COMPRESSIVE SAMPLING OF NON-NEGATIVE SIGNALS

Paul D. O'Grady *Scott T. Rickard*

Complex & Adaptive Systems Laboratory,
University College Dublin,
Belfield, Dublin 4, Ireland.

ABSTRACT

Traditional Nyquist-Shannon sampling dictates that a continuous time signal be sampled at twice its bandwidth to achieve perfect recovery. However, It has been recently demonstrated that by exploiting the structure of the signal, it is possible to sample a signal below the Nyquist rate and achieve perfect reconstruction using a random projection, sparse representation and an ℓ_1-norm minimisation. These methods constitute a new and emerging theory known as Compressive Sampling (or Compressed sensing).

Here, we apply Compressive Sampling to non-negative signals, and propose an algorithm—Non-negative Underdetermined Iteratively Reweighted Least Squares (NUIRLS)—for signal recovery. NUIRLS is derived within the framework of Non-negative Matrix Factorisation (NMF) and utilises Iteratively Reweighted Least Squares as its objective, recovering non-negative minimum ℓ_p-norm solutions, $0 \leq p \leq 1$. We demonstrate that—for sufficiently sparse non-negative signals—the signals recovered by NUIRLS and NMF are essentially the same, which suggests that a non-negativity constraint is enough to recover sufficiently sparse signals.

1. INTRODUCTION

The Nyquist-Shannon sampling theorem states that in order for a continuous-time signal to be represented without error from its samples, the signal must be sampled at a rate that is at least twice its bandwidth. In practice, signals are often compressed soon after sampling, trading off perfect recovery for some acceptable level of error. Clearly, this is a waste of valuable sampling resources. In recent years, a new and exciting theory of *Compressive Sampling* (CS) [1, 2] (also known as compressed sensing among other related terms) has emerged, in which a signal is sampled and compressed simultaneously using sparse representations at a greatly reduced sampling rate. The central idea being that the number of samples needed to recover a signal perfectly depends on the structural content of the signal—as captured by a sparse representation that parsimoniously represents the signal—rather than its bandwidth.

More formally, CS is concerned with the solution, $\mathbf{x} \in \mathbb{R}^N$, of an under-determined systems of equations of the form $\mathbf{\Phi x} = \mathbf{y}$, where the *sampling matrix* $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ has fewer rows than columns, *i.e.*, $M < N$. Consequently, there are fewer equations than unknowns, which implies that there are infinitely many solutions. In order to regularise the problem, *i.e.*, specify a solution, an assumption is made about the properties of the solution, which usually involves specifying a norm constraint. Critical to the theory of CS is the assumption that the solution $\mathbf{x}$ is *sparse*, *i.e.*, $\mathbf{y}$ has a parsimonious representation in a fixed basis $\mathbf{\Phi}$ such that the distribution of coefficients in $\mathbf{x}$ resemble a Laplacian distribution, which leads to efficient compression. The most natural norm constraint for this assumption is the ℓ_0 (pseudo-)norm, as it indicates the number of nonzero coefficients. However, minimisation of the ℓ_0 norm is a non-convex optimisation, which is *NP-complete* and cannot be computed in polynomial time. For these reasons the ℓ_1 norm is usually specified, as it is computationally tractable and also recovers sparse solutions. The first algorithm proposed to solve an under-determined system of equations by specifying that the solution be the minimum ℓ_1-norm solution is *basis pursuit* [3]:

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1, \text{ subject to } \mathbf{\Phi x} = \mathbf{y}, \tag{1}$$

where the recovered signal, $\mathbf{x}$, is such a solution.

Eq. 1 assumes that $\mathbf{y}$ is sparse in the canonical basis of $\mathbb{R}^N$, *i.e.*, the identity basis. However, many signals of practical interest are not sparse in the canonical basis, *e.g.*, speech, and an appropriately sparse transformation basis $\mathbf{A} \in \mathbb{R}^{N \times N}$, *e.g.*, wavelet basis, may be required, where $\mathbf{\Phi}$ in Eq. 1 is replaced by $\mathbf{\Phi A}$. Furthermore, the selection of the sampling matrix is dependent on the sparse basis, which brings to attention the other critical insight in the theory of CS—that of *incoherence*: In order to specify the minimal number of measurements, M, required to achieve perfect recovery, $\mathbf{\Phi}$ needs to be maximally incoherent with $\mathbf{A}$ (or the canonical basis) *i.e.*, have a non-parsimonious representation in $\mathbf{A}$—a notion which is contrary to sparseness. In

This material is based upon works supported by the Science Foundation Ireland under Grant No. 05/YI2/I677.

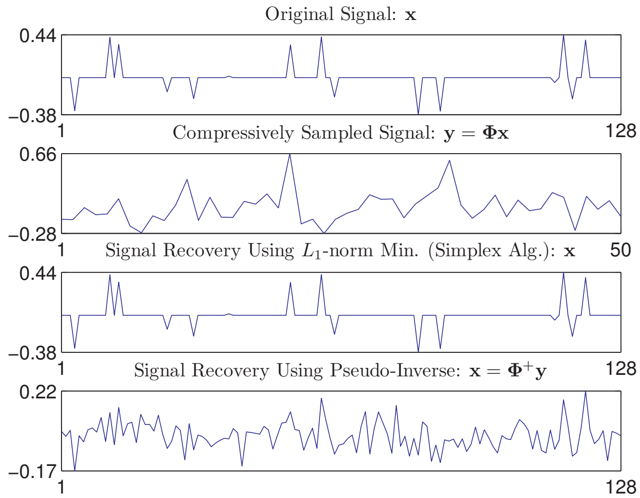

Fig. 1. Compressive sampling of a real-valued signal: The original signal $\mathbf{x}$ is of length $N = 128$ with support $K = 15$ and is sparse in the canonical basis. $\mathbf{x}$ is projected onto the sampling matrix $\mathbf{\Phi}$, which has entries drawn from a Gaussian distribution, resulting in a compressively sampled signal $\mathbf{y}$ of length $M = 50$. Recovery of $\mathbf{x}$ is performed using ℓ_1-norm minimisation and the pseudo-inverse. It is evident that the minimum ℓ_1-norm solution perfectly recovers the original signal, as it fits with the assumption of a sparse $\mathbf{x}$, while the pseudo-inverse, *i.e.*, minimum ℓ_2-norm solution, recovers a denser solution, which differs considerably to the original $\mathbf{x}$.

the case where $\mathbf{y}$ is sparse in the canonical basis, the rows of $\mathbf{\Phi}$ may be constructed by randomly selecting rows of the Fourier basis, since the Fourier basis is maximally incoherent with the canonical basis. Although, if $\mathbf{y}$ is sparse in a wavelet basis, our constructed $\mathbf{\Phi}$ will be coherent with such, and M will increase. Therefore, in order to simplify the selection of $\mathbf{\Phi}$, its row entries are typically drawn from a random Gaussian distribution, as it is universally incoherent with sparse transformations, and performs exact recover with the minimal number of measurements with high probability. The apparent duality between sparseness and incoherence, is conceptually similar to the more familiar duality between frequency and time in Fourier analysis, where an impulse function in the time domain is spread out in the frequency domain—random measurements of which will lead to a better reconstruction of the impulse function as the representation is dense.

So if the entries of $\mathbf{\Phi}$ are drawn from a random Gaussian distribution, what dimension must M be in order to reliably achieve perfect recovery? Candès and Tao [4] present an important result that answers this question: For a K-sparse signal, *i.e.*, a signal with K support ($\|\mathbf{x}\|_0 = K$), the lower bound on M for perfect recovery is

$$M \geq CK \log(N), \tag{2}$$

where C depends on the desired probability of success, which tends to one as $N \rightarrow \infty$. This remarkable result specifies, with high probability, the required number of measurements for a K-sparse signal. Furthermore, the result highlights the importance of the sparseness of the transformations achieved by $\mathbf{A}$, as the lower bound on M decreases as K decreases. An example of the compressive sampling of a real-valued signal is presented in Figure 1.

In this paper, we apply CS to non-negative signals, $\mathbf{x}$, and require both $\mathbf{\Phi}$ and $\mathbf{y}$ to be non-negative also,

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_p, \text{ subject to } \mathbf{\Phi x} = \mathbf{y};\ \mathbf{\Phi}, \mathbf{x}, \mathbf{y} \geq 0, \tag{3}$$

where $0 \leq p \leq 1$, and propose Non-negative Under-determined Iteratively Reweighted Least Squares (NUIRLS) for this purpose. Furthermore, in light of recent developments in uniqueness results for non-negative under-determined systems of equations [5], we investigate when algorithms that recover different minimum ℓ_p-norm solutions, recover similar signals, and demonstrate that the standard Non-negative Matrix Factorisation (NMF) algorithm of Lee and Seung [6] can be used to recover sufficiently sparse signals from compressively sampled non-negative data.

This paper is organised as follows: We overview Iteratively Reweighted Least Squares in Section 2 and NMF in Section 3. We present a nonnegative under-determined IRLS algorithm with multiplicative updates in Section 4, and perform signal recovery from compressively sampled non-negative data using the proposed algorithm in Section 5. We finish with a discussion and conclusion in Section 6 & Section 7 respectively.

2. ITERATIVELY REWEIGHTED LEAST SQUARES

For our purposes, we desire the minimum ℓ_1-norm solution, Eq. 1, and require an objective function that recovers such solutions. However, the ℓ_1-norm objective has a discontinuity at the origin, and is therefore non-differentiable and cannot be minimised using standard gradient methods. Typically, the ℓ_1-norm objective is approximated by a function such as the *Huber* M-estimator [7], where the function penalises reconstruction error linearly for large error and behaves quadratically when the error falls beneath some small threshold close to the discontinuity. Another approach is to use *Iteratively Reweighted Least Squares* (IRLS), which approximates the ℓ_1-norm objective by reweighting the differentiable Least Squares objective, where the residual error e, as specified by the ℓ_p norm, is computed by reweighting the minimum ℓ_2-norm solution: $|e|^p \equiv |e|^{p-2}e^2$, where $p = 1$.

In the context of CS, an IRLS algorithm specific to under-determined systems of equations is required,

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{Q}^{-1}\mathbf{x}\|_2, \text{ subject to } \mathbf{\Phi x} = \mathbf{y}. \tag{4}$$

IRLS algorithms, unlike the pseudo-inverse, have no closed-form solution, as $\mathbf{Q}$ is dependent on the previous ℓ_p-norm solution estimate, $\mathbf{x}$. Therefore, in order to improve the estimate of the ℓ_p-norm solution, the procedure is repeated for a number of iterations.

A popular algorithm for performing under-determined IRLS is the FOCUSS algorithm [8], which performs non-convex ℓ_p-norm minimisations, *i.e.*, $0 \leq p \leq 1$, and recovers sparse solutions. Furthermore, the FOCUSS algorithm is used in the compressive sampling of real-valued data [9]. The update equations for FOCUSS are:

$$\mathbf{x}_{k+1} = \mathbf{Q}_k \mathbf{\Phi}^\mathsf{T} (\mathbf{\Phi} \mathbf{Q}_k \mathbf{\Phi}^\mathsf{T})^{-1} \mathbf{y}, \tag{5a}$$

$$\mathbf{Q}_k = \operatorname{diag}(|\mathbf{x}_k|^{(2-p)}), \tag{5b}$$

where $\mathbf{Q}_k$ is initialised with the identity matrix, $\mathbf{Q}_0 = \mathbf{I}$, resulting in the initial solution estimate, $\mathbf{x}_1$, being the minimum ℓ_2-norm solution. On first inspection of Eq. 5a it may be tempting to suggest that the update equation is derived using normal equations, as is the case with the pseudo-inverse and over-determined IRLS. However, this is not the case, the FOCUSS algorithm is derived by solving for the Lagrange multipliers of Eq. 4, then substituting the result into the solution, which gives a fixed point iteration for recovering minimum ℓ_p-norm solutions [8].

Alternatively, Eq. 4 can be restated as

$$\min_{\mathbf{g} \in \mathbb{R}^N} \|\mathbf{g}\|_2, \text{ subject to } \mathbf{\Phi} \mathbf{Q} \mathbf{g} = \mathbf{y}, \tag{6}$$

where the new problem is to find $\mathbf{x} = \mathbf{Q}\mathbf{g}$, which results in the following algorithm:

$$\mathbf{g}_{k+1} = \mathbf{Q}_k \mathbf{\Phi}^\mathsf{T} (\mathbf{\Phi} \mathbf{Q}_k \mathbf{\Phi}^\mathsf{T})^{-1} \mathbf{y}, \tag{7a}$$

$$\mathbf{x}_{k+1} = \mathbf{Q}_k \mathbf{g}_{k+1}, \tag{7b}$$

$$\mathbf{Q}_k = \operatorname{diag}(|\mathbf{x}_k|^{(1-(p/2))}), \tag{7c}$$

where Eq. 7b computes the reweighting of the minimum norm solution Eq. 7a.

To explain how iterative reweighting results in a minimum ℓ_p-norm solution consider the objective minimised at each iteration for Eq. 4 & Eq. 5b,

$$\|\mathbf{Q}_k^{-1} \mathbf{x}_{k+1}\|^2 = \sum_{i=1}^{N} \Big(\frac{x_{k+1}^i}{|x_k^i|^{(2-p)}} \Big)^2. \tag{8}$$

The relatively large entries in $\mathbf{Q}$ deemphasise the contribution of the corresponding entries of $\mathbf{x}$ to the objective (Eq. 8), and vice versa. Therefore, large entries in $\mathbf{x}_k$ result in larger corresponding entries in $\mathbf{x}_{k+1}$, if the respective columns in $\mathbf{\Phi}$ are significant in fitting $\mathbf{y}$, implying that once a favourable weighting is obtained, the weighting at the next iteration continues to be favourable resulting in convergence to a minimum ℓ_p-norm solution. For $p = 1$, reweighting by $\mathbf{Q}$ deemphasises signal outliers in the much the same way as they are by ℓ_1-norm regression methods. Furthermore, the solutions recovered by the FOCUSS algorithm correspond to those recovered by ℓ_1-norm minimisation using the simplex algorithm, where each solution has at most M non-zero entries, ensuring that the recovered signal is sparse.

3. NON-NEGATIVE MATRIX FACTORISATION

A popular method for the analysis of non-negative matrices is *Non-negative Matrix Factorisation* [10, 6], which approximates a non-negative matrix, $\mathbf{V}$, as a product of two non-negative matrices, $\mathbf{V} \approx \mathbf{W}\mathbf{H}$; decomposing the matrix into a non-negative basis, $\mathbf{W}$, and associated coefficients $\mathbf{H}$. NMF is a *parts-based* approach that makes no statistical assumption about the data. Instead, it assumes for the domain at hand, *e.g.* grey-scale images, that negative numbers are physically meaningless—which is the foundation for the assumption that the search for a decomposition should be confined to the non-negative orthant, *i.e.*, the nonnegativity assumption. The lack of statistical assumptions makes it difficult to prove that NMF will give correct decompositions. However, it has been shown in practice to give correct results.

NMF, and its extensions, has been applied to a wide variety of problems including brain imaging [11], tensor factorisation [12], and even automatic ASCII Art conversion [13]. Furthermore, in combination with a magnitude spectrogram representation, NMF has been applied to audio processing tasks such as speech separation [14, 15, 16].

3.1. Standard NMF Algorithm

The NMF algorithm, presented using standard notation, is as follows: Given a non-negative matrix $\mathbf{V} \in \mathbb{R}^{\geq 0, M \times T}$, the goal is to approximate $\mathbf{V}$ as a product of two non-negative matrices $\mathbf{W} \in \mathbb{R}^{\geq 0, M \times R}$ and $\mathbf{H} \in \mathbb{R}^{\geq 0, R \times T}$,

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}. \tag{9}$$

The parameter R, which specifies the number of columns in $\mathbf{W}$ and rows in $\mathbf{H}$, determines the rank of the approximation. Typically $R \leq M$, where $\mathbf{W}$ is over-determined and NMF reveals low-rank features of the data. The selection of an appropriate value for R usually requires prior knowledge, and is important to obtaining a satisfactory decomposition. An important consideration in the derivation of the NMF algorithm is the selection of the objective function, Lee and Seung [6] proposed Least Squares (LS),

$$D_{\text{LS}}(\mathbf{V}, \mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{W}\mathbf{H}\|^2. \tag{10}$$

NMF minimises Eq. 10 while enforcing a non-negativity constraint on the resulting factors:

$$\min_{\mathbf{W}, \mathbf{H}} D_{\text{LS}}(\mathbf{V}, \mathbf{W}, \mathbf{H}) \qquad \mathbf{W}, \mathbf{H} \geq 0,$$

resulting in a parts-based decomposition, where the basis in $\mathbf{W}$ resemble parts of the input data, which can only be summed together to approximate $\mathbf{V}$.

The NMF objective (Eq. 10) is convex in $\mathbf{W}$ and $\mathbf{H}$ individually but not together. Therefore NMF algorithms usually alternate updates of $\mathbf{W}$ and $\mathbf{H}$. The objective is minimised using a diagonally rescaled gradient descent algorithm [6], which enforces the non-negativity constraint and leads to the following multiplicative update equations,

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{\mathbf{V}\mathbf{H}^{\mathsf{T}}}{\mathbf{W}\mathbf{H}\mathbf{H}^{\mathsf{T}}}, \tag{11a}$$

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^{\mathsf{T}}\mathbf{V}}{\mathbf{W}^{\mathsf{T}}\mathbf{W}\mathbf{H}}, \tag{11b}$$

where $\otimes$ denotes an element-wise (also known as Hadamard or Schur product) multiplication, and division is also element-wise. As the NMF algorithm iterates, its factors converge to a local optimum of Eq. 10.

For the interested reader, illustrative examples of the factors obtained by NMF when applied to synthetic data are presented in [17].

4. NON-NEGATIVE UNDER-DETERMINED IRLS

We propose an algorithm for the recovery of sparse signals from compressively sampled non-negative data. The algorithm performs under-determined IRLS, as stated in Eq. 6, with an additional non-negativity constraint, *i.e.*, $\mathbf{\Phi}, \mathbf{Q}, \mathbf{g}, \mathbf{y} \geq 0$. We refer to the algorithm as *Non-negative Under-determined Iteratively Reweighted Least Squares* (NUIRLS), which performs non-convex minimisations recovering non-negative ℓ_p-norm, $0 \leq p \leq 1$, solutions. Furthermore, the algorithm is derived within the framework of NMF, resulting in multiplicative update equations. We would also like to note that unlike standard NMF, since we employ under-determined IRLS, $R \geq M$.

By comparing Eq. 3, which describes compressive sampling of non-negative data, to Eq. 10, which describes NMF with a minimum ℓ_2-norm objective, it is easy to see that in the context of CS, the NMF factors $\mathbf{W}$ & $\mathbf{H}$ correspond to $\mathbf{\Phi}$ & $\mathbf{x}$, while $\mathbf{V}$ corresponds to $\mathbf{y}$. We will use the CS notation in the exposition of the proposed algorithm.

NUIRLS employs the following objective function,

$$D_{\text{NUIRLS}}(\mathbf{y}, \mathbf{\Phi}, \mathbf{Q}_k, \mathbf{g}) = \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi}\mathbf{Q}_k\mathbf{g}\|^2, \tag{12}$$

which when minimised performs the minimisation problem stated in Eq. 6. For our purposes, unlike NMF, we do not require a matrix factorisation but a non-negative ℓ_p-norm solution, $\mathbf{x}$, given a fixed sampling matrix, $\mathbf{\Phi}$, which results in the following minimisation problem for NUIRLS:

$$\min_{\mathbf{g}} \; D_{\text{NUIRLS}}(\mathbf{y}, \mathbf{\Phi}, \mathbf{Q}_k, \mathbf{g}) \qquad \mathbf{\Phi}, \mathbf{Q}, \mathbf{g}, \mathbf{y} \geq 0,$$

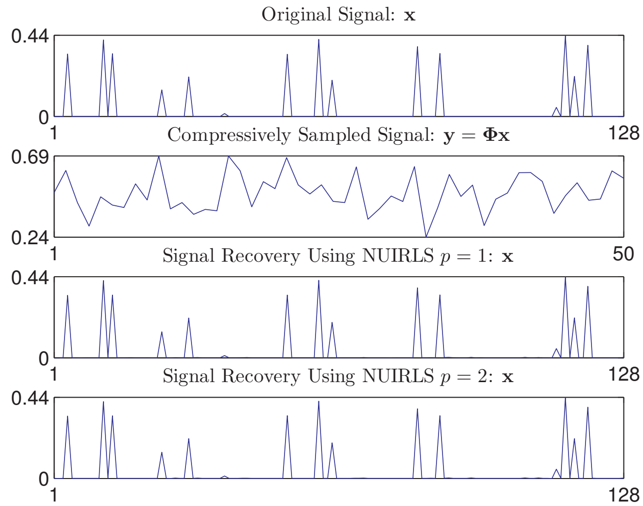

Fig. 2. Compressive sampling of a non-negative signal: The original signal $\mathbf{x}$ is of length $N = 128$ with support $K = 15$ and is sparse in the canonical basis. $\mathbf{x}$ is projected onto the sampling matrix $\mathbf{\Phi}$, which has non-negative entries drawn from a rectified Gaussian distribution, resulting in a compressively sampled signal $\mathbf{y}$ of length $M = 50$. Recovery of $\mathbf{x}$ is performed by the proposed NUIRLS algorithm with $p = 1$ & $p = 2$, and recovers minimum ℓ_1-norm & ℓ_2-norm solutions respectively. It is evident that both achieve perfect recovery for non-negative data—which contrasts with compressive sampling of real-valued data where the pseudo-inverse (minimum ℓ_2-norm solution) fails to recover sparse signals—suggesting that a non-negativity constraint may be all that is needed to recover compressively sampled non-negative sparse signals.

The update equations for the standard NMF algorithm (Eq. 11) are derived using the diagonally rescaled gradient descent algorithm, the NUIRLS update equations are also derived in this way, resulting in the following algorithm:

$$\mathbf{g}_{k+1} = \mathbf{g}_k \otimes \frac{\mathbf{Q}_k\mathbf{\Phi}^{\mathsf{T}}\mathbf{y}}{\mathbf{Q}_k\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}\mathbf{Q}_k\mathbf{g}_k}, \tag{13a}$$

$$\mathbf{x}_{k+1} = \mathbf{Q}_k\mathbf{g}, \tag{13b}$$

$$\mathbf{Q}_k = \operatorname{diag}((\mathbf{g}_k)^{(1-(p/2))}). \tag{13c}$$

NUIRLS is initialised with $\mathbf{Q}_0 = \mathbf{I}$, resulting in the initial signal estimate, $\mathbf{x}_1$, being the minimum non-negative ℓ_2-norm solution. Furthermore, NUIRLS resembles Eq. 7, the key difference being the preservation of non-negativity through a multiplicative update equation.

As discussed in Section 3, NMF is an iterative algorithm, IRLS is also an iterative algorithm, combining both results in NUIRLS being a two-step iterative algorithm, where the minimum ℓ_2-norm solution at each NMF iteration is iteratively reweighted to recover a minimum ℓ_p-norm solution, which is used in the next NMF iteration and so on. As this process is repeated, NUIRLS converges to a local optimum of Eq. 12. Furthermore, unlike the standard NMF update equations (Eq. 11), which have update equations for each matrix $\mathbf{W}$ & $\mathbf{H}$, the update equation for $\mathbf{x}$ is restricted to column vectors, as $\mathbf{Q}$ is specific to each $\mathbf{x}$.

We present an example of the compressive sampling of

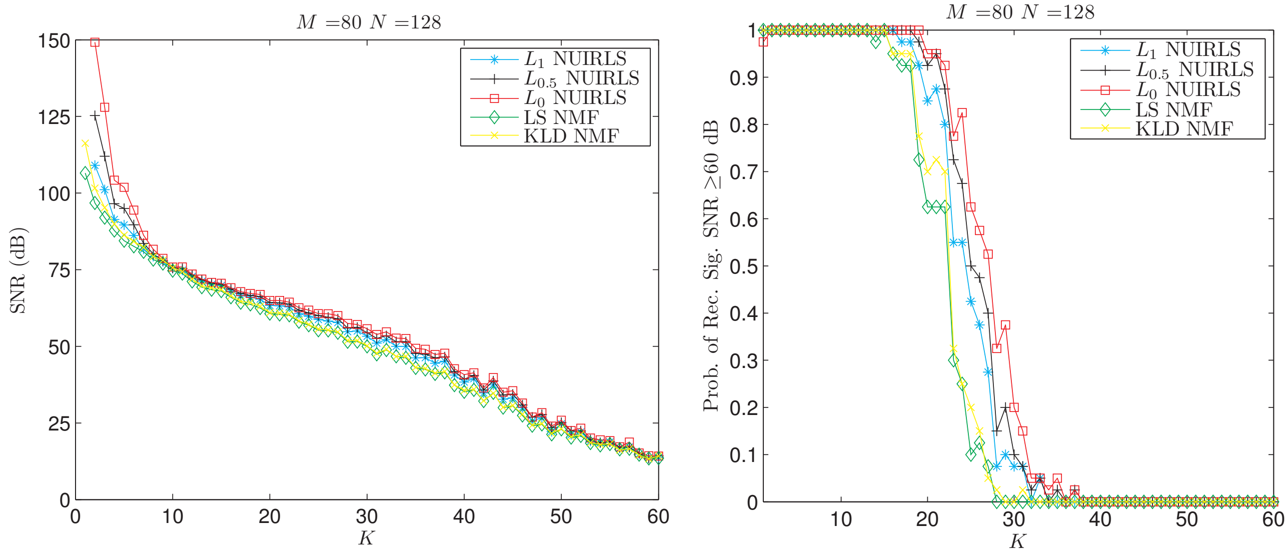

Fig. 3. Non-negative signals with random support and magnitude of increasing K-sparseness are compressively sampled using a non-negative $\mathbf{\Phi}$ of fixed size, $M = 80$ & $N = 128$. The compressively sampled signals are recovered using NUIRLS and the standard Lee and Seung update equations for Least Squares and Kullback-Leibler Divergence NMF. The experiment is repeated for 40 Monte Carlo runs and the Signal-to-Noise Ratio (SNR) of the original and recovered signals, averaged over all runs, are presented (left). The probability of recovering a signal with SNR $\geq$ 60 dB, which indicates a high quality signal, as a function of K is also presented (right). It is evident for $K \leq 15$ that standard NMF achieves a SNR $\geq$ 60 dB, while for NUIRLS $K \leq 19$ achieves the same SNR. The plots indicating that a non-negativity constraint is enough to recover non-negative signals that are sufficiently sparse. For the specified experiment, signals with K-sparseness ≤ 15 fit this category.

non-negative data using NUIRLS in Figure 2, where the $\mathbf{x}$ and $\mathbf{\Phi}$ used in the example are the absolute value of those used in Figure 1. We run NUIRLS for 350 NMF iterations with each having 100 IRLS iterations, and test NUIRLS using $p = 1$ & $p = 2$, where $p = 2$ results in solutions that correspond to Least Squares NMF. Returning to Figure 1, it is evident that the pseudo-inverse, *i.e.*, minimum ℓ_2-norm solution, fails to recover the original sparse signal from compressively sampled real-valued data. However, for non-negative signals, Figure 2 demonstrates that it is possible to recover sparse solutions from the compressively sampled signals by selecting the minimum ℓ_2-norm solution, $p = 2$; suggesting that, for this example, the non-negativity constraint of NUIRLS is enough to recover the original signal.

5. NUMERICAL EXPERIMENTS

We use NUIRLS to recover sparse signals from compressively sampled non-negative data, and compare the recovered signals to those recovered by the standard Lee and Seung update equations for Least Squares and Kullback-Leibler Divergence NMF [6].

We perform compressive sampling where $\mathbf{\Phi}$ and $\mathbf{x}$ are of fixed dimension, $M = 80$ & $N = 128$, and test for a number of signals with increasing K-sparseness, with $K = 60$ being the maximum. Since $\mathbf{\Phi}$ must be non-negative, its entries are drawn from a rectified Gaussian Distribution, where the absolute value of negative values is used. We run NUIRLS for 1500 NMF iterations, each having 150 IRLS iterations, and specify $p = \{0, 0.5, 1\}$. In order to keep both NUIRLS and standard NMF in an even setting, NMF is run for 225000 (1500×150) iterations. The experiment is repeated for 40 Monte Carlo runs, where a new $\mathbf{\Phi}$ is constructed for each run. The Signal-to-Noise Ratio (SNR) of the recovered signals are averaged over all Monte Carlo runs, and are plotted in Figure 3.

The plots indicate that for recovered signals with a required SNR of 60 dB, standard NMF successfully achieves the desired SNR for $K \leq 15$, while NUIRLS achieves the same SNR for $K \leq 19$, which demonstrates that if the compressively sampled signal is sufficiently sparse, in this case $K \leq 15$, a non-negativity constraint is enough to recover the signal [5]. Therefore, standard NMF can be employed in the recovery of sufficiently sparse signals from compressively sampled non-negative data. Unlike, signal recovery from compressively sampled real-valued data, where the ability to recover signals is dependent on the selection of a norm constraint that fits with the sparseness assumption of the signal.

It is evident that NUIRLS produces better quality signals as K approaches 40. However, it is important to note that as K approaches N the signal can no longer be consider sparse in the canonical basis, and CS fails with minimum ℓ_2-norm solutions recovering the best, although of very poor SNR, signals. Furthermore, it is evident that SNR performance decreases as K increases, which indicates that NUIRLS and NMF need to be run for more iterations as the signal becomes less sparse in order to achieve the required SNR.

6. DISCUSSION

For the recovery of sparse signals from compressively sampled real-valued data, if $\mathbf{\Phi}$ has entries drawn from a Gaussian distribution and the signal is sparse in a given sparse basis, $\mathbf{A}$, then $\mathbf{\Phi}$ is maximally incoherent with the basis with high probability. Following this reasoning, for our non-negative signal experiments we select the entries of $\mathbf{\Phi}$ from a rectified Gaussian distribution, as the signals under consideration are also sparse. However, it may be that our non-negative $\mathbf{\Phi}$ is not optimally incoherent with $\mathbf{A}$, which will have an effect on the quality of the recovered signals. Furthermore, there exists preconditioning methods that improve the incoherence properties of $\mathbf{\Phi}$ by reducing correlations between its constituent columns [5], thereby reducing the required number of measurements, M. The proposed algorithm may also benefit from preconditioning $\mathbf{\Phi}$. However, the preconditioned matrix must also be non-negative.

It has been demonstrated in the CS literature that employing sophisticated regularisation strategies in combination with IRLS, which manifests as a modification of Eq. 13c, improves the ability of the algorithm to recover sparse signals [9]. It is possible that NUIRLS may also benefit from such an extension.

Through the presented illustrative examples of CS on both real-valued and non-negative signals, we endeavour to dispel any norm-centric preconceptions that may arise when transposing the theory of compressive sampling to the non-negative domain, specifically that non-convex minimisations are necessary to recover sparse signals.

Finally, in this paper we set the stage for our main motivation in this direction: An NMF algorithm that discovers a sparse basis, $\mathbf{W}$, and corresponding activations, $\mathbf{H}$, for non-negative data, $\mathbf{V}$, in a compressively sampled domain, $\mathbf{\Phi V}$. Work related to this algorithm will be presented soon.

7. CONCLUSION

In this paper, we applied the theory of compressive sampling to non-negative signals, where the sampling matrix and basis are also non-negative, and proposed an algorithm—Non-negative Under-determined Iteratively Reweighted Least Squares (NUIRLS)—for signal recovery. The proposed algorithm is derived within the framework of Non-Negative Matrix Factorisation resulting in multiplicative update equations.

We apply the standard least squares NMF update and the proposed algorithm to a compressive sampling non-negative signal recovery task and demonstrated that, unlike the compressive sampling of real-valued data, sparse signals may be recovered from minimum ℓ_2-norm solutions using NMF. However, we demonstrated that the choice of norm becomes more important as the sparseness of the signal decreases, with the NUIRLS algorithm recovering the best signals. Demonstrating that, for sufficiently sparse non-negative signals, the norm used to specify a solution becomes less important as a non-negativity assumption is all that is required for signal recovery.

8. REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[2] David L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[4] Emmanuel Candès and Terence Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, pp. 5406–5425, 2006.

[5] Alfred M. Bruckstein, Michael Elad, and Michael Zibulevsky, "A non-negative and sparse enough solution of an underdetermined linear system of equations is unique," Submitted to IEEE Transactions on Information Theory, 2008.

[6] Daniel D. Lee and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," in *Adv. in Neu. Info. Proc. Sys. 13*. 2001, pp. 556–62, MIT Press.

[7] P. J. Huber, *Robust Statistics*, Wiley, 1981.

[8] Bhaskar D. Rao and Kenneth Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 187–200, January 1999.

[9] Rick Chartrand and Wotao Yin, "Iteratively reweighted algorithms for compressive sensing," Submitted, 2008.

[10] P. Paatero and U. Tapper, "Positive matrix factorization: A nonnegative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, pp. 111–26, 1994.

[11] M. N. Schmidt and H. Laurberg, "Non-negative matrix factorization with gaussian process priors," *Computational Intelligence and Neuroscience*, 2008.

[12] Amnon Shashua and Tamir Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in *ICML '05: Proceedings of the 22nd international conference on Machine learning*, New York, NY, USA, 2005, pp. 792–799, ACM.

[13] Paul D. O'Grady and Scott T. Rickard, "Automatic ASCII art conversion of binary images using non-negative constraints," in *Proceedings of the Irish Signal and Systems Conference*, 2008, pp. 186–191.

[14] Paris Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *Fifth International Conference on Independent Component Analysis*, Granada, Spain, Sept. 22–24 2004, LNCS 3195, pp. 494–9, Springer-Verlag.

[15] D. FitzGerald, M. Cranitch, and E. Coyle, "Sound source separation using shifted non-negative tensor factorisation," in *Proceedings, IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006.

[16] Paul D. O'Grady and Barak A. Pearlmutter, "Discovering convolutive speech phones using sparseness and non-negativity," in *Seventh International Conference on Independent Component Analysis*, London, UK, Sept. 2007, LNCS 4666, pp. 520–7, Springer-Verlag.

[17] Paul D. O'Grady, *Sparse Separation of Under-Determined Speech Mixtures*, Ph.D. thesis, National University of Ireland Maynooth, 2007.