APPROXIMATE LOWER BOUNDS FOR RATE-DISTORTION IN COMPRESSIVE SENSING SYSTEMS

B. Mulgrew and M.E. Davies

Institute for Digital Communications
School of Engineering and Electronics
The University of Edinburgh
Edinburgh, EH9 3JL, UK
Email: B.Mulgrew@ed.ac.uk

ABSTRACT

We attempt to quantify the possible gains that can be achieved by examining a rate-distortion competition between a conventional and a compressive sampling solution to data rate reduction. Simple approximate expression are developed for the minimum bit rate required to obtain the best achievable average performance from the compressive sensing system and the performance that would be achieved if that rate requirement was met. An example of a signal that contains a small number of phasors in Gaussian white noise is used to validate these results and to compare the degradation in performance of the 2 systems when lower bit rates than those required by the theory are employed.

Index Terms**—** compressive sensing, rate-distortion

1. INTRODUCTION

The original compressed sensing concepts of [1][2] have been extended to handle noisy environments as in [3] and [4]: the driver in [3] has been to provide upper bounds on the achievable distortion in the reconstruction i.e. what is the worst performance that can be guaranteed while in [5] rate distortion theory is used to set lower bounds on the distortion in reconstruction for both low and high signal to noise conditions. In [6] expressions for the asymptotic bounds on the distortion at a particular rate are developed. While in [7], a lower bound on the minimum signal-to-noise ratio to detect a sparse signal is derived. The objective here is to develop approximate expressions for the minimum bit rate required to obtain the best achievable average performance from a compressive sensing system and the performance that would be achieved if that rate requirement was met.

The paper is organised as follows: in section 2 a rate-distortion competition is defined; assumptions are set out in section 3; section 4 provides expressions for the bit rate required to obtain the best achievable performance from a compressive sensing system as well as an estimate for the reconstruction distortion when that rate is satisfied; these expressions are validated for a phasors in white noise problem in section 5 using the orthogonal least squares (OLS) algorithm [8] (with an appropriately chosen stopping criteria) as the reconstruction algorithm; brief conclusions are drawn in section 6.

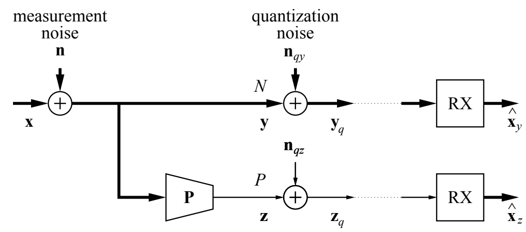

Fig. 1. Block diagram: rate-distortion competition

2. PROBLEM DEFINITION

An N-vector, $\mathbf{y}$, of complex measurements is assumed to be a sparse combination of M N-vectors such as $\underline{\phi}_i$ and additive white Gaussian noise $\mathbf{n}$. Thus:

$$\mathbf{y} \quad = \quad \mathbf{\Phi h} + \mathbf{n} = \mathbf{x} + \mathbf{n} \tag{1}$$

where: $\mathbf{\Phi} = [\underline{\phi}_1 \underline{\phi}_2 ... \underline{\phi}_M]$. The majority of the elements of M-vector $\mathbf{h}$ are zero. The matrix of atoms $\mathbf{\Phi}$ is $(N \times M)$ where $M > N$ to allow for frames and overcomplete representations. The “clean” or noise free signal is $\mathbf{x}$. The signal to noise ratio is: $S_{yn} = |\mathbf{\Phi h}|^2 / N\sigma_n^2$ where $\sigma_n^2 = E[n_i n_i^*]$ is the variance of each element of the white noise vector. The classical or conventional approach is to quantize $\mathbf{y}$ using simple scalar quantization of its elements and transmit these quantized samples as in the upper branch of Fig. 1.

The lower branch of Fig 1 illustrates the compressive sensing approach. To reduce the amount of data and/or compress the data a further matrix multiplication is applied to $\mathbf{y}$ using a $P \times N$ matrix $\mathbf{P}$

$$\mathbf{z} \quad = \quad \mathbf{Py} = (\mathbf{P\Phi})\mathbf{h} + \mathbf{Pn} \tag{2}$$

where $P < N$. To ease the burden of coding, $\mathbf{P}$ is chosen to be computationally simple. Equation (2) is a similar problem to (1) with $(\mathbf{P\Phi})$ replacing $\mathbf{\Phi}$. Thus Fig. 1 illustrates a rate distortion competition between a compressive sensing system and a conventional scalar quantizer (upper branch). Whether we reconstruct from $\mathbf{y}$ or from $\mathbf{z}$ the distortion measure, measure of the signal reconstruction quality or mean squared error (MSE) is:

$$D_r = \frac{E[|\mathbf{x} - \hat{\mathbf{x}}|^2]}{|\mathbf{x}|^2} \tag{3}$$

Quantized data is transmitted over identical communications channels to receivers (RX) which perform the reconstruction. The effects of the communications channels are not considered here.

Consider the case were $\mathbf{P}$ is a Bernoulli type matrix whose elements are drawn at random from $\pm 1 \pm j$ we have the identity:

$$\mathrm{tr}(\mathbf{P}^H\mathbf{P}) = 2PN \tag{4}$$

and the approximation:

$$\mathbf{P}\mathbf{P}^H \approx 2N\mathbf{I}_P \tag{5}$$

the latter is similar to estimating the covariance matrix of a white random process and is reasonable provided $P < N/2$. The other matrix that is significant is $\mathbf{P}^H\mathbf{P}$ for which:

$$E[\mathbf{P}^H\mathbf{P}] = 2P\mathbf{I}_N \tag{6}$$

Lastly, the measurement noise, quantization noise and projection matrices are mutually independent random variables.

3. RATE AND DISTORTION

First quantize and transmit the elements of $\mathbf{z}$

$$\begin{aligned} \mathbf{z}_q &= q((\mathbf{P\Phi})\mathbf{h} + \mathbf{Pn}) \\ &\approx (\mathbf{P\Phi})\mathbf{h} + \mathbf{Pn} + \mathbf{n}_{qz} \end{aligned} \tag{7}$$

where $\mathbf{n}_{qz}$ is the usual additive i.i.d. approximation with signal to quantization noise, S_{zq} defined as:

$$S_{qz} = 10\log_{10}\left(\frac{E[\mathbf{h}^H\mathbf{\Phi}^H\mathbf{P}^H\mathbf{P\Phi h}] + E[\mathbf{n}^H\mathbf{P}^H\mathbf{Pn}]}{E[\mathbf{n}_{qz}^H\mathbf{n}_{qz}]}\right) \tag{8}$$

The number of bits required to transmit P elements of $\mathbf{z}$ is proportional to: $PS_{zq}/6$. To transmit the signal and measurement noise without further distortion the quantization noise level should be set below that of the thermal noise, i.e.:

$$E[\mathbf{n}_{qz}^H\mathbf{n}_{qz}] < E[\mathbf{n}^H\mathbf{P}^H\mathbf{Pn}] \tag{9}$$

giving a lower bound on S_{zq} of:

$$\begin{aligned} S_{qz} &> 10\log_{10}\left(1 + \frac{\mathrm{tr}(E[\mathbf{P}^H\mathbf{P}]\mathbf{x}\mathbf{x}^H)}{\mathrm{tr}(\mathbf{P}^H\mathbf{P})\sigma_n^2}\right) \\ &> 10\log_{10}\left(1 + \frac{|\mathbf{x}|^2}{N\sigma_n^2}\right) \end{aligned} \tag{10}$$

Thus by drawing a projection matrix $\mathbf{P}$ at random the average signal to noise ratio in each element of the vector $\mathbf{z}$ is almost identical to the signal to noise ratio in each element of the vector $\mathbf{y}$. As indicated earlier this signal to noise ratio indicates the minimum number of bits per element required to represent these vectors. Allocating many more bits than this is not likely to provide much in the way of performance improvement. *This does not, however, tell us how much performance loss we might incur by using* $\mathbf{z}$ *rather than* $\mathbf{y}$. To get the best performance out of $\mathbf{z}$, we need to transmit *more than*:

$$\frac{P}{6N}\left(1 + \frac{|\mathbf{x}|^2}{N\sigma_n^2}\right)_{dB} \text{bits/sample} \tag{11}$$

of $\mathbf{y}$. Thus to reach the *best achievable performance* we might pragmatically expect to have to transmit one extra bit per element of $\mathbf{z}$ in which case the rate expression becomes:

$$R_z \approx \frac{P}{6N}\left(1 + \frac{|\mathbf{x}|^2}{\sigma_n^2 N}\right)_{dB} + \frac{P}{N} \tag{12}$$

Given P samples in $\mathbf{z}$, this is the minimum bit rate required to provide the best achievable reconstruction performance from these samples. Ideally we desire a relationship between quality of reconstruction and bit rate. Recalling the defining equation (1) and *adopting an "oracle assumption"*, i.e.: assuming that a greedy algorithm such as OLS [8] or orthogonal matching pursuit (OMP) [9] has correctly selected the N_a non-zero or "active" elements in $\mathbf{h}$ to form a N_a-vector $\mathbf{h}_a$, we have:

$$\mathbf{z} = \mathbf{Py} = (\mathbf{P\Phi}_a)\mathbf{h}_a + \mathbf{Pn}$$

where $\mathbf{\Phi}_a$ is a $N \times N_a$ matrix constructed from the appropriate columns of $\mathbf{\Phi}$ where $N > N_a$. With these assumptions we now have a classical overdetermined set of equations with a least squares solution. There is considerable literature and analysis available under the heading of "recursive least squares" adaptive filtering [10]. The LS estimate is given by:

$$\hat{\mathbf{h}}_a = \left(\mathbf{\Phi}_a^H\mathbf{P}^H\mathbf{P\Phi}_a\right)^{-1}\mathbf{\Phi}_a^H\mathbf{P}^H\mathbf{z} \tag{13}$$

and the estimation error $\mathbf{h}_e = \mathbf{h}_a - \hat{\mathbf{h}}_a$ is given by:

$$\hat{\mathbf{h}}_e = \left(\mathbf{\Phi}_a^H\mathbf{P}^H\mathbf{P\Phi}_a\right)^{-1}\mathbf{\Phi}_a^H\mathbf{P}^H\mathbf{Pn}$$

Using (5) the error covariance matrix is approximated:

$$E[\mathbf{h}_e\mathbf{h}_e^H] \approx 2N\sigma_n^2 E_p\left[\left(\mathbf{\Phi}_a^H\mathbf{P}^H\mathbf{P\Phi}_a\right)^{-1}\right] \tag{14}$$

The order of the expectation operator and the matrix inversion can be interchanged if the row vectors of $\mathbf{P\Phi}_a$ are i.i.d. Gaussian (Appendix G of [10]). With this in mind, consider the ith row vector of $\mathbf{P}$ as $\mathbf{p}_i^H$, the ith row vector of $\mathbf{P\Phi}_a$ is thus $\mathbf{p}_i^H\mathbf{\Phi}_a$. Hence if the row vectors of $\mathbf{P}$ are independent so will the row vectors of $\mathbf{P\Phi}_a$. Further, since each row $\mathbf{p}_i^H\mathbf{\Phi}_a$ is itself constructed from N_a linear combinations of N i.i.d. random variables, the central limit theorem would suggest that this row vector would be well approximated by a Gaussian distribution. With this Gaussian i.i.d. assumption, (14) can be simplified for $P > N_a + 1$:

$$\begin{aligned} E[\mathbf{h}_e\mathbf{h}_e^H] &\approx \frac{2N\sigma_n^2}{P - N_a - 1}\left(\frac{E_p\left[\mathbf{\Phi}_a^H\mathbf{P}^H\mathbf{P\Phi}_a\right]}{P}\right)^{-1} \\ &= \frac{N\sigma_n^2}{P - N_a - 1}\left(\mathbf{\Phi}_a^H\mathbf{\Phi}_a\right)^{-1} \end{aligned} \tag{15}$$

The estimation error we seek is: $\mathbf{e}_x = \mathbf{x} - \hat{\mathbf{x}}_a = \mathbf{\Phi}_a\mathbf{h}_e$ Thus we have an expression for the numerator of the performance metric:

$$\begin{aligned} E[\mathbf{e}_x^H\mathbf{e}_x] &= \mathrm{tr}\left\{\mathbf{\Phi}_a^H\mathbf{\Phi}_a E[\mathbf{h}_e\mathbf{h}_e^H]\right\} \\ &\approx \frac{N}{P - N_a - 1}\sigma_n^2 N_a \end{aligned} \tag{16}$$

In decibels the distortion is:

$$(D_{rz})_{dB} \approx -\left(\frac{\sigma_n^2 N}{|\mathbf{x}|^2}\right)_{dB} + \left(\frac{N_a}{P - N_a - 1}\right)_{dB} \tag{17}$$

Thus, given P and knowledge of N_a, we can evaluate the best achievable reconstruction performance. In particular (17) and (12) define how the best achievable reconstruction performance and the minimum bit rate to achieve that performance vary with P.

4. RESULTS

To illustrate these ideas we construct an example that is simple enough to remove any problems of computational complexity but rich enough to expose strengths and weaknesses in practical application. The discrete-time signal in question consists of a small number, N_f, of complex phasors of unknown frequencies $\{\omega_i\}_{i=1}^{N_f}$ and unknown complex amplitudes $\{A_i\}_{i=1}^{N_f}$ in additive Gaussian white noise. Thus:

$$y(t) = \sum_{i=1}^{N_f} A_i e^{j\omega_i t} + n(t) \tag{18}$$

where t is the integer time index and the frequencies are normalized such that the sampling rate is unity, i.e.: $0 \leq \omega_i < 1$. The signal/noise ratio of $y(t)$ is: $S_y = \sum_{i=1}^{N_f} |A_i|^2 / E[|n|^2]$. The nature of the signal (18) suggests that it could be represented in a sparse manner using Fourier functions. However, by definition of (1) we can only accommodate a finite number of them, which suggests a sampling processes in the frequency domain similar to the discrete Fourier transform (DFT). For a data record of N-samples i.e. : $0 \leq t < N$ a phasor has the form $\phi_m(t) = e^{j\Delta_\omega m t}$ where $\Delta_\omega = \frac{2\pi}{ND}$ and D is the oversampling factor, typically $D = 2$. For a results shown in this section: a projection matrix $\mathbf{P}$ is drawn at random; for each point on a graph measurements are made by averaging over an ensemble of 20 runs at which: the frequencies $\{\omega_i\}_{i=1}^{3}$ are drawn from a uniform distribution; the power in the phasors in deciBels are drawn uniformly in the range 0 to 50 dB; the phase shift of each phasor is drawn uniformly over 360 degrees; thermal noise is added to give an indicated signal to noise ratio and Gaussian noise is added to model quantization noise.

Both basis and matching pursuit algorithms are available to solve the problems of equations such as (1) and (2). Here we choose the greedy forward regression technique, orthogonal least squares (OLS) [8] because: (i) it is also based strongly on LS estimation and thus it is a reasonably choice to validate the performance trade-off's predicted in section 4; (ii) it has been shown to function well in noisy environments [11]; (iii) it has a convenient stopping criteria and thus has some capability to self-select N_a. The technique is described in detail in [8] in the context of basis function selection for radial basis function neural networks. However it is not restricted to that application and it can be applied to any regression problem formulated as per equation (1). It is a foward selection technique known as order recursive matching pursuit (ORMP) in the approximation literature. A convenient mechanism for stopping is to use an Akaike-type criteria which provides a compromise between the number of regressors N_r and the quality of the estimate as in [8]. The parameter χ used in AIC is the critical value of the chi-squared distribution with one degree of freedom for a given level of confidence. Chen et al. [8] suggest a choice might be $\chi = 4$ for a confidence level of ~ 0.05. However here we use values of 6 and 8 for the conventional and compressive cases respectively. As a second stage of refinement we prune the weights selected by the OLS/AIC combination based on an estimate of the error covariance of the weights themselves. *An explicit expression is given by (14) in terms of the variance of the noise* σ_n^2 *and the expectation over random projection matrices, neither of which we have access to in practice.* If we replace the variance with an estimate of the variance of the residuals and the expectation with one realisation of its argument we obtain a crude estimate of the error covariance:

$$E[\mathbf{h}_e \mathbf{h}_e] \approx 2\mathbf{e}^H \mathbf{e} \left(\mathbf{\Phi}^H \mathbf{P}^H \mathbf{P} \mathbf{\Phi} \right)^{-1} \tag{19}$$

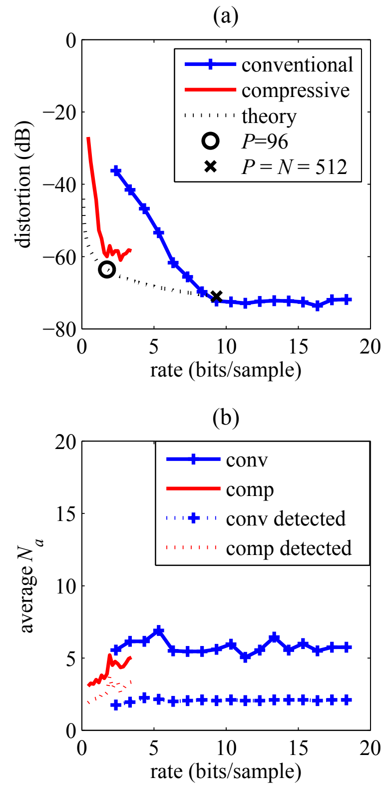

Fig. 2. Sparse signal consisting of 2 phasors with dynamic range 50 dB; signal/noise of 50 dB: (a) rate distortion curves; (b) average number of atoms versus rate.

The diagonal of the error covariance matrix contains the individual variances associated with the estimate of each coefficient which can be used to set a threshold for accepting of rejecting a particular coefficient - the argument being that a particular coefficient estimate needs to be significantly larger than its own estimation variance to have any value. Once the weights have been pruned the LS estimate of the remaining weights are recomputed.

For the initial result of Fig. 2 a sparse signal is used by restricting the selection of frequencies to two that lie exactly on the oversampled Fourier grid used for reconstruction. Thus we can say precisely that $N_a = N_f = 2$. The signal to noise ratio S_y is 50 dB, $N = 512$ and $P = 96$. Fig. 2(a) shows how the performance in terms of reconstruction MSE, D_r of (3), varies with the transmitted bit rate in term of bits per element of the raw measurement vector $\mathbf{y}$. Quantization noise is simulated by adding appropriate levels of Gaussian white noise to the elements of both $\mathbf{y}$ and $\mathbf{z}$. The curve marked "theory" shows the predicted lower bound on distortion (17) varies with rate (12) as P increases. Two particular cases are highlighted: $P = 96$ and $P = N = 512$.

The curve labeled "conventional" is obtained by applying the OLS algorithm with AIC and pruning directly to $\mathbf{y}$ after quantization. At first sight this may not appear to be a conventional algorithm to apply to this problem but it is a useful benchmark of performance as it will be close to the maximum likelihood estimate $E[\mathbf{x}|\mathbf{y}]$ and it is capable of exploiting the sparse nature of the signal. However it is conventional in the sense that the rate is controlled simply by allocating bits on a per element basis to $\mathbf{y}$. The curve marked compressive is obtained by applying the same algorithm to the vector $\mathbf{z}$ after quantization. In using AIC a value of $\chi = 6$ is used for the conventional case and $\chi = 8$ for the compressive. The selection of χ is a well known difficulty with AIC and the subsequent use of the pruning technique alleviates most of the difficulties associated with it. The curves labeled "detected" on Fig. 2(b) indicate that the error variances on the diagonal of (19) were used to set a thresold to accept or reject elements of $\mathbf{h}$ i.e. a detection threshold for pruning. In all cases examined here the detection threshold was set 10 dB above the estimated error variance. It is clear that in the conventional case the correct number of atoms have been chosen after pruning. AIC over-estimates this values. For the compressive case, even with pruning, an average of 3 rather than 2 atoms are selected.

It is clear that equation (12) reliably predicts the minimum rate at which the best achievable performance is obtained for a given value of P: 1.75 bits/sample at $P = 96$ and 9.33 bits/sample at $P = N = 512$. Equation (17) provides a reasonable estimate of the minimum distortion that is achievable at this bit rate and above for a particular value of P e.g.: -64 dB for $P = 96$ and -71 dB for $P = N = 512$. At rates below the minimum the performance degrades monotonically as the level of the quantization noise swamps the additive noise.

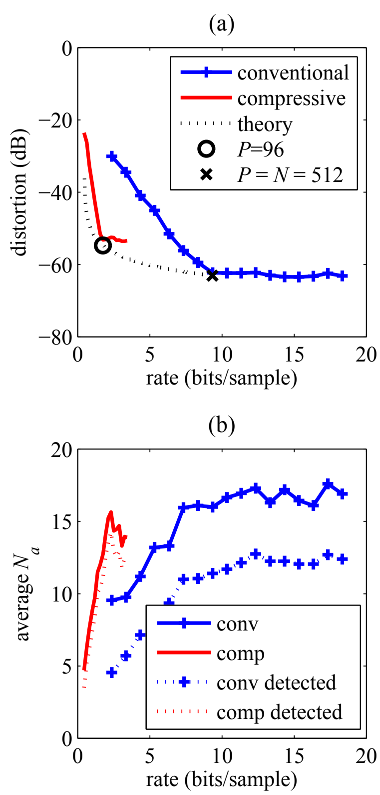

Fig. 3. Compressive signal consisting of 2 phasors with dynamic range 50 dB; signal/noise of 50 dB: (a) rate distortion curves; (b) average number of atoms versus rate.

To warrant the use of a compressive sensing system we must have some degree of confidence from the outset that a basis or frame such as Φ exists which permits a sparse of compressive representation of the observable vector $\mathbf{y}$. In addition we might need an estimate of N_a to assess what degree of compression could be achieved. One way of way of addressing both of these issues is to apply the conventional scheme described above to typical signals to provide an estimate of N_a. The OLS/AIC plus pruning technique provides reliable estimates of N_a as shown above. This estimate can then be used in equation (17) to predict the minimum distortion that can be achieved against P and hence rate.

For Fig. 3, the two phasors can have any possible frequency and hence will not fall exactly onto the Fourier grid. The signal is thus no longer sparse with respect to the frame $\mathbf{\Phi}$ but rather compressive but admits a sparse approximation. From Fig. 3(b) we conclude that a value of $N_a = 12$ is sufficient to represent the signal under these conditions as this is the value used for the conventional system after pruning. This value is used for the theory curve of Fig. 3(a). In common with Fig. 2(a), equation (17) predicts the lower bound on performance at a particular rate.

5. SUMMARY AND CONCLUSIONS

Because the expectation of the matrix $\mathbf{PP}^H$ is diagonal, the signal to noise ratio before and after projection are approximately the same. Hence the number of bits required to extract the maximum performance from each sample of $\mathbf{y}$ and $\mathbf{z}$ is the same. Having done the projection we are left with a standard least squares parameter estimation problem but with fewer ,i.e. P, observations than the original problem, i.e. N, and hence degraded performance. LS convergence analysis can be applied to this problem to relate distortion to P and hence rate.

6. REFERENCES

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

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

[3] J. Haupt and P. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions Information Theory*, 2007, To appear.

[4] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[5] Alyson K. Fletcher, Sundeep Rangan, Vivek K Goyal, and Kannan Ramchandran, "Denoising by sparse approximation: Error bounds based on rate-distortion theory," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 26318, 19 pages, 2006.

[6] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On the rate-distortion performance of compressed sensing," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, Apr. 2007, vol. 3.

[7] Alyson K. Fletcher, Sundeep Rangan, and Vivek K Goyal, "Rate-distortion bounds for sparse approximation," in *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, Aug. 2007, pp. 254–258.

[8] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[9] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers*, Nov. 1993.

[10] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, fourth edition, 2002.

[11] B. Mulgrew, "Algorithms for passive SAR imaging," in *IEEE/IET Waveform Diversity and Design*, Jan. 2006.