–110.