KNOWLEDGE-ENHANCED MATCHING PURSUIT

Yuejie Chi

Electrical and Computer Engineering
The Ohio State University
chi@ece.osu.edu

Robert Calderbank

Electrical and Computer Engineering
Duke University
robert.calderbank@duke.edu

ABSTRACT

Compressive Sensing is possible when the sensing matrix acts as a near isometry on signals of interest that can be sparsely or compressively represented. The attraction of greedy algorithms such as Orthogonal Matching Pursuit is their simplicity. However they fail to take advantage of both the structure of the sensing matrix and any prior information about the sparse signal. This paper introduces an oblique projector to matching pursuit algorithms to enhance detection of a component that is present in the signal by reducing interference from other candidate components based on prior information about the signal as well as the structure of the sensing matrix. Numerical examples demonstrate that performance as a function of SNR is superior to conventional matching pursuit.

Index Terms— matching pursuit, sparsity, support recovery, oblique projection

1. INTRODUCTION

Compressive Sensing (CS) [1] employs algorithms such as ℓ_1 minimization [2] or greedy pursuits [3] to recover signals that are sparse or compressible. The attraction of greedy algorithms is their simplicity, but the performance is strongly dependent on the worst case coherence of the sensing matrix (the maximal off-diagonal entry in the Gram matrix).

There are important applications such as microscopy [4] and high resolution spectrum analysis where the sensing matrix is determined by the point spread function of the underlying physics and cannot be optimized to satisfy near-isometry properties. In these applications, Orthogonal Matching Pursuit (OMP) [3] is known to perform poorly [5]. There is however a long tradition in radar receiver design of using a mismatched filter [6, 7] rather than a matched filter to modulate the ambiguity function of a transmitted waveform. A recent example is a passive radar for which the illuminating waveform is the DVB-T television signal [8]. It is natural to connect the Gram matrix in sparse recovery with the ambiguity function in radar signal processing. Motivated by this connection, we propose a new algorithm called Oblique Matching Pursuit (ObMP) where an oblique projector manages the Gram of the sensing matrix.

We also propose to optimize the oblique projector to take advantage of signal priors. We note that it is common practice in radar signal processing to optimize the waveform so the sidelobes of the ambiguity function are suppressed where the target is most likely to occur [9]. We optimize the oblique projector in ObMP to shape the expected Gram of the sensing matrix, for example to suppress the

off-diagonal entries in the Gram matrix with higher nonzero probabilities. However, the signal-to-noise ratio (SNR) is traded off by not correlating with the original sensing matrix, in order to obtain gain in desired interference suppression, which is confirmed in both theoretical analysis and numerical examples.

Note that while Lee et al. [10] also introduce oblique projectors, their focus is preservation of the restricted isometry property, whereas our focus is on coherence properties. More importantly, we propose to find an oblique projector via optimization which allows incorporation of prior information about the signal and the structure of the sensing matrix.

The rest of the paper is organized as follows. Section 2 presents the framework where an oblique projector is introduced in matching pursuit. Section 3 discusses optimization of the oblique projector with respect to the structure of the sensing matrix and the signal prior. Section 4 provides numerical examples and we conclude in Section 5.

2. FRAMEWORK

Given a measurement vector
$$z = X\beta + \eta \in \mathbb{C}^n, \tag{1}$$

where X is an $n \times p$ unit-column sensing matrix, $\beta \in \mathbb{C}^p$ is a k-sparse vector, and $\eta \in \mathbb{C}^n \sim \mathcal{N}(0, \sigma^2)$ is the white additive Gaussian noise, our goal is to reconstruct the sparse vector β from noisy measurements z. If the support of β is correctly detected, the signal β can be reconstructed by least-squares estimation using the restricted sensing matrix with columns corresponding to the nonzero entries, hence we restricted ourselves to support recovery, i.e. finding the support of β.

We now introduce the oblique operator Y with respect to X, which is a second $n \times p$ matrix, also with unit column norms. Algorithm 1 and Algorithm 2 describe the proposed Sorted Oblique Thresholding (SObT) and Oblique Matching Pursuit (ObMP) algorithms. When $Y = X$, they are equivalent to the conventional one-step thresholding (OST) [11] and OMP respectively.

Algorithm 1 Sorted Oblique Thresholding (SObT)

1: $f := Y^H z$;
2: Denote the ordering of the entries in $|f|$ by $\mathcal{I}$;
3: $\hat{\mathcal{S}} := \mathcal{I}(1 : k)$.

Instead of correlating the received signal with X, it is now correlated with Y in Algorithm 1 and 2, yielding the output
$$f = Y^H z = Y^H X\beta + Y^H \eta \triangleq G\beta + Y^H \eta, \tag{2}$$

Algorithm 2 Oblique Matching Pursuit (ObMP)

1: Initialization: $\hat{\mathcal{S}}_0 :=$ empty set, residual $r_0 = z$
2: **for** $t := 1 : k$ **do**
3:     $f := Y^H r_{t-1}$
4:     $i := \arg\max_j |f_j|$
5:     $\hat{\mathcal{S}}_t := \hat{\mathcal{S}}_{t-1} \cup \{j\}$
6:     $r_t := z - X_{\hat{\mathcal{S}}_t}(Y_{\hat{\mathcal{S}}_t}^H X_{\hat{\mathcal{S}}_t})^{-1} Y_{\hat{\mathcal{S}}_t}^H z$
7: **end for**
8: $\hat{\mathcal{S}} := \hat{\mathcal{S}}_t$

---

where $G = Y^H X = [g_{ij}]$ is the new Gram matrix, and G is not Hermitian in general. Algorithm 2 further assumes an oblique projection in step 6, so that the residual at the tth step r_t is orthogonal to the selected columns in $Y_{\hat{\mathcal{S}}_t}$, i.e. $Y_{\hat{\mathcal{S}}_t}^H r_t = 0$. This guarantees a new atom is selected in the next step.

3. KNOWLEDGE-ENHANCED MATCHING PURSUIT

Since X and Y have unit norm columns, any diagonal entry of G satisfies $|g_{ii}| = |\langle y_i, x_i \rangle| = |y_i^H x_i| \leq 1$. We single out the smallest diagonal entry of G by defining

$$\alpha(X, Y) = \min_i |g_{ii}|;$$

and single out the largest off-diagonal entry of G by defining the worst case mutual coherence

$$\mu(X, Y) = \max_{i \neq j} |g_{ij}|.$$

When $X = Y$ we denote the worst-case mutual coherence by $\mu(X)$, which degenerates to the coherence of X.

These two quantities are sufficient to provide the performance guarantees given in Theorem 1 below.

Theorem 1. *The SObT algorithm for the measurement in* (1) *succeeds with probability at least* $1 - (\pi p)^{-1}$ *as long as*

$$\beta_{\min} \geq \frac{2\mu(X, Y)}{\alpha(X, Y)}\|\beta\|_1 + \frac{2\sigma\sqrt{\log p}}{\alpha(X, Y)}, \tag{3}$$

where $\beta_{\min} = \min_i |\beta_i|$ *the minimum nonzero entry of* β.

Proof: We first use Proposition 5 in [12] which shows with probability at least $1 - (\pi p)^{-1}$, the noise is bounded as

$$\|Y^H \eta\|_\infty \leq \sigma\sqrt{2\log p}. \tag{4}$$

Define the event when (4) happens as $\mathcal{G}$. Let Π be the support of β, and $\Pi^c = \{1, \cdots, p\} \backslash \Pi$. Under $\mathcal{G}$, for $i \in \Pi$,
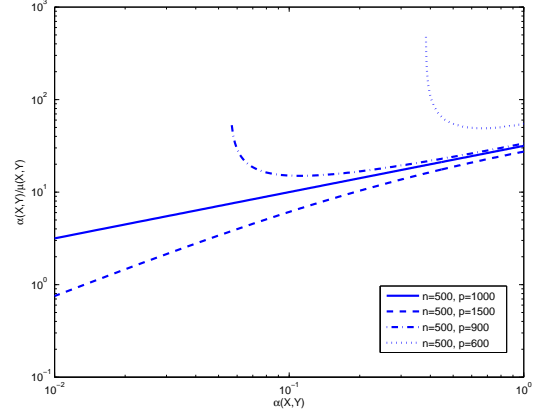
$$|y_i^H z| = |y_i^H x_i \beta_i + \sum_{j \neq i, j \in \Pi} y_i^H x_j \beta_j + y_i^H \eta|$$
$$\geq \alpha(X, Y)|\beta|_{\min} - \mu(X, Y)\|\beta\|_1 - \sigma\sqrt{2\log p}, \tag{5}$$

and for $i \in \Pi^c$,

$$|y_i^H z| = |\sum_{j \in \Pi} x_i^H x_j \beta_j + x_i^H \eta|$$
$$\leq \mu(X, Y)\|\beta\|_1 + \sigma\sqrt{2\log p}, \tag{6}$$

The SObT algorithm succeeds when

$$\min_{i \in \Pi} |y_i^H z| \geq \max_{i \in \Pi^c} |y_i^H z|, \tag{7}$$

Fig. 1. $\alpha(X, Y)$ versus $\alpha(X, Y)/\mu(X, Y)$ under the special case when $g = g_v(1 + \epsilon r)$, where r has an even number of 1's and -1's.

which is equivalent to (3) combining (5) and (6). $\qquad\square$

Since $\alpha(X, Y) \leq 1$, the second term in (3) is larger than for OST, so we lose noise resilience. Consider the first term in (3), and assume that the worst-case coherence $\mu(X)$ of X is on the order of $n^{-1/\gamma}$ where $\gamma \geq 2$ from the Welch bound. When $\gamma > 2$, it is possible to make $\mu(X, Y)/\alpha(X, Y)$ much smaller than $\mu(X)$, therefore we gain in interference suppression. This trade-off can be quantified using the following generalized Welch bound.

Theorem 2 (Generalized Welch bound). *Let X and Y be unit-column matrices of dimension $n \times p$, and $G = Y^H X = [g_{ij}]$, then*

$$\max_{i \neq j} |g_{ij}|^2 \geq \frac{1}{p(p-1)} g^T\left(\frac{J}{n} - I\right)g \tag{8}$$

where g is the vector composed of the diagonal terms of G, and J is the $p \times p$ matrix with every entry 1.

Proof: The Frobenius norm of G is given as

$$\|G\|_F^2 = \sum_{i=1}^p \sum_{j=1}^p |\langle x_i, y_j \rangle|^2 = \sum_{i=1}^p \sum_{j=1}^p |g_{ij}|^2 = \sum_{i=1}^p \lambda_i^2,$$

where λ_i's are singular values of G. We write the trace of G as

$$\text{tr}\{G\}^2 = g^T J g = \left(\sum_{i=1}^r \lambda_i\right)^2 \leq r\sum_{i=1}^r \lambda_i^2 \leq n\sum_{i=1}^p \lambda_i^2,$$
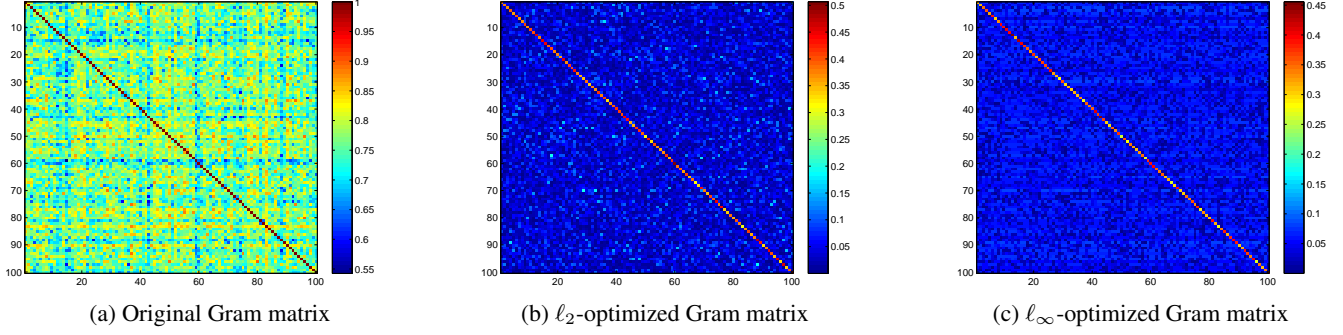
where the first inequality follows from the Cauchy-Schwartz inequality, and r is the number of non-zero singular values which is upper bounded by n. Therefore

$$\|G\|_F^2 = g^T g + \sum_{i \neq j} |g_{ij}|^2 \geq \frac{\text{tr}\{G\}^2}{n} = \frac{1}{n}g^T J g,$$

then we have

$$\max_{i \neq j} |g_{ij}|^2 \geq \frac{1}{p(p-1)}\sum_{i \neq j}|g_{ij}|^2 \geq \frac{1}{p(p-1)}g^T\left(\frac{J}{n} - I\right)g.$$

$\qquad\square$

Discussions: When $p \leq n$, we may force $\mu(X, Y) = 0$ by choosing Y to be the column normalized version of $X^\dagger = X(X^H X)^{-1}$, but the cost is reduction in $\alpha(X, Y)$ that may result in diminished performance.

| (a) Original Gram matrix | (b) ℓ_2-optimized Gram matrix | (c) ℓ_∞-optimized Gram matrix |

Fig. 2. The original and optimized Gram matrix: (a) original, (b) ℓ_2 optimized and (c) ℓ_∞ optimized.

| (a) SNR = 20dB | (b) SNR = 30dB | (c) SNR = 40dB |
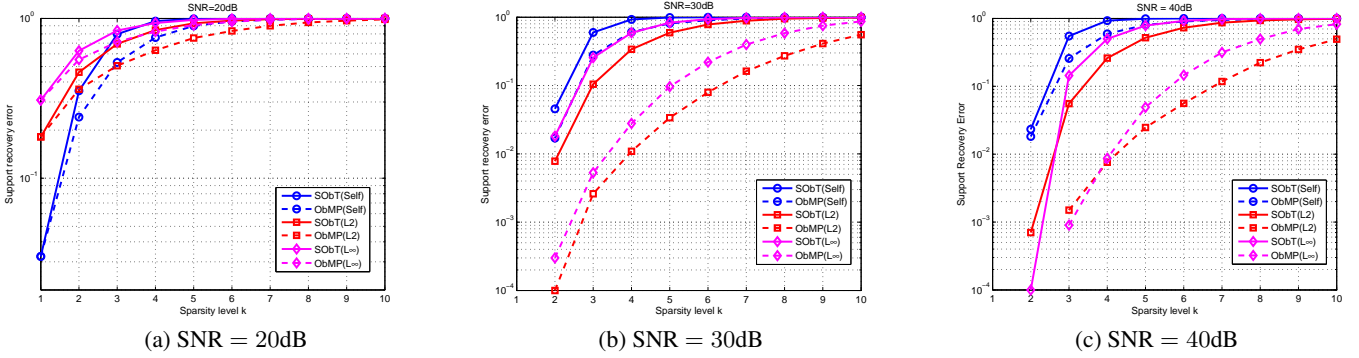
Fig. 3. The probability of error for support recovery with respect to the sparsity level of different optimized oblique projectors under (a) SNR = 20dB, (b) SNR = 30dB and (c) SNR = 40dB.

When $p > n$, we consider a special case using Theorem 2 to exemplify the trade-off between $\alpha(X, Y)/\mu(X, Y)$ and $\alpha(X, Y)$. Choose Y such that g can be written as $g = g_v(1 + \epsilon r)$, where r is an vector with even number of 1's and -1's, so $1^T r = 0$. Moreover, $\alpha(X, Y)$ is upper bounded by $\alpha(X, Y) = g_v(1 - \epsilon) \leq \frac{1-\epsilon}{1+\epsilon}$, and from Theorem 2 we have

$$\max_{i \neq j} |g_{ij}|^2 \geq \frac{1}{p(p-1)} g_v^2 (\frac{p^2}{n} - p - \epsilon^2 p) = \frac{g_v^2(p - n - n\epsilon^2)}{n(p-1)},$$

Hence $\alpha(X, Y)/\mu(X, Y)$ is upper bounded by

$$\frac{\alpha(X, Y)}{\mu(X, Y)} \leq (1 - \epsilon) \sqrt{\frac{n(p-1)}{p - n - n\epsilon^2}}.$$

Fig. 1 shows the upper bound of $\alpha(X, Y)$ versus the upper bound of $\alpha(X, Y)/\mu(X, Y)$ when $n = 500$ and various p. An interesting phase transition occurs when $p = 2n$. As the upper bound of $\alpha(X, Y)/\mu(X, Y)$ decreases, the upper bound of $\alpha(X, Y)$ decreases monotonically when $p > 2n$; and it first decreases and then increases dramatically when $p < 2n$, indicating the potential gain of exploring this trade-off at high SNR regime.

3.1. Optimization with respect to the sensing matrix

We propose two methods of optimizing Y, based on minimization of the ℓ_2 distance and the ℓ_∞ distance between the Gram matrix G and the identity matrix respectively, ie.

$$Y = \underset{Y}{\operatorname{argmin}} \ \|Y^H X - I\|_p^2, \tag{9}$$

which can be decomposed for each column y_i of Y as

$$y_i = \underset{y_i}{\operatorname{argmin}} \ \|y_i^H X - e_i^T\|_p^2, \tag{10}$$

where $p = 2$ or ∞, e_i is the ith column of the identity matrix. When $p = 2$, (10) can be regarded as optimization of the spectral norm of $(Y^H X - I)$, and $y_i = (XX^H)^{-1} x_i$. When $p = \infty$, (10) can be regarded as optimization of the worst-case mutual coherence of Y and X, and y_i can be found via linear programming. Then the normalized solution is given as $y_i^* = y_i / \|y_i\|$ for $i = 1, \cdots, p$.

3.2. Optimization with respect to the signal prior

We consider a model for the sparse signal β in which a non-zero entry appears in position i with probability w_i, and $\sum_{i=1}^p w_i = 1$. This model is different from block or group sparsity and is motivated in part by sensing of frequency sparse signals where we have prior information about the transmitted spectrum [13].

It is desirable to optimize the Gram matrix $G = Y^H X$ to match the probability distribution of the signal support, so that the cross-terms corresponding to indices with higher probability are smaller than those corresponding to indices with lower probability. To this end, we propose the following optimization problem, given as

$$Y = \underset{Y}{\operatorname{argmin}} \ \sum_{i=1}^p \sum_{j=1}^p w_j((Y^H X)_{ij} - \delta_{ij})^2, \tag{11}$$

where we assume the prior probability is used directly as the weight.

This can again be decomposed for each column y_i of Y as

$$y_i = \underset{y_i}{\arg\min} \sum_{j=1}^{p} w_j((y_i^H X)_j - \delta_{ij})^2$$

$$= \underset{y_i}{\arg\min} \ (X^H y_i - e_i)^H W_i (X^H y_i - e_i), \qquad (12)$$

where $W_i = \text{diag}\{w_1, \cdots, w_p\}$. Then y_i can be written explicitly as

$$y_i = w_i (X W_i X^H)^{-1} x_i. \qquad (13)$$

Then the normalized column is given as $y_i^* = y_i / \|y_i\|$.

4. NUMERICAL EXAMPLES

4.1. Optimization with respect to sensing matrix

We first consider a case when the design matrix X doesn't satisfy the required near isometry property. Let $n = 40$, $p = 100$, and the design matrix X is generated with i.i.d. uniform random variables in $[0, 1]$. The support of the signal is uniformly sampled with the amplitude of the nonzero entries given as 1, and the noise is generated as $\eta \sim \mathcal{N}(0, \sigma^2)$. The SNR is calculated as $-20 \log_{10} \sigma$ dB. Fig. 2 (a) shows the Gram matrix of the original matrix $G = X^H X$, and Fig. 2 (b) and (c) show the optimized Gram matrix using the oblique operator from (9) for $p = 2$ and $p = \infty$. The optimized Gram matrices have lower cross-terms but the diagonal entries are less than one.
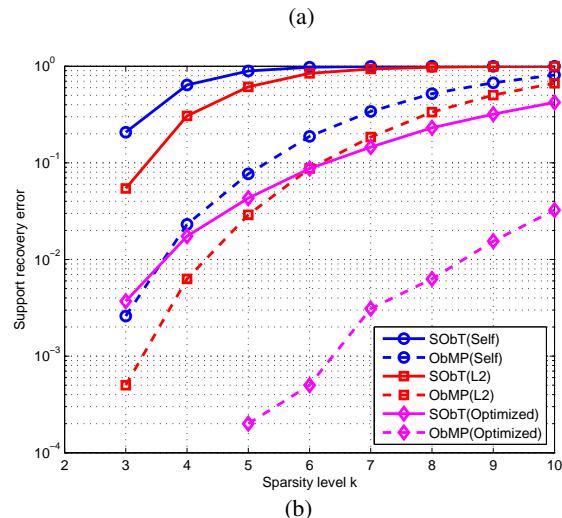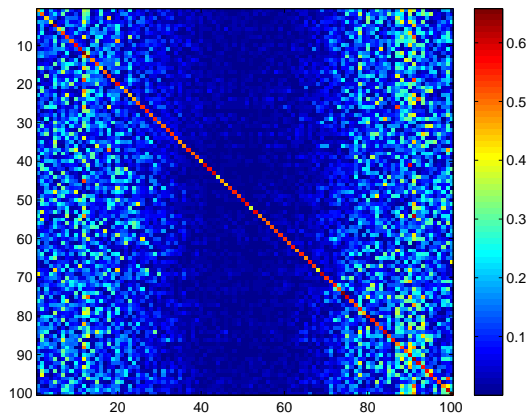
Fig. 3 shows the probability of error for support recovery using Algorithm 1 and Algorithm 2 against the sparsity level when SNR = 20dB, 30dB and 40dB. It is evident that the performance gain is significant in the high SNR regime, as suggested by our discussions of the trade-off of noise resilience and interference suppression. It is also worth noting that ℓ_2-optimized oblique projector outperforms the ℓ_∞-optimized one in most scenarios.

4.2. Optimization with respect to signal prior

In the second example, we assume the sparse signal follows a certain prior distribution. Let $n = 40$, $p = 100$, and the design matrix X is generated with i.i.d. standard Gaussian entries. The probability w_i that the ith entry of the signal β is nonzero is proportional to $e^{-(i-50)^2/2}$, and the amplitude is 1. Fig. 4 (a) shows the optimized Gram matrix obtained via (13), where the cross terms are particularly lower around the center due to higher probability of these entries being nonzero. Fig. 4 (b) shows the probability of error for support recovery using Algorithm 1 and Algorithm 2 against the sparsity level, where the sparse signal is generated with replacement following the above distribution when SNR is 40dB. The optimized oblique operator performs significantly better than the ℓ_2 optimized operator without considering prior information and the original choice.

5. CONCLUSIONS

In this paper we considered optimization of OST and OMP for sparse support recovery based on oblique projections. Motivated by the mismatched filter design in radar, the introduction of an oblique projector optimizes the Gram matrix of the sensing matrix in particular when it does not satisfy the required near isometry properties of CS, and allows explicit exploitations of signal priors into the recovery algorithms. We also explained the trade-off between noise resilience and interference suppression, hence the approach is more favorable in the high SNR regime. Numerical examples are provided to show

(a)

(b)

Fig. 4. When the support is generated following a probability distribution: (a) the optimized Gram matrix; (b) the probability of error for support recovery using oblique projectors.

the proposed approach achieves better performance by exploring the knowledge at hand. Future work includes exploitation of more sophisticated signal priors and extensions to other algorithms.

6. REFERENCES

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

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[3] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[4] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley, "Compressed sensing and electron microscopy," in *Imaging at the Edge - Nanoscale Imaging in Electron Microscopy*. Springer, 2011.

[5] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[6] C. Stutt and L. Spafford, "A 'best' mismatched filter response for radar clutter discrimination," *IEEE Trans. Inf. Theor.*, vol. 14, no. 2, pp. 280–287, 1968.

[7] R. McAulay and J. Johnson, "Optimal mismatched filter design for radar ranging, detection, and resolution," *IEEE Trans. Inf. Theor.*, vol. 17, no. 6, pp. 696–701, 1971.

[8] H. Harms, L. Davis, and J. Palmer, "Understanding the signal structure in DVB-T signals for passive radar detection," in *Proceedings of 2010 IEEE Radar Conference*, May 2010.

[9] Y. Chi, A. Pezeshki, and R. Calderbank, "Complementary waveforms for sidelobe suppression and radar polarimetry," in *Applications and Methods of Waveform Diversity*. SciTech Publishing, 2010.

[10] K. Lee, Y. Bresler, and M. Junge, "Oblique pursuits for compressed sensing," *Arxived*, 2012.

[11] W. U. Bajwa, R. Calderbank, and S. Jafarpour, "Why gabor frames? two fundamental measures of coherence and their role in model selection," *J. Commun. Netw.*, vol. 12, no. 4, pp. 289–307, Aug. 2010.

[12] Y. Chi and R. Calderbank, "Coherence-based performance guarantees of orthogonal matching pursuit," in *Proc. 50th Allerton Conference on Communication, Control and Computing*, Monticello, Illinois, Sep. 2012.

[13] A. Harms, W. U. Bajwa, and R. Calderbank, "A constrained random demodulator for sub-Nyquist sampling," *accepted, IEEE Trans. on Signal Processing*, 2012.