The performance of linear programming techniques that are applied in the signal identification and reconstruction process in compressed sensing (CS) is governed by both the number of measurements taken and the number of nonzero coefficients in the discrete basis used to represent the signal. To enhance the capabilities of CS, we have developed a technique called Variable Projection and Unfolding (VPU). VPU extends the identification and reconstruction capability of linear programming techniques to signals with a much greater number of nonzero coefficients in the basis in which the signals are compressible with significantly better reconstruction error.

1. INTRODUCTION

CS leverages the property that a small number of random linear projections of a sparse signal contain most of the signal’s salient information [1]. To reconstruct the original signal, linear programming is used to find a sparse signal that matches the random linear projections of the original signal. CS linear programming techniques used to reconstruct sparse signals are referred to in the literature as basis pursuit (BP) [2]. Proposed hardware realizations of CS receivers employing BP for backend processing include random RF/IF filters and sub-Nyquist sampling analog-to-digital converters (ADCs) [3] as illustrated in Figure 1, enabling sensors to operate with lower power and higher dynamic range and to transmit or receive raw data at higher information rates.

BP is remarkably efficient in reconstructing signals that occupy a small fraction of the basis in which the signals have a sparse representation. However, we have found that in dense signaling environments, i.e., environments in which the signals occupy more than 15–20 percent of the downsampled basis, BP is unable to effectively reconstruct the original signal. To reconstruct sparse and not-so-sparse signals that occupy up to nearly 100 percent of the downsampled basis, we have developed VPU, which is a maximum likelihood (ML) sequential detection and estimation technique that searches for and identifies the contiguous columns in the basis matrix that the signals span. VPU is similar to orthogonal matching pursuit (OMP) [3]; however, unlike OMP, which searches over rank-1 subspaces to identify the basis support, VPU is capable of searching over subspaces of any rank. In this paper, we present the VPU algorithm and its computational complexity and highlight the results of sensor network simulations comparing VPU and BP performance. Finally, we derive the ML/minimum mean square error (MMSE) bounds and compare them to BP bounds to support the simulation results.

2. VARIABLE PROJECTION AND UNFOLDING

VPU is similar to OMP with the principal exception that the search process is not restricted to identify basis support only over rank-1 subspaces. VPU sequentially searches over rank-$n$ subspaces to identify the basis support. As opposed to correlation, VPU projects the signal onto the subspace spanned by the candidate $n$ columns. If the magnitude of the projection is above a prescribed threshold, the candidate columns are added to the reconstruction basis set, and this set will be used to form new projections when the search commences (there is no subtraction). VPU proceeds by sequentially moving to a new rank-$n$ subspace (e.g., adjacent or overlapping basis columns) and periodically incrementing the subspace size. Unlike OMP, VPU exploits the correlation between coefficients to enable it to significantly outperform either OMP or BP in reconstructing signals that occupy more than 15 per-
Consider a signal \( x \in \mathbb{C}^{N \times 1} \) occupying unknown frequency locations in some bandwidth \( B \) that is digitized at a rate \( 2B/d \), with \( A \equiv \Theta \Psi \) where \( \Psi \in \mathbb{C}^{N \times N} \) is the basis matrix in which \( x \) has a sparse representation and \( \Theta \in \mathbb{R}^{[N/2] \times N} \) is a randomizing downsample matrix \([3]\). The digitized signal can then be described by \( y = A(x + v) = A x + \omega \) where the noise \( v \in \mathbb{R}^{N \times 1} \) and signal \( x \) are normally distributed as \( v \sim \mathcal{N}(0, C_v) \) and \( x \sim \mathcal{N}(\bar{x}, C_x) \), and the symbol \( [z] \) represents the greatest integer that is less than or equal to \( z \). VPU is capable of reconstructing not-so-sparse signals in the presence of noise, where the term ‘not-so-sparse’ refers to signals that occupy up to \( \frac{N}{2} \) elements in the basis in which the signal is sparse. VPU is a sequential estimation and detection process that searches for the columns \( s = \{s_1, s_2, \cdots, s_q\} \) of \( A \) that minimize the reconstruction error, i.e., it finds

\[
\arg \min_{s} \|e_{\text{MMSE}}\|_2^2 = \arg \min_{s} \|(I - A(s)) y\|_2^2, \quad \text{where} \quad A(s) = (A_s)((A_s^H C_v^{-1}(A_s) + C_{x_1})^{-1}(A_s^H C_v^{-1})
\]

(1) where \((\cdot)_s\) represents the columns indexed by \( s \), and \( x_s \) is the vector spanning the columns of the matrix indexed by \( s \). The VPU algorithm is shown in Figure 2. VPU sequentially searches for the columns that minimize in equation (1); the columns of \( A \) that minimize the error in (1) are then fixed and the search continues for other signals present.

1. Set \( \hat{s} = \emptyset, n = 0 \)
2. Solve equation (1) for all \( s = \hat{s} \cup \hat{s}, \) where \( \hat{s} \subseteq \{1,\ldots,N\} \) is a contiguous sequence of integers of the form \((B \cdot j) + \{1,\ldots,B\}\), where \( B \leq \text{max swath} \) and \( B = \min\text{swath} + \beta \cdot i \), for any nonnegative integers \( i \) and \( j \).
3. Update \( \hat{s} \) to be the \( s \) found in step 2.
4. \( n = n + 1 \). If \( n < \text{signals present} \), go to step 2.

![Fig. 2. The VPU algorithm for identifying and reconstructing signals using a coarse search.](image)

Note that the coarseness of the search is controlled by the three parameters: \( \min\text{swath}, \max\text{swath} \) (which is typically set equal to \( \frac{N}{2}\)), and \( \beta \), which controls how many excess locations in the signal basis are used in the detection and estimation process. A large \( \beta \) translates into reduced sensitivity due to the inclusion of excess noise, while a small \( \beta \) improves detection and estimation performance at the expense of computational complexity. Once all the signals are found, a fine search is conducted to excise unnecessary columns that were included as an artifact of the swath size in coarse search.

For cases where the channel coefficients and signal are unknown and for relatively high signal-to-noise ratios (SNRs), equation (1) can be simplified using

\[
A(s) = (A_s)((A_s^H C_v^{-1}(A_s) + C_{x_1})^{-1}(A_s^H C_v^{-1}) \Rightarrow \arg \min_{s} \|e_{\text{MMSE}}\|_2^2 \approx \arg \min_{s} y^T P_{(A_s)}(s)y
\]

(2) where \( P_{(A_s)}(s) = (I - A_s)((A_s^H C_v^{-1}(A_s) + C_{x_1})^{-1}(A_s^H C_v^{-1}) \) is the projector onto the null space of the columns of \((A_s).\) Because the projector in (2) is varied over the columns of the matrix \( A(s), \) the signal is detected and reconstructed via variable projection and unfolding by locating the column positions in the basis matrix that the signal vector spans prior to downsampling and aliasing. Once the column locations that the signal(s) span have been identified, equation (2) can be reformulated for joint synchronous center frequency estimation and baseband signal reconstruction. As an example, consider the simple case of a single signal that is sparse in the frequency basis, then

\[
y = (A(\omega_k))_s \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v \quad \text{given} \begin{bmatrix} \Theta & \Theta \end{bmatrix}^T \begin{bmatrix} \cos \omega_k n \\ -\sin \omega_k n \end{bmatrix} = (\Psi)_s
\]

(3) where the carrier phase offset \( \phi \), baseband signal \( x \) and center frequency \( \omega_k \) are simultaneously recovered using fine search VPU over \( \omega_k \), where the initial estimate of \( \omega_k \) is the frequency associated with center column of \((\Psi)_s \). Extending (3) to handle multiple signals is a straightforward extension of the formulation above.

The computational complexity of VPU is dominated by iterative matrix inversion requiring \( O(k \cdot q^2) \) operations per iteration, with \( k = \sum_{i=1}^{N/\beta} \max\text{swath} - (\beta \cdot i) \cdot \min\text{swath} \) \( \times \) signals present iterations. The computational complexity of VPU is higher than both MP (\( k_B P \cdot N^2 \)) and matching pursuit (MP) O(\( \frac{N}{\beta} \cdot Nq \)) [3], where \( k_B P \) is the number of iterations in the basis pursuit optimization. However, we show that VPU is capable of reconstructing signals that occupy up to \( \frac{N}{2} \) positions of the discrete signal basis while both MP and BP fail far before reaching this limit.

### 2.1. VPU Performance

Two simulation scenarios were used to test the efficacy of VPU and to compare its basis support identification and reconstruction performance against BP. In the first test scenario, VPU and BP detected and reconstructed a random signal without noise that occupied from 8 to 62 contiguous positions in a 520-dimensional frequency basis, i.e., with \( \Theta \in \mathbb{R}^{64 \times 520}, \Psi_{\text{IDFT}} \in \mathbb{C}^{520 \times 520}. \) The second simulation scenario modeled a signal intelligence (SIGINT) Ku-band receiver operating in an environment with one to three binary phase-shift keyed (BPSK) signals present and power levels at the receiver
that ranged from -67 dBm to -77 dBm in a 500 MHz band, yielding roughly a 10 dB-20 dB SNR. The receiver employed an ADC with a sampling rate of 125 MHz; in all cases the aggregate BPSK signal spectral occupancy was 52 MHz. For all scenarios we ran one hundred thousand Monte Carlo simulations and measured the performance of both VPU and BP; for BP we used an $\ell_2-\ell_1$ mixed-norm optimization
\begin{equation}
\arg\min_x \|y - \Theta \Psi_{IDFT} x \|_2^2 + \lambda \|x\|_1,
\end{equation}
where the regularization parameter $\lambda$ was chosen to optimize BP performance [4]. In all cases the randomizing downsampling matrix $\Theta$ had eight unique elements per row that were chosen from an i.i.d. Gaussian distribution.

2.1.1. Test Scenario 1: Identification and Reconstruction of Noiseless Wideband Signal

Identification performance in Table 1 corresponds to the percentage of times in our simulations that BP and VPU correctly identified all of the locations in the frequency basis that that the signal randomly occupied. Note that in all cases VPU was able to simultaneously identify and reconstruct the signal with virtually perfect performance. As predicted in the forthcoming section on MMSE/ML and BP bounds, BP never was able to perfectly identify the frequency support and reconstruct signals that occupied any more than $\frac{1}{4}$ of the downsampled Nyquist band. VPU is able to reconstruct wideband signals that occupied just up to $\left\lfloor \frac{N}{2} \right\rfloor$ positions of the basis.

<table>
<thead>
<tr>
<th></th>
<th>8 loc.</th>
<th>16 loc.</th>
<th>24 loc.</th>
<th>62 loc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>100%</td>
<td>62%</td>
<td>17%</td>
<td>0%</td>
</tr>
<tr>
<td>VPU</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1. BP and VPU identification performance of a noiseless signal occupying 8, 16, 24 and 62 frequency locations in a 520-point frequency basis with a downsampling factor of 8.

2.1.2. Test Scenario 2: SIGINT

The identification, reconstruction and demodulation performance of the SIGINT receiver is illustrated in Figures 3 and 4 below. In Figure 3, we illustrate the basis support misidentification rate of VPU for the cases where one, two or three BPSK signals were present and whose aggregate spectral occupancy prior to the 3 dB pulse-shaping filter roll-off was approximately 52 MHz. Given correct frequency support identification, the measured demodulation bit error rate (BER) after reconstruction was compared to the theoretical optimum performance of an ML soft-decision demodulator with known support. Figure 4 illustrates the difference between theoretical optimum and VPU performance with the gap graphically illustrating the effects of the folding loss, which will be quantified in Section 3. In Figure 4, VPU BER performance is averaged over reconstructing from one to three BPSK signals, while for BP the average is only over one reconstructed BPSK signal; this is because after one signal BP performance was not noticeably better than random selection.

3. CS PERFORMANCE BOUNDS

The following subsections compare the BP and ML/MMSE bound for the identification and reconstruction of sparse and not-so-sparse signals. The analog receiver configuration used for the derivation of the ML/MMSE bound and the description of the BP bound presented below are illustrated in Figure 1, and the details of the hardware implementation are presented in [3].

3.1. CS via Basis Pursuit

In the following paragraphs we will outline the basis pursuit performance bound as first presented in [1] and [5], and then offer an interpretation of the bound in the context of identify-
ing not-so-sparse signals at the output of a downsampling CS receiver.

Consider a sparse signal $x$ occupying unknown frequency locations in some bandwidth $B$ that is digitized at a rate $2B/d$, with $A \equiv \Theta \Psi$ as presented in Section 2. The digitized signal can then be described by

$$y = Ax + \omega,$$

where $\omega \in \mathbb{R}^{\frac{N}{d}} \times 1$ is the noise with power inequality $E\{\omega^T\omega\} \leq \eta$, $E\{\cdot\}$ is the expectation operator and the symbol $\lfloor z \rfloor$ represents the greatest integer that is less than or equal to $z$. In [2] it was demonstrated that it is possible to identify and reconstruct the sparse signal $x$ by solving the basis pursuit (BP) problem formulated as

$$\min \|x\|_1 \text{ s.t. } \|y - Ax\|^2_2 \leq \eta.$$  

(6)

The successful recovery of $x$ by BP is guaranteed [5] if the matrix $A$ obeys the uniform uncertainty principle. Defining $\Lambda \subset \{1, \cdots, N\}$ and $(A)_\Lambda$ to consist of the columns of $A$ indexed by $\Lambda$, then the local isometry constant $\delta_\Lambda(A)$ is the smallest number satisfying

$$\left(1 - \delta_\Lambda(A)\right)\|x_\Lambda\|^2_2 \leq \|(A)_\Lambda x_\Lambda\|^2_2 \leq \left(1 + \delta_\Lambda(A)\right)\|x_\Lambda\|^2_2,$$

(7)

where $x_\Lambda \in \mathbb{C}^{|\Lambda|}$ and the global restricted isometry constant is defined as

$$\delta_q(A) \equiv \sup_{|\Lambda| = q} \delta_\Lambda(A),$$

(8)

with $|\Lambda|$ denoting the cardinality of the set $\Lambda$ and $q$ denoting the number of nonzero values in the vector $x$. One can show that it is possible to recover an estimate of signal $x$ to within a Euclidean distance no greater than $CN[1]$ where $C$ is a constant if

$$\delta_q(A) + 3\delta_q(A) < 2.$$  

(9)

An interpretation of equations (7) through (9) above is that if downsampling $d$ is increased additively by some number $\alpha$, then the recovery of the signal $x$ to within some Euclidean distance $CN$ is now only possibly guaranteed if the maximum number of non-zero coefficients in $x$ span $4\alpha$ fewer positions in the basis $\Psi$. As an example, consider the toy problem of a signal received without noise, i.e., $\omega = 0$, where it should theoretically be possible to perfectly recover the signal $x$ if the conditions as expressed in (9) are met. With $\lambda$ denoting an eigenvalue of the matrix $(A)_\Lambda^H(A)_\Lambda$, then from (7) it is easy to see that

$$\lambda_{\min} \leq \frac{\|(A)_\Lambda x_\Lambda\|^2_2}{\|x_\Lambda\|^2_2} \leq \lambda_{\max},$$

(10)

so that $\delta_q(A)$ is bounded from below by 1 when $4q > \lfloor \frac{N}{d} \rfloor$ (since $\lambda_{\min} = 0$), which violates the inequality of equation (9). In other words, even with no noise present there is no guarantee that BP can identify and reconstruct the signal $x$ if the number of nonzero values in the basis $\Psi$ is greater than $\frac{1}{4} \left\lfloor \frac{N}{d} \right\rfloor$ under the most ideal of conditions. This bound is consistent with empirical BP performance as tabulated in Table 1 in Section 2.

3.2. CS via Maximum Likelihood Techniques (VPU)

In this section we will derive the maximum likelihood bound for VPU processing as presented in Section 2.

Let $s = \{s_1, s_2, \cdots, s_N\}, s^e = \{s_{q+1}, s_{q+2}, \cdots, s_N\}$, where $s_i \in \{1, 2, \cdots, N\}, s_i \neq s_j$, and $A = [(A)_s] (A)_s^T$ where $(A)_s \in \mathbb{C}^{\frac{N}{d}} \times q$ is the matrix whose columns are spanned by the signal(s) in the passband with $\lfloor \frac{N}{d} \rfloor \geq q$. Then the received signal vector can be expressed as

$$y = (A)_s x_s + (A)_s^T x_{s^e} + Av,$$

(11)

where $x_s \in \mathbb{C}^{q \times 1}$ is the signal energy in the passband, $x_{s^e} \in \mathbb{C}^{(N-q) \times 1}$ is the signal energy in the stopband, and $v \in \mathbb{C}^{N \times 1}$ is the receiver noise. It is easy to see from equation (11) that if the matrix $A$ is orthonormal, i.e., perfect mutual incoherence holds, then the signal $x_s$ may be recovered from $y$ in the presence of noise by matched filtering, i.e., $(A)_s^H y = x_s + v_s$. However, downsampling will generate a matrix $A$ with more columns than rows; therefore, all of the columns will never be perfectly orthogonal. In [6] and [7], reconstruction performance is directly related to the degree of coherence between the columns in $A$. We will take a statistical signal processing viewpoint of reconstruction performance under the following conditions: the noise is independent and normally distributed as $v \sim \mathcal{N}(0, \sigma_v^2 I)$, and the stopband signal energy is negligible, e.g., $x_{s^e} \approx 0$. Then the maximum likelihood estimator (MLE) is given by

$$x_{s}^{MLE} = \left[(A)_s^H C_v^{-1}(A)_s\right]^{-1}(A)_s^H C_v^{-1} y = x_s + v_s + \left[(A)_s^H C_v^{-1}(A)_s\right]^{-1}(A)_s^H C_v^{-1}(A)_s^T v_{s^e},$$

(12)

where the covariance $C_v = A (\sigma_v^2 I) A^H$. Equation (12) consists of three terms: the desired signal, the noise at the locations in the basis that the signal occupies, and a folding loss which corresponds to noise outside these locations folding back in. The folding loss is the penalty that is paid in CS for using reduced-rate sampling and ML reconstruction. Notice that the degree of the folding loss is directly related to the coherence between the columns in the matrix $A$; as the cosines of the angles between the columns of $(I - \Theta\Psi)$ and $(A)_s$ approaches zero, so does the folding loss. The MLE formulation in equation (12) may significantly amplify some of the noise components. Another cost function, the minimum mean square error (MMSE) [8], is often used to balance the quality of the estimate against the potential for noise amplification. Using equation (11) with $x_{s^e} \approx 0$ and $x_s \sim \mathcal{N}(0, C_v)$, e.g., the signal is subject to frequency selective Rayleigh fading,
then the MMSE estimator is given by
\[
\hat{s}_{\text{MMSE}} = (A_s^H C_v^{-1} (A)_s + C_{x_s}^{-1})^{-1} (A_s^H C_v^{-1} y) - \hat{A}^{-1} (A_s^H C_v^{-1} (A)_s x_s + v),
\]
where \(\hat{A} = (A_s^H C_v^{-1} (A)_s)\).

The folding loss associated with ML/MMSE identification and reconstruction is illustrated in Figure 5. The scenario illustrated in Figure 5 employed a measurement matrix \(\Theta\) with entries randomly chosen from a \(\pm\) Bernoulli distribution with signals whose basis coefficients covered random but contiguous positions in frequency. The folding loss may be explained in two ways. First, as explained previously, there is a loss due to noise from ambiguous basis locations folding back onto the basis locations that the signal occupies. This is reflected by the fact that the folding loss is greater when the downsampling is more aggressive, e.g., from a factor of 2 to a factor of 8. Second, as the bandwidth of the signal increases the variability in the frequency response of the random linear projections makes it more likely that there will be noise amplification in the identification and reconstruction process. This is illustrated in Figure 5, which shows that as the bandwidth increases, noise amplification, as reflected by the change in folding loss, also increases. The choice of an ML or MMSE estimator is dependent on the signal bandwidth; at one extreme narrowband spikes (tones) are best re-

constructed using an ML estimator, while a wider-band signal should be identified and reconstructed using an MMSE estimator to obtain the best trade between noise amplification and reconstruction performance. Although the scenario illustrated in Figure 5 plots folding loss versus the fraction of nonzero coefficients in the downsampled frequency basis, the results hold for any orthonormal basis in which the signal has a sparse representation.

4. SUMMARY

In this paper we have described and demonstrated VPU’s capability of reconstructing signals that can occupy up to nearly 100 percent of the downsampled basis. We have also derived an ML/MMSE bound on VPU performance as a function of the number of nonzero basis coefficients and downsampling factor. It is interesting to note that although VPU is capable of reconstructing not-so-sparse signals where BP fails, there is an exponential loss in sensitivity in the reconstruction process as the number of nonzero basis coefficients increases.

5. REFERENCES