Nonlinear Sampling for Sparse Recovery

Seyed Amir-Hossein Hosseini, Mahdi Barzegar Khalilsarai, Arash Amini and Farokh Marvasti
Advanced Communications Research Institute (ACRI)
Electrical Engineering Department, Sharif University of Technology, Tehran, Iran
Email: a_h_hosseini@ee.sharif.ir, m_barzegar@ee.sharif.ir, aamini@sharif.ir, marvasti@sharif.ir

Abstract—Linear sampling of sparse vectors via sensing matrices has been a much investigated problem in the past decade. The nonlinear sampling methods, such as quadratic forms are also studied marginally to include undesired effects in data acquisition devices (e.g., Taylor series expansion up to two terms). In this paper, we introduce customized nonlinear sampling techniques that provide possibility of sparse signal recovery. The main advantage of the nonlinear method over the conventional linear schemes is the reduction in the number of required samples to \(2k\) for recovery of \(k\)-sparse signals. We also introduce a low-complexity reconstruction method similar to the annihilating filter in the sampling of signals with finite rate of innovation (FRI). The disadvantage of this nonlinear sampler is its sensitivity to additive noise; thus, it is suitable in scenarios dealing with noiseless data such as network packets, where the data is either noiseless or it is erased. We show that by increasing the number of samples and applying denoising techniques, one can improve the performance. We further introduce a modified version of the proposed method which has strong links with spectral estimation methods and exhibits a more stable performance under noise and numerical errors. Simulation results confirm that this method is much faster than \(\ell_1\)-norm minimization routines, widely used in linear compressed sensing and thus much less complex.

I. INTRODUCTION

A large body of problems in engineering fields such as signal processing can be stated as inverse problems, where some information about an unknown but desired object is available, but in general, the data is not sufficient to uniquely determine this object. In many cases, among possible candidates that are consistent with the data, there is only one that looks reasonable; i.e., it matches our prior knowledge about the object. A famous example is the sampling problem where the content of a continuous-domain signal is condensed into a set of discrete samples. An early breakthrough in sampling theory was the famous Nyquist-Shannon sampling theorem: a signal can be perfectly recovered from its samples, if its highest frequency component is less than half the sampling rate. The more recent example of compressed sensing aims for finding a high-dimensional vector from its projections onto lower-dimensional surfaces. The prior information here is that the original vector has a sparse representation in some transform domain, which is a generalization of bandlimitedness in the Fourier domain (traditional sampling theory). The incoherency of the projection surfaces with the latter transform domain is a fundamental constraint to ensure recovery. [1]. The interested readers are referred to [2], [3] for a further detailed review of the literature.

In conventional compressed sensing, the measurements are obtained by linear combination of the unknowns. However, in some applications, the physics of the problem impose certain types of nonlinearity. As an example, in X-Ray Crystallography, the measurements contain only the information related to the intensity of the diffraction pattern and not its phase, violating the linearity condition [4], [5], [6], [7]. In order to transpose a nonlinear system into the standard linear framework of compressed sensing, the Taylor series expansion is proposed in [8], [9]. This technique achieves a linear matrix model subject to non-convex constraints after applying a lifting technique. It is observed that the input can be recovered by a convex relaxation of this problem, given that the number of samples is sufficiently large. The results presented in [8], [9] convey the message that the nonlinearity of the sampler is an undesired phenomenon that degrades the performance, and should be avoided as much as possible. Our goal in this paper is to show that this is not the general case, and if one is allowed to design and customize the nonlinear sampler, it might even be advantageous in terms of the number of samples. Although, the literature on nonlinear compressed sensing is relatively limited, the interested readers may find valuable information on nonlinear sampling in [10], [11].

In this paper, we propose a specific form of nonlinear sampling that uniquely identifies \(k\)-sparse vectors with as few as \(2k\) noiseless samples. Furthermore, we introduce a reconstruction method without applying the lifting technique that perfectly recovers all sparse vectors (instead of recovery with high probability) below a sparsity level with considerably less computational complexity compared to \(\ell_1\)-minimization approaches. The proposed nonlinear samples are such that the degree of nonlinearity increases as the number of samples increases. As a consequence, our attempts towards implementing the technique of Nonlinear Basis Pursuit (NLBP) developed for general nonlinear samplers in [8], [9] were all unsuccessful. In fact, the lifting technique in NLBP results in dramatic increase in the size of the model when the degree of nonlinearity increases. Hence, it asks for more memory as well as computational power. The alternative recovery method used in this paper directly works with nonlinear forms and avoids linearization. However, it is tailored for the proposed sampler and cannot be adapted to other types of nonlinearities.

The proposed nonlinear method, which has significant connections with the common methods in spectral estimation and coding [12], is sensitive to noise. This issue was already observed in other areas including the nonlinear reconstruction...
technique in FRI models [12]. Because of strong similarities between our reconstruction technique and the annihilating filter in FRI sampling, we exploit the iterative denoising algorithm of [13] developed in this context to mitigate the effect of noise. Although increasing the number of samples followed by denoising greatly improves the performance, it is preferred to employ the nonlinear sampling in applications with controlled noise effects. For instance, it can be used as a network coding strategy to sample and encode received packets of all input links simultaneously, where in each time interval most of the input links are inactive. As the packets are protected by strong channel codes, they are usually modeled as either noiseless or erased.

The rest of the paper is organized as follows: our nonlinear sampling and recovery approaches are discussed in section II. We investigate the recovery problem in presence of noise in section III. The simulation results are presented in section IV and finally, section V concludes our work.

II. NONLINEAR SAMPLING FOR SPARSE RECOVERY

Let $x_{n\times 1}$ be a $k$-sparse vector; i.e., except for $k$ of its elements, the rest are zero. Our goal is to represent $x_{n\times 1}$ using few [nonlinear] samples, such that $x_{n\times 1}$ could be tractably recovered. As there are $2k$ unknowns in $x_{n\times 1}$, namely, the location and value of non-zero elements, we intuitively expect to have a unique representation of $x$ with $2k$ samples. Indeed, here we introduce a scheme which employs as few as $2k$ measurements of $x$ to uniquely identify it.

**Sampler 1.** Let $a_1, \ldots, a_n$ be real/complex numbers and let $x_{n\times 1}$ be a given [sparse] vector. We define the power sampling set of order $m$ by $\{y_1, \ldots, y_m\}$ as follows:

$$
\begin{align*}
    y_1 &= a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \\
    y_2 &= a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2, \\
    &\vdots \\
    y_m &= a_1 x_1^m + a_2 x_2^m + \cdots + a_n x_n^m,
\end{align*}
$$

where $x_i$ is the $i^{th}$ element in $x_{n\times 1}$.

For a $k$-sparse vector $x_{n\times 1}$ whose support is $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$, the $q^{th}$ power sample can be rewritten as

$$
y_q = \sum_{j=1}^{k} a_{i_j} x_{i_j}^{q},
$$

Similar to the annihilating filter, we define $H(z)$ as the polynomial that has non-zero elements of $x_{n\times 1}$ as its roots:

$$
H(z) = \sum_{r=0}^{k} h_r z^{-r} = \prod_{j=1}^{k} \left(1 - x_{i_j} z^{-1}\right).
$$

By interpreting $H(z)$ as the $Z$-transform of a discrete filter $h_q$ and applying it to the power samples $y_q$, we obtain that

$$
y_q * h_q = \sum_{r=0}^{k} h_r y_{q-r}
$$

$$
= \sum_{r=0}^{k} h_r \left(\sum_{j=1}^{k} a_{i_j} x_{i_j}^{q-r}\right)
$$

$$
= \sum_{j=1}^{k} a_{i_j} x_{i_j}^{q} \left(\sum_{r=0}^{k} h_r x_{i_j}^{-r}\right) = 0
$$

(3)

Annihilating the measurements, the filter $h$ is called the annihilating filter in the literature of [13] and [14]. Since the roots of $H(z)$ uniquely specify the nonzero elements of the unknown signal $x$, our goal shall be to find the annihilating filter coefficients according to (3) to determine the nonzero elements. A useful reformulation of (3) in the matrix form is

$$
Ah = 0,
$$

(4)

where $h = [h_0, \ldots, h_k]^T$ and $A$ is a Toeplitz matrix based on the power samples given by

$$
A = \begin{bmatrix}
    y_{k+1} & y_k & \cdots & y_1 \\
    y_{k+2} & y_{k+1} & \cdots & y_2 \\
    \vdots & \vdots & \ddots & \vdots \\
    y_m & y_{m-1} & \cdots & y_{m-k}
\end{bmatrix}
$$

(5)

To derive $h$, we can use the singular-value decomposition of $A$ to form its null-space (zero singular-values). It is clear that any vector in the null-space of $A$ satisfies (4) as $h$. With no priority which one to choose, we can determine the nonzero elements of $x$ by finding the roots of $H(z) = \sum_{r=0}^{k} h_r z^{-r}$.

Although the roots of $H(z)$ reveal the nonzero values in $x$, their locations are not determined yet. In simple words, unlike conventional sparse recovery techniques, our nonlinear recovery method works by first evaluating the nonzero elements and then, locating them. For this purpose, we consider the following system of linear equations:

$$
\begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_m
\end{bmatrix} =
\begin{bmatrix}
    x_1^{a_1} & x_1^{a_2} & \cdots & x_1^{a_k} \\
    x_2^{a_1} & x_2^{a_2} & \cdots & x_2^{a_k} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_m^{a_1} & x_m^{a_2} & \cdots & x_m^{a_k}
\end{bmatrix}
\begin{bmatrix}
    \hat{a}_1 \\
    \hat{a}_2 \\
    \vdots \\
    \hat{a}_k
\end{bmatrix}
$$

(6)

and solve it to estimate the coefficients $\hat{a}_1, \ldots, \hat{a}_k$ associated with nonzero elements. Ideally, the set $\{\hat{a}_j\}_j$ is just a permutation of $\{a_{i_j}\}_j$. Now, if $a_{i_j}$ are all distinct, i.e. $a_{i_j} \neq a_{i_l}, \forall i_l \neq j$, then, there is a unique matching between $\hat{a}_1, \ldots, \hat{a}_k$ and the subset $\{a_{i_j}\}_j$ of $\{\hat{a}_j\}_j$, which also determines the support of the vector.

For (6) to have a unique solution, $A$ needs to have at least $k$ linearly independent rows. This in turns requires at least $2k$ measurements, i.e. $m \geq 2k$. It is not difficult to see that is the non-zero elements of the spares vectors are pairwise distinct, then, $m = 2k$ also works:
Theorem 1. Let \( x_{n \times 1} \) be a \( k \)-sparse vector with no repeated nonzero element. In the absence of noise, \( x_{n \times 1} \) can be perfectly recovered using merely \( 2k \) power sampling measurements as described in sampler 1.

In the case of noise-contaminated measurements, we generally need to increase the number of samples. In Section III we describe how we can alleviate the noise effect by having more samples.

III. Noise Effect

Nonlinear methods are inherently more sensitive to noise than linear alternatives. To achieve robustness against noise, we have to increase the number of measurements to \( m > 2k \). To efficiently take advantage of larger sample set, we adopt the denoising algorithm presented in [13] to decrease the noise level before following with the previous nonlinear scheme. Redefining the matrix \( A \) we have that

\[
A = \begin{bmatrix} y_{L+1} & y_L & \cdots & y_1 \\ y_{L+2} & y_{L+1} & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_m & y_{m-1} & \cdots & y_{m-L} \end{bmatrix},
\]

where \( L \) is an integer satisfying \( m - k \geq L \geq k \). The idea is quite simple: we first enhance the measurements using a composite property mapping algorithm [15]. Then, using the denoised measurements, the annihilating filter coefficients are calculated according to (4). In the absence of noise, \( A \) is a Toeplitz matrix whose rank never exceeds \( k \). In the noisy setup, \( A \) remains Toeplitz, however its rank likely exceeds \( k \). This information can be used to denoise the matrix \( A \) obtained by measurements and consequently the measurements themselves. In each iteration of the mentioned algorithm, \( A \) is mapped to the nearest matrix lying in the appropriate matrix spaces [14]. Simulation results indicate a more accurate recovery, if \( L \) is set to \( \lfloor \frac{m}{2} \rfloor \) in (7) for denoising. Although, in the case of \( L > k \) the rank of \( A \) remains unchanged (\( \text{rank}(A) = k \)), the dimensionality of its null-space increases from 1 to \( L - k + 1 \) (\( \dim \mathcal{N}(A) = L - k + 1 \)). This implies that there are \( L - k + 1 \) different annihilating filters for matrix \( A \) whose degrees are \( L > k \). As a result, the set of roots in each polynomial contains the desired nonzero elements as well as some other irrelevant ones; coming up with an ambiguity in distinguishing the nonzero elements from improper values. To sum it up, \( L = \lfloor \frac{m}{2} \rfloor \) is better in defining matrix of measurements \( A \) for denoising part, while in the calculation of filter coefficients \( h \), we have to rearrange \( A \) setting \( L = k \). Algorithm 1 provides an overall view of the procedure for denoising and finding both the value and position of the nonzero entries in \( x_{n \times 1} \).

The other practical issue becoming more important in presence of noise is the amplitudes of unknowns. Stemming from the nature of our sampling approach, the last measurements have a much greater amplitude than the first ones, if the amplitudes of all or some of the elements are greater than 1. At a fixed SNR, this makes the first measurements very vulnerable against noise. On the other hand, if the amplitudes of some of the elements are less than 1, their contribution to the last measurements gradually fades. To avoid these problems, we highly recommend that the nonzero elements be mapped to a set of corresponding values on the unity circle using a one-to-one function just before their exponentiation. The mentioned function should be such that \( f(0) = 0 \), to keep the sparsity of the original signal. Supplying the readers with an example of such functions, we introduce \( \sin \) for signals with real values.

Sampler 2. Let \( a_1, \ldots, a_n \) be real/complex numbers and let \( x_{n \times 1} \) be a given real/sparse vector of which entries lie in the interval \([-N, N]\). We define the sin sampling set of order \( m \) by \( \{y_1, \ldots, y_m\} \) as follows:

\[
\begin{align*}
y_1 &= a_1 \sin \left( \frac{\pi x_1}{2N} \right) + \cdots + a_n \sin \left( \frac{\pi x_n}{2N} \right), \\
y_2 &= a_1 \sin \left( \frac{\pi x_1}{2N} \right) + \cdots + a_n \sin \left( \frac{\pi x_n}{2N} \right), \\
&\vdots \\
y_m &= a_1 \sin \left( \frac{\pi x_1}{2N} \right) + \cdots + a_n \sin \left( \frac{\pi x_n}{2N} \right),
\end{align*}
\]

where \( x_i \) is the \( i \)th element in \( x_{n \times 1} \).

Since \( \sin(x) = \frac{e^{ix} - e^{-ix}}{2i} \), each \( \sin(\cdot) \) function in the previous measurements may be decomposed into two exponential functions on the unity circle. This doubles the number of nonzero elements, requiring at least \( 4k \) measurements, if

Algorithm 1 Proposed Algorithm

1. **Input**: Measurements \( y = \{y_1, \ldots, y_m\} \) from sampler 1 and coefficients \( a = \{a_1, \ldots, a_n\} \)
2. procedure **DENISOING(y)**
3. Set \( L = \left\lfloor \frac{m}{2} \right\rfloor \) and form \( A \) according to (7)
4. while \( \left( \frac{\sum_{i=k+1}^n a_i}{\sum_{i=k}^n a_i} \right) \geq \alpha \) do
5. \[ [U, \Sigma, \Sigma(m-L)\times(L+1), V^*] = SVD(A) \]
6. Build \( \Sigma^* \) keeping \( k \) largest diagonal elements of \( \Sigma \)
7. \( A \leftarrow U \times \Sigma^* \times V^* \)
8. Average over diagonals of \( A \) to form a Toeplitz matrix
9. end while
10. procedure **NONZERO VALUES ESTIMATION(y)**
11. Set \( L = k \) and form \( A \) according to (7)
12. \( h_{(k+1) \times 1} \leftarrow \) The right eigen-vector corresponding to the zero eigen-value of \( A \)
13. Form \( H(z) \) from \( h \)
14. Nonzero \( \leftarrow \) roots of \( H(z) \)
15. end procedure
16. procedure **SUPPORT DETECTION(a)**
17. Estimate coefficients using (6)
18. Compare estimated coefficients to \( a_1, \ldots a_n \) and specify the support
19. end procedure
a method like the one already explained in section II is employed for recovery. Of course, it can be easily seen that every $k$ nonzero elements in this scheme uniquely determine the other $k$ values. This information can be exploited to reduce the number of required measurements for a perfect recovery; However we will not investigate this problem in the current paper and defer it to a future work. The other benefit of the sinusoidal scheme is its easy-to-implement property. Employing a simple FM modulator and demodulator, one can actualize the theoretic material provided in sections II and III with respect to the following equation:

$$y(t) = \sum_{i=1}^{n} a_i \sin(2\pi f_i t), \quad f_i = x_i$$

(8)

Obviously, sampling from $y(t)$ at times $t = \frac{1}{4N}, \ldots, \frac{m}{4N}$ yields the measurements of sampling method 2. As an application of the proposed scheme, we can think of a communication system in which there are $n$ nodes sending and receiving data at arbitrary times. If there is a guarantee that not more than $k$ nodes are busy at each time, then receiver does not have to wait to take $n$ samples to decode the raw data, increasing the rate of data transmission by a factor of $(\frac{n}{2})$. In Figure 1, the block diagram of a FM modulator is depicted.

**IV. SIMULATION RESULTS**

To validate the practical ability of our proposed method in recovering a sparse signal from its noisy measurements, we present the results of several numerical experiments in this section. In each experiment, we employ the sampling and reconstruction method, explained in section III and our goal is to investigate its performance in terms of: 1) Probability of support recovery and 2) SNR of reconstructed signal in various additive noise levels. To that end, we generate a signal of length $n = 50$ with a random sparsity pattern, where the amplitude of any nonzero entry is set to 1 as justified earlier. Then we use our nonlinear sampling scheme (Sampler 1) to form the measurements, which will next be contaminated with white gaussian noise with a specific variance. The recovery phase consists of two steps: a denoising part to ensure a robust recovery, followed by the previously mentioned algorithm for value and support recovery of nonzero coefficients.

Note that employing the lifting operator proposed by [8], [9] is ineffective in the sense that it both reduces output SNR and is also much more complex in implementation, and its simulation results are not included here for brevity.

Figure 2 demonstrates the frequency of successful support recovery against SNR of measurements for sparsity orders $k = 1$ to $k = 6$. Here, as a minimum number of measurements, we have set $m$ to be two times the sparsity order. A more reliable performance criterion is the SNR of reconstructed signal, which is defined as the power of the reconstructed signal divided by the power of its error compared to the main measured signal presented in dBs.

Figure 3 indicates the value of this criterion against different measurements’ SNRs for sparsity orders $k = 1$ to $k = 6$.

Increasing the number of measurements improves the quality of reconstructed signal in a considerable manner. To show this behavior we introduce the OverSampling Rate ($OSR$) as the ratio of available number of observations to the minimum of required measurements $(2k)$. So when $OSR = 2$ it implies that $m = 2 \times (2k) = 4k$. In the following simulations, the sparsity order ($k$) is set to be 6 and then $OSR$ is swept from 1 to 4. As shown in Figure 4, the probability of support recovery is dramatically increased when $OSR$ rises from 1 to 4.

In Figure 5 output SNR is plotted against SNR of measure-
Fig. 4. Probability of support recovery vs SNR of measurements for different values of OSR.

Fig. 5. SNR of reconstructed signal vs SNR of measurements for different values of OSR.

ments for different OSRs. It is clearly seen that when OSR changes from 1 to 4 the output SNR is averagely improved by 50 dBs which is a quite desirable property.

In the end, we would like to mention that the coefficients in the simulations provided here are chosen to be equally located on the unity circle; however there is no serious limitation for that and it merely promotes the recovery quality slightly.

V. Conclusion

In this paper, we introduced a nonlinear sampling and reconstruction method recovering sparse vectors. The fundamental theme of our approach was to recover the unknown signal from a set of observations, each being a linear combination of the powers of vector elements. In the absence of noise, we showed that a sparse signal of order $k$ can be exactly recovered using merely $2k$ of such measurements. This number of measurements is much less than what is required in conventional linear compressed sensing. In the presence of noise, however, our technique proves to be more vulnerable against noise than linear methods. To compensate for noise, we opt for increasing the number of samples, which provides us with the possibility of denoising the measurements. Simulation results confirm improvement in the stability of this technique at higher sampling rates.

REFERENCES