Enveloped Sinusoid Parseval Frames

Geoff Goehle*, Benjamin Cowen†, J. Daniel Park‡, and Daniel C. Brown§

Applied Research Laboratory,
Pennsylvania State University,
State College, PA

Email: *goehle@psu.edu, †bmc6220@psu.edu, ‡jdp971@psu.edu, §dcb19@psu.edu

Abstract—This paper presents a method of constructing Parseval frames from any collection of complex envelopes. The resulting Enveloped Sinusoid Parseval (ESP) frames can represent a wide variety of signal types as specified by their physical morphology. Since the ESP frame retains its Parseval property even when generated from a variety of envelopes, it is compatible with large scale and iterative optimization algorithms. The ESP construction provides an analysis and synthesis transform pair that can be tuned according to the characteristics of the signal to which it is applied. This work provides examples of ESP frame generation for both synthetic and experimentally measured signals. Furthermore, the frame’s compatibility with distributed sparse optimization frameworks is demonstrated. Numerical experiments on acoustics data reveal that the flexibility of this method allows it to be simultaneously competitive with a signal agnostic short-time Fourier Transform in time-frequency processing and also with Prony’s Method for time-constant parameter estimation, surpassing the shortcomings of each individual technique.

I. INTRODUCTION

The problem of decomposing a digital signal into oscillating components captures a broad spectrum of applications. Basic signal processing tools such as the Fourier Transform, the short-time Fourier Transform (STFT), wavelets, empirical mode decomposition, and others, generally seek to rewrite the signal as a linear combination of elementary signals or atoms [1], [2], [3]. Examination of the signal in the alternative representation can reveal important properties [4], enable specialized denoising [5], [6], improve detection [7] performance, and make parameter estimation easier [8].

However, physical signals often violate the assumptions underlying these techniques, making the resulting decompositions difficult to interpret. Researchers have developed a wide variety of approaches for representing complicated and nuanced signals in terms of elementary components, ranging from common, generic textures [9], [10] to tunable wavelets [2], to fully data-driven methods [11], [12]. In some work a superset of multiple transforms is employed to capture temporally overlapping components of the signal [13], [14].

The modern extension of this approach is to fix an over-complete basis (or frame) of elementary atoms, then employ convex optimization to infer a weighting of the atoms that best fits the data [15]. Regularization plays an important role in characterizing the solution to overdetermined systems with a common choice being sparsity. Sparse optimization can produce coefficients that are mostly zero, so that only a small subset of available atoms are actually used to compose the signal’s representation. Sparse representations are both efficient and interpretable [16].

A different type of approach is to approximate the signal with a physics-based model whose parameters are inferred. These parameters may indirectly define a superposition of atoms, but atom coefficients are not optimized directly. Prony’s Method is a classical example that seeks to compute the poles of an infinite impulse response filter whose impulse response best matches the given data [1], indirectly representing the data as a superposition of exponentially decaying sinusoids. This technique is still employed in practice [17] because its model is explainable in terms of the signal morphology. However, the underlying assumption that the signal has a decaying exponential envelope and identification of the signal start time are critical to its performance.

This paper presents a procedure for generating a STFT frame using any number of complex signal envelopes, called the Enveloped Sinusoid Parseval (ESP) frame. The ESP frame uses the same underlying structure as a traditional STFT, but while the usual STFT is largely signal agnostic, an ESP frame can be easily tuned to represent a wide variety of signals by supplying relevant envelopes. It is compatible with modern convex analysis techniques, and because it retains the Parseval property, it is efficient and practical to deploy in large-scale and distributed iterative optimization algorithms.

Section II presents the ESP approach, including formulas for accelerated computation, and implementation details for incorporating it into regularized least squares problems. Section III provides an ESP frame construction for theoretical and experimental time series, and demonstrates ESP as a nonlinear denoising filter in comparison with an STFT frame. Section IV demonstrates the ESP frame’s utility in parameter estimation in comparison with Prony’s Method.

II. ENVELOPED SINOUSOID PARSEVAL FRAMES

Section II-A defines the ESP frame for a general set of complex envelopes, and provides formulas for accelerated computation. Optimization algorithms for $L_1$-regularized least-squares based coefficient inference are presented in Section II-B.
A. Definition

Consider the complex finite-dimensional Hilbert space \( \mathbb{C}^N \). All norms \( (\| \cdot \|) \) are computed in the \( \ell_2 \) sense unless otherwise stated. A tight frame is a collection of vectors \( \{ a_k \}_{k=0}^{K-1} \) in \( \mathbb{C}^N \) and \( \alpha > 0 \) such that

\[
\| w \|^2 = \alpha \sum_{k=0}^{K-1} |\langle w, a_k \rangle|^2 \text{ for all } w \in \mathbb{C}^N. \tag{1}
\]

A Parseval frame is a tight frame with \( \alpha = 1 \). Given a tight frame we define the analysis operator to be the linear map from vectors \( w \in \mathbb{C}^N \) to frame coefficients \( A w \in \mathbb{C}^K \) such that

\[
A w[k] = \langle w, a_k \rangle. \tag{2}
\]

Elements of \( A \) are given by \( A[m,n] = \overline{a_m[n]} \). The defining characteristic of tight frames is that the original vector can be reconstructed from the frame coefficients via the formula \([18, \text{Prop. 3.11}]

\[
w = \frac{1}{\alpha} A^* A w = \frac{1}{\alpha} \sum_{k=0}^{K-1} \langle w, a_k \rangle a_k. \tag{3}
\]

The synthesis operator \( A^* \) is the conjugate transpose of the analysis operator, and maps frame coefficients back to the signal space. In the case that \( A \) is a Parseval frame, \( A^* \) serves as the frame’s left-inverse.

We define the class of Enveloped Sinusoid Parseval Frames by applying enveloping functions to the (non-unitary) Discrete Fourier Transform (DFT) basis. First, specify the envelopes as a collection of vectors \( \{ e_l \}_{l=0}^{L-1} \) which are not identically zero. We then have the following from [19].

**Theorem 1.** Given a set of nonzero \( N \)-dimensional vectors \( \{ e_l \}_{l=0}^{L-1} \), the vectors \( \{ a_{l,k,m} \} \) defined by

\[
a_{l,k,m}[n] = e_l[n - m \mod N] \exp(2\pi j k(n - m)/N) \tag{4}
\]

for \( l = 0, \ldots, L - 1 \) and \( k, m, n = 0, \ldots, N - 1 \) form a tight frame with \( \alpha = N \sum_l |e_l|^2 \).

Notably Theorem 1 also follows from viewing the frame as a multi-window STFT. The frame vectors \( a_{l,k,m} \) have indices corresponding to the envelope \( l \), the modulation frequency \( k \), and the circular shift \( m \).

For ESP frames \( A \) and \( A^* \) are very large, so direct computation of matrix products is expensive. However both analysis and synthesis can be sped up significantly using the Fast Fourier Transform (FFT). If we define the vector \( c_{k,l} \) as \( c_{k,l}[m] = A^* w[k, l, m] \) one can show, after some matrix algebra, that

\[
c_{k,l} = F^* D(S^k F H(e_l)) F w, \tag{5}
\]

\[
w = \frac{1}{\alpha} F^* \sum_{k,l} D(S^{-k} F e_l) F c_{k,l}, \tag{6}
\]

where \( F \) represents the DFT operator, \( S \) right circular shifting, \( D(v) \) the diagonal matrix with entries given by \( v \), and \( H \) the (conjugate linear) operator \( H w[n] = w[N - n \mod N] \).

In practical terms, the formulations in (5) and (6) admit parallelization of the time dimensions, and the underlying computations can be sped up using GPU parallelization.

B. Coefficient Inference

Frames, by their nature, do not uniquely represent vectors and for any given vector \( w \) there will be a linear subspace of frame coefficients \( c \) such that \( \frac{1}{\alpha} A^* c = w \). Sparse frame representations are important in the ESP setting because the canonical frame coefficients \( A w \) are computed independently, and thus for a general set of envelopes are expected to contain redundant information. Through application of \( L_1 \)-regularization, atoms can be made to “compete” with each other and a sparse coefficient vector, with much less redundancy, can be computed that maintains exact reconstruction of the input signal (or in the case of noisy data, maintains some allowable error). In the ideal scenario, regularization can be used to identify superimposed components of a signal formed from a linear combination of frame vectors. This is most effective with highly distinct envelopes.

In formal terms, \( L_1 \)-regularization entails finding frame coefficients that solve either the Basis Pursuit (BP) problem

\[
\arg \min_c \| D(\lambda) c \|_1 \text{ such that } \frac{1}{\alpha} A^* c = w \tag{7}
\]

or the Basis Pursuit Denoising (BPD) problem

\[
\arg \min_c \| D(\lambda) c \|_1 + \frac{1}{2} \| \frac{1}{\alpha} A^* c - w \|^2, \tag{8}
\]

where \( \lambda > 0 \) is a weight vector that allows the user to control relative penalization between coefficients. Note that \( D(\lambda) c \) is equivalent to element-wise multiplication. A constant parameter \( \lambda \) is often used in place of a vectorized \( \lambda \).

These convex optimization problems can be solved by the Split Augmented Lagrangian Shrinkage Algorithm (SALSA), which is an instance of the Alternating Direction Method of Multipliers (ADMM) for which convergence is proved [20]. The steps of SALSA [21, Algorithm 4] applied to the BP problem (7) are written in Algorithm 1, where soft(\cdot) denotes the soft-thresholding function

\[
\text{soft}(x, T) = \begin{cases} 
\frac{|x|-T}{|x|} x & |x| > T \\
0 & |x| \leq T.
\end{cases}
\]

Note that for a Parseval frame \( \alpha = 1 \) in Algorithm 1.

When the frame does not admit a sparse representation of the signal, as is the case with noisy data, BPD may be used to search for a sparse solution at the cost of reconstruction error. In this case, SALSA is also performed using Algorithm 1 with the addition of a \((1 + \mu/\alpha)^{-1}\) coefficient in the computation of \( x_n \). Both algorithms can be sped up by the use of (5) and (6) for analysis and synthesis, as well as the addition of the predictor-corrector-type acceleration described in [22, Algorithm 8].

For BPD, note that there exists \( \lambda_{\text{max}} \) given by

\[
\lambda_{\text{max}} = \| A w \|_\infty \tag{9}
\]
such that for all $\lambda \geq \lambda_{\text{max}}$ the BPD solution is zero [23, Section V.B]. The subsequent BPD experiments set $\lambda$ as a percentage of $\lambda_{\text{max}}$ with $\lambda = 0.1\lambda_{\text{max}}$ as a common choice.

At convergence, the choice of $\mu$ does not impact the solution for Algorithm 1, but it impacts the convergence rate. In the subsequent experiments $\mu$ is set to $\lambda/p$ where $p$ is the 99th percentile of the initial coefficient magnitudes $|x_0[k, l, m]|$. This causes the first soft threshold of either the BP or BPD algorithm to zero-out 99% of the coefficients. In the case of vector-weighted BPD $\mu = \text{mean}(\lambda)/p$ is selected.

### III. NOISE DENOISING

In this section the robustness of the ESP frame approach to noise is evaluated and BPD is used to filter noisy time series, both synthetic generated and experimentally collected. Notionally, if the envelopes are chosen so that the ESP frame vectors are a good model for the signal, then the $L_1$-regularized representation of the clean signal will be sparse. Then, when BPD is applied to the noisy signal, the sparsification of the signal will preferentially remove noise and increase SNR. The ESP frame denoising performance is compared to the same of an STFT-based frame. The selected STFT is windowed with cosine functions so that it is also a Parseval frame as described in [21]. The STFT frame uses a window length of 64, resulting in $2,432$ frame coefficients. Finally, the tradeoffs between sparsity and the reconstruction error for BPD regularized ESP and STFT frame coefficients are compared.

For the following sections we will start with either a synthetically or experimentally generated signal $h$ and will form a noisy signal $h_N$ by adding white Gaussian noise at some specified SNR (measured against the power of $h$). We do not expect BPD to exactly reconstruct the original signal, even in the noise-free case, and we track the reconstruction error using the relative error of the reconstructed signal $h_R$.

$$ E = \frac{\|h_R - h\|}{\|h\|}. $$

The residual $h_R - h$ can be used to compute the reconstructed SNR via

$$ \text{SNR} = 20 \log \left( \frac{\|h\|}{\|h_R - h\|} \right). $$

We will look for this reconstructed SNR to produce a gain over the SNR of the added Gaussian noise as an indication that the BPD noise reduction process has been successful.

Section III-A presents denoising applied to a synthetically generated time series using a specially constructed ESP frame. Section III-B presents a similar denoising analysis, but applied to experimentally collected time series. Finally Section III-C describes a comparative analysis between ESP frame denoising and STFT based denoising.

### A. Synthetic Data

Consider an ESP frame engineered to detect resonance frequencies, such as those from the transfer function $H : \mathbb{C} \to \mathbb{C}$ defined by

$$ H(z) = \frac{z - 1}{(z - \alpha)(z - \beta)(z - \beta)}, $$

$$ \alpha = -1/\tau_a + 2\pi j f_a, \quad \beta = -1/\tau_b + 2\pi j f_b, $$

where $f_a = 5\text{kHz}$, $\tau_a = 3\text{ms}$, $f_b = 13\text{kHz}$, and $\tau_b = 0.8\text{ms}$. The synthetic signal $h$ was generated with $N = 1000$ samples and sampling frequency $f_s = 100\text{kHz}$ by applying this transfer function to the Kronecker delta vector $v$ where $v_0 = 1$ and $v_i = 0$ for $i \neq 50$ (visualized in the top-left subplot of Figure 2). The impulse is chosen so that the synthetic signal starts at 0.5ms. It can be shown via partial fractions [24] that $h$ is a combination of shifted exponentially decaying sinusoids.

This suggests that an ESP frame constructed from exponential envelopes would be appropriate for analysis of this signal.

Given a sampling frequency $f_s$ define the ESP envelopes $e_l$ such that

$$ e_l[n] = \exp \left( -n/(f_s \tau_l) \right) $$

for $\tau_l > 0$. These envelopes are parameterized by the time constants $\tau_l$ and we will use them to construct an ESP frame with atoms consisting of shifted, exponentially decaying, sinusoids. If

$$ \tau_l = 10^l/5 - 4 \text{ for } l = 0, \ldots , 8 $$

then $L = 9$ and the time constants range from 0.1ms to 10ms.

From (4) we construct the ESP frame atoms

$$ a_{l,k,m}[n] = c_l \exp \left( -\frac{n - m \mod N}{f_s \tau_l} + 2\pi j k(n - m) \right) \frac{1}{N} $$

where the $c_l$ are chosen so that $\|e_l\| = (NL)^{-1/2}$. The vectors $a_{l,k,m}$ form a Parseval frame by Theorem 1. With this configuration there are $N^2L = 11$ million frame functions parameterized by time constant (0.1ms - 10ms), modulation frequency ($-50\text{kHz} - 50\text{kHz}$), and circular time shift (0ms - 5ms).

Since ESP frame coefficients are three dimensional they can be difficult to visualize. In order to produce something similar to a time/frequency analysis (TFA) plot we take the Maximum Intensity Projection (MIP) along the envelope axis and plot the magnitude of the coefficients as functions of the time shift and frequency modulation parameters. As an example, the resulting image for the unregularized ESP frame coefficients of $h$ is...
shown in Figure 1. The most important fact about the ESP envelope MIP is that, unlike a traditional TFA, it does not show the amount of signal energy present at a particular time and frequency. Instead it shows the amplitude of a coefficient associated to a particular time shift and frequency modulation of a frame envelope. Indeed, the unregularized ESP frame coefficients behave similarly to a matched filter bank in that the coefficient power will generally have a maximum when the time shift and modulation frequency causes an envelope to correlate to a component in the underlying signal. For instance, the two main frequency peaks of $h$ are clearly visible in Figure 1, while the time shift peaks are present but broader. As expected, the frequency with the smaller time constant has less power than the frequency with the longer time constant.

**BPD on Synthetic Data:** Moving on to sparse representations of $h$, Figure 2 shows the BPD coefficient MIP and corresponding reconstruction of a noisy version $h_N$ of the synthetic signal. The noisy signal was generated using an SNR of 10dB, and 1000 iterations of BPD were computed with $\lambda = 0.1\lambda_{\text{max}}$. The computed sparse coefficient vector has 2008 nonzero ESP frame coefficients with a sparsity of 99.98%, all clustered around the correct frequencies and time shift. The relative error between the regularized reconstructed signal and the pure signal is 17.4% with a reconstructed SNR of 15.2dB, a 5.2dB gain from the initial SNR. The left subplot of Figure 2 reveals that the reconstructed signal has less additive noise, particularly later on in the time series.

Using Algorithm 1 with the STFT frame in place of the ESP frame produces the coefficients shown in Figure 3. The $L_1$-parameter $\lambda = 0.1\lambda_{\text{max}}$ was set using the $\lambda_{\text{max}}$ corresponding to the STFT, and 1000 iterations were used. The solution has 58 nonzero coefficients with a sparsity of 97.41%. The relative error between the regularized reconstructed signal and the pure signal is 24.8% with a reconstructed SNR of 12.8dB. The increase in reconstructed SNR is 3dB smaller than the ESP frame case. Of course, this is for single noise level and choice of $\lambda$. A more detailed comparative analysis is presented in Section III-C

**B. Experimental Data**

The datasets for this section were collected by tapping an 8-inch long, 2-inch diameter steel cylinder and wooden cylinder with an impact hammer and recording the emitted sound. The recordings were taken with a sampling frequency of $f_s = 16$ kHz, after downsampling. The hammer was outfitted with a force sensor which triggered the time series to start recording at the moment the hammer impacted the cylinder. Overall, 2,200 samples were taken over 0.55 seconds for each tap. The normalized spectral power densities for both cylinders are displayed in Figure 4. The densities were computed using the first 250ms of the corresponding time series.

In order to avoid any transient effects of the impact, the signal used for analysis starts at the 100th sample, with $N = 1024$ samples from the starting point. Since the signal is decaying, the same exponential envelopes $e_t$ defined in (11)
are appropriate for this analysis, with time constants
\[ \tau_l = 10^{l/4-3} \text{ for } l = 0, \ldots, 10, \]
so \( L = 11 \) and the \( \tau_l \) range from 1ms to 316ms. About 11.5 million ESP frame vectors follow via (12), and are parameterized by time constant (1ms - 316ms), modulation frequency (−8kHz - 8kHz), and time shift (0ms - 70.25ms).

The unregularized steel and wood cylinder ESP coefficient envelope MIPs are displayed in Figure 5. The main response in the 4.5kHz frequency range is clearly visible in both sets of coefficients, as is the secondary frequency response in the wood cylinder time series. However, the coefficients are spread across the entire time shift axis, which is in conflict with the physical setup of the experiment.

**BPD on Experimental Data:** Next we apply regularization to the experimental time series. Similar to the synthetic case, noisy signals \( \mathbf{h}_N \) were generated by adding white Gaussian noise at 10dB SNR. Figure 6 shows the resulting BPD regularized ESP coefficients, using 1000 iterations and \( \lambda = 0.1\lambda_{\text{max}} \) as before. The resulting coefficients for the steel cylinder data have a relative reconstruction error of 17.6% with a reconstructed SNR of 15.1dB, a 5.1dB gain. The regularized coefficients have a sparsity of 99.95% with 5745 nonzero coefficients. The coefficients for the wood cylinder data have a relative reconstruction error of 27.6% with a reconstructed SNR of 11.2dB. The regularized coefficients have a sparsity of 99.97% with 2998 nonzero coefficients. The primary frequency responses for both objects are visible, as is the secondary frequency response for the wood cylinder. For the steel cylinder data the frame coefficients are spread across the entire time shift axis, which is somewhat inconsistent with our expectations for the physics of the experiment and may represent intrinsic (non-Gaussian) experimental noise.

Since the regularization process prioritizes sparsity, it tends to filter out both the noise and low magnitude resonant components. This can be seen for both data sets. All of the frequency components are either clustered around the main frequency peaks or are very low frequency, so that quieter frequency responses (such as the small peak at 2.43kHz in the steel cylinder data) have been filtered out. A promising observation is that the reconstructed signal for the wood cylinder data seems to have very few high frequency components after 30ms. This is consistent with the physics of the experiment since the wood cylinder resonant frequencies have short time constants.

For comparison the regularized STFT frame coefficients are displayed in Figure 7. They are computing using 1000 iterations of BPD and \( \lambda = 0.1\lambda_{\text{max}} \). The STFT coefficients exhibit very little power outside the primary frequency responses. The steel cylinder coefficients have a reconstruction error of 17.3% with a reconstructed SNR of 15.2dB and a sparsity of 97.3% with 68 nonzero coefficients. The wood cylinder data has a reconstruction error of 38.6% with a reconstructed SNR of 8.2 dB, a 1.8 dB loss, and a sparsity of 99.1% with 24 nonzero coefficients.

For the steel cylinder data, the STFT frame performance is comparable to that of the ESP frame, with the caveat that the actual number of nonzero coefficients is much smaller. This is less true for the wood cylinder data, where the reconstructed SNR goes from a small gain for the ESP frame to a small loss for the STFT. The STFT frame also visually does a poorer job of reconstructing the time series, particularly the late-time wave shape (see bottom left plot in Figure 7).

### C. Denoising Analysis

In the previous sections \( \lambda \) and SNR were fixed. In order to illustrate a broader picture of the sparsity/reconstruction tradeoffs for ESP and STFT frames Figure 8 displays BPD results applied to a variety of initial SNRs and \( \lambda \).s. The marker style denotes the algorithm, and colors denote a fixed SNR level (−30dB, −15dB, 0dB, 15dB, 30dB). A range of percentages of \( \lambda_{\text{max}} \) were computed for each algorithm. Each resulting coefficient vector is a point on the graph, where its position along the \( x \)-axis corresponds to its percentage of nonzero coefficients. Note \( \lambda_{\text{max}} \) is computed separately for each frame using (9), and the \( y \)-axis is the reconstructed SNR.
Reconstructed SNR (dB) vs. Frequency (kHz) for steel cylinder (top) and wood cylinder (bottom) tap data after signal (left) and scaled color plots of the STFT frame coefficients (right) for synthetic data.

Fig. 7. Plots of $b, b_N$ with 10dB SNR and the reconstructed regularized signal (left) and scaled color plots of the STFT frame coefficients (right) for steel cylinder (top) and wood cylinder (bottom) tap data after 1000 iterations with $\lambda = 0.1\lambda_{\text{max}}$. Intensities are shown on a dB scale relative to the maximum frame coefficient amplitude.

Reconstructed SNR vs. Percent Nonzero Coefficients for synthetic data.

Fig. 8. Reconstructed SNR vs. percentage of nonzero coefficients for the synthetic signal (top), steel cylinder time series (bottom left) and wood cylinder time series (bottom right) using both an ESP frame and a STFT frame over SNR ranging from −30dB to 30dB. The BPD algorithm was applied to the noisy time series using 1000 iterations and $\lambda$ ranging logarithmically from 0.00001$\lambda_{\text{max}}$ to $\lambda_{\text{max}}$. Larger $\lambda$-values occur to the left of smaller $\lambda$-values.

|                     | 0dB SNR | 15dB SNR | 30dB SNR |
|---------------------|---------|----------|----------|
| Synthetic ESP Gain  | 11.9    | 9.5      | 8.5      |
| Synthetic STFT Gain | 8.8     | 5.4      | 3.4      |
| Steel ESP Gain      | 13.2    | 7.4      | 1.8      |
| Steel STFT Gain     | 8.8     | 5.5      | 1.3      |
| Wood ESP Gain       | 9.5     | 4.7      | 1.0      |
| Wood STFT Gain      | 8.3     | 3.7      | 0.7      |

Table I. Maximum Reconstructed SNR gain for signals denoised using BPD with ESP and STFT frames.

In general, the STFT frame has a greater percentage of nonzero coefficients than the corresponding ESP frame coefficients, generally falling in the 50-60% range. This is expected because the ESP frame is highly overdetermined and therefore has many more coefficients overall. For small $\lambda$ the percentage of nonzero ESP frame coefficients ranges from approximately 15% for large SNR to approximately 30% for low SNR.

For both frames, when the SNR of the noisy signal is less than 0dB the reconstructed SNR increases (and the number of nonzero coefficients decreases) as $\lambda$ grows until the BPD solution becomes zero at $\lambda_{\text{max}}$. This indicates the denoising process was unsuccessful. Conversely, when the SNR of the noisy signal is greater than or equal to 0dB the reconstructed SNR increases as $\lambda$ grows until it hits some maximum, and then decreases until it reaches zero. In this case, the regularization process is initially removing more noise power than signal power, resulting in an SNR gain, until it hits some optimal $\lambda$ that depends on the frame and the signal. After this optimal value, further increases in $\lambda$ cause the regularization process to reduce both noise and signal power until the BPD solution becomes zero at $\lambda_{\text{max}}$.

The gain of the optimal reconstructed SNR over the SNR of the noisy signal is indicative of the frame’s ability to denoise and is shown in Table I. The key takeaway is that the optimal SNR gain is consistently larger for the ESP frame than it is for the STFT frame. This is true across all three signals for the range of noise SNR’s which produced successful denoising. This maximum SNR gain is greater than indicated by the examples in Sections III-A and III-B, particularly in the case of the wood cylinder data.

Overall we find that the ESP and STFT frames have similar performance with regards to denoising. However, the ESP frame based denoising produces gains which range from 0.3dB to 5.1dB higher than the STFT frame. The ESP frame gain is better for the synthetic signal than for the experimental signals as the ESP frame vectors are more closely aligned to the time series in the synthetic case.

IV. PARAMETER ESTIMATION

In addition to denoising, another important application of ESP frames is parameter estimation. Since the unregularized ESP frame coefficients are correlation based we expect the coefficients to have peaks when the frame vector is well matched with the signal in question. Our approach will be to use the parameter values associated to these peaks as estimates for feature parameters in the underlying signal. While this
produces an unbiased estimation in the case of a single atom by the Cauchy-Schwartz inequality, in the case of signals with multiple components this estimate is not necessarily unbiased. If the frame is constructed so that the signal is known to have a sparse frame representation then this issue can be addressed via $L_1$-regularization. Importantly, if the signal does not have a sparse frame representation then the $L_1$-regularization process can introduce significant bias into the parameter estimation process, as we see below.

For this section’s analysis we will focus on the identification of resonances in the signals presented in Section III. We use the frames discussed in that section to estimate resonance frequencies and time constants with both unregularized and sparse coefficients. Since the resolution on the time constant axis is poor we will interpolate to get more precise estimates. As the time constants are sampled on a logarithmic scale we use a geometric average weighted by the coefficient amplitudes

$$\tilde{\tau} = \left(\frac{\sum_{m}^{} |c_{k,l,m}| \tau_{l_{1}}}{\sum_{m}^{} |c_{k,l,m}| \tau_{l_{2}}}ight) \beta \text{ s.t.} \quad (13)$$

We compare the performance of our ESP frame based estimates to Prony’s Method [1], a least squares regression based approach for estimating decaying resonances. Prony’s Method assumes the signal can be modeled as a sum of exponentially decaying sinusoids which start at time zero. When this assumption holds, and the order of the least squares regression matches the number of poles in the signal, Prony’s Method is capable of producing extremely accurate estimates. However, it is also known that noise and late starting signals can adversely affect Prony’s Method. To address these issues, in the case of noise we will be utilizing the SVD-based noise reduction techniques described in [1, Section 11.9]. For late starting signals we will utilize a time shift to ensure the exponential decay starts at time zero. Unlike the ESP frame approach this requires us to know, or estimate, the number of poles and the start of the exponential decay. It is not expected that ESP frames will outperform Prony’s Method in terms of accuracy in optimum conditions. Instead however, the intention is to utilize ESP frames on signals where the number of poles or the start of the resonance component is not known a priori.

Section IV-A presents examples of estimating resonance parameters using synthetic time series, as well as a comparison between the parameter estimation performance of ESP frames and Prony’s Method. Section IV-B presents a similar analysis using experimental time series and also discusses the use of weighted BPD to encode prior knowledge when generating sparse coefficients.

A. Synthetic Time Series

For the synthetic time series presented in Section III-A we know that $h$ contains two resonances with frequencies and time constants of 5kHz and 3ms, and 13kHz and 0.8ms that both start at 0.5ms. Using Prony’s Method on a shifted version of the clean signal with 4 poles we can recover the parameters for each resonance exactly. However, if we apply Prony’s Method directly to $h$ without the shift we find estimates of 5.03kHz and 4.46ms, and 12.86kHz and 1.21ms. For comparison if we use the unregularized ESP frame coefficients we get 5kHz and 13kHz and (13) we estimate the frequency and time constants of the two resonances to be 5kHz and 2.52ms and 13kHz and 0.63ms. Here we have recovered the frequency components exactly and the error in the time constants is better than when Prony’s Method is applied without a shift. That being said, for all future Prony’s Method estimates we will apply any shifts necessary to ensure optimum performance.

While the estimates above were taken from a clean signal, we are generally interested in parameter estimation in the presence of noise. If we add noise at 10dB SNR, as described in Section III, and use Prony’s Method (with shifting and 30 poles filtered to 4 using SVD) to estimate the resonance parameters we get 5.01kHz and 2.50ms, and 12.95kHz and 0.55ms. For comparison if we compute the BPD regularized ESP frame coefficients as in Figure 2 we obtain estimates of 5kHz and 2.48ms, and 13kHz and 0.95ms. In this case the regularized BPD approach does a slightly better job of estimating the frequency parameters and is about as accurate as Prony’s Method at estimating the time constants.

Noise Analysis: While the examples above indicate that ESP frames can be reasonably utilized as a parameter estimation tool, a further comparative analysis with Prony’s Method is warranted. Specifically we wish to compare the bias and variance of ESP frame and Prony’s Method based parameter estimates in the presence of added noise. To this end noise was added to the synthetic time series at levels ranging from −15dB SNR, resulting in predominantly noise, to 30dB SNR, resulting in predominantly signal. At each noise level the frequency and time constant of both resonance peaks was estimated using unregularized ESP frame coefficients, sparse ESP frame coefficients, and Prony’s Method. For the unregularized ESP frame coefficients the parameter estimates were generated using the coefficient peaks near 5kHz and 13kHz and (13). The BPD regularized coefficients were generated using $\lambda = 0.1 \lambda_{\text{max}}$. 1000 iterations, and the same parameter estimation process as the unregularized case. Finally for Prony’s Method we use a shifted version of the noisy time series with 30 poles filtered to 4 via SVD. Each of these estimates was generated for 100 noise realizations and the resulting estimation mean and standard deviation are plotted in Figure 9 with bias and standard deviation values at the 30dB level shown in Table II.

The frequency estimates have similar performance, for both bias and standard deviation, across all three estimation techniques and produce quality estimates at or above 0dB SNR. The one exception is the Prony’s Method 13kHz peak frequency estimate at 0dB SNR, which has a notably larger bias and standard deviation. There is comparatively more variability in the time constant estimates. None of the methods produces viable estimates at the -15dB SNR level. At 0dB SNR the ESP frame based estimates are significantly better.
Mean and Std. Dev. for Synthetic Data 5kHz Peak Frequency Estimates

Mean and Std. Dev. for Synthetic Data 13kHz Peak Frequency Estimates

Mean and Std. Dev. for Synthetic Data 5kHz Peak TC Estimates

Mean and Std. Dev. for Synthetic Data 13kHz Peak TC Estimates

Fig. 9. Means and standard deviations for ESP frame based and Prony’s Method based estimates of resonance frequency (top) and time constant (bottom) for the synthetic time series 5kHz peak (left) and 13kHz peak (right). Mean is indicated by plotted point and standard deviation by the length of the whiskers. True parameter values are indicated by the dashed line. The mean and standard deviation were computed using 100 estimates from signals with added noise at the indicated SNR.

|                | 5kHz Res. Freq. | 13kHz Res. Freq. | 5kHz Res. TC | 13kHz Res. TC |
|----------------|-----------------|-----------------|--------------|--------------|
| **ESP Bias**   | 0.00116629      | 0.00116629      | 0.00116629   | 0.00116629   |
| **ESP Std. Dev.** | 0.00705890       | 0.00705890       | 0.00705890   | 0.00705890   |
| **BPD ESP Bias** | 0.00034626       | 0.00034626       | 0.00034626   | 0.00034626   |
| **BPD ESP Std. Dev.** | 0.0002189        | 0.0002189        | 0.0002189    | 0.0002189    |
| **Prony Bias** | -0.000141        | -0.000141        | -0.000141    | -0.000141    |
| **Prony Std. Dev.** | 0.0003243         | 0.0003243         | 0.0003243    | 0.0003243    |

Table II. Bias and standard deviations for ESP frame and Prony’s Method parameter estimates at 30dB SNR. Frequency values are in kHz and time constant (TC) values are in ms.

than the Prony’s Method based estimate. At 15dB SNR all three methods have similar performance while the Prony’s Method estimates are significantly better at the 30dB SNR level, as can be seen in Table II.

B. Experimental Time Series

While we know the true values of the resonant frequencies and time constants for the synthetic data, and were able to precisely control the amount of added noise, we do not have this luxury for the experimental time series. Instead we will use a Prony’s Method based estimate created using the original experimental time series as the “true” value for the primary resonance time constant for the steel and wood cylinder data. Using Prony’s Method with 16 poles filtered to 8 using SVD we estimate the steel cylinder resonance at 4.80kHz has a time constant of 166.8ms and the wood cylinder resonance at 4.32kHz has a time constant of 5.22ms. As a point of comparison, if we use the unregularized ESP frame coefficients, Figure 5, to estimate the frequency and time constant we get 4.80kHz and 177.87ms for the steel cylinder data and 4.32kHz and 5.66ms for the wood cylinder data. There is good agreement between the frequency estimates and the time constant estimates are reasonably close. This is a positive indication that we will be able to utilize the unregularized coefficients for parameter estimation. On the other hand, if we apply the same estimation process to the BPD regularized coefficients, Figure 6, we get an estimate of 4.80kHz and 54.95ms for the steel cylinder and 4.33kHz and 5.72ms for the wood cylinder. While the wood cylinder estimate is similar in both cases, the time constant for the steel cylinder estimate is significantly biased as a result of the regularization.

Weighted Basis Pursuit Denoising: One potential method for dealing with the bias introduced by regularization, and to account for the fact that the regularization does not produce coefficients consistent with our understanding of the physics of the experiment, is to utilize a vectorized λ in Algorithm 1. We know from the physical setup of the experiment that, outside of noise components, the signal should have a zero time shift. We can encode this prior knowledge into the BPD process by nonconstant λ. Specifically define w such that

\[
  w_{l,k,m} = \begin{cases} 
  0.1 & m < 10 \\
  0.2 & m \geq 10.
  \end{cases}
\]

This vector places extra weight on coefficients with a time shift greater than or equal to 0.156ms, increasing the cost of using these coefficients in the optimization. We allow time shifts of up to \( m = 10 \) so that the signal may start with varying phase. We then define \( \lambda = \lambda_{\text{max}} w \) and apply BPD. Using 1000 iterations produces the set of coefficients shown in Figure 10.

It is clear that the weighting is successful at concentrating the frame coefficients at smaller time shifts. For the wood cylinder most of the power is in the first millisecond, however there are some coefficients at 30ms and 40ms which correspond to similar late time energy seen in the unregularized STFT (not shown). The reconstruction accuracy of...
| Dataset       | Recon. Error (%) | Res. Freq. (kHz) | Res. TC (ms) |
|---------------|------------------|------------------|--------------|
| Steel         | 0.0              | 4.80             | 177.9        |
| BP Steel      | 17.6             | 4.80             | 54.95        |
| WBP Steel     | 24.0             | 4.80             | 177.63       |
| Wood          | 0.0              | 4.32             | 5.66         |
| BP Wood       | 27.6             | 4.33             | 5.72         |
| WBP Wood      | 34.7             | 4.33             | 5.84         |

Table III. Reconstruction error and estimated resonance frequency and time constant for unregularized, BP regularized and weighted BP regularized ESP frame coefficients for steel and wood cylinder time series data.

| Method       | Wood Res. Freq | Steel Res. Freq | Wood Res. TC | Steel Res. TC |
|--------------|----------------|-----------------|--------------|---------------|
| ESP Bias     | 0.000811189    | 0.00081189      | -0.00089914  | 0.026285      |
| ESP Std. Dev.| 0.0            | 0.0             | 0.0016578    | 0.004194      |
| BPD ESP Bias | 0.005283       | 0.0051199       | -0.1796      | -0.21042      |
| BPD ESP Std. Dev | 0.0          | 0.0             | 0.0068052    | 0.030472      |
| Prony Bias   | -0.35157 - 5   | 0.0058036       | -0.98415     | -1.76042      |
| Prony Std. Dev | 0.00052806     | 3.9251e-5       | 0.068351     | 5.3159        |

Table IV. Bias and standard deviations for ESP frame and Prony’s Method parameter estimates at 30dB SNR. Frequency values have are in kHz and time constant (TC) values are in ms.

the weighted basis pursuit coefficients is worse than for the unweighted basis pursuit coefficients, see Table III, due to the increase in the average value of $\lambda$. More importantly the weighting improves the time constant parameter estimation for the steel cylinder to 177.63ms, while the estimate for the wood cylinder resonance time constant is consistent across all estimation techniques.

Noise Analysis: As part our investigation of ESP frames as a parameter estimation tool we performed the same analysis presented in Section IV-A on the experimental time series. Specifically we added noise to the experimental time series at levels ranging from −15dB SNR to 30dB SNR. At each noise level the frequency and time constant of each resonance was estimated using unregularized ESP frame coefficients, weighted BPD regularized ESP frame coefficients, and Prony’s Method. For the unregularized ESP frame coefficients the parameters were estimated using (13). The regularized coefficients were generated using $\lambda = \lambda_{\text{max}} w$ and 1000 iterations, as in the previous section. Finally for Prony’s Method we use 16 poles filtered to 8 via SVD. Each of these estimates was generated for 100 noise realizations and the resulting estimated means and standard deviations are plotted in Figure 11 with the bias and standard deviations at 30dB SNR shown in Table IV. Recall the “true” parameter values were generated using Prony’s Method as described at the start of Section IV-B.

As in Section IV-A all three methods have similar behavior with regards to estimation of the frequency parameter, producing quality estimates at or above the 0dB SNR level. None of the methods produces a viable time constant estimate at -15dB SNR. At 0dB SNR the ESP frame based estimates are reasonably close to the true value while the Prony’s Method estimates are near zero. At the 15dB and 30dB SNR levels the ESP frame based estimates have a significantly smaller bias than Prony’s Method, see Table IV. Overall we find that ESP frames can be used to estimate resonance parameters and are competitive with Prony’s Method, particularly in the presence of noise.

![Fig. 11. Means and standard deviations for ESP frame based and Prony’s Method based estimates of resonance frequency (top) and time constant (bottom) for the steel cylinder (left) and wood cylinder (right) experimental time series. Mean is indicated by plotted point and standard deviation by the length of the whiskers. The best measured value is indicated by the dashed line. The −15dB SNR BPD ESP estimate mean for the wood cylinder data is outside the scale of the plot. Mean and standard deviation were computed using 100 estimates from signals with added noise at the indicated SNR.](image)

V. Discussion

This paper presented a method of constructing Parseval frames from any collection of complex envelopes. The resulting ESP frames can represent a wide variety of signal types as specified by their physical morphology. Since the ESP frame retains its Parseval property it is compatible with large scale and iterative optimization algorithms such as SALSA, and sparse sets of ESP frame coefficients can be generated using traditional convex optimization. This work presented examples of ESP frame generation, as well as $L_1$-regularized coefficient generation, for both synthetic and experimentally measured signals. The use of sparse coefficients for both denoising and parameter estimation was also demonstrated.

When seeking sparse sets of ESP frame coefficients we generally expect the signal will not exactly equal a small linear combination of frame vectors. This can be due to the presence of noise or because of poor resolution in the envelope parameter. While noise can be mitigated using any number of techniques, including BPD, the fact that the dimension of the ESP frame is given by $N^2 L$, where $L$ is the number of envelopes, means that achieving a very fine resolution along the envelope axis can be computationally infeasible. In either case, we expect that the optimal set of BP coefficients may not be particularly sparse and that high levels of reconstruction error may be needed to produce sparse coefficients using BPD. For many applications, though, it is not necessary to achieve true sparsity and instead we simply desire the ESP frame coefficients to produce discrete peaks.

With regards to denoising, we found that ESP frames are competitive with the STFT as a noise reduction tool, producing larger SNR gains over a range of noise levels. While in terms
of percentages the ESP Frame representations were sparser than the STFT representations, because the ESP frame is so large the STFT frame regularization ends up producing significantly fewer nonzero coefficients. Additionally the ESP frame approach also takes longer to converge and is more computationally intensive. In the ideal scenario we expect that ESP frame based denoising will outperform STFT based denoising since the ESP frame can be used to encode a desired signal model while the STFT is signal agnostic.

We also found that ESP frames are competitive with Prony’s Method when applied to resonance parameter estimation. The unregularized ESP frame coefficients perform about as well as the regularized ESP frame based estimates across all test cases while being less computationally intensive. The ESP frame approaches produced viable estimates of the time constant parameter at a lower SNR than Prony’s Method. At very high SNR Prony’s Method produced better estimates in the synthetic time series case, while the ESP frame estimates were better for the experimental time series. This is consistent with the fact that Prony’s Method is very accurate when its underlying signal model is a good match for the time series. It is not thought that ESP frames will outperform Prony’s Method in terms of accuracy in optimal conditions. Instead however, the intention is to utilize ESP frames on signals where the number of poles or the start of the resonance component is not known *a priori*.

There are a number of possible future applications for ESP frames, ranging from Morphological Component Analysis to generating feature sets for use in signal classification [19]. Another potential avenue of investigation is to try and allow the envelope parameter to vary as part of the $L_1$-regularization procedure. This could enable the ESP frame envelopes to be more data informed and may further enhance sparsity. Overall, Enveloped Sinusoid Parseval frames are a flexible signal analysis tool, particularly when combined with convex optimization, and offer a wide range of applications. The ESP frame can be easily tuned to represent a wide variety of signals simply by providing their relevant envelopes. It is compatible with modern convex analysis techniques, and is efficient and practical to deploy.

**ACKNOWLEDGMENT**

This work was sponsored in part by the Department of the Navy, Office of Naval Research under ONR award numbers N00014-18-1-2820 and N00014-19-1-2221.

**REFERENCES**

[1] S. L. J. Marple, *Digital Spectral Analysis*. Mineola, New York: Dover Publications, Inc., 2019.

[2] I. W. Selesnick, “Wavelet transform with tunable Q-factor,” IEEE Transactions on Signal Processing, vol. 59, no. 8, pp. 3560–3575, 2011.

[3] P. Flandrin, G. Rilling, and P. Goncalves, “Empirical mode decomposition as a filter bank,” IEEE Signal Processing Letters, vol. 11, no. 2, pp. 112–114, 2004.

[4] G. Cai, I. W. Selesnick, S. Wang, W. Dai, and Z. Zhu, “Sparsity-enhanced signal decomposition via generalized minimax-concave penalty for gearbox fault diagnosis,” Journal of Sound and Vibration, vol. 432, pp. 213–234, 2018.

[5] H. Guo, J. E. Odegaard, M. Lang, R. A. Gopinath, I. W. Selesnick, and C. S. Burrus, “Wavelet based speckle reduction with application to SAR based ATDR,” in Proceedings of 1st international conference on image processing, vol. 1. IEEE, 1994, pp. 75–79.

[6] I. W. Selesnick and M. A. Figueiredo, “Signal restoration with complete wavelet transforms: Comparison of analysis and synthesis priors,” in Wavelets XIII, vol. 7446. International Society for Optics and Photonics, 2009, p. 74460D.

[7] A. Parekh, I. W. Selesnick, D. M. Rapoport, and I. Ayappa, “Detection of k-complexes and sleep spindles (DETKS) using sparse optimization,” Journal of Neuroscience methods, vol. 251, pp. 37–46, 2015.

[8] P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, “Multiple and single snapshot compressive beamforming,” The Journal of the Acoustical Society of America, vol. 138, no. 4, pp. 2003–2014, 2015.

[9] G. Easley, D. Labate, and W. Q. Lim, “Sparse directional image representations using the discrete shearlet transform,” Applied and Computational Harmonic Analysis, vol. 25, no. 1, pp. 25–46, 2008.

[10] J. Ma and G. Plonka, “The curvelet transform,” IEEE signal processing magazine, vol. 27, no. 2, pp. 118–133, 2010.

[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding,” in Proceedings of the 26th annual international conference on machine learning, 2009, pp. 689–696.

[12] B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” Nature, vol. 381, no. 6583, pp. 607–609, 1996.

[13] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA),” Applied and Computational Harmonic Analysis, vol. 19, no. 3, pp. 340–358, 2005.

[14] B. Singh and H. Watanumaga, “A removal of eye movement and blink artifacts from EEG data using morphological component analysis,” Computational and Mathematical methods in medicine, vol. 2017, 2017.

[15] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, vol. 57, no. 11, pp. 1413–1457, 2004.

[16] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell_1$ minimization,” Proceedings of the National Academy of Sciences, vol. 100, no. 5, pp. 2197–2202, 2003.

[17] N. Karpushkin, F. Guber, and A. Ivashkin, “Application of the prony least squares method for fitting signal waveforms measured by sampling ADC,” AIP Conference Proceedings, vol. 2163, no. 030006, 2019.

[18] D. Han, K. Kornelson, D. Larson, and E. Weber, Frames for Undergraduates. Providence, Rhode Island: American Mathematical Society, 2007.

[19] G. Goehle, B. Cowen, T. E. Blanford, J. D. Park, and D. C. Brown, “Approximate extraction of late-time returns via morphological component analysis,” 2022, Journal of the Acoustical Society of America, Submitted.

[20] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, “Fast image recovery using variable splitting and constrained optimization,” IEEE Transactions on Image Processing, vol. 19, no. 9, pp. 2345–2356, 2010.

[21] I. Selesnick, “$L_1$-norm penalized least squares with SALSA,” 2014, [Online]. Available: http://cnx.org/content/m48933/.

[22] T. Goldstein, B. O’Donoghue, S. Setzer, and R. Baraniuk, “Fast alternating direction optimization methods,” SIAM Journal of Imaging Sciences, vol. 7, no. 3, pp. 1588–1623, 2014.

[23] A. Xenaki and Y. Paibhus, “Compressive synthetic aperture sonar imaging with distributed optimization,” Journal of the Acoustical Society of America, vol. 146, no. 3, pp. 1839–1850, 2019.

[24] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1989.