SPARSE RECOVERY USING AN SVD APPROACH TO INTERFERENCE REMOVAL AND PARAMETER ESTIMATION

Charles Ethan Hayes, James H. McClellan, Waymond R. Scott, Jr.

Georgia Institute of Technology
Electrical and Computer Engineering
777 Atlantic Drive NW Atlanta, GA 30332-0250

ABSTRACT

This work focuses on parametric sparse sensing models and looks to improve ℓ_1 regularization results when the model dictionary is strongly coherent and/or regularization parameters are unknown. The singular value decomposition (SVD) of the model's dictionary matrix is used to construct signal and noise subspaces. A method that uses the measurements to automatically optimize the subspace division along with a way to estimate the noise level is introduced. The signal-noise subspace decomposition is then extended to deal with an interfering signal that lies in a known linear subspace by modifying the SVD and performing the sparse recovery in the modified signal subspace. The proposed technique is applied successfully to the Discrete Spectrum of Relaxation Frequencies (DSRF) extraction problem for Electromagnetic Induction (EMI) underground sensing where a strong interference from the soil is a significant concern.

***Index Terms*—** Compressed Sensing, Interference Elimination, Parameter Estimation, Coherent Dictionary

1. INTRODUCTION

This work has been developed in an attempt to better understand and address the Discrete Spectrum of Relaxation Frequencies (DSRF) inversion problem that is used to detect and estimate metallic targets with Electromagnetic Induction (EMI) sensors [1]. Although this is the initial application, the discovered methods appear to have far broader implications and are presented here in a generic format to be adapted to any similar problem. The underlying problem being addressed is that of finding a sparse vector x from $m = Ax + Gz + \epsilon$, where $m \in \Re^M$ is a set of measurements made. The vector $x \in \Re^N$ represents the linear combination coefficients of the possible signals in the measurements, $A \in \Re^{M\times N}$ is the dictionary of signals that is expected, $M < N$ such that A is an under-determined matrix, $G \in \Re^{M\times L}$ is the dictionary of interference signals possible, $z \in \Re^L$ is the vector of coefficients for how the interference signals are added in, and $\epsilon \in \Re^M$ is a vector of $i.i.d.\ \mathcal{N}(0, \sigma^2)$ noise terms added to the measurements. This problem is expected to have a sparse x since only a few of the signals in the dictionary will be contained in the measurements, but z is not sparse since L is expected to be a small number of basis vectors that describe known interference. In order to solve this problem, a simplified model of $m = Ax + \epsilon$ is addressed without the interference signal, and the interference will then be addressed in Section 4. The goal of this problem is to recover the original x that created the measurements. In other words, the goal is to minimize the error when fitting the measurements subject to a sparsity constraint on x. It is well understood that the ℓ_1 norm is a useful convex regularization term that can be added to enforce sparsity. This can lead to the optimization problem (1) which can be efficiently solved using methods developed around the LASSO problem[2, 3].

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \|m - Ax\|_2 \le b \tag{1}$$

When solving this problem, it is helpful to have prior knowledge about the noise level in order to set the value of b in (1). When a parameter cannot be estimated, it is common to use methods such as cross-validation to find an optimal range of the parameter. Difficulty arises when the optimization problem is very sensitive to this parameter and there is no clear choice without prior knowledge of x. The DSRF example used throughout this paper is a strong example of this since the columns of A are highly coherent, causing the optimization to be extremely sensitive to the parameter choice. The DSRF example also has little useful *a priori* knowledge, since the $\|x_0\|_1$ and number of supports of x_0 are constantly changing between targets, where x_0 is the true x being solved. The noise term can be estimated from surrounding measurements when the sensor is moving, but the estimate is inaccurate because the true noise term includes an interference signal from the soil that is not guaranteed to be consistent between measurements.

This work provides a method to improve the results from

This material is based upon work supported in part by the US Army REDCOM CERDEC Night Vision and Electronic Sensors Directorate, Science and Technology Division, Countermine Branch, and in part by the U. S. Army Research Office under grant number W911NF-11-1-0153.

978-1-4673-9169-6/15/$31.00 ©2015 IEEE

(1) by using a singular value decomposition (SVD) in order to create signal and noise subspaces. This allows for an optimization where the highly coherent columns of A are converted into the orthonomal columns of the SVD, and allows for improved recovery of x_0 in a lower dimensional space by ensuring greater signal power over the noise. An added benefit from this method is it inherently creates a method of estimating the necessary parameter b needed for (1). Finally, this work will point out an easy extension of the proposed algorithm to handle linear parametric interference models and will discuss some further benefits from this expansion.

2. THE MODEL

The model being addressed is (2a) where $m(\theta)$ are M measurements being obtained, θ is some parameter based on how the measurements are made, and $S(\theta)$ is the desired signal. $S(\theta)$ is assumed to be a linear combination of known signals modeled by $S(\theta) = \sum_{h=1}^{H} c_h \mathfrak{f}_h(\theta)$. This parametric model can then be expanded into matrix notation as described by (2b), where there are more possible signals than measurements, or $M < N$. The matrix A is $M \times N$ and the recovery problem $m = Ax + \epsilon$ is an under-determined system.

$$m(\theta) = S(\theta) + \epsilon \tag{2a}$$

$$\begin{bmatrix} S(\theta_1) \\ \vdots \\ S(\theta_M) \end{bmatrix} = \begin{bmatrix} \mathfrak{f}_1(\theta_1) & \cdots & \mathfrak{f}_N(\theta_1) \\ \vdots & \ddots & \vdots \\ \mathfrak{f}_1(\theta_M) & \cdots & \mathfrak{f}_N(\theta_M) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = Ax \tag{2b}$$

In the DSRF example problem, θ is defined as a frequency measurement where the frequency range spans 300 Hz to 100 kHz. The hardware makes 21 complex measurements and the real and imaginary parts are separated and concatenated on top of each other to form $m \in \Re^{42}$. The DSRF basis functions are the Fourier transform of exponential relaxations at different relaxation frequencies, so the model can be described as $\mathfrak{f}_l(\theta_i) = \dfrac{1 - j\theta_i/\zeta_l}{1 + j\theta_i/\zeta_l}$ where ζ_l is the relaxation frequency. There is also a $\mathfrak{f}_0(\theta_i) = 1$ basis vector to allow for constant offsets. Like the measurements, the basis vectors are also separated between the real and imaginary parts and concatenated on top of each other. The signals are chosen by log-sampling the frequency range of ζ_l, and typically have around 100 samples. This creates an $A \in \Re^{42\times 100}$ for the DSRF example.

3. SINGULAR VALUE DECOMPOSITION APPROACH

3.1. Breakdown

Insight into the structure of the dictionary A can be found by applying a singular value decomposition (SVD) on A. This decomposition states that $A = U\,\Sigma\,V^T$, where U and V are both square, unitary matrices such that $U^TU = I = UU^T$ and $V^TV = I = VV^T$. By performing the SVD on the dictionary A and ordering the singular values in descending order, the problem can be broken down into the subspaces seen in (3) where the division is based on the magnitudes of the singular values. $\mathrm{U_S}$ spans the measurement space that is deemed to have 'good' SNR because the singular values are large, and the measurements are made up of mostly signal. $\mathrm{U_N}$ spans the measurement space that is deemed to have 'bad' SNR because the singular values are below some threshold, and is characterized by the noise overpowering the signal. $\mathrm{V_S}$ spans the coefficient space that projects into signal space spanned by $\mathrm{U_S}$, $\mathrm{V_N}$ spans the coefficient space that projects into the noise space that is spanned by $\mathrm{U_N}$, and V_0 spans the null space of A which cannot affect the measurements at all. It is useful to note that a projection into any of these subspaces is simply $P_{\mathrm{V_S}}(y) = \mathrm{V_S}\,\mathrm{V_S}^T\,y$ because $\mathrm{V_S}^T\,\mathrm{V_S} = I$.

$$A = \begin{bmatrix} \mathrm{U_S} & \mathrm{U_N} \end{bmatrix} \begin{bmatrix} \Sigma_\mathrm{S} & 0 & 0 \\ 0 & \Sigma_\mathrm{N} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{V_S}^T \\ \mathrm{V_N}^T \\ V_0^T \end{bmatrix} \tag{3}$$

The decision boundary between the signal and noise space is determined by the singular values. If there is a clear drop off in the singular values between two different groupings, then these can naturally be chosen as the signal and noise subspaces. If such a clear delineation is not evident from the singular values, such as in the DSRF case (figure 1(a)), then there is an intuitive approach of automatically picking the subspaces that will be discussed in Section 3.4.

3.2. Solution

By using this SVD decomposition, it can be directly surmised that the solution to x is primarily devoted to constraining the error in the signal subspace. This comes from the fact that $m = Ax + \epsilon = P_{\mathrm{U_S}}(Ax) + P_{\mathrm{U_N}}(Ax) + \epsilon$ and $P_{\mathrm{U_N}}(Ax) = \mathrm{U_N}\,\Sigma_\mathrm{N}\,\mathrm{V_N}^T\,x$. Because Σ_N was chosen to contain the singular values smaller than the noise level of ϵ, the contribution from $P_{\mathrm{U_N}}(Ax)$ is expected to be hidden by the noise and can be safely ignored in fitting to the model. This directly leads to rewriting equation (1) as (4a). Equation (4b) can be obtained by applying the projection operator $P_{\mathrm{U_S}}(y) = \mathrm{U_S}\,\mathrm{U_S}^T\,y$, and noting that $\mathrm{U_S}^T\,\mathrm{U_S} = I$, so $\mathrm{U_S}$ does not effect the ℓ_2-norm and can be ignored. The final optimization equation in (4c) can be arrived at by distributing the $\mathrm{U_S}^T$ and noting that $\mathrm{U_S}^T\,\mathrm{U_N} = 0$ by definition.

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \|P_{\mathrm{U_S}}(m - Ax)\|_2 \le b \tag{4a}$$

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \|\,\mathrm{U_S}^T(m - Ax)\|_2 \le b \tag{4b}$$

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \|\,\mathrm{U_S}^T\,m - \Sigma_\mathrm{S}\,\mathrm{V_S}^T\,x\|_2 \le b \tag{4c}$$

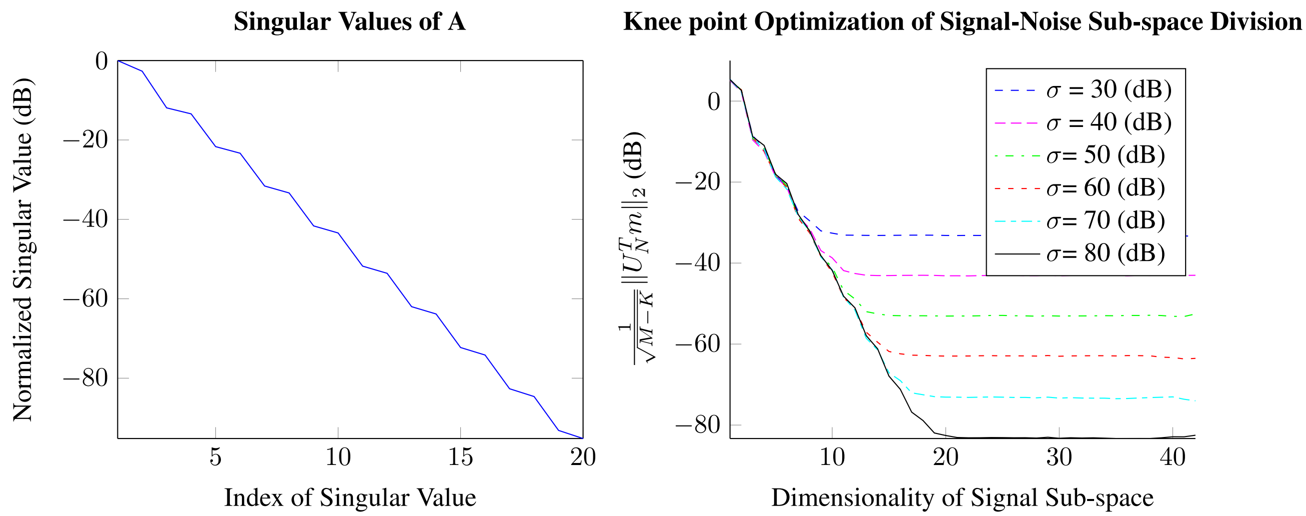

Fig. 1: (a) Singular values of A. (b) Depicting the knee point for the Subspace Division optimization

3.3. Accurate Noise Estimation

An important result of using the SVD approach is that it allows for accurate estimation of the noise variance in the current measurements, from the current measurements.

Lemma 1. *By assuming the noise is independent between each measurement and choosing a $\Sigma_N \in \Re^{M-K \times M-K}$ that consists of sufficiently small enough singular values, then*

$$\sigma \geq \tfrac{1}{\sqrt{M-K}} \| \mathrm{U_N}^T m\|_2$$

Proof. Let us define $x^* = \mathrm{V_S}\, \Sigma_\mathrm{S}^{-1}\, \mathrm{U_S}^T\, m$ as the least square solution to m in the signal subspace.
$\| \mathrm{U_N}^T m\|_2 = \|m - Ax^*\|_2$ can be derived by showing x^* perfectly fits the signal space, but $P_{\mathrm{U_N}}(Ax^*) = 0$.
$\| \mathrm{U_N}^T m\|_2 = \|P_{\mathrm{U_N}}(Ax + \epsilon - Ax^*)\|_2$ can then be achieved by replacing m with its model and noting the error is in $\mathrm{U_N}$.
$\| \mathrm{U_N}^T m\|_2 = \| \mathrm{U_N}^T(Ax + \epsilon - Ax^*)\|_2$ is achieved by replacing the projection operator with its equation and noting that $\mathrm{U_N}$ is orthonormal so it can be ignored.
$\| \mathrm{U_N}^T m\|_2 = \| \Sigma_\mathrm{N}\, \mathrm{V_N}^T(x - x^*) + \mathrm{U_N}^T\, \epsilon\|_2$ is from distributing $\mathrm{U_N}^T$.
$\| \mathrm{U_N}^T m\|_2 \leq \| \Sigma_\mathrm{N}\, \mathrm{V_N}^T(x - x^*)\|_2 + \| \mathrm{U_N}^T\, \epsilon\|_2$ is from using the triangle inequality to separate the terms.
$\| \mathrm{U_N}^T m\|_2 \leq \| \Sigma_\mathrm{N}\, \mathrm{V_N}^T\, x\|_2 + \sigma\sqrt{M-K}$ is from noting that $\mathrm{V_N}^T x^* = 0$ and that the second term was white noise multiplied by a unitary matrix. K has been defined as the number of singular values in Σ_S.
$\| \mathrm{U_N}^T m\|_2 \leq \| \Sigma_\mathrm{N}\, \|_2 \|x\|_2 + \sigma\sqrt{M-K}$ is achieved by breaking apart the ℓ_2 norm.
$\| \mathrm{U_N}^T m\|_2 \leq \Sigma_\mathrm{N}[1,1]\|x\|_2 + \sigma\sqrt{M-K}$ is from evaluating the ℓ_2 norm of Σ_N.
Finally, by choosing a small enough Σ_N[1,1] that is significantly smaller than σ, it shows that $\| \mathrm{U_N}^T m\|_2 \leq sigma\sqrt{M-K}$, which finishes the proof. □

The estimate from Lemma 1 can then be placed back into (4c) to create (5) which is a constrained optimization problem with all of its parameters estimated from the current measurements.

$$\min_x \|x\|_1$$
$$\text{s.t.} \ \| \mathrm{U_S}^T\, m - \Sigma_\mathrm{S}\, \mathrm{V_S}^T\, x\|_2 \leq \sqrt{\frac{K}{M-K}} \| \mathrm{U_N}^T\, M\|_2 \tag{5}$$

3.4. Subspace Division Optimization

By taking a closer look at (6) which can be derived from the proof of Lemma 1, it can be noted that the size of the measurements projected onto the noise subspace is dominated by either the noise in the subspace or the signal projected onto the subspace.

$$\frac{1}{M-K}\| \mathrm{U_N}^T\, m\|_2^2 \leq \frac{1}{M-K}\| \Sigma_\mathrm{N}\, \mathrm{V_N}^T\, x\|_2^2 + \sigma^2 \tag{6}$$

This is easily seen in figure 1(b) where the larger K values remain constant since the noise is dominating the term, but before the clear knee point, the term is dominated by the singular values. This fact can be used to optimize where the division between $\mathrm{U_S}$ and $\mathrm{U_N}$ should lie. By starting at a singular value that is known to be smaller than any expected noise term, the $\mathrm{U_S}$ subspace can be decreased one dimension at a time until the knee point is reached where the signal term starts to overpower the noise. This optimization guarantees that all of the measurements in the $\mathrm{U_S}$ space chosen have a stronger signal component in them than the random noise affecting them. A short cut to this optimization is noticing that the normalized singular value that is closest to the σ tends to be where the location of the knee point is located, where the normalized singular values are the singular values divided by the largest one.

3.5. Results

The new optimization problem in (4c) is used on the DSRF problem with b being estimated as described in (5). It is compared to a minimization using the full dictionary in figure 2, where an estimate, $\hat{\sigma}$, is used in b. The results for various signal to noise ratios (SNR) are shown where a Monte Carlo was run 100 times for each SNR on synthetic data. The Earth Mover's Distance (EMD) is used as a distance measure of how close the inverted DSRF is to the original DSRF.

As can be seen from the results, the SVD approach is able to out perform even the ideal case of the straight forward minimization problem given the exact b term. It also out performs a $\hat{\sigma} \approx 0.56\sigma$, which is equivalent to estimating $\widehat{\text{SNR}} = \text{SNR} + 5$ where the SNR is in dB and the signal strength is equal to 1. If $\widehat{\text{SNR}} = \text{SNR} - 5$, which is equivalent to $\hat{\sigma} \approx 1.78\sigma$, then the direct LASSO gives even poorer results. This shows that the SVD implementation is preferred over a direct application of (1), and gains even more benefit when the b cannot be exactly calculated.

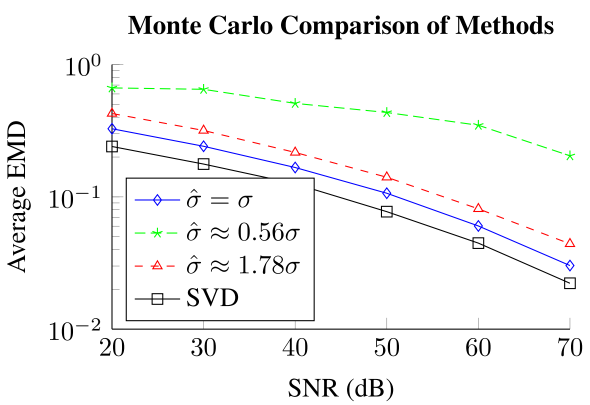

Fig. 2: Monte Carlo simulation comparing new approach with standard LASSO

4. INCORPORATING LINEAR INTERFERENCE

As mentioned before, EMI sensing has an expected interference source in the measurements due to the soil response. This has led to the expansion of the SVD approach to handle interference that can be modeled with a linear combination of known basis vectors. This approach allows for a simple expansion to remove linear interference through the use of a projection operator. Some examples of undesired linear interference could be sensor drift or an expected interfering response that can be modeled, such as the soil response in the DSRF example.

4.1. Modeling the Interference

The first step is to extend the model being used for the measurements. An interference term is added back into the model in (7a) where $G(\theta)$ is the new interference term. In order to remove the interference in the described way, it must be able to be modeled as $G(\theta) = \sum_{l=1}^{L} c_l \mathfrak{g}_l(\theta)$. Once it is modeled in this way, a G matrix can be designed as shown in (7b)

$$m(\theta) = S(\theta) + G(\theta) + \epsilon \tag{7a}$$

$$\begin{bmatrix} G(\theta_1) \\ \vdots \\ G(\theta_M) \end{bmatrix} = \begin{bmatrix} \mathfrak{g}_1(\theta_1) & \cdots & \mathfrak{g}_L(\theta_1) \\ \vdots & \ddots & \vdots \\ \mathfrak{g}_1(\theta_M) & \cdots & \mathfrak{g}_L(\theta_M) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_L \end{bmatrix} = Gz \tag{7b}$$

where G is created from the basis vectors of the interference subspace. This allows for the creation of a projection matrix $P_G = G(G^T G)^{-1} G^T$ that projects onto the interference subspace, as well as $P_G^{\perp} = I - P_G$ that projects onto the subspace orthogonal to the interference. If it is possible to choose the vectors of G to be orthonormal, then $G^T G = I$ which simplifies the projection operator, but this is not necessary.

For the DSRF example, the linear parametric interference model comes from empirical studies of the soil response. The soil response can be modeled with two basis vectors: $\mathfrak{g}_1(\theta_l) = 1$ and $\mathfrak{g}_2(\theta_l) = \log\left(\frac{\theta_l}{\theta_C}\right) + \frac{j\pi}{2}$. Just like with the signal, the real and imaginary parts are separated and concatenated on top of each other to form $G \in \Re^{42\times 2}$. The constant θ_C is a chosen frequency parameter; when chosen as $\theta_C = \prod_{i=1}^{M} \sqrt[M]{\theta_i}$, the two vectors will be orthogonal[4].

4.2. Interference Restructuring

In order to restructure the problem to account for the interference, the projection operator can be used to split the dictionary into two parts as seen in (8), where $\mathrm{A_G}$ has become the part of the dictionary that includes the interference and $\mathrm{A_G^{\perp}}$ is the other part that contains the information perpendicular to the interference. By design, these sub-dictionaries have the following features: $rank(\mathrm{A_G}) = L$, $rank(\mathrm{A_G^{\perp}}) = M - L$, and $\mathrm{A_G}^T \mathrm{A_G^{\perp}} = 0$ which implies that their ranges are orthogonal.

$$A = P_G A + P_G^{\perp} A = \mathrm{A_G} + \mathrm{A_G^{\perp}} \tag{8}$$

From here, an SVD can be run on each sub-dictionary and the signal-noise separation can be performed on $\mathrm{A_G^{\perp}}$ just like before to create (9). When an SVD is performed on $\mathrm{A_G}$, it creates the structure of (9a). This creates a subspace in the measurement domain, $\mathrm{U_G}$, that contains the interference and is represented by an orthonormal set of vectors that span the subspace of G. It is also known that $\hat{U}_{G(S+N)}$ has to span the subspace of $\mathrm{A_G^{\perp}}$ in the measurement domain that is perpendicular to the interference since the range spaces must be orthogonal. $\mathrm{V_G}$ describes the subspace of coefficients that project onto the measurements affected by the interference.

When the SVD is applied to $\mathrm{A_G^{\perp}}$, a signal-noise separation can be performed exactly like in Section 3.1 and the result looks like (9b). This creates a signal subspace, $\mathrm{U_{GS}}$, and a noise subspace, $\mathrm{U_{GN}}$, just like before. The only difference now is that $\mathrm{A_G^{\perp}}$ is no longer full rank, so there is now a null space of $\mathrm{A_G^{\perp}}^T$ that is described by $\hat{U}_G$. Because $\mathrm{A_G}^T \mathrm{A_G^{\perp}} = 0$, it is known that $\hat{U}_G$ spans the same space as

$\mathrm{U_G}$ and G, but there are no guarantees of it being identical to either of these matrices. Finally, there are the coefficient subspaces of $\tilde{V}_0$ and $\hat{V}_0$ that represent the null spaces of the corresponding sub-dictionaries, but nothing can be assumed about their structure.

$$\mathrm{A_G} = \begin{bmatrix} \mathrm{U_G} & \hat{U}_{G(S+N)} \end{bmatrix} \begin{bmatrix} \Sigma_\mathrm{G} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathrm{V_G}^T \\ \tilde{V}_0^T \end{bmatrix} \tag{9a}$$

$$\mathrm{A_G^\perp} = \begin{bmatrix} \mathrm{U_{GS}} & \mathrm{U_{GN}} & \hat{U}_G \end{bmatrix} \begin{bmatrix} \Sigma_\mathrm{GS} & 0 & 0 \\ 0 & \Sigma_\mathrm{GN} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathrm{V_{GS}}^T \\ \mathrm{V_{GN}}^T \\ \hat{V}_0^T \end{bmatrix} \tag{9b}$$

By using the fact that an SVD can also be written as $A = \sum_{i=1}^{M} u_i s_i v_i^T$, the sub-dictionaries can be viewed as

$$\mathrm{A_G} = \sum_{i=1}^{L} u_{g_i} s_{g_i} v_{g_i}^T \tag{10a}$$

$$\mathrm{A_G^\perp} = \sum_{i=1}^{K} u_{gs_i} s_{gs_i} v_{gs_i}^T + \sum_{k=1}^{M-K-L} u_{gn_k} s_{gn_k} v_{gn_k}^T \tag{10b}$$

Then by adding the sub-dictionaries back together,

$$A = \sum_{j=1}^{L} u_{g_j} s_{g_j} v_{g_j}^T + \sum_{i=1}^{K} u_{gs_i} s_{gs_i} v_{gs_i}^T + \sum_{k=1}^{M-K-L} u_{gn_k} s_{gn_k} v_{gn_k}^T \tag{10c}$$

which can be written in matrix notation as (11) to resemble the SVD.

$$\begin{aligned} A &= \check{U}\check{\Sigma}\check{V}^T \\ &= \begin{bmatrix} \mathrm{U_G} & \mathrm{U_{GS}} & \mathrm{U_{GN}} \end{bmatrix} \begin{bmatrix} \Sigma_\mathrm{G} & 0 & 0 \\ 0 & \Sigma_\mathrm{GS} & 0 \\ 0 & 0 & \Sigma_\mathrm{GN} \end{bmatrix} \begin{bmatrix} \mathrm{V_G}^T \\ \mathrm{V_{GS}}^T \\ \mathrm{V_{GN}}^T \end{bmatrix} \end{aligned} \tag{11}$$

Although (11) has the same form as an SVD, the only guarantee about its structure is that $\check{U}^T\check{U} = I = \check{U}\check{U}^T$ which comes from the fact that the sub-dictionaries' ranges are orthogonal. $\check{V}^T\check{V} \neq I$ because $\mathrm{V_G}^T \mathrm{V_{GS}}$ and $\mathrm{V_G}^T \mathrm{V_{GN}}$ are no longer known.

If it is desired to keep $\check{V}$ orthonormal while maintaining the interference in a specific range of the SVD, then a new dictionary $\bar{A} = A - P_G^\perp A P_{\mathrm{V_G}}$ can be defined, where $P_{\mathrm{V_G}} = \mathrm{V_G}\,\mathrm{V_G}^T$ is the projection operator onto $\mathrm{V_G}$. This can be seen by simplifying $A = (P_G + P_G^\perp)A(P_{\mathrm{V_G}} + P_{\mathrm{V_G}}^\perp)$ and setting $P_G^\perp A P_{\mathrm{V_G}} = 0$. $\bar{A}$ is the dictionary closest to the original A whose SVD automatically separates the interference from the rest of the measurements.

4.3. Improved Solver

Combining the interference removal with the solver obtained from the SVD approach gives (12). The main difference to note is that the constraint is now in the subspace that has a 'good' SNR and is also unaffected by the interference.

In addition, the random error estimation should also improve since it no longer contains the interference.

$$\begin{aligned} &\min_x \|x\|_1 \quad \text{s.t.} \\ &\| \mathrm{U_{GS}}^T m - \Sigma_\mathrm{GS}\, \mathrm{V_{GS}}^T x\|_2 \leq \sqrt{\tfrac{K}{M-K-L}}\| \mathrm{U_{GN}}^T m\|_2 \\ &\text{(opt)}\ \| \mathrm{U_G}^T m - \Sigma_\mathrm{G}\, \mathrm{V_G}^T x\|_2 \leq q \end{aligned} \tag{12}$$

There is also now an optional constraint that can be used. This optional constraint allows for a smooth transition between when to use information in the interference subspace or not depending on how strong the interference plus noise term is, since both will be competing with the signal in this subspace. In order to use the optional constraint, it is necessary to have an accurate estimate of the interference that would be application specific. It is also not worth attempting to add the optional constraint if the interference is always expected to cover the signal and thus no useful information could be obtained for the optimization. It is also worth noting that if the optional constraint is included, then this problem no longer represents the LASSO problem since there are two quadratic constraints, and more work would have to be done to find an efficient solver. It is still solvable since it remains a convex problem. In the DSRF problem the optional constraint would be a useful addition since a large portion of the signal strength tends to be contained in the ground subspace and the ground interference does not always overwhelm the signal. Currently the use of this constraint hinges on whether there is an accurate method of predicting the strength and variance of the ground response in order to assign q a meaningful value.

4.4. Results

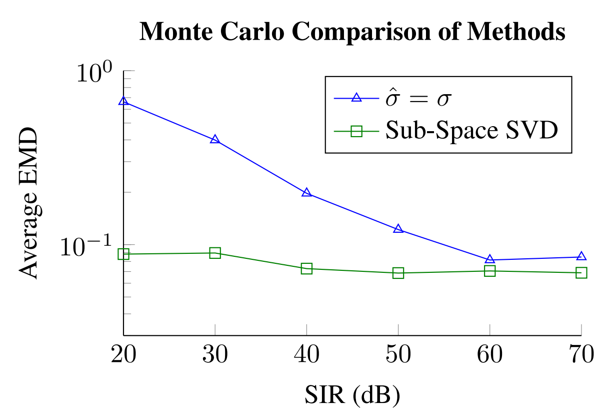

Fig. 3: Monte Carlo simulation comparing Interference removal with standard LASSO

The improved solver in (12) is applied to the DSRF problem without using the optional constraint. The solutions are compared to the minimization problem from (1) with the full dictionary that uses the exact b. Figure 3 shows these results for different signal to interference ratios (SIR) where a Monte Carlo was run 100 times for each SIR on synthetic data. Each test case also contained random noise at 50 dB below the signal, which is typical for high-precision EMI sensors. The EMD is used again to measure the distance of the inverted DSRF to the original DSRF.

As can be seen in figure 3, when the SIR>50 the results are still similar to figure 2 for SNR=50. As SIR decreases though, the standard LASSO begins to diminish in accuracy due to it not being able to overcome the interference. As expected, the improved solver is able to invert the DSRF consistently over a wide SIR range because the interference has been completely isolated from the problem.

4.5. Signal Indicator

By performing the interference subspace separation, there is another very useful by-product that falls out of this technique whenever its application has the goal of inverting multiple sets of measurements. This is due to the fact that the subspace SVD has created the matrix $\check{U}$ which can be used in (13) to obtain a decomposition of m. The component of greatest importance is m_{GS}, which is known to be isolated from the interference as well as have a large SNR where the signal is designed to be above the noise floor. These properties allow m_{GS} to be an ideal signal indication system that can be checked in order to see if there is a signal present in the current set of measurements before even calculating the inversion. This signal indicator can also be used in an application where there is no interference since the $\mathrm{U_S}^T\, m$ has the same properties.

$$\check{U}^T m = \begin{bmatrix} \mathrm{U_G}^T \\ \mathrm{U_{GS}}^T \\ \mathrm{U_{GN}}^T \end{bmatrix} m = \begin{bmatrix} m_G \\ m_{GS} \\ m_{GN} \end{bmatrix} \tag{13}$$

This is useful in the DSRF example because the EMI application is aimed at taking measurements throughout time and location as it is scanned over the ground. This signal indicator can be a quick and easy way to indicate when a target is present and the DSRF inversion should be processed. By using this technique, the DSRF inversion could be run only when a target is present, which would result in far fewer computations.

5. CONCLUSION

In conclusion, this paper has introduced a new method to separate the measurements made for a sparse recovery system into different subspaces. The subspaces are developed through the use of the SVD to create a signal and noise space. This division has a natural separation that can be found as described in Section 3.4 based on the noise level in the measurements and its relationship to the singular values. The noise subspace can be used to find a robust estimate for the noise level in the measurements, as discussed in Section 3.3. Figure 1 clearly shows how consistently the noise space can estimate the noise level and how close this estimate is to the true value. It also shows how the singular values relate to the separation between the signal and noise subspace. Combining all of these elements to create a solver in the signal space leads to equation (5). This technique has been successfully implemented to recover the DSRF for EMI sensing and the results from a Monte Carlo over multiple SNR are shown in figure 2, where it can be clearly seen that the SVD approach provides better results than a standard sparsity minimization using the exact parameter information.

This decomposition of the measurements can also be naturally extended to incorporate a linear parametric model of an interference signal that is introduced into the measurements. This scenario can be addressed by using a projection matrix from the interference model to split the dictionary into two separate dictionaries: the interference dictionary and the perpendicular dictionary. An SVD can then be applied to both of these dictionaries as described in Section 4.2, and then recombined to create a dictionary decomposition that has a signal subspace, noise subspace, and an interference subspace. This leads to a solver as described in (12) that is unaffected by the interference. The subspace decomposition has been successfully applied to the DSRF problem and figure 3 shows how the proposed technique successfully decouples the problem from the interference.

Another benefit from creating such a decomposition is that it creates a method to project the measurements into a signal subspace that is isolated from interference and is also expected to have signals that are stronger than the noise level. This subspace can be used to test whether a signal is expected to be present, even before any recovery attempts are made. Such a feature can be very useful for continuous streams of data in order to reduce computation by implementing a signal indication system as described in Section 4.5.

To summarize, this work introduces a new way to transform a sparse recovery problem. The transformation can be applied to any dictionary matrix and provides a method to automatically detect optimization variables from the measurements so that no *a priori* knowledge is needed to perform the sparse recovery. It also allows for the removal of interference signals that can corrupt the measurements. It is likely that other sensing scenarios that require interference nulling could exploit this approach.

6. REFERENCES

[1] Mu-Hsin Wei, W. R. Scott, and J. H. McClellan, "Robust estimation of the discrete spectrum of relaxations for electromagnetic induction responses," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 3, pp. 1169–1179, March 2010.

[2] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 04 2004.

[3] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, pp. 1–22, 2010.

[4] W. R. Scott, G. D. Larson, C. E. Hayes, and J. H. McClellan, "Experimental detection and discrimination of buried targets using an improved broadband CW electromagnetic induction sensor," 2014.