WAVE PHYSICS INFORMED DICTIONARY LEARNING IN ONE DIMENSION

Harsha Vardhan Tetali
Electrical & Computer Engineering
University of Florida
Gainesville
vardhanh71@ufl.edu

K. Supreet Alguri
Electrical & Computer Engineering
University of Utah
Salt Lake City
kssupreet@gmail.com

Joel B. Harley
Electrical & Computer Engineering
University of Florida
Gainesville
joel.harley@ufl.edu

ABSTRACT

Detecting and locating damage information from waves reflected off damage is a common practice in non-destructive structural health monitoring systems. Yet, the transmitted ultrasonic guided waves are affected by the physical and material properties of the structure and are often complicated to model mathematically. This calls for data-driven approaches to model the behaviour of waves, where patterns in wave data due to damage can be learned and distinguished from non-damage data. Recent works have used a popular dictionary learning algorithm, K-SVD, to learn an overcomplete dictionary for waves propagating in a metal plate. However, the domain knowledge is not utilized. This may lead to fruitless results in the case where there are strong patterns in the data that are not of interest to the domain. In this work, instead of treating the K-SVD algorithm as a black box, we create a novel modification by enforcing domain knowledge. In particular, we look at how regularizing the K-SVD algorithm with the one-dimensional wave equation affects the dictionary learned in the simple case of vibrating string. By adding additional non-wave patterns (noise) to the data, we demonstrate that the "wave-informed K-SVD" does not learn patterns which do not obey the wave equation hence learning patterns from data and not the noise.

Index Terms— Dictionary learning, K-SVD, regularization, wave equation, theory-guided data science

1. INTRODUCTION

Strong sensitivity of ultrasonic guided waves to defects in structures and their ability to interrogate across large distances make ultrasonic guided waves a great tool for researchers in structural health monitoring (SHM) [1]. Most SHM systems rely on the theoretical solution of elastic waves propagation in a solid medium. One example of such waves are Lamb waves [2]. SHM systems are generally applied to regular and simple geometries and adapting to more complex geometries has always remained a problem. Irregular and complex geometries pose problems of ultrasonic guided wave reflections that are often very difficult to model. Therefore, traditional localization and characterization techniques[3, 4, 5, 6] that depend on efficient numerical models do not perform well in these circumstances. To address this, a well-followed paradigm is to use numerical techniques to simulate guided waves and create a "predictive" model mimicking the experimental setup of a guided wave system. This numerically simulated data can then be compared with the data obtained from experiments to identify any damage features. This is the standard strategy for matched field processing localization methods, which are also widely used in underwater acoustics [7] and RADAR [8]. Various numerical techniques have been proposed to simulate the experimental data. However, the mismatch between a "predictive" model and experimental data is a large gap to fill in, which consequently makes the model "unpredicitve" and therefore useless for damage prediction.

In recent work, the emergence of data-driven models find a role to play in filling this vacuum of experimental uncertainty. Popular among the prior works is the sparse wavenumber analysis (SWA) [9] which creates a predictive model by combining experimental data with the analytic solutions for Lamb waves in various types of media (isotropic and anisotropic waves [10], for example). SWA also relies on an analytic model, thus predicting reflections from complex geometries cannot be reliable. In other work, a completely data-driven approach of creating a dictionary (a set of signals whose linear combinations can recreate the guided waves) based on numerical simulations is studied for a planar metallic plate [11]. This is accomplished with dictionary learning algorithms [12]. This dictionary is then used with experimental data to create a predictive model that extrapolates wave behaviour within a structure. It has also been observed that the dictionary can in the right circumstances match the theoretical dictionary that theory predicts.

Given the developments to date, we have two extrema: at one end, there is a complete theoretical dictionary with no role of data in determining the dictionary, and on the other end a

978-1-7281-0824-7/19/$31.00 ©2019 IEEE

dictionary is learned from data with no role of theory. In this work, we take a step to gain understanding of the stages where a dictionary can be developed with utilization of both theory and data. We take a step in unwrapping the popular dictionary algorithm, K-SVD [13], and integrate specific domain knowledge, in this case, the wave equation, to create a dictionary that is restricted to the particular domain. We do this by adding a wave equation based regularizer. Ref. [14] has an interesting survey of methods to combine domain knowledge with data for theory-guided data science. Ref. [15] also discusses methods for introducing prior knowledge in dictionary learning algorithms. We do this for the simple case of a fixed string. In doing so, we contrast the results obtained using K-SVD algorithm with the modified algorithm, which we call the wave-informed K-SVD. We train dictionaries using both the algorithms: K-SVD and wave-informed K-SVD on wave data corrupted by white Gaussian noise. We observe that wave-informed K-SVD does not learn noisy atoms, while simple K-SVD learns atoms corrupted by noise.

2. NOTATION

In this paper, boldfaced letters $\mathbf{Y}$ represent a matrix and the representation $\mathbf{Y}_i$ represents the i-th columns and $\widehat{\mathbf{Y}}_j^T$ represents the j-th row of the matrix $\mathbf{Y}$. A matrix in the t-th iteration in a sequence of matrices updating through an algorithm is denoted as $\mathbf{D}^{(t)}$. $\|\mathbf{Y}\|_F$ denotes the Frobenius norm of the matrix $\mathbf{Y}$. Similarly $\|\mathbf{x}\|_p$ of a vector $\mathbf{x}$ represents the number of the p-th norm of $\mathbf{x}$.

3. DICTIONARY LEARNING: A BRIEF OVERVIEW

The role of dictionary learning in machine learning and signal processing can be seen to be similar to that of the role of Fourier analysis. Decomposing a function into a scaled sum of basic functions helps in obtaining insights not visible from the function itself. Incidentally, one such example is the Fourier transform.

Dictionary learning differs from ordinary Fourier analysis in the sense that the basis vectors are not fixed (as they are in the case of discrete Fourier transform), and are learned from the data. Another added feature is sparse representations. In dictionary learning, the bases are learned in a way that the linear combination utilizes the least number of bases for the representation of each signal. This added flexibility comes from the fact that the dictionary matrix need not be a square matrix and is also not orthogonal. We can generalize to the idea of dictionary learning from the discrete Fourier transform.

Let columns of the matrix $\mathbf{Y}$ contain a particular signals of interest. Let $\mathbf{W}$ be the DFT matrix compatible in size with $\mathbf{Y}$. Also let,

$$\mathbf{X} = \mathbf{W}\mathbf{Y} \tag{1}$$

Thus, $\mathbf{Y} = \mathbf{W}^{-1}\mathbf{X}$, since the DFT matrix is invertible. In the dictionary learning perspective, the matrix $\mathbf{W}^{-1}$ can be considered a dictionary and the matrix $\mathbf{X}$ as the corresponding coefficient matrix.

When a dictionary is learned from data, through a dictionary learning algorithm, to satisfy the sparsity conditions, the dictionary is often overcomplete, thus non-square. In summary, a general dictionary learning algorithm decomposes a data matrix $\mathbf{Y}$ into a product of matrices $\mathbf{D}$ and $\mathbf{X}$ (mathematically, $\mathbf{Y} = \mathbf{D}\mathbf{X}$), where $\mathbf{D}$ is the named as the dictionary matrix and $\mathbf{X}$, with a desired level of sparsity, is called the coefficient matrix.

4. THE K-SVD ALGORITHM

In general, dictionary learning is the process of learning an overcomplete dictionary to represent a signal as sparsely as possible. The elements of the dictionary are called atoms and they need not be orthogonal. The signal on which the dictionary is trained can be represented as sparse combination of the atoms in the dictionary. The most common application of dictionary learning is in compressed sensing or signal recovery. Various algorithms have been developed to solve dictionary learning. Below, we have the dictionary learning algorithm setup as an optimization problem:

$$\min_{\mathbf{X},\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \text{ subject to } \|\mathbf{X}_i\|_0 \leq s \tag{2}$$

One such algorithm is K-SVD [13]. In various applications (e.g. [16], [17]), better results have been obtained from this algorithm as compared to any other static dictionary. The K-SVD algorithm comprises of two steps:

1. The sparse coding step
2. Dictionary update step

These two steps are performed consecutively for each iteration, either until convergence or up to a specified number of iterations. We follow the later in this work. In this paper, we use Orthogonal Matching Pursuit (OMP) in the sparse coding step as it is usually used when the sparsity is known beforehand and incidentally assuming sparsity serves our purpose. We refer the reader to [18] and [19] for OMP. Below, we review the dictionary update step in detail. Note that we rigorously review K-SVD to provide the necessary background to understand our derivation of wave informed K-SVD.

4.1. Dictionary update step

In this part of the algorithm, we estimate a new $\mathbf{D}$ based on the computed $\mathbf{X}$ and the data $\mathbf{Y}$. We consider updating the kth column of the dictionary matrix $\mathbf{D}$ with all other columns undisturbed, and we iteratively do this for each k. We write the objective function so that the kth column of $\mathbf{D}$ is updated keeping the rest of the columns constant, we refer the reader to [13] for more details

$$\|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 = \|\mathbf{Y}-\sum_{j=1}^{K}\mathbf{D}_j\widehat{\mathbf{X}}_j^T\|_F^2 \quad (3)$$

$$\|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 = \|\left(\mathbf{Y}-\sum_{j\neq k}\mathbf{D}_j\widehat{\mathbf{X}}_j^T\right)-\mathbf{D}_k\widehat{\mathbf{X}}_k^T\|_F^2 \quad (4)$$

Now letting $\left(\mathbf{Y}-\sum_{j\neq k}\mathbf{D}_j\widehat{\mathbf{X}}_j^T\right)=\mathbf{E}_k$, we have:

$$\|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 = \|\mathbf{E}_k-\mathbf{D}_k\widehat{\mathbf{X}}_k^T\|_F^2 \quad (5)$$

Since the columns of $\mathbf{X}$ are sparse, we expect the vector $\widehat{\mathbf{X}}_k$ to have zeros and the product $\mathbf{D}_k\widehat{\mathbf{X}}_k^T$ will have zero columns. We form a new matrix $\widetilde{\mathbf{E}}_k$ from the matrix $\mathbf{E}_k$ by removing columns of $\mathbf{E}_k$ whose corresponding columns in $\mathbf{D}_k\widehat{\mathbf{X}}_k^T$ are zero. We do this to ensure the sparse structure of matrix $\mathbf{X}$ obtained from orthogonal matching is not lost. Similarly we also remove the zeros in $\widehat{\mathbf{X}}_k$ and call the new vector $\widetilde{\mathbf{X}}_k$, this ensures that $\widetilde{\mathbf{E}}_k$ and $\mathbf{D}_k\widetilde{\mathbf{X}}_k^T$ have the same dimensions, and that the subtraction is possible. We again refer the reader to [13] for more details. Note that the dictionary update needs to be solved for many iterations. Consider the the t-th iteration. Thus, we write the objective as:

$$\left[\mathbf{D}_k^{(t+1)},\widetilde{\mathbf{X}}_k^{(t+1)}\right] = \underset{d,u}{\text{arg min}}\|\widetilde{\mathbf{E}}_k^{(t)}-\mathbf{d}\mathbf{u}^T\|_F^2 \quad (6)$$

such that $\|\mathbf{d}\|_2 = 1$. Constraining the norm of each dictionary column reduces the search space without loss of generality because an appropriate scaling can be introduced in the corresponding coefficient. Also note that we update the coefficient row vector in the same step as a by-product of dictionary update.

Since constrained optimization problems are easily solved in the dual domain (with Lagrange multipliers), we pose the minimization objective in the dual form. We introduce a Lagrange Multiplier ν for the constraint $\|\mathbf{d}\|_2 = 1$ and form the Lagrangian

$$L(\mathbf{d},\mathbf{u},\nu) = \|\widetilde{\mathbf{E}}_k^{(t)}-\mathbf{d}\mathbf{u}^T\|_F^2+\nu\left(\|\mathbf{d}\|_2^2-1\right) \quad (7)$$

Now the expression above can be reduced to other forms to completely describe the solution of the optimization problem. After using properties of trace of a matrix on the first term in the RHS of (7) and the constraint $\|\mathbf{d}\|_2 = 1$, we obtain

$$\|\widetilde{\mathbf{E}}_k^{(t)}-\mathbf{d}\mathbf{u}^T\|_F^2 = \|\widetilde{\mathbf{E}}_k^{(t)}\|_F^2-2\mathbf{d}^T\widetilde{\mathbf{E}}_k^{(t)}\mathbf{u}+\mathbf{u}^T\mathbf{u} \quad (8)$$

We use (8) in (7) and set $\frac{\partial}{\partial\mathbf{u}}L(\mathbf{d},\mathbf{u},\nu)=0$. By incorporating the constraint, $\|\mathbf{d}\|_2 = 1$, we get

$$\frac{\partial L(\mathbf{d},\mathbf{u},\nu)}{\partial\mathbf{u}} = -2\widetilde{\mathbf{E}}_k^{(t)T}\mathbf{d}+2\mathbf{u}=0 \Rightarrow \mathbf{u}=\widetilde{\mathbf{E}}_k^{(t)T}\mathbf{d} \quad (9)$$

Re-substituting $\mathbf{u}=\widetilde{\mathbf{E}}_k^{(t)T}\mathbf{d}$ into (15) we get the following minimization problem:

$$\underset{\mathbf{d},\|\mathbf{d}\|_2=1}{\text{arg min}}\|\widetilde{\mathbf{E}}_k^{(t)}-\mathbf{d}\mathbf{d}^T\widetilde{\mathbf{E}}_k^{(t)}\|_F^2 \quad (10)$$

We now use some algebra and the constraint to get,
$\|\widetilde{\mathbf{E}}_k^{(t)}-\mathbf{d}\mathbf{d}^T\widetilde{\mathbf{E}}_k^{(t)}\|_F^2 = \text{tr}\left(\widetilde{\mathbf{E}}_k^{(t)T}\widetilde{\mathbf{E}}_k^{(t)}\right)-\mathbf{d}^T\widetilde{\mathbf{E}}_k^{(t)}\widetilde{\mathbf{E}}^{(t)T}\mathbf{d}$
and thus pose the optimization problem to be,

$$\underset{\mathbf{d},\|\mathbf{d}\|_2=1}{\text{arg max}}\;\mathbf{d}^T\widetilde{\mathbf{E}}_k^{(t)}\widetilde{\mathbf{E}}_k^{(t)T}\mathbf{d} \quad (11)$$

The solution to which is readily found in optimization literature (e.g., [20]) as the eigenvector corresponding to the largest eigenvalue of $\widetilde{\mathbf{E}}_k^{(t)}\widetilde{\mathbf{E}}_k^{(t)T}$. This is equivalent to finding the top singular vector of the matrix $\widetilde{\mathbf{E}}_k^{(t)}$. Combining this result with (9), get an expression for $\widetilde{\mathbf{X}}_k^{(t+1)}$. Thus, $\mathbf{D}_k^{(t+1)}$ is the top singular vector of $\widetilde{\mathbf{E}}_k^{(t)}$ and $\widetilde{\mathbf{X}}_k^{(t+1)}=\widetilde{\mathbf{E}}_k^{(t+1)T}\mathbf{D}_k^{(t+1)}$. This minimization problem can be equivalently seen as the rank-1 approximation of $\widetilde{\mathbf{E}}_k^{(t)}$ (which uses singular value decomposition for the approximation) or equivalently, in each iteration $\widetilde{\mathbf{E}}_k^{(t)}$ but we derived it because this will come handy in enforcing the physics constraint. The exact steps are described in Algorithm 1.

Algorithm 1 K-SVD, **Input**: $\mathbf{Y}\in R^{m\times n}, K\in N$

Intialize $\mathbf{D^{(0)}}$, $iter$ (no. of iterations)
Set $t=0$
repeat
  Sparse Code Stage:
  $i=1,2,...,N$; $\min_{\mathbf{X}_i}\{||\mathbf{Y}_i-\mathbf{D}^{(t)}\mathbf{X}_i||_F^2\}$ subject to $||\mathbf{X}_i||_0\leq s$
  Dictionary Update Stage:
  k = 1,2,...,K; $\mathbf{E}_k^{(t)}=\mathbf{Y}-\sum_{j\neq k}\mathbf{D}_j^{(t)}\widehat{\mathbf{X}}_j^{(t)T}$
  Let S contain indices of columns that are non-zero. Now $\widetilde{\mathbf{E}}_k^{(t)}$ is formed from $\mathbf{E}_k^{(t)}$ by selecting columns indicated by S.
  Singular Value Decomposition $\widetilde{\mathbf{E}}_k^{(t)}=\mathbf{U\Delta V}$
  Choose column $\mathbf{D}_k^{(t)}$ to be first column of U
  Update: $\widetilde{\mathbf{X}}_k^{(t)}=\widetilde{\mathbf{E}}_k^{(t)T}\mathbf{D}_k^{(t)}$
  $\widehat{\mathbf{X}}_k^{(t)}$ is constructed from $\widetilde{\mathbf{X}}_k^{(t)}$ by placing the elements of the latter at the indices indicated by S, zeros otherwise.
  $t\leftarrow t+1$
until $t==iter$

5. ENFORCING THE WAVE EQUATION INTO THE K-SVD ALGORITHM

In this work, we assume that data is obtained from a one dimensional medium, such as string. Thus, the one-dimensional wave equation is the physics model that is enforced into the algorithm. The one dimensional wave equation and its Fourier transform in time (if it exists) are given by,

$$\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f(x,t)}{\partial t^2} \iff \frac{\partial^2 F(x,\omega)}{\partial x^2} = \frac{-\omega^2}{v^2}F(x,\omega) \tag{12}$$

where v is the velocity of the wave.

We then discretize everything above and write (12) as a discrete-space matrix form, where $\mathbf{L}$ is the second difference matrix

$$\mathbf{L} = \begin{bmatrix} -2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -2 \end{bmatrix} \tag{13}$$

Thus, for each ω, we can concisely write,

$$\mathbf{Lf} = g\mathbf{f} \tag{14}$$

for a suitable constant g. To enforce structure into the atoms of the dictionary, we impose that each atom of the dictionary approximately satisfies the wave (14), thus we now have an objective function with an added regularization term to the main dictionary learning objective. We choose regularization constants γ_i associated with each dictionary atom. Thus the modified objective function now turns out to be,

$\min\limits_{\mathbf{X},\mathbf{D}}\{||\mathbf{Y}-\mathbf{DX}||_F^2 + \sum_{j=1}^{K}\gamma_j\|\mathbf{LD}_j - g_j\mathbf{D}_j\|_2^2\}$
subject to $||\mathbf{X}_i||_0 \le s$

Note that the sparse coding step does not change while implementing the algorithm. The only change is in the dictionary update step. We can also write the k-th dictionary atom update step in the t-th iteration as,

$$\begin{aligned}&\left[\mathbf{D}_k^{(t+1)}, \widehat{\mathbf{X}}_k^{(t+1)}, g_k^{(t+1)}\right] \\ &\quad = \arg\min_{\mathbf{d},\mathbf{u},g_k}\|\mathbf{E}_k^{(t)} - \mathbf{du}^T\|_F^2 + \gamma_k\|\mathbf{Ld} - g_k^{(t)}\mathbf{d}\|_2^2\end{aligned} \tag{15}$$

such that $\|\mathbf{D}_k^{(t+1)}\|_2 = 1$. We first derive an update rule for each $g_k^{(t)}$. For this we differentiate the function to be minimized in (15) with respect to the scalar $g_k^{(t)}$ at $\mathbf{d} = \mathbf{D}_k^{(t)}$ and $\mathbf{u} = \widehat{\mathbf{X}}_k^{(t)}$ and set it equal to zero to find the updated value $g_k^{(t+1)}$. For ease of differentiation, we choose λ to represent $g_k^{(t)}$, i.e., $\lambda = g_k^{(t)}$. Thus, we have,

$$\frac{\partial}{\partial\lambda}\left(\|\mathbf{E}_k^{(t)} - \mathbf{du}^T\|_F^2 + \gamma_k\|\mathbf{Ld} - \lambda\mathbf{d}\|_2^2\right)\Big|_{\lambda = g_k^{(t+1)}} = 0 \tag{16}$$

Differentiating at $g_k^{(t+1)}$, we obtain,

$$g_k^{(t+1)} = \mathbf{D}_k^{(t)T}\mathbf{LD}_k^{(t)} \tag{17}$$

We next try to optimize for $\mathbf{u}$ and $\mathbf{d}$, differentiating the Lagrangian formed by the objective function defined in (15) with respect to $\mathbf{u}$ and setting it to zero gives the same solution as obtained from (9). Now re-substituting this back in the objective and after dropping out terms that do not depend on our varying quantity, $\mathbf{D}_k^{(t)}$, and rearranging we get the objective to be

$$\arg\min_{\mathbf{D}_k, \|\mathbf{D}_k\|_2=1} \mathbf{D}_k^T\mathbf{BD}_k \equiv \arg\max_{\mathbf{D}_k, \|\mathbf{D}_k\|_2=1} -\mathbf{D}_k^T\mathbf{BD}_k \tag{18}$$

where,

$$B = \gamma_k\left(\mathbf{L} - g_k^{(t+1)}\mathbf{I}\right)\left(\mathbf{L} - g_k^{(t+1)}\mathbf{I}\right)^T - \widehat{\mathbf{E}}^{(t)}\widehat{\mathbf{E}}^{(t)T}$$

Now this implies that the K-SVD algorithm gets modified in the SVD step where we now have to take the top eigenvector of the matrix $\widetilde{\mathbf{E}}_k^{(t)}\widetilde{\mathbf{E}}_k^{(t)T} - \gamma_k(\mathbf{L} - g_k\mathbf{I})(\mathbf{L} - g_k\mathbf{I})^T$ instead of $\widetilde{\mathbf{E}}_k^{(t)}\widetilde{\mathbf{E}}^{(t)T}$ (which is the same as the top left singular vector of $\widetilde{\mathbf{E}}^{(t)}$). We now summarize this in Algorithm 2.

Algorithm 2 wave-informed K-SVD, **Input**: $\mathbf{Y} \in R^{m\times n} K \in N$

Intialize $\mathbf{D}^{(0)}$, $\mathbf{g}^{(0)} = (g_1^{(0)}, g_2^{(0)}, \cdots, g_K^{(0)})$ and $iter$ (no. of iterations)
Set $t = 0$
repeat
  Sparse Code Stage:
  i = 1,2,...,N ; $\min\limits_{\mathbf{X}_i}\{||\mathbf{Y}_i - \mathbf{D}^{(t)}\mathbf{X}_i||_F^2\}$ subject to $||\mathbf{X}_i||_0 \le s$
  Dictionary Update Stage:
  $g_k^{(t)} = \mathbf{D}_k^{(t)T}\mathbf{LD}_k^{(t)}; k = 1, 2, ..., K$
  $\mathbf{E}_k^{(t)} = \mathbf{Y} - \sum_{j\neq k}\mathbf{D}_j^{(t)}\widehat{\mathbf{X}}^{(t)T}_j; k = 1, 2, ..., K$
  Let S contain indices of columns that are non-zero. Now $\widetilde{\mathbf{E}}_k^{(t)}$ is formed from $\mathbf{E}_k^{(t)}$ by selecting columns indicated by S.
  Eigen Value Decomposition of $\widetilde{\mathbf{E}}_k^{(t)}\widetilde{\mathbf{E}}_k^{(t)T} - \gamma_k(\mathbf{L} - g_k\mathbf{I})(\mathbf{L} - g_k\mathbf{I})^T = \mathbf{U\Delta U}^{-1}$
  Choose column $\mathbf{D}_k^{(t)}$ to be first column of U
  Update $\widetilde{\mathbf{X}}_k^{(t)} = \widetilde{\mathbf{E}}_k^{(t)T}\mathbf{D}_k^{(t)}$
  $\widehat{\mathbf{X}}_k^{(t)}$ is constructed from $\widetilde{\mathbf{X}}_k^{(t)}$ by placing the elements of the latter at the indices indicated by S, zeros otherwise.
  $t \leftarrow t + 1$
until $t == iter$

6. SIMULATION RESULTS

In this simulation, we have synthesized data of a string, fixed at both ends, oscillating in a combination of 4 different modes and a single velocity. We also impose an exponential reduction of the wave amplitude with time. Additionally, we corrupt the data with white Gaussian noise, $n(x,t)$, which does not obey the wave equation (12) for real wave parameters. We observe the effect of noise of different SNRs. We also note that the noise introduced in the data is both time and space variant. While SNR is defined in the usual way, $10\log(P_s/P_n)$, where P_s is the signal power and P_n is the noise power, we calculate the power of signal and noise over all space and time. A continuous version of this data is represented in (19). Note that w_k is calculated using $w_k = ck$, where c is the velocity of the wave.

$$y(x,t) = \sum_{k=1}^{4} \sin(kx)\sin(w_k t)e^{-4t} + n(x,t) \quad (19)$$

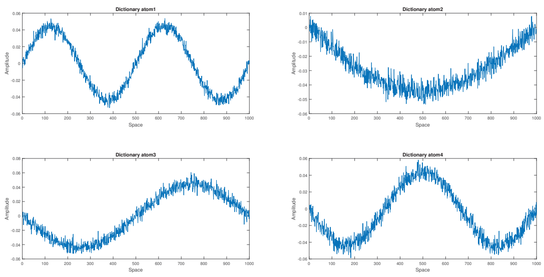

Fig. 1. Dictionary atoms from K-SVD - Experiment where the wave data is corrupted by Gaussian noise of SNR -13 dB

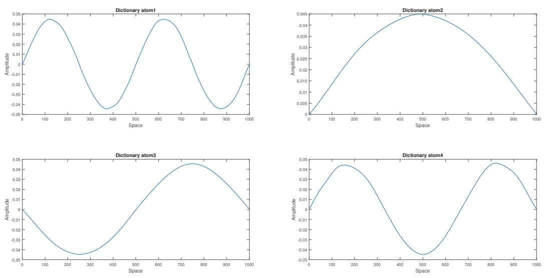

Fig. 2. Dictionary atoms from wave-informed K-SVD - Experiment where the wave data is corrupted by Gaussian noise of SNR -13 dB

A sampled version of $y(x,t)$ defined in (19) is the matrix $\widetilde{\mathbf{Y}}$. The columns of the data $\widetilde{\mathbf{Y}}$ represent the string along space whereas rows represent the string along time. We take the discrete Fourier transform on each row of $\widetilde{\mathbf{Y}}$ to form $\mathbf{Y}$. We now perform K-SVD and wave-informed K-SVD (with number of dictionary elements $K = 4$) and a sparsity of $s = 1$ (for the coefficient matrix $\mathbf{X}$) on $\mathbf{Y}$. Each dictionary atom has a different regularization, γ_k. We chose the $\gamma_k \propto 1/g_k^2$ with a proportionality constant of around 10^5. This proportionality constant is observed to depend on the power of noise present.

Comparing Figure 3 with Figure 4 and Figure 1 with Figure 2, it is clear that K-SVD learns noisy atoms from data whereas the wave-informed K-SVD learns non-noisy versions from data. This justifies the enforcement of wave physics into the algorithm as it signifies that wave-informed K-SVD did not learn noise that does not obey the wave equation (12).

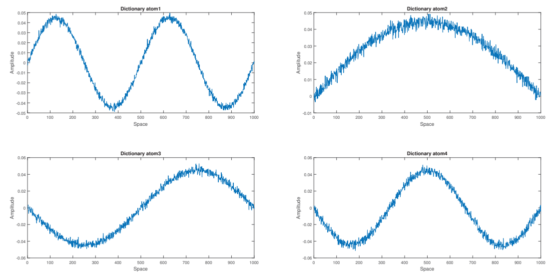

Fig. 3. Dictionary atoms from K-SVD - Experiment where the wave data is corrupted by Gaussian noise of SNR -7 dB

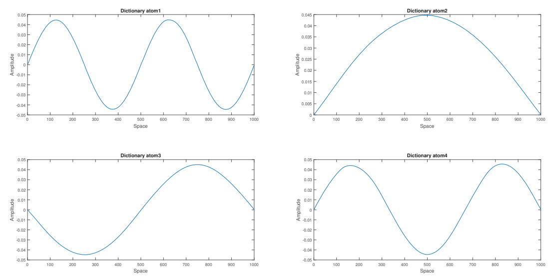

Fig. 4. Dictionary atoms from wave-informed K-SVD - Experiment where the wave data is corrupted by Gaussian noise of SNR -7 dB

7. CONCLUSIONS

In this paper, we showed how physics can be enforced into the popular dictionary learning algorithm, K-SVD, and developed the wave-informed K-SVD algorithm. We can look at this algorithm as a filter that filters signals based on the physical domain they are described by. In future work, we want to reason out the requirement of the large value of regularization constant used. Also, as a next step we would like to verify if wave-informed KSVD works better than KSVD for data undersampled over space. We also want to develop more theoretical understanding of the result being produced by the algorithm and see how it compares with the theoretical solution of the wave equation (a partial differential equation).

8. ACKNOWLEDGEMENTS

This research is supported by the Air Force Office of Scientific Research under award number FA9550-17-1-0126 and the National Science Foundation under award number 1839704.

9. REFERENCES

[1] Joseph L Rose, "Ultrasonic guided waves in structural health monitoring," in *Key Engineering Materials*. Trans Tech Publ, 2004, vol. 270, pp. 14–21.

[2] Horace Lamb, "On waves in an elastic plate," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 93, no. 648, pp. 114–128, 1917.

[3] Victor Giurgiutiu, Jingjing Bao, and Wei Zhao, "Active sensor wave propagation health monitoring of beam and plate structures," in *Smart Structures and Materials 2001: Smart Structures and Integrated Systems*. International Society for Optics and Photonics, 2001, vol. 4327, pp. 234–246.

[4] Seth S Kessler, S Mark Spearing, and Constantinos Soutis, "Damage detection in composite materials using lamb wave methods," *Smart Materials and Structures*, vol. 11, no. 2, pp. 269, 2002.

[5] Jennifer E Michaels, "Detection, localization and characterization of damage in plates with an in situ array of spatially distributed ultrasonic sensors," *Smart Materials and Structures*, vol. 17, no. 3, pp. 035035, 2008.

[6] Joel B Harley and José MF Moura, "Data-driven matched field processing for lamb wave structural health monitoring," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1231–1244, 2014.

[7] Arthur B Baggeroer, William A Kuperman, and Peter N Mikhalevsky, "An overview of matched field methods in ocean acoustics," *IEEE Journal of Oceanic Engineering*, vol. 18, no. 4, pp. 401–424, 1993.

[8] Wen-Qin Wang and Huaizong Shao, "Range-angle localization of targets by a double-pulse frequency diverse array radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 1, pp. 106–114, 2013.

[9] Joel B Harley and José MF Moura, "Sparse recovery of the multimodal and dispersive characteristics of lamb waves," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2732–2745, 2013.

[10] Bertram Alexander Auld, *Acoustic Fields and Waves in Solids*, New York: Wiley, 1973.

[11] K Supreet Alguri, Chen Ciang Chia, and Joel B Harley, "Model-driven, wavefield baseline subtraction for damage visualization using dictionary learning," vol. 2, pp. 2276–2283.

[12] K Supreet Alguri, Joseph Melville, and Joel B Harley, "Baseline-free guided wave damage detection with surrogate data and dictionary learning," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3807–3818, 2018.

[13] Michal Aharon, Michael Elad, Alfred Bruckstein, et al., "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311, 2006.

[14] Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2318–2331, 2017.

[15] Abd-Krim Seghouane and Asif Iqbal, "Basis expansion approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE transactions on medical imaging*, vol. 36, no. 9, pp. 1796–1807, 2017.

[16] Ya Lei, Yong Fang, and Liming Zhang, "A weighted k-svd-based double sparse representations approach for wireless channels using the modified takenaka-malmquist basis," *IEEE Access*, vol. 6, pp. 54331–54342, 2018.

[17] Zhuolin Jiang, Zhe Lin, and Larry S Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[18] Yagyensh Chandra Pati, Ramin Rezaiifar, and Perinkulam Sambamurthy Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*. IEEE, 1993, pp. 40–44.

[19] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[20] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.