

Applications of sparse approximation in communications

A. C. Gilbert

Department of Mathematics
University of Michigan
Ann Arbor, MI 48109
Email: annacg@umich.edu

J. A. Tropp

Department of Mathematics
University of Michigan
Ann Arbor, MI 48109
Email: jtropp@umich.edu

Abstract—Sparse approximation problems abound in many scientific, mathematical, and engineering applications. These problems are defined by two competing notions: we approximate a signal vector as a linear combination of elementary atoms and we require that the approximation be both as accurate and as concise as possible. We introduce two natural and direct applications of these problems and algorithmic solutions in communications. We do so by constructing enhanced codebooks from base codebooks. We show that we can decode these enhanced codebooks in the presence of Gaussian noise. For MIMO wireless communication channels, we construct simultaneous sparse approximation problems and demonstrate that our algorithms can both decode the transmitted signals and estimate the channel parameters.

I. INTRODUCTION

Sparse approximation problems arise in a host of scientific, mathematical, and engineering settings. Statisticians concern themselves with subset selection in regression [1], approximation theorists address m -term approximation of functional spaces [2], [3], and electrical engineers use sparse representations for the compression and analysis of audio, image, and video signals [4], [5], [6], [7], [8], [9], [10]. Sparsity criteria also arise in deconvolution [11], signal modeling [12], machine learning [13], denoising [14], and regularization [15]. While each application calls for a slightly different problem formulation, the overall goal is to identify a good approximation involving a few elementary signals—a *sparse* approximation. Sparse approximation problems have two characteristics.

- The signal vector is approximated with a linear model of elementary signals (drawn from a fixed collection). This collection is large and linearly dependent.
- We seek a compromise between approximation error (usually measured with Euclidean norm) and the number of elementary signals in the linear combination.

The main focus of attention has been on algorithms for solving the variants of sparse approximation, as well as the geometric properties of overcomplete dictionaries that guarantee success of these algorithms.

Communications is one natural application area for sparse approximation which has received little attention. This is surprising as there is a natural interpretation of the dictionaries as spherical codes. The geometric properties of these dictionaries that make feasible efficient algorithms are the same geometric

features that make these dictionaries good codes. Furthermore, the problem of finding a short linear combination of codewords that lies close to a given input signal is just one level of abstraction away from the problem of classical decoding.

We present two natural and direct applications of sparse approximation in communications. The first application is a classical decoding problem for the Gaussian channel with a twist on the codebook; we use the notions in sparse approximation to enhance the base codebook and to build an exponentially larger codebook. The algorithms for sparse approximation provide provable decoding algorithms for this problem. Second, we build a simultaneous sparse approximation problem and show that it arises in MIMO communication channels. Again, the algorithms for simultaneous sparse approximation present provable decoding algorithms.

In the first section we review sparse approximation and the algorithms for solving the m -term problem. Next, we define enhanced codebooks and show how sparse approximation is simply decoding over an enhanced codebook. We give two simple examples to illustrate our problem. For the second application, we define simultaneous sparse approximation, show how one version of it is equivalent to decoding MIMO channels, and then present several experiments to demonstrate the effectiveness of sparse approximation.

II. SPARSE APPROXIMATION

To set the stage for sparse approximation, we define its essential characters in this section. We state formally the one variant which we apply to communications and, finally, we give the algorithms which solve this problem variant provably well.

The signal vector s is an element of the signal space \mathbb{C}^d endowed with the complex inner product. A dictionary for the signal space is a finite collection \mathcal{D} of unit-norm elementary signals. These elementary signals are called atoms, and each atom is denoted by φ_ω , where the parameter ω ranges over an index set Ω . It is clear that $N = |\mathcal{D}| = |\Omega|$ denotes the size of the dictionary or the number of atoms it contains. From the dictionary, we form a matrix whose ω -th column is the vector φ_ω . We denote this matrix Φ and refer to this matrix as the dictionary synthesis matrix.

There are many ways to describe the geometry of the dictionary \mathcal{D} . One useful approach is to examine how the atoms are correlated with each other. To that end, we define the coherence μ of the dictionary as

$$\mu = \max_{\lambda \neq \omega} |\langle \varphi_\lambda, \varphi_\omega \rangle|.$$

A less pessimistic measure and a generalization is the cumulative coherence function $\mu_1(\cdot)$. It is defined for positive integers m by the formula

$$\mu_1(m) = \max_{|\Lambda| \leq m} \max_{\omega \notin \Lambda} \sum_{\lambda \in \Lambda} |\langle \varphi_\omega, \varphi_\lambda \rangle|$$

where the index set $\Lambda \subset \Omega$. The cumulative coherence measures the maximum total correlation between a fixed atom and m distinct atoms. It can be bounded above by the coherence as $\mu_1(m) \leq m\mu$. The definition of coherence and its importance are discussed in [16], [17], [18], [19], [20].

A representation of a signal is a linear combination of atoms that equals the signal. Every representation is parameterized by a list of coefficients that we collect into a coefficient vector, which formally belongs to \mathbf{C}^Ω . Given a coefficient vector c , observe that the product $s = \Phi c$ yields a signal vector as

$$s = \Phi c = \sum_{\omega \in \Omega} c_\omega \varphi_\omega.$$

In other words, the signal is synthesized as a linear combination of atoms whose coefficients are given by the vector c .

A sparse approximation problem seeks an approximation that can be represented with a low cost. There are many ways to calculate this cost. One natural method is to charge a representation for the number of atoms that participate. We define the sparsity of a coefficient vector to be the number of nonzero entries in the vector. We calculate the sparsity via the ℓ_0 quasi-norm $\|\cdot\|_0$, defined as

$$\|c\|_0 = |\text{supp}(c)| = |\{\omega \in \Omega | c_\omega \neq 0\}|.$$

While there are many variants of sparse approximation, we stick with one particular formulation, known in the approximation theory community as m -term approximation. Using a linear combination of m atoms or fewer from the dictionary, find the best approximation of the target signal.

$$\min_{c \in \mathbf{C}^\Omega} \|s - \Phi c\|_2 \quad \text{subject to} \quad \|c\|_0 \leq m. \quad (\text{SPARSE})$$

There are a number of algorithms which solve the m -term sparse approximation problem. They are Orthogonal Matching Pursuit (OMP) and Basis Pursuit (an algorithm that falls under the rubric of convex relaxation). In keeping with the spirit of the analysis of approximation algorithms, we refer to the optimal coefficient vector as c_{opt} , the optimal approximation a_{opt} , and we index the sub-dictionary of vectors that make up a_{opt} as Λ_{opt} . Let assume that the number of vectors m in our optimal solution is less than $1/(3\mu)$. The following theorem tells us that OMP is an approximation algorithm for SPARSE.

Theorem 1: Assume that $\mu_1(T) < 1/2$. After T iterations, OMP will produce an approximation a_t that satisfies the error bound

$$\|s - a_T\|_2 \leq \left(1 + T \frac{1 - \mu_1(T)}{(1 - 2\mu_1(T))^2}\right)^{1/2} \|s - a_{\text{opt}}\|_2.$$

Observe that if the optimal solution a_{opt} captures signal completely (i.e., the error $\|s - a_{\text{opt}}\|_2 = 0$), then OMP will recover the optimal solution exactly. We can also solve SPARSE by proxy with convex relaxation but we do not go into details here. We also know that the average case problem is solvable by convex relaxation [21], [22] and preliminary numerical evidence suggests the same is true of OMP [23].

III. ENHANCED CODEBOOKS

The m -term approximation problem, at first glance, looks like a classical decoding problem except we want to find the closest (in Euclidean distance) m -term linear combination of codewords to the received vector rather than the closest codeword. In fact, we need only adjust our perspective to build a new decoding problem.

Let Φ be the base codebook of size $|\Phi| = N$. We assume that Φ is a spherical code that spans \mathbf{C}^d and that the coherence of Φ is μ . Let \mathcal{C} be a packing in \mathbf{C}^m of either the unit sphere or the unit cube. The size of the packing we denote by $|\mathcal{C}|$. We generate an enhanced codebook (\mathcal{C}, Φ, m) by forming all m -term linear combinations of codewords in Φ and using the points in the packing \mathcal{C} as coefficients. More formally, each codeword s in (\mathcal{C}, Φ, m) is of the form

$$s = \sum_{j=1}^m c_j \varphi_{\omega_j} \quad \text{where } c = (c_1, \dots, c_m) \in \mathcal{C} \text{ and } \varphi_{\omega_j} \in \Phi.$$

The size of the enhanced codebook is $\binom{N}{m} \cdot |\mathcal{C}|$ which is exponentially larger than the size of the base codebook. Observe that we can scale the packing appropriately to meet or to achieve any power constraints. A second variation of this construction is to use linear combinations of no more than m codewords, rather than exactly m codewords.

Let us assume that our base codebook has relatively low coherence, that $\mu_1(m) < 1/3$. Suppose that our transmitted signal consists of $x \in (\mathcal{C}, \Phi, m)$ plus white Gaussian noise w ; i.e., $s = x + w$. We also assume that the noise conforms to a hard power constraint $\|w\|_2 < \nu$. Then, our theoretical results guarantee that OMP will find a codeword $a \in (\mathcal{C}, \Phi, m)$ that is within a factor of $(\sqrt{1 + 6m})\nu$ of the transmitted codeword x . In this form, these bounds do not promise recovery of x (just approximate recovery); rather, they raise intriguing questions about the geometry of the enhanced codebook and its dependence on the base codebook and the packing. Preliminary evidence [23] suggests that, in the presence of no or little noise, for most codewords x we can exactly decode s and recover x using OMP and [22] demonstrates that convex relaxation decodes most of the codewords in the enhanced codebook without any added noise.

Let us illustrate our enhanced codebook with two examples. In Figure 1 we start with the canonical basis (and the negatives

of the basis vectors e_1 and e_2) in \mathbf{R}^2 . Note that the packing radius of this base codebook is $r = \sqrt{2}/2$, that we can transmit 1 bit per time unit with this code, and that the average power (or squared norm of each codeword) is 1. In Figure 2 we take as a different base codebook that formed by two unit vectors v_1 and v_2 which are separated by an angle of $\pi/3$. Note that this base codebook has a larger coherence $\mu = 1/2$ than the canonical basis (which has a coherence of zero). Furthermore, the packing radius of this code is less $r = 1/2$. We can transmit 1 bit per time unit and the average power per codeword is also 1, as in the canonical case.

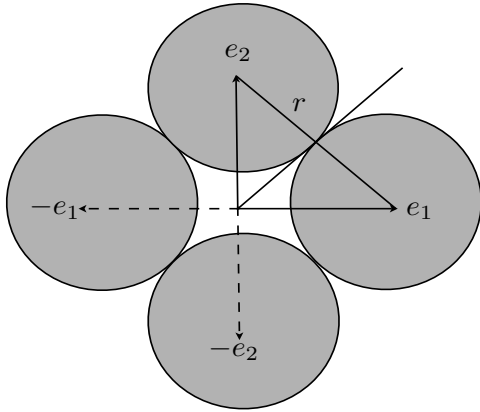


Fig. 1. Canonical basis as base codebook.

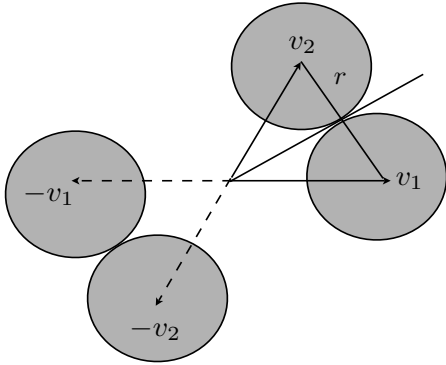


Fig. 2. Two unit vectors separated by a $\pi/3$ angle as base codebook.

In Figures 3 and 4 we demonstrate the enhanced codebooks obtained from the canonical ones by taking linear combinations of the base codevectors. In this case, we allow the coefficients $\{1, -1\}$ only. These choices correspond to the set of all possible coefficients $\mathcal{A} = \{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$ which form a packing (in the unit square). In the canonical basis example, the enhanced codebook consists of eight codewords each with average power $(4 \cdot 1 + 4 \cdot 2)/8 = 3/2$ per

codeword. The packing radius decreases from $r = \sqrt{2}/2$ to $r_e = 1/2$ and we can transmit $3/2$ bits per time unit with this code. In the second case, the enhanced codebook also consists of eight codewords, each with an average power of $(6 \cdot 1 + 2 \cdot 3)/8 = 3/2$ per codeword. We observe that both the base codebook and the enhanced version are not symmetric; there are more codewords clustered about the diagonal. We can expand these two examples by taking packings of coefficients in larger cubes centered at the origin. All of the resulting enhanced codebooks have the same structural relationships as the simple examples we illustrate above. The rates of the two enhanced codebooks are equal, as are the average powers. The hexagonal enhanced codebook retains its asymmetry with the same fixed aspect ratio.

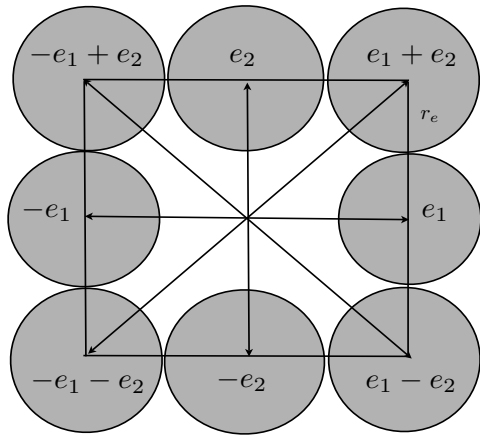


Fig. 3. Enhanced canonical basis using combinations of \pm canonical vectors.

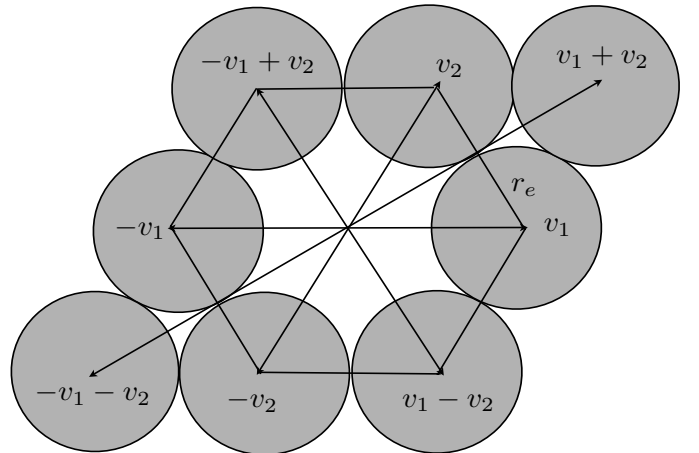


Fig. 4. Enhanced codebook using combinations of \pm vectors separated by a $\pi/3$ angle.

IV. SIMULTANEOUS SPARSE APPROXIMATION

One generalization of simple sparse approximation has received some attention in the literature [24], [23], [25], [26].

Suppose that we have several observations of a signal that has a good sparse approximation. Each view is contaminated with noise, which need not be statistically independent. It seems clear that we should be able to use the additional samples to produce a superior estimate of the underlying signal. This intuition leads to the simultaneous sparse approximation problem. Given several input signals, approximate all these signals at once using different linear combinations of the same elementary signals, while balancing the error in approximating the data against the total number of elementary signals that are used.

In this case a signal S is a matrix of K columns, each column has d rows. Given a coefficient matrix C , the matrix product $S = \Phi C$ yields the signal matrix. That is, the k -column of the signal matrix is synthesized as a linear combination of atoms whose coefficients are listed in the k -th column of the coefficient matrix,

$$\mathbf{s}_k = \Phi \mathbf{c}_k = \sum_{\omega \in \Omega} c_{\omega k} \varphi_{\omega}$$

for each k . We measure the approximation error with the Frobenius norm and constrain the sparsity to a total of m atoms. We extend both SPARSE to the simultaneous setting and, concomitantly, we modify each of the algorithms OMP and convex relaxation appropriately for the simultaneous sparse approximation problem [25], [23]. Both algorithms return not only the m atoms φ_{ω} in the sparse representation but also estimate the coefficient matrix C .

We have the following provable result for the simultaneous OMP (S-OMP) algorithm.

Theorem 2: Assume that $\mu_1(m) < \frac{1}{2}$. After m iterations, S-OMP will produce an approximation A_m with m atoms that satisfies the error bound

$$\|S - A_m\|_F^2 \leq \left[1 + \frac{K m (1 - \mu_1(m))}{(1 - 2 \mu_1(m))^2} \right] \|S - A_{\text{opt}}\|_F^2. \quad (1)$$

In words, S-OMP is an approximation algorithm for SPARSE. We have a similar result for convex relaxation but do not state it as we use S-OMP only in the sequel.

V. MIMO CHANNELS

One mathematical view of slowly varying wireless channels is as a sparse approximation problem. Consider the MIMO channel $\mathbf{S} = H\mathbf{x} + \mathbf{w}$ with m transmit antennas and K receive antennas. At each symbol time, $\mathbf{x} \in \mathbf{C}^m$, $\mathbf{y} \in \mathbf{C}^K$, and \mathbf{w} is a complex Gaussian vector in \mathbf{C}^K . We take the matrix $H \in \mathbf{C}^{K \times m}$ comprised of i.i.d. circular symmetric complex Gaussian entries so that our MIMO model is consistent with the i.i.d. Rayleigh fading model. We use a block length d and assume that the channel properties do not change over the block.

For each block, the received signal in the MIMO channel is a matrix of K signals, each of which is a (random) linear combination of m atoms corrupted by additive white Gaussian noise. Each transmit antenna sends one codeword

φ_{ω_j} . More formally, each row \mathbf{s}_k of S is synthesized as a linear combination of codewords plus noise

$$\mathbf{s}_k = \sum_{j=1}^m H_{jk} \varphi_{\omega_j} + \mathbf{w}_k.$$

It is a trivial matter to transpose the definition of the simultaneous sparse approximation problem to reflect the MIMO configuration. Note that the S-OMP algorithm is agnostic about the type of coefficient matrix H ; it need not be random, its only feature is that it is unknown. The S-OMP algorithm does assume that the receive antennas are coordinated in some fashion in that S-OMP requires full knowledge of each receive antenna's signal.

We perform several numerical experiments to validate our approach. We take as our codebook the (complex) Kerdock code in 8 dimensions. See [27], [28] for details. This code is a particularly useful code in that it consists of 64 codewords (up to multiplication by a power of i) in dimension 8 and it has a coherence of $\mu = 1/2\sqrt{2}$. We build a received signal matrix S as follows. We choose m atoms at random and form a linear combination with random coefficients H_{jk} i.i.d. complex Gaussian. To these combinations we add white complex Gaussian noise scaled appropriately to achieve the desired SNR. For these experiments, we vary m from 2 to 7; we vary K from 1 to 8; and we examine SNR values of 10, 11.2, 13, 16, and 20 dB. For each parameter set, we perform 1000 independent trials and take an average over these trials.

Because the S-OMP algorithm returns both the (estimated) transmitted codewords and the (estimated) channel coefficients, we examine the performance of algorithm on both outputs. To determine how well S-OMP does in recovering the transmitted codewords, we compute the Hamming distance between the set of recovered atoms and the core set. (Hamming distance zero means that we recover the entire set, while distance one means that we fail to recover any of the core atoms.) We do not strive to measure how far the recovered codewords are from the transmitted ones, only that we have failed. In Figure 5, we plot the average Hamming distance (over 1000 trials) as a function of the number of receive antennas for different SNR levels. In this experiment, the number of transmit antennas is fixed at four. Observe that as we add more receive antennas, we are able to recover more of the transmitted codewords. Using only two antennas, we are able to recover three out of four codewords (except at the lowest SNR levels).

In a second experiment, we fix the number of receive antennas at two and test the channel estimation performance of the S-OMP algorithm. We compute the average error (in Frobenius norm) that each receive antenna makes in estimating the channel parameters (the entries in the matrix H). In Figure 6 we plot the average error as a function of the number of transmit antennas at different SNR levels. As the number of transmit antennas increases, the average error actually rises and then decreases until the transmit antennas number seven. With more than five transmit antennas, some averaging effects

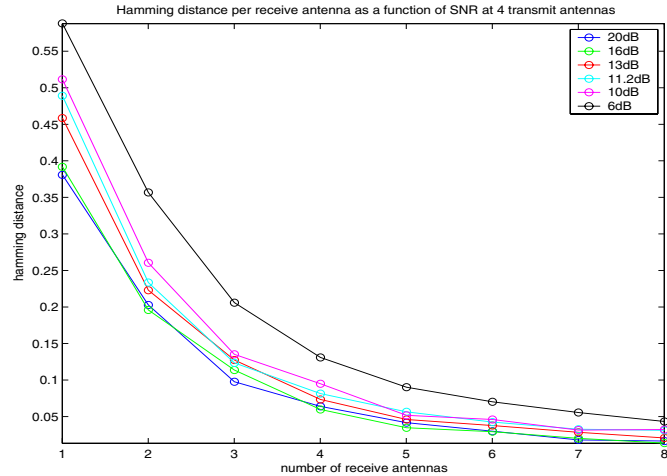


Fig. 5. Using four transmit antennas, the Hamming distance between the recovered set of codewords and the transmitted set, as a function of the number of receive antennas for different SNR levels.

may aid in the channel estimation as each receive antenna gathers more information about the channels.

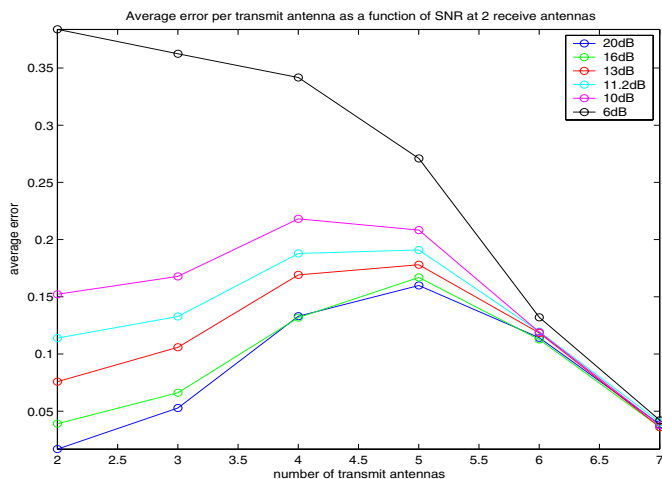


Fig. 6. With two fixed receive antennas, the average channel estimation error as a function of the number of transmit antennas at different SNR levels.

VI. CONCLUSION

We introduce two applications of sparse approximation in communications. Our enhanced codebook construction provides a host of alternate codes based upon well-studied ones and suggests interesting geometric coding theory questions. In addition, we show that the simultaneous versions of sparse approximation arise naturally in MIMO wireless channels. This framework offers potential for MIMO systems. We demonstrate that S-OMP is a useful decoding algorithm.

ACKNOWLEDGMENT

A.C.G. is supported by NSF DMS-0354600. The authors thank Robert Calderbank and Martin Strauss for helpful conversations.

REFERENCES

- [1] A. J. Miller, *Subset selection in regression*, Chapman and Hall, London, 2nd edition, 2002.
- [2] R. A. DeVore, "Nonlinear approximation," *Acta Numerica*, pp. 51–150, 1998.
- [3] Vladimir Temlyakov, "Nonlinear methods of approximation," *Foundations of Computational Math.*, 2002.
- [4] S. Mallat and Z. Zhang, "Matching pursuits with time frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [5] G. Davis, S. Mallat, and M. Avellaneda, "Greedy adaptive approximation," *Journal of Constructive Approximation*, vol. 13, pp. 57–98, 1997.
- [6] L. F. Villemoes, "Best approximation with walsh atoms," *Constr. Approx.*, vol. 13, pp. 329–355, 1997.
- [7] Z. Zhang, *Matching Pursuit*, Ph.D. thesis, New York University, 1993.
- [8] R. Gribonval E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Processing*, pp. 101–111, 2003.
- [9] P. Frossard P. Vandergheynst R. M. Figueras i Ventura M. Kunt, "A posteriori quantization of progressive matching pursuit streams," *IEEE Trans. Signal Processing*, pp. 525–535, 2004.
- [10] T. Nguyen A. Zakhor, "Matching pursuits based multiple description video coding for lossy environments," *ICIP 2003*, 2003.
- [11] H. L. Taylor S. C. Banks J. F. McCoy, "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 1, no. 44, pp. 39–52, 1979.
- [12] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1979.
- [13] F. Girosi, "An equivalence between sparse approximation and support vector machines," *Neural Comput.*, pp. 1455–1480, 1998.
- [14] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999, electronic.
- [15] Ingrid Daubechies, Michel DeFRise, and Christine De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications in Pure and Applied Math*, vol. 57, pp. 1413–1457, 2004.
- [16] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *The 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, Jan. 2003.
- [17] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," In preparation.
- [18] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, Nov. 2001.
- [19] D. L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization," Draft, Dec. 2002.
- [20] J. A. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," ICES Report 04-04, The University of Texas at Austin, 2004.
- [21] D. Donoho, "For most large underdetermined systems of linear equations, the minimal l^1 -norm near-solution approximates the sparsest," 2004.
- [22] D. Donoho, "For most large underdetermined systems of linear equations, the minimal l^1 -norm solution is the sparsest," 2004.
- [23] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation part i: Greedy pursuit," *submitted*, 2004.
- [24] D. Leviatan and V. N. Temlyakov, "Simultaneous approximation by greedy algorithms," IMI Report 2003:02, Univ. of South Carolina at Columbia, 2003.
- [25] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous approximation by greedy pursuit," in *ICASSP 2005*, 2005.
- [26] Xiaoming Huo, "private communication," 2005.
- [27] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, Elsevier/North-Holland, 1977.
- [28] A. R. Hammons Jr. P. Vijay Kumar A. R. Calderbank N. J. A. Sloane P. Sole, "The z_4 -linearity of kerdock, preparata, goethals and related codes," *IEEE Trans. Inform. Theory*, pp. 301–319, 1994.