

Wave Physics-Informed Matrix Factorizations

Harsha Vardhan Tetali, Joel B. Harley, Benjamin D. Haeffele

Abstract—With the recent success of representation learning methods, which includes deep learning as a special case, there has been considerable interest in developing techniques that incorporate known physical constraints into the learned representation. As one example, in many applications that involve a signal propagating through physical media (e.g., optics, acoustics, fluid dynamics, etc), it is known that the dynamics of the signal must satisfy constraints imposed by the wave equation. Here we propose a matrix factorization technique that decomposes such signals into a sum of components, where each component is regularized to ensure that it nearly satisfies wave equation constraints. Although our proposed formulation is non-convex, we prove that our model can be efficiently solved to global optimality. Through this line of work we establish theoretical connections between wave-informed learning and filtering theory in signal processing. We further demonstrate the application of this work on modal analysis problems commonly arising in structural diagnostics and prognostics.

Index Terms—physics-guided machine learning, dictionary learning, modal decomposition, unsupervised learning

I. INTRODUCTION

REPRESENTATION learning has gained importance in fields that utilize data generated through physical processes, such as weather forecasting [1], manufacturing [2], structural health monitoring [3], acoustics [4], and medical imaging [5]. Moreover, learning physically consistent and interpretable features can improve our understanding of the physically viable information about the data and the composition of the system or process generating it. Unfortunately, however, features of the data learned through generic machine learning algorithms typically do not correspond to physically interpretable quantities.

For example, physically consistent solutions are important to spatio-temporal modal analysis, where recovering modes of vibration yield important information about a physical system or process (e.g., buildings, bridges, space shuttles, satellites, acoustic instruments) [6]–[10]. Furthermore, these problems are also relevant to frequency-wavenumber spectral analysis [11], which is the basis for many advances in RADAR [11], seismic [12], and acoustic [13] beamforming and signal processing, where each sinusoid may correspond to a different source or target. Physically consistent solutions provide a trustworthy basis in terms of the physics of the system in contrast to any other basis, which might be obtained by generic machine learning algorithms. This suggests the need for learning algorithms which choose bases that are physically consistent.

A. Related Work

Modified learning paradigms, suited to different physical application domains, have begun to draw interest [14]. For

example, several researchers have designed physics-informed neural networks to learn approximate solutions to a partial differential equation [15]–[19]. These have been recently applied to ultrasonic surface waves to extract velocity parameters and denoise data [20]. Similar physics-guided neural networks [14], [21]–[23] use physics-based regularization to improve data processing, demonstrating that regression tasks (such as estimating the temperature throughout a lake) can be more accurate and robust when compared with purely physics-based solutions or purely data-driven solutions. These approaches demonstrate the strong potential for physics-informed machine learning but remain early in their study.

Representation theory has been applied to other wave-based problems. For example, generative models have been studied for applications in electroencephalography [24] and seismology [25], although these learning systems do not assume physical knowledge. Other prior work has considered sparse generative representation of ultrasonic waves through the use of sparse signal processing [26], dictionary learning [27]–[29], and neural network sparse autoencoders [30]. These methods utilize the fact that many waves can be represented as a small, sparse sum of spatial modes and attempt to extract these modes within the data [27]. These dictionary learning algorithms have further been combined with wave-informed regularizers to incorporate physical knowledge of wave propagation into the solution [31], [32].

One of the key challenges for many of these problems is that they typically require solving a non-convex optimization problem, and as a result the learned representation can be highly dependent on details such as initialization or choice of optimization algorithm. Attention has been devoted recently to the non-convex optimization problems that arise in representation learning, and positive results have begun to emerge in certain settings. For example, within the context of low-rank matrix recovery, it has been shown that gradient descent is guaranteed to converge to a local minimum and all local minima will be globally optimal under certain circumstances [33]–[36]. Further, other recent work has studied ‘structured’ matrix factorization problems, where one promotes properties in the matrix factors by imposing regularization on the factorized matrices (e.g., if one desires a factor to be sparse an ℓ_1 norm penalty could be imposed on one of the matrix factors). It has been shown that such problems can also be solved to global optimality in certain circumstances, but whether such guarantees can be made depends critically on how the regularization is formulated [37]–[39].

B. Contributions

In this work we make the following contributions. We first introduce a **wave-informed matrix factorization model**,

which ensures that the recovered components minimize wave equation constraints, allowing for direct interpretability of the factors based on assumed physics. This is distinguished from traditional factorization methods, such as principal component analysis (PCA), which are not constrained by physical knowledge as well as from models which explicitly parameterize wave-like factors, and are often too constrained for a problem, particularly in the presence of noise. In other words, our approach balances the utilization of both prior assumptions from physics and fills the gap between purely data driven methods [40] and purely physics based methods.

Second, although our resulting formulation requires solving a *non-convex* optimization problem, we identify its form amenable to be provably solved to global optimality [38]. This differs from prior wave-physics informed dictionary learning [31], which provides no such guarantees.

Third, unlike physics-informed neural networks, we show that our wave physics-informed matrix factorization has a satisfying interpretation as an optimal filter problem. Specifically, we show that our optimization algorithm based on [38] reveals a solution is that of an optimal bank of 1st-order Butterworth filters. This interpretability enables confidence in the algorithm and intelligent selection of hyper-parameters.

In summary, this work's novel contributions are in demonstrating the first framework for matrix factorization that is constrained solely on the wave equation. Unlike wave-informed dictionary learning [31], we show this algorithm is both highly interpretable from a signal processing perspective and satisfies global optimality conditions based on results the literature [38]. Finally, we evaluate the effectiveness of the algorithm in four scenarios – homogeneous vibrations, non-homogeneous vibrations, travelling plane waves, and segmented vibrations. We show that this framework outperforms most other methods used the literature for wave mode decomposition, particularly in the presence of high noise and segmented, spatially varying wave behavior.

II. WAVE-INFORMED MATRIX FACTORIZATION

In this section, we develop a novel matrix factorization framework that allows us to decompose a matrix $\mathbf{Y}$, with each column representing sampling along one dimension (e.g., space), as the product of two matrices $\mathbf{D}\mathbf{X}^\top$, where one of the matrices will be softly constrained to nearly satisfy wave equation constraints. Additionally, our model will learn the number of modes (columns in $\mathbf{D}$ and $\mathbf{X}$) directly from the data without specifying this *a priori*.

In subsection II-A, we introduce our objective function. This subsection can be seen as an extension of the previously proposed wave-informed dictionary learning framework to a matrix factorization framework. In II-B, we go over a few ideas from the optimization literature that are useful to solve the formulated problem. In II-C and II-D, we describe the main portions of the wave-physics informed matrix factorization algorithm. In II-E, we describe an interesting connection of this framework to signal processing. We emphasize that the novel parts of this work is in subsections II-C and II-E. The work optimization discussions in II-B are specific cases of

previous work and are shown for the benefit of the reader, particularly those in wave physics who may not be familiar with the optimization literature.

A. Problem Formulation

The problem formulation described in this section is inspired from the wave-physics regularizer proposed in [31]. In this work, we additionally justify the wave-physics informed problem formulation in [31] as a naturally arising idea from the theory of solutions of partial differential equations. A key concept in the solution of partial differential equations is the notion of separation. For example, given a PDE in space ℓ and time t , we can often assume the solution is of the form

$$f(\ell, t) = d(\ell)x(t). \quad (1)$$

In the case of linear PDEs, we find that if $d_i(\ell)x_i(t)$ are solutions for all $i \in [1, N]$, then

$$f_i(\ell, t) = \sum_{i=1}^N \alpha_i d_i(\ell)x_i(t) \quad (2)$$

is also a solution for any arbitrary choice of linear coefficients, $\{\alpha_i\}_{i=1}^N$. For a discrete approximation of this solution, we note that the product of continuous functions $d_i(\ell)x_i(t)$ on $\ell \in [0, L]$ and $t \in [0, T]$ can be approximated as an outer product of two vectors, leading directly to a matrix factorization model, where we further note that without loss of generality the scaling coefficients (α_i) can be absorbed into the factorized matrices.

Specifically, we consider matrices ($\mathbf{D} \in \mathbb{R}^{N_d \times N}$, $\mathbf{X} \in \mathbb{R}^{N_t \times N}$) defined such that

$$\begin{aligned} \mathbf{d}_i &= [d_i(0), d_i(\Delta\ell), \dots, d_i(N_d\Delta\ell)]^\top \\ \mathbf{x}_i &= [x_i(0), x_i(\Delta t), \dots, x_i(N_t\Delta t)]^\top \\ \mathbf{D}\mathbf{X}^\top &= \sum_{i=1}^N d_i(\ell)x_i(t) = \sum_{i=1}^N \mathbf{d}_i\mathbf{x}_i^\top, \end{aligned} \quad (3)$$

where $\mathbf{d}_i$ and $\mathbf{x}_i$ denote columns of $\mathbf{D}$ and $\mathbf{X}$, respectively, and $N_d = \lfloor L/\Delta\ell \rfloor$ and $N_t = \lfloor T/\Delta t \rfloor$. Recall, we also wish to enforce physical consistency in our model, which we accomplish through theory-guided regularization (see [14]) in a cost function which 1) captures how our matrix factorization model matches the observed samples and 2) incorporates a regularization term which enforces physical consistency in the matrix factors. Specifically, we consider a model of the form

$$\min_{\mathbf{D}, \mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}^\top\|_F^2 + \lambda \Theta(\mathbf{D}, \mathbf{X}) \quad (4)$$

where $\|\mathbf{Y} - \mathbf{D}\mathbf{X}^\top\|_F^2$ is a data consistency loss, $\Theta(\mathbf{D}, \mathbf{X})$ is our physics-informed regularizer, and $\lambda > 0$ is a hyper-parameter to balance the trade-off between fitting the data and satisfying our model assumptions. When $\Theta(\mathbf{D}, \mathbf{X})$ is defined by the wave equation, we call this *wave-informed matrix factorization*.

We derive a wave-informed $\Theta(\mathbf{D}, \mathbf{X})$ by considering the one-dimensional wave equation evaluated at $d_i(\ell)x_i(t)$

$$\frac{\partial^2 [d_i(\ell)x_i(t)]}{\partial \ell^2} = \frac{1}{c^2} \frac{\partial^2 [d_i(\ell)x_i(t)]}{\partial t^2} \quad (5)$$

The above equation constrains that each $d_i(\ell)x_i(t)$ component ($i \in [1, N]$) is a wave. This can be seen as decomposing the data into a superposition of simpler waves. Furthermore, the Fourier transform in time on both sides of (5) yields

$$\frac{\partial^2 [d_i(\ell)]}{\partial \ell^2} X_i(\omega) = \frac{-\omega^2}{c^2} d_i(\ell) X_i(\omega) \quad (6)$$

$$\frac{\partial^2 [d_i(\ell)]}{\partial \ell^2} = \frac{-\omega^2}{c^2} d_i(\ell) \quad (7)$$

at points where $X_i(\omega) \neq 0$. This is specifically known as the Helmholtz equation. The Helmholtz equation is analytically solvable for constant ω/c and the solution takes the form $d_i(\ell) = A \sin(\omega\ell/c) + B \cos(\omega\ell/c)$, and the constants A and B can be determined by initial conditions. While the Helmholtz equation is often applied in the frequency domain, it is also applicable in time, where the relationship must be true for every frequency-velocity (ω, c) mode in the data. Hence, if the data is sparse in these domains (as is often the case in modal analysis), then we can use the Helmholtz constraint to estimate and decompose all (ω, c) modes in $\mathbf{Y}$.

We discretize the Helmholtz equation such that

$$\mathbf{L}\mathbf{d}_i = -k_i^2 \mathbf{d}_i \quad (8)$$

where $k_i = \omega_i/c_i$, also known as the angular wavenumber, and $\mathbf{L}$ is a discrete second derivative operator. The specific form of $\mathbf{L}$ depends on the boundary conditions (e.g., Dirichlet, Dirichlet-Neumann, etc), which can be readily found in the numerical methods literature (see for example [41], [42]). In this paper, we use the Laplacian matrix ($\mathbf{L}$) defined by

$$\mathbf{L} = \frac{1}{(\Delta l)^2} \begin{bmatrix} -2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -2 \end{bmatrix}. \quad (9)$$

Note that to reduce notational complexity, we assume $\Delta l = 1$ without any loss of generality, as any change in Δl can be absorbed into the value of k_i after plugging (9) into (8).

We also specify the dependence on i owing to the fact that ω/c and $d_i(l)$ determine each other. For the factorization to satisfy the wave equation, we desire that (8) is satisfied for some value of k_i , which we promote with the regularizer

$$\min_{k_i} \|\mathbf{L}\mathbf{d}_i + k_i^2 \mathbf{d}_i\|_F^2. \quad (10)$$

Note that for the above regularizer it is easily shown that the optimal value for k_i^2 must lie in the range $[0, \frac{4}{\Delta l}]$, where $\frac{4}{\Delta l}$ is the smallest eigenvalue of $-\mathbf{L}$ [43]. Since we specifically chose $\Delta l = 1$ (without loss of generality), we have that the optimal value of k_i^2 must lie in the range $[0, 4]$.

In addition to requiring that the columns of $\mathbf{D}$ satisfy the wave equation, we also would like to constrain the number of modes to be minimal. We accomplish this by adding the squared Frobenius norms to both $\mathbf{D}$ and $\mathbf{X}$, which is known to induce low-rank solutions in the product $\mathbf{D}\mathbf{X}^\top$ due to

connections with the variational form of the nuclear norm [37], [38], [44]. Hence, our complete regularization is defined by

$$\Theta(\mathbf{D}, \mathbf{X}, \mathbf{k}) = \frac{1}{2} \sum_{i=1}^N \bar{\theta}(\mathbf{d}_i, \mathbf{x}_i, k_i) \quad (11)$$

$$\begin{aligned} \bar{\theta}(\mathbf{d}_i, \mathbf{x}_i, k_i) &= (\|\mathbf{x}_i\|_F^2 + \|\mathbf{d}_i\|_F^2) + \gamma \min_{k_i} \|\mathbf{L}\mathbf{d}_i + k_i^2 \mathbf{d}_i\|_F^2 \\ &= \|\mathbf{x}_i\|_F^2 + \min_{k_i} \mathbf{d}_i^\top (\mathbf{I} + \gamma(\mathbf{L} + \mathbf{I}k_i^2)) \mathbf{d}_i \\ &= \|\mathbf{x}_i\|_F^2 + \min_{k_i} \mathbf{d}_i^\top \mathbf{A}(k_i) \mathbf{d}_i \end{aligned} \quad (12)$$

$$\text{where } \mathbf{A}(k_i) = \mathbf{I} + \gamma(\mathbf{L} + \mathbf{I}k_i^2). \quad (13)$$

With this regularization function, the complete cost function is defined by

$$\min_{\mathbf{D}, \mathbf{X}, \mathbf{k}, N} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}^\top\|_F^2 + \frac{\lambda}{2} \|\mathbf{X}\|_F^2 + \frac{\lambda}{2} \sum_{i=1}^N \mathbf{d}_i^\top \mathbf{A}(k_i) \mathbf{d}_i \quad (14)$$

where note that we are also optimizing over the number of modes (N) and the wavenumber $\mathbf{k} = [k_1, \dots, k_N]^\top$ for each mode in the data.

B. Model Optimization

Our model in (14) is inherently non-convex in $(\mathbf{D}, \mathbf{X}, \mathbf{k})$ due to the matrix factorization model and further complicated by the fact that we are additionally searching over the number of columns/modes N . However, despite the challenge of non-convex optimization, we can solve (14) by leveraging prior results from optimization theory for structured matrix factorization [37], [39]. Note that, for the remainder of this subsection we reiterate the concepts and framework described in [37], [38] leading to adapting the algorithm mentioned therein to our specific model and finally obtaining an optimal solution of the objective function (14). For a matrix factorization problem of the form

$$\min_N \min_{\mathbf{D}, \mathbf{X}} L(\mathbf{D}\mathbf{X}^\top) + \lambda \sum_{i=1}^N \bar{\theta}(\mathbf{d}_i, \mathbf{x}_i) \quad (15)$$

where $L(\hat{\mathbf{Y}})$ is any function which is convex and once differentiable in $\hat{\mathbf{Y}}$ and $\bar{\theta}(\mathbf{d}, \mathbf{x})$ is any function that satisfies the following three conditions

- 1) $\bar{\theta}(\alpha \mathbf{d}, \alpha \mathbf{x}) = \alpha^2 \bar{\theta}(\mathbf{d}, \mathbf{x})$, $\forall (\mathbf{d}, \mathbf{x})$, $\forall \alpha \geq 0$.
- 2) $\bar{\theta}(\mathbf{d}, \mathbf{x}) \geq 0$, $\forall (\mathbf{d}, \mathbf{x})$.
- 3) For all sequences $(\mathbf{d}^{(n)}, \mathbf{x}^{(n)})$ such that $\|\mathbf{d}^{(n)}(\mathbf{x}^{(n)})^\top\| \rightarrow \infty$ then $\bar{\theta}(\mathbf{d}^{(n)}, \mathbf{x}^{(n)}) \rightarrow \infty$,

an algorithm exists that can obtain the global minimum solution. In supplementary material VII-A, we show that the wave-informed cost function satisfies all three conditions, where the optimization over the k_i parameters is done inside $\bar{\theta}$ as in (12).

Moreover, in [38] it is shown that a given point $(\tilde{\mathbf{D}}, \tilde{\mathbf{X}})$ is a globally optimal solution of (15) iff the following two conditions are satisfied:

- 1) $\langle -\nabla L(\tilde{\mathbf{D}}\tilde{\mathbf{X}}^\top), \tilde{\mathbf{D}}\tilde{\mathbf{X}}^\top \rangle = \lambda \sum_{i=1}^N \bar{\theta}(\tilde{\mathbf{d}}_i, \tilde{\mathbf{x}}_i)$
- 2) $\Omega_{\tilde{\theta}}^\circ(-\frac{1}{\lambda} \nabla L(\tilde{\mathbf{D}}\tilde{\mathbf{X}}^\top)) \leq 1$

where in our scenario

$$-\frac{1}{\lambda} \nabla L(\mathbf{D}\mathbf{X}^\top) = \frac{1}{\lambda} (\mathbf{Y} - \mathbf{D}\mathbf{X}^\top)$$

denotes the gradient w.r.t. the matrix product and $\Omega_{\bar{\theta}}^{\circ}(\cdot)$ is referred to as the polar problem, which is defined as

$$\Omega_{\bar{\theta}}^{\circ}(\mathbf{Z}) \equiv \sup_{\mathbf{d}, \mathbf{x}} \mathbf{d}^{\top} \mathbf{Z} \mathbf{x} \quad \text{s.t.} \quad \bar{\theta}(\mathbf{d}, \mathbf{x}) \leq 1. \quad (16)$$

It is further shown in [38] (Proposition 3) that the first condition above will always be satisfied for any first-order stationary point $(\tilde{\mathbf{D}}, \tilde{\mathbf{X}})$, and that if a given point is not globally optimal then the objective function (15) can always be decreased by augmenting the current factorization by a solution to the polar problem as a new column:

$$\begin{aligned} (\mathbf{D}, \mathbf{X}) &\leftarrow \left(\begin{bmatrix} \tilde{\mathbf{D}}, & \tau \mathbf{d}^* \end{bmatrix}, \begin{bmatrix} \tilde{\mathbf{D}}, & \tau \mathbf{x}^* \end{bmatrix} \right) \\ \mathbf{d}^*, \mathbf{x}^* &\in \arg \max_{\mathbf{d}, \mathbf{x}} \frac{1}{\lambda} \mathbf{d}^{\top} (\mathbf{Y} - \mathbf{D} \mathbf{X}^{\top}) \mathbf{x} \quad \text{s.t.} \quad \bar{\theta}(\mathbf{d}, \mathbf{x}) \leq 1 \end{aligned} \quad (17)$$

for an appropriate choice of step size $\tau > 0$.

Despite the fact that [38] gives a meta-algorithm that is guaranteed to solve (15) to global optimality, one needs to solve the polar problem in (16) to both verify global optimality and escape non-optimal stationary points, which is itself another optimization problem. This presents a potentially significant computational challenge in practice, as for many choices of regularization function, solving the polar problem is NP-hard [39]. For example, if the regularization on the matrix columns is simply the sum of the ℓ_2 norms of the columns, then the polar problem is easily solved as a maximum eigenvalue problem, but if we either replace one of the ℓ_2 norms by the ℓ_{∞} norm, the problem becomes NP-hard [39]. Fortunately, however, in the next section we show that the polar problem that results from our wave-informed factorization model is tractable and leads to polynomial time global optimality.

C. Solving the Polar Problem

Specifically, for our regularization function (11), we show the following result which allows the polar problem to be solved efficiently, enabling efficient and guaranteed optimization (14).

Theorem 1. For (14), the polar problem in (16) is

$$\begin{aligned} \Omega_{\bar{\theta}}^{\circ}(\mathbf{Z}) &= \max_{\mathbf{d}, \mathbf{x}, k} \mathbf{d}^{\top} \mathbf{Z} \mathbf{x} \\ \text{s.t.} \quad &\mathbf{d}^{\top} \mathbf{A}(k) \mathbf{d} \leq 1, \|\mathbf{x}\|_F^2 \leq 1, 0 \leq k \leq 2. \end{aligned}$$

where $\mathbf{A}(k)$ is as defined in (13). Further, if we define k^* as

$$k^* = \arg \max_{k \in [0, 2]} \|\mathbf{A}(k)^{-1/2} \mathbf{Z}\|_2. \quad (18)$$

Then the optimal values of $\mathbf{d}, \mathbf{x}, k$ are given as $\mathbf{d}^* = \mathbf{A}(k^*)^{-1/2} \tilde{\mathbf{d}}$, $\mathbf{x}^* = \tilde{\mathbf{x}}$, and k^* . Where $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{x}}$ are the left and right singular vectors, respectively, associated with the largest singular value of $\mathbf{A}(k^*)^{-1/2} \mathbf{Z}$.

The proof of Theorem 1 is detailed in supplementary material VII-B. The result implies that we can solve the polar problem by performing a one-dimensional line search over k . Due to the fact that the largest singular value of a matrix is a Lipschitz continuous function, this line search can be solved efficiently by a variety of global optimization algorithms. In particular, the above line search over k is Lipschitz continuous (with

respect to k^2) with a Lipschitz constant, L_k , which is bounded by

$$L_k \leq \begin{cases} \frac{2}{3\sqrt{3}} \sqrt{\gamma} \|\mathbf{Y} - \mathbf{D} \mathbf{X}^{\top}\|_2 & \gamma \geq \frac{1}{32} \\ 4\gamma(1 + 16\gamma)^{-\frac{3}{2}} \|\mathbf{Y} - \mathbf{D} \mathbf{X}^{\top}\|_2 & \gamma < \frac{1}{32} \end{cases} \quad (19)$$

$$\leq \frac{2}{3\sqrt{3}} \sqrt{\gamma} \|\mathbf{Y} - \mathbf{D} \mathbf{X}^{\top}\|_2$$

This is formally stated and proven in supplementary material VII-C as Theorem 2, where we also give a formal result, in Corollary 1, which guarantees that the line search over k (and hence the overall polar problem) can be solved in polynomial time. Building from this key result, next we discuss an algorithm to provably solve our wave-informed factorization model in polynomial time.

D. Wave-Informed Factorization Algorithm

Algorithm 1 defines our wave-informed matrix factorization algorithm. The algorithm has three components that are iterated: (1) perform gradient descent with a fixed number of modes/columns, N , to reach a first order stationary point for $\mathbf{D}$, $\mathbf{X}$, and k , (2) solve the polar problem via Theorem 1 to obtain $\mathbf{d}^*$, $\mathbf{x}^*$, and k^* , and (3) if the stopping condition is not met, update $\mathbf{D}$, $\mathbf{X}$, and k by appending the polar solution $(\mathbf{d}^*, \mathbf{x}^*, k^*)$ to the solution (scaled by an optimal step size τ); otherwise, terminate the algorithm.

Algorithm 1 Wave-Informed Matrix Factorization

- 1: Input $\mathbf{D}_{init}, \mathbf{X}_{init}, \mathbf{k}_{init}$, stopping tolerance: ϵ
 - 2: Initialize $(\mathbf{D}, \mathbf{X}, \mathbf{k}) \leftarrow (\mathbf{D}_{init}, \mathbf{X}_{init}, \mathbf{k}_{init})$
 - 3: Initialize $(\tilde{\mathbf{D}}, \tilde{\mathbf{X}}, \tilde{\mathbf{k}}) \leftarrow (\mathbf{D}_{init}, \mathbf{X}_{init}, \mathbf{k}_{init})$
 - 4: **while** $\left| \Omega_{\bar{\theta}}^{\circ}(\frac{1}{\lambda}(\mathbf{Y} - \tilde{\mathbf{D}} \tilde{\mathbf{X}}^{\top})) - 1 \right| \geq \epsilon$ **do**
 - 5: Perform gradient descent on (14) to obtain $(\tilde{\mathbf{D}}, \tilde{\mathbf{X}}, \tilde{\mathbf{k}})$
 - 6: Calculate $\Omega_{\bar{\theta}}^{\circ}(\frac{1}{\lambda}(\mathbf{Y} - \tilde{\mathbf{D}} \tilde{\mathbf{X}}^{\top}))$ and obtain $\mathbf{d}^*, \mathbf{x}^*, k^*$
 - 7: **if** $\Omega_{\bar{\theta}}^{\circ}(\frac{1}{\lambda}(\mathbf{Y} - \tilde{\mathbf{D}} \tilde{\mathbf{X}}^{\top})) - 1 > \epsilon$ **then**
 - 8: $(\mathbf{D}, \mathbf{X}, \mathbf{k}) \leftarrow \left(\begin{bmatrix} \tilde{\mathbf{D}}, & \tau \mathbf{d}^* \end{bmatrix}, \begin{bmatrix} \tilde{\mathbf{X}}, & \tau \mathbf{x}^* \end{bmatrix}, \begin{bmatrix} \tilde{\mathbf{k}}^{\top}, & k^* \end{bmatrix}^{\top} \right)$
 - 9: **end if**
 - 10: **end while**
-

1) *Gradient Descent Update:* We begin our algorithm by performing block coordinate descent with respect to $\mathbf{D}$ and $\mathbf{X}$ along with a block minimization step with respect to k on the objective in (4), which we note is guaranteed to reach a first-order stationary point [45]. The number of columns of $\mathbf{D}$ and $\mathbf{X}$ (denoted by N in (14)) is fixed and does not change during this step. Specifically, we use the following update equations (for a step size α)

$$\begin{aligned} \mathbf{d}_i^+ &= \mathbf{d}_i - \alpha \left((\mathbf{D} \mathbf{X}^{\top} - \mathbf{Y}) \mathbf{x}_i + \lambda \mathbf{d}_i \right. \\ &\quad \left. + 2\gamma \lambda (\mathbf{L} + k_i^2 \mathbf{I})^2 \mathbf{d}_i \right) \end{aligned} \quad (20)$$

$$\mathbf{x}_i^+ = \mathbf{x}_i - \alpha \left((\mathbf{D}^{\top} \mathbf{X}^{\top} - \mathbf{Y})^{\top} \mathbf{d}_i^+ + \lambda \mathbf{x}_i \right) \quad (21)$$

$$k_i^+ = \sqrt{-\frac{(\mathbf{d}_i^+)^{\top} \mathbf{L} \mathbf{d}_i^+}{\|\mathbf{d}_i^+\|_2^2}} \quad \text{for each } i = 1, \dots, N \quad (22)$$

where $\mathbf{D}^+$, $\mathbf{X}^+$, and k_i^+ are the updated values and the i subscript denotes a column of the respective matrix.

2) *Solve Polar Problem*: After finding a first order stationary point $(\tilde{\mathbf{D}}, \tilde{\mathbf{X}}, \tilde{\mathbf{k}})$, we next solve the polar problem according to Theorem 1 with $\mathbf{Z} = \frac{1}{\lambda}(\mathbf{Y} - \tilde{\mathbf{D}}\tilde{\mathbf{X}}^\top)$. In particular, we first solve the one-dimensional line search for the optimal value of k in (18), where this line search can be solved using many approaches. For example, the LIPO algorithm given in [46] can provably solve the line search in polynomial time. Once the optimal k^* is obtained, the values of $\mathbf{d}^*$ and $\mathbf{x}^*$ are then computed directly from the eigen-decomposition of $\mathbf{A}(k^*)$ as described in Theorem 1.

3) *Stopping Condition and Growing the Factorization*: After solving the polar problem as described above, the value of the polar gives a test for checking global optimality, and if this test is not met then we can escape the first order stationary point by appending the polar solution to our factorization. Specifically, once we evaluate the polar $\Omega_\theta(\frac{1}{\lambda}(\mathbf{Y} - \tilde{\mathbf{D}}\tilde{\mathbf{X}}^\top))$, then in [38] (Prop. 4) it is shown that the distance to the global optimum (in objective value) is directly proportional to the value of the polar value minus 1 (and that the value of the polar is always ≥ 1), so choosing to stop the algorithm when the value of polar takes a value $\leq 1 + \epsilon$ also guarantees optimality to within $\mathcal{O}(\epsilon)$. If the value of the polar does not meet our stopping condition, then we continue the optimization by appending the polar solution to our factorization after an optimal scaling. Specifically, let

$$(\mathbf{D}_\tau, \mathbf{X}_\tau, \mathbf{k}_\tau) = \left(\begin{bmatrix} \tilde{\mathbf{D}} & \tau \mathbf{d}^* \end{bmatrix}, \begin{bmatrix} \tilde{\mathbf{X}} & \tau \mathbf{x}^* \end{bmatrix}, \begin{bmatrix} \tilde{\mathbf{k}}^\top & k^* \end{bmatrix}^\top \right) \quad (23)$$

be our optimization variables with the new polar solution appended to the factorization scaled by τ , then we find the optimal value of τ to minimize the original objective function:

$$\min_{\tau} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}_\tau \mathbf{X}_\tau^\top\|_F^2 + \frac{\lambda}{2} \|\mathbf{X}_\tau\|_F^2 + \frac{\lambda}{2} \sum_{i=1}^N \mathbf{D}_\tau^\top \mathbf{A}(k_{i,\tau}) \mathbf{D}_\tau \quad (24)$$

It is shown in supplementary material VII-D that the optimal τ is given by

$$\tau = \frac{\sqrt{(\mathbf{d}^*)^\top (\mathbf{Y} - \tilde{\mathbf{D}}\tilde{\mathbf{X}}^\top) \mathbf{x}^* - \lambda}}{\|\mathbf{d}^*\|_2 \|\mathbf{x}^*\|_2} \quad (25)$$

After updating the variables as in (24) with the optimal choice of τ , we then return to gradient descent updates until reaching a first order stationary point.

Taken together, our proposed algorithm results in a polynomial-time algorithm for solving our main objective (14). We do not give a formal result as the algorithm given in Algorithm 1 is simply one choice of optimization algorithm which is possible due to the polynomial time solution for the polar problem that we give in Theorem 1 and the overall convergence of the algorithm is largely known from prior work, provided the polar problem can be solved. For example, the authors of [45] prove that the block coordinate gradient descent portion of the algorithm will converge to a stationary point at a rate of $\mathcal{O}(1/t)$ (with t being the number of iterations), which can be improved to $\mathcal{O}(1/t^2)$ with Nesterov

acceleration. Likewise, the authors of [39] show that a Frank-Wolfe/conjugate-gradient step (i.e., when we escape poor stationary points by solving the polar problem) will reduce the objective with a linear convergence rate. From this it is straight-forward to combine these arguments to give an overall guarantee of convergence, but we do not do so here to limit the scope and length of the manuscript.

E. Signal Processing Interpretation of the Regularizer

We now discuss an interesting interpretation of the above algorithm from a signal processing perspective. Specifically, if we let $\mathbf{L} = \Gamma \Lambda \Gamma^\top$ denote a singular value decomposition of $\mathbf{L}$, then note that when identifying the optimal k value in the polar program, we solve for

$$\arg \max_{k \in [0, 2]} \|\Gamma(\mathbf{I} + \gamma(k^2 \mathbf{I} + \Lambda)^2)^{-1/2} \Gamma^\top \mathbf{Z}\|_2. \quad (26)$$

This optimization has an intuitive interpretation from signal processing. Given that Γ contains the eigenvectors of a Toeplitz matrix, those eigenvectors have spectral qualities similar to the discrete Fourier transform (the eigenvectors of the related circulant matrix would be the discrete Fourier transform [47], [48]). As a result, $\Gamma^\top$ transforms the data $\mathbf{Z}$ into a spectral-like domain and Γ returns the data back to the original domain. Since the other terms are all diagonal matrices, they represent element-wise multiplication across the data in the spectral domain. This is equivalent to a filtering operation, with filter coefficients given by the diagonal entries of $(\mathbf{I} + \gamma(k^2 \mathbf{I} + \Lambda)^2)^{-1/2}$. Figure 1 shows two such filter responses at $k^2 = 2.5, \gamma = 1000$ and $k^2 = 1.5, \gamma = 10000$. Observe that, an increased value of γ has reduced the bandwidth of the filter.

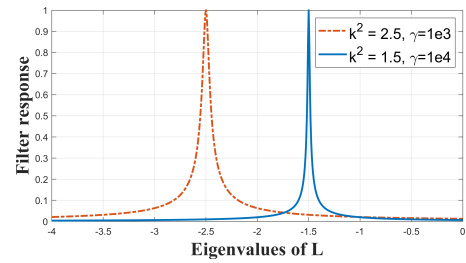


Fig. 1: Butterworth Filter on the spectral-like domain

Furthermore, recall that the transfer function of a 1st-order Butterworth filter is given by:

$$T(\omega) = \frac{1}{\sqrt{1 + \gamma(\omega_0 + \omega)^2}} \quad (27)$$

where ω_0 is the center frequency of the passband of the filter and $1/\sqrt{\gamma}$ corresponds to the filter's -3dB cut-off frequency. Comparing this to the filter coefficients from (26), we note that the filter coefficients are identical to those of the 1st-order band-pass Butterworth filter, where Λ corresponds to the angular frequencies.

As a result, we can consider this optimization as determining the optimal filter center frequency $\omega_0 = k^2$ with fixed bandwidth ($1/\sqrt{\gamma}$) that retains the maximum amount of signal power from $\mathbf{Z}$. Likewise, the choice of the γ

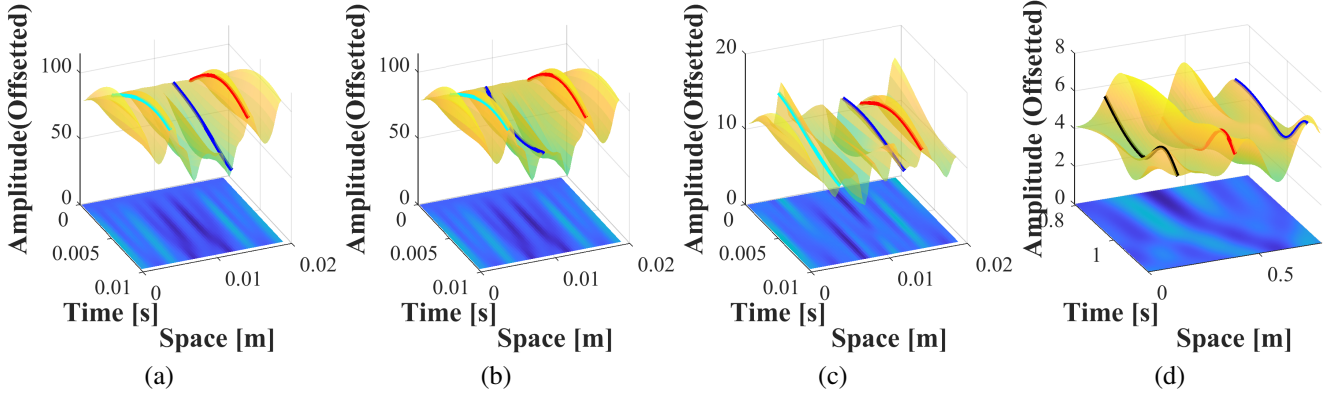


Fig. 2: (a) Homogeneously vibrating medium marked with different colors at three different time instances for a short spatial interval at the start. (b) Non-Homogeneously (exponentially decaying over space) vibration data marked with different colors at three different time instances for a short spatial interval at the start. (c) Traveling wave data marked with different colors at three different time instances for a short spatial interval at the start. (d) Data of two mediums joined end-to-end and vibrating together marked with different colors at different time instances for a short spatio-temporal interval around the joint.

hyperparameter sets the bandwidth of the filter. As $\gamma \rightarrow \infty$, the filter bandwidth approaches 0 and thereby restricts us to a single-frequency (i.e., Fourier) solution. Furthermore, we can provide a recommended lower bound for γ according to $\gamma > 1/k_{bw}^2$, where k_{bw} is the bandwidth of the signal within this spectral-like domain. We would like to emphasize that this work introduces a unique theme within the works of scientific machine learning. Through means of wave-informed learning, we obtain a new method to mathematically derive the well-known *digital Butterworth filter*.

III. EXPERIMENTAL EVALUATION

We demonstrate the results from wave-informed matrix factorization on four simulated vibration datasets, each characterized by increasing complexity: (1) a homogeneous vibration with fixed boundary conditions, (2) an non-homogeneous vibration with fixed boundary conditions and spatial amplitude decay, (3) a plane wave traveling with temporal decay, and (4) a vibration in a segmented medium, where each segment has an independent wavenumbers and amplitude. In all four datasets, one dimension of the matrix represents space and the other dimension represents time, as is common in modal analysis [49], [50]. Multiple modes share the same spatial locations. Fig. 2 illustrates these four datasets.

For each method and dataset, we study five different signal-to-noise ratio (SNR) levels. Each signal is defined by

$$\hat{y}(\ell, t) = y(\ell, t) + \eta(\ell, t), \quad (28)$$

where $y(\ell, t)$ represents our data and $\eta(\ell, t)$ represents our noise. The noise is assumed to be additive white Gaussian noise across both space and time. The SNR is defined by

$$\text{SNR} = \frac{\sum_{\ell, t} |y(\ell, t)|^2}{\sum_{\ell, t} |\eta(\ell, t)|^2}. \quad (29)$$

The SNR is assumed to be unknown to each of the algorithms.

The following subsections describe how each of our datasets are computed. Table I lists the parameters for each of these

scenarios. Note that each case can have potentially an infinite number of modes N . In practice, these modes usually attenuate as the associated frequency / wavenumber increases. To simplify this, we simulate a finite number of modes for each dataset. For the segmented data, k_n represents the first 8 theoretical modes (four in each segment) for vibration across two connected segments.

To solve the polar problem and estimate k^* in Algorithm 1, many approaches can be taken. In this paper, we solve for k^* by first coarsely searching for a maximum among coarsely uniformly placed points. For a predetermined region around the maximum, we perform a fine search by taking the maximum over finely uniformly placed points in that region

A. Homogeneous Vibration Data with Fixed Boundaries

Our first dataset emulates vibrations in a medium with fixed (Dirichlet) boundary conditions such that $y(0, t) = 0$ and $y(L, t) = 0$, where L is the length of the medium. The sinusoids reflect from each end and create a standing wave in space. Such data is defined by

$$y(\ell, t) = \sum_{n=1}^N e^{-\alpha_n t} \sin(nk\ell) (a_n \sin(n\omega t) + b_n \cos(n\omega t)). \quad (30)$$

This type of data is commonly found in vibrations and modal analysis problems [51], [52]. For example, the modes and their parameters can be used to determine material or structural characteristics [8]–[10]. Apart from the added noise, we assume a damping over time to introduce a non-ideality. For this illustration we choose $L = 1$.

B. Inhomogeneous Vibration Data With Spatial Decay

The previous dataset assumed spatial homogeneity. That is, the behavior of the medium (i.e., the wavenumber and amplitude) did not vary across space ℓ . This assumption may not always be true. As a result, we would like to extract modes

TABLE I: Table of parameters used to generate data, with additive noise $\eta_n \sim \mathcal{N}(0, 1)$ and $u_n \sim \mathcal{U}[0, 1]$. The random variables are independently sampled for each parameter and $1 \leq n \leq N$.

Data	$(\Delta\ell, L)$	$(\Delta t, T)$	α_n	β_n	k_n	ω_n	N	δ
Homogeneous Vibration	(0.0901, 1)	(0.0005, 2)	$\frac{n}{2}$	0	$n\pi$	$106k_n$	6	100
Non-homogeneous Vibration	(0.901, 10)	(0.0005, 2)	0	$\frac{4+n}{10} + u_n$	$n\pi$	$106k_n$	6	1
Travelling Plane Wave	(0.901, 10)	(0.0005, 2)	$\frac{n}{2}$	0	$n\pi$	$106k_n$	6	1
Segmented Vibration	(0.01, 2)	(0.02, 10)	-	-	$2\pi \pm \arctan\left(\sqrt{\frac{1}{3}(13 \pm 4\sqrt{10})}\right)$	$3k_n$	8	1

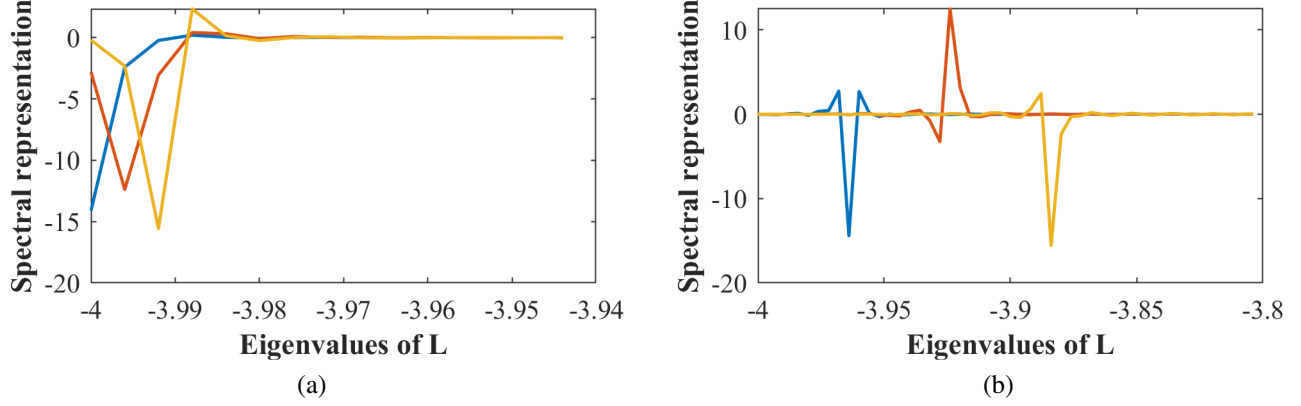


Fig. 3: Spectra (with respect to Eigen-basis of $\mathbf{L}$) of decaying exponentials $e^{-\beta_1\ell} \sin(\pi\ell)$, $e^{-\beta_2\ell} \sin(2\pi\ell)$ and $e^{-\beta_3} \sin(3\pi\ell)$ (indicated using blue, red and yellow colors respectively) for (a) $\ell \in [0, 1]$, since here $L=1$ (b) $\ell \in [0, 10]$, since here $L=10$

with inhomogenities. We consider the same fixed boundary data but with a spatial decay such that

$$y(\ell, t) = \sum_{n=1}^N e^{-\beta_n\ell} \sin(k_n\ell) (a_n \sin(n\omega t) + b_n \cos(n\omega t)) . \quad (31)$$

In this expression, we have introduced a new parameter β_n that causes the amplitude to decay over space, similar to an evanescent wave in electro-magnetics and is usually meant to persist over longer regions of space. This example is meant to explore if each algorithm addresses amplitude non-homogeneities. In contrast to the previous case, where we considered measuring vibrations along a string of length 1, we consider a scenario of measuring waves for a longer length. In this illustration, we fix a length of 10 for measuring waves decaying exponentially over space. As an illustration for the advantage of length, the spectrum (computed by multiplying data with the transpose of the eigenvectors Γ) of three exponentially decaying sinusoids (for length 1 and length 10) corresponding to $k_1 = \pi$, $k_2 = 2\pi$, and $k_3 = 3\pi$ are shown in Figure 3(a) and Figure 3(b), respectively. In each plot, the peaks correspond to the same three values of k , but due to the 10 times increases in length, Figure 3(b) exhibits a 10 times improvement in resolution in the wavenumber domain. In the presence of spectral spread (e.g., caused by a spatial decay), this can improve our ability to distinguish each mode. The parameters generating the data are given in Table I.

C. Travelling Plane Wave Data

We previously considered vibration-like modes with fixed boundaries. This dataset considers sinusoids of uniformly

spaced wavenumbers travelling in time and space without explicit boundary conditions, defined by

$$y(\ell, t) = \sum_{n=1}^N e^{-\alpha_n t} \sin(k_n\ell + \omega_n t) . \quad (32)$$

This expression resembles a travelling plane wave rather than a vibrational wave. These types of data are commonly found in beamforming problems for RADAR [11], [53], ultrasound [13], [54], audio processing [55] and other fields using spatial signal processing. In beamforming, the parameter k_n is usually proportional to the direction of arrival. In ultrasound and geophysics, k_n may represent a transverse or longitudinal modes of vibration [56]. We also have the same assumption of measuring the traveling wave data for a longer length (here 10 times the homogeneous case). Note that time and space are not immediately separable in this equation, as observed in the other datasets. However, (32) can be expressed as

$$\begin{aligned} y(\ell, t) &= \sum_{n=1}^N e^{-\alpha_n t} \sin(k_n\ell - \phi_n + \phi_n + \omega_n t) \\ &= \sum_{n=1}^N e^{-\alpha_n t} [\sin(k_n\ell - \phi_n) \cos(\omega_n t + \phi_n) \\ &\quad + \cos(k_n\ell - \phi_n) \sin(\omega_n t + \phi_n)] \end{aligned} \quad (33)$$

for an any arbitrarily chosen ϕ_n . Hence, a travelling plane wave mode can be expressed as the sum of two standing wave modes, similar to our second dataset, with 90 degree phase differences. In our results, we assume $\phi_n = 0$ so that errors in Table II will be small when the extracted components match $\phi_n = 0$ while errors in Table III will be invariant to ϕ_n since the magnitude of the Fourier transform ignores phase delay.

D. Multi-Segment Data

Our final dataset considers a problem with both amplitude and wavenumber spatial inhomogeneities. This occurs when there are media adjacent to each other. As a signal travels from one medium to the next, the amplitude varies according to the interface's reflection and transmission coefficients and the wavenumber varies according to Snell's law. We generate vibration data in two media connected end-to-end that is continuous and once differentiable at the point where the two different they connect (see [56] for related theory). We assume the two mediums to have a density ratio of 9 to 1, respectively. The velocity of the wave traveling in the string is inversely proportional to the square root of density, thus the velocities must be in the ratio 3 to 1. The mathematical expression generating the data is:

$$y(\ell, t) = \begin{cases} \sum_{i=1}^N \sin(3k_i \ell) \sin(\omega_i(t/c)), & 0 \leq \ell < 1 \\ \sum_{i=1}^N \sin(k_i \ell)(\ell) \sin(\omega_i(t/3c)), & 1 < \ell \leq 2 \end{cases} \quad (34)$$

where c is the velocity of propagation in the first medium. The wavenumbers for this problem take a specific solution based on the condition, shown in the Table I, assuming fixed, zero-valued boundaries at $\ell = 0$ and $\ell = 2$ and the continuity condition at $\ell = 1$.

E. Choice of Regularization Parameters

To provide a fair and informative comparison between each method, we assume oracle knowledge of N throughout our results. As previously mentioned, deviation from 1 in the polar problem in wave informed wave factorization can be used as a metric for monitoring convergence and as a stopping condition for the algorithm (and by extension the final choice of N), but the appropriate threshold is unclear in the presence of noise and a similar convergence property does not exist for all of our comparison methods. While there are a number of statistical methods for estimating N based on singular value decomposition that could be applied to every method [57], [58], we want our performance to reflect choice of algorithms rather than errors in the estimation of N . Hence using oracle knowledge of N to stop the algorithm as appropriate.

For wave-informed matrix factorization, we select λ to be $3/4\sigma(N)$, where $\sigma(N)$ is the N -th singular value of our data matrix $\mathbf{Y}$. This choice of λ is intuitively chosen based on the interpretation of low rank factorization as a soft thresholding operation over the singular values [37], [38]. The $3/4$ factor was obtained empirically based on our examples. However, reducing the factor is reasonable given the additional constraint provided by the Helmholtz equation.

We select γ in wave-informed matrix factorization based on our intuition of $1/\sqrt{\gamma}$ being bandwidth of a filter. In general, the bandwidth of exponentially decaying sinusoids discussed here is small, so γ is expected to be relatively large. Since our modes are constrained to a finite length in space, we know the bandwidth of a pure sinusoid will be one sample of the discrete Fourier transform. Therefore, we choose our bandwidth based on this assumption so that

$$\gamma = \delta \left(\frac{M}{\pi} \right)^2, \quad (35)$$

where M is the number of samples in the spatial domain. While this bandwidth is slightly smaller than some of the data due to spatial variations, this difference should be corrected for in the gradient descent step of wave-informed matrix factorization. An analytical way to motivate the same would be to observe the filter coefficients.

$$\frac{1}{\sqrt{1 + \gamma(k^2 + \Lambda_{ii})^2}}$$

is the i -th filter coefficient. Assuming pure Dirichlet boundary conditions, we have from [43] that,

$$\Lambda_{ii} = -4 \sin^2 \left(\frac{\pi i}{2(M+1)} \right)$$

For M large enough compared to i , we have that,

$$\Lambda_{ii} = -4 \sin^2 \left(\frac{\pi i}{2(M+1)} \right) \approx -\frac{\pi^2 i^2}{(M+1)^2}$$

Again, for large enough M , we can approximate the filter coefficients as below by substituting γ from (35),

$$\frac{1}{\sqrt{1 + \delta \frac{M^2}{\pi^2} \left(k^2 - \frac{\pi^2 i^2}{(M+1)^2} \right)^2}} \approx \frac{1}{\sqrt{1 + \delta \left(\frac{M^2 k^2}{\pi^2} - i^2 \right)^2}}$$

where i represents the actual (not angular) wavenumber scale. For instance, when $\delta = 1$, the bandpass Butterworth filter has a -3 dB cut-off wavenumber of 1 m^{-1} in the actual wavenumber scale. For larger values of δ , the -3dB cut-off wavenumber reduces to $1/\delta \text{ m}^{-1}$ and is useful to pick isolated wavenumbers. The value of δ used for each case is mentioned in Table I. Observe that homogeneous vibrations requires a larger value of $\delta (= 100)$ to pick isolated wavenumbers whereas others still work with a value of $\delta = 1$.

IV. RESULTS & DISCUSSIONS

For each dataset, we compare our approach with five other non-parametric matrix factorization / blind source separation methods. The algorithms we compare with include independent component analysis (**ICA**) [59], dynamic mode decomposition (**DMD**) [60], multidimensional empirical mode decomposition (**EMD**) [61]–[63], principal component analysis (**PCA**) [64], and wave-informed K-singular value decomposition (**WIKSVD**) [31]. We choose these method because they are widely used in modal analysis throughout the literature or represent a precursor to wave-informed matrix factorization (**WIMF**). Each method places different assumptions on the extracted modes. Briefly, ICA extracts components components that are statistically independent, PCA extracts components that are orthogonal to each other, DMD extracts components with a set frequency and decay based on estimating a state transition matrix, EMD extracts components by iteratively interpolating between maxima and minima, and WIKSVD extracts components that assume spatial components satisfy the wave equation and are sparse in the frequency domain. By comparison, WIMF extracts spatial components that satisfy the wave equation and minimizes the number of extracted modes.

Based on the filter interpretation of WIMF, it has characteristics similar to DMD since the filter's center frequency and bandwidth identifies the frequency and possibly decay of the extracted modes.

A. General Performance

In Table II, we show the mean squared errors for each of the recovered modes obtained by each method,

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^N \left\| \frac{\mathbf{d}_n}{\|\mathbf{d}_n\|_2} - \frac{c_n \mathbf{d}_n^{(\text{true})}}{\|\mathbf{d}_n^{(\text{true})}\|_2} \right\|_2^2 \quad (36)$$

where $\mathbf{d}_n$ is extracted mode and $\mathbf{d}_n^{(\text{true})}$ is the true mode. In our analysis, $\mathbf{d}_n^{(\text{true})}$ is chosen (without replacement) to be the mode with the maximal correlation (in absolute value) with $\mathbf{d}_n$ and $c_n = \text{sign}(\mathbf{d}_n^T \mathbf{d}_n^{(\text{true})})$. When computing the mean squared error, we normalize each mode since the amplitude is not uniquely determined in matrix decompositions.

Table III shows the mean squared error between the Fourier magnitudes true and extracted modes, using the same definition of (36) but replacing $\mathbf{d}_n$ and $\mathbf{d}_n^{(\text{true})}$ with the magnitude of their Fourier transforms, respectively. These results provide a different perspective, identifying if the methods obtain the correct frequencies ignoring the effect of phase shifts.

Fig. 4 further illustrates examples of two modes extracted by each method for each dataset at an SNR of ∞ . Across all of the datasets, ICA has the overall poorest performance, followed by EMD, PCA, DMD, WIKSVD, and WIMF. We hypothesize ICA performs poorly since wave modes are not random signals, and while the modes are orthogonal, they are not statistically independent. EMD performs poorly since mixtures of sinusoids where the amplitudes and/or frequencies are too similar cannot be recovered by the algorithm [65]. DMD and WIKSVD performs most poorly in high noise scenarios and with spatially segmented wavenumbers. DMD performs poorly at high noise because the estimation of the state transition matrices does not consider noise. Though a similar wave-informed regularizer is introduced for WIKSVD, it does not extract the basis as well as WIMF we believe due to the lack of filtering behavior in WIKSVD. This filtering effect particularly improves WIMF over WIKSVD for high noise regimes. The following subsections study the performance with each dataset in greater detail.

B. Homogeneous Vibration Data With Fixed Boundaries

Table II and III show that this relatively simple dataset achieves the overall best performance among the datasets. Wave-informed matrix factorization, wave-informed K-SVD, and dynamic mode decomposition perform best, albeit under different conditions. Wave-informed K-SVD and dynamic mode decomposition perform the best in no-noise conditions. Yet, dynamic mode decomposition is more sensitive to noise. This is evident in Tables II, III as with increasing noise the errors indicated in the tables increase. Wave-informed K-SVD performs best in low to mild noise (≤ -3 dB SNR) whilst wave-informed matrix factorization performs best in heavy noise (> -3 dB SNR).

C. Inhomogeneous Vibration Data With Spatial Decay

These results follow a similar trend as the homogeneous vibration data. In the inhomogeneous data, the spatial decay spreads the signal across the wavenumber domain, resulting in generally less compressible data that is more sensitive to noise. As a result, we see a reduction in performance for WIMF, WIKSVD, DMD in Table II and III. In case of infinite SNR, DMD performs best as it is designed to extract damped sinusoids [66]. However, it remains sensitive to noise.

PCA and EMD appear unstable with this dataset. This occurs because the parameter β_n varies across each method and SNR. This randomness in conjunction with the incorrect extraction of multiple wavenumbers in a single component, as previously described for the homogeneous data, is the source for this error. As a result, EMD and PCA are to be sensitive to the variation in spatial decay.

D. Traveling Plane Wave Data

Due to the expression derived in (33), we extract $2N$ modes rather than N modes. This is seen in Fig. 4(c), where the two most dominant modes have the same frequencies. It is also important to note that the decomposition is not unique due to the arbitrary phase term ϕ_n . This discrepancy seen in the difference in PCA's relatively poor performance in Table II (where ϕ_n is assumed to be 0) and the excellent performance in Table III, which ignores the phase in its computation. We further observe in Fig. 4(c) that PCA does not have precise phases of 0 and 90 degrees. Similarly, DMD incorrectly extracts modes with equal phases. This is because DMD extracts a complex-valued factorization. This reconstructs the data correctly but creates a large error when compared with the true real-valued modes. Note that the Fig. 4(c) shows the real part of the DMD result but the result in Table II and III compute the error with the complex values. In no other datasets does DMD produce a strong imaginary component.

In addition, WIKSVD appears to have difficulty extracting a mode with a 90 degree phase based on the errors in Table II and that the second most prominent mode in Fig. 4(c) has the wrong frequency. This can be explained due to our choice of Laplacian matrix $\mathbf{L}$, which inherently assumes fixed, zero-valued boundary conditions. Similarly, we observe that WIMF in Fig. 4(c) forces zero-valued boundaries, but the second mode immediately adjusts to have an effective 90 degree phase. This shows that WIMF is more adaptable than WIKSVD. As a result, WIMF overall achieves the best results, but has a larger error relative to the previous two datasets.

E. Multi-Segment Data

Tables II and III show that this dataset is the most difficult of the four. Under low noise conditions, WIMF achieves the best performance. From Fig. 4(d), we observe the relatively large error is due to wavenumbers information "leaking" from one segment to the other. In particular, WIKSVD and DMD include the the modal behavior both regions in their components. Hence, these solution incorrectly include two different wavenumbers in their components.

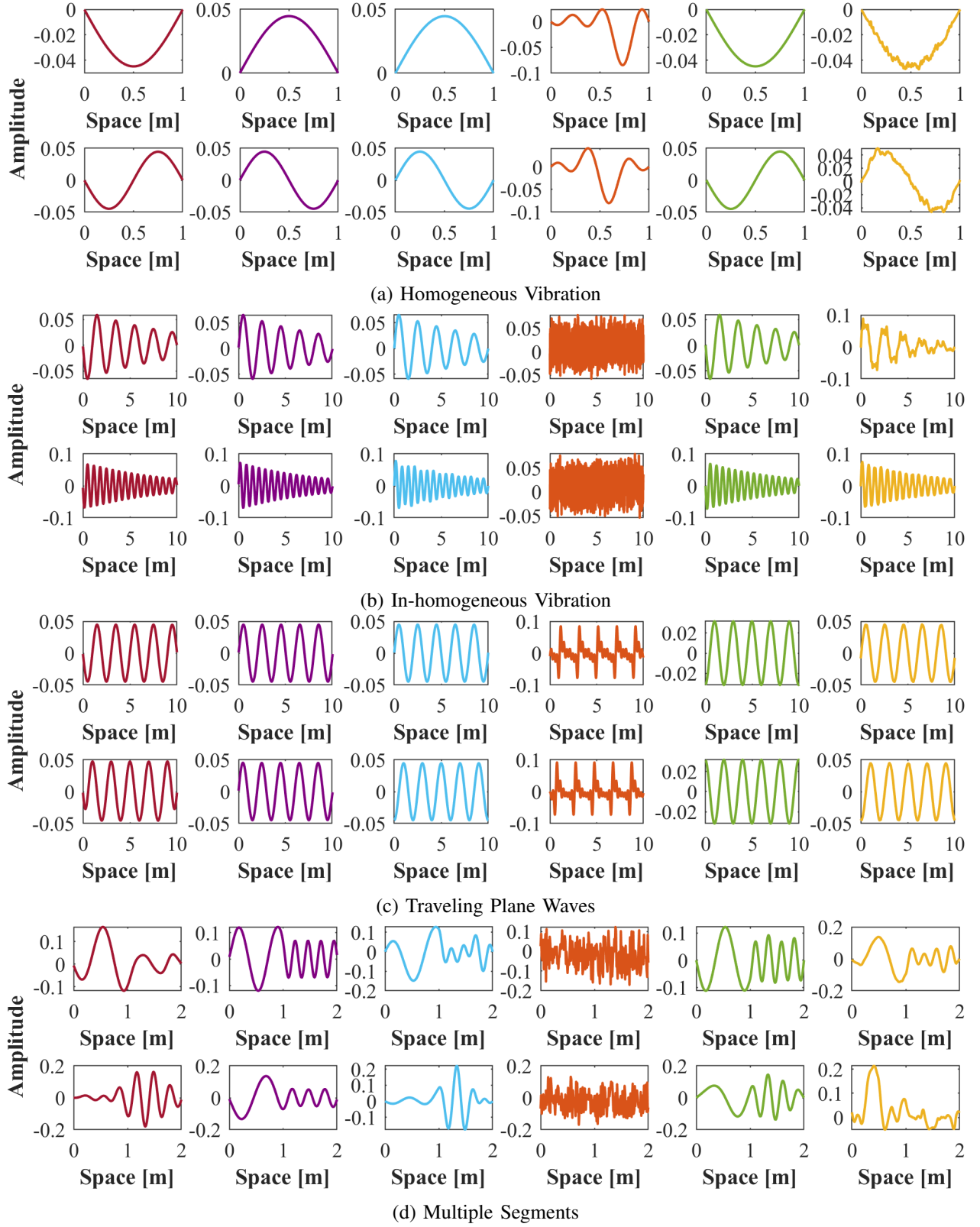


Fig. 4: Two recovered modes (rows) obtained from (WIMF), (WIKSVD), (PCA), (ICA), (DMD), (EMD) for SNR = ∞ .

TABLE II: Mean squared errors (times 10^3) between actual basis generating the data for each case (averaged over 20 Monte-Carlo experiments) and the matrix $\mathbf{D}$ obtained from the algorithms (**WIMF**), (**WIKSVD**), (**ICA**), (**DMD**), (**EMD**) and (**PCA**). Highlight values show the best performance.

Data	SNR [dB]	WIKSVD	ICA	DMD	EMD	PCA	WIMF
Homogeneous	-11.8 $\pm$ 0.4	0.8 $\pm$ 0.1	1827.0 $\pm$ 14.2	233.9 $\pm$ 18.1	1173.4 $\pm$ 159.5	61.8 $\pm$ 8.2	0.7 $\pm$ 0.5
Homogeneous	-5.9 $\pm$ 0.6	0.2 $\pm$ 0.1	1823.3 $\pm$ 11.2	63.9 $\pm$ 8.0	1132.6 $\pm$ 130.5	13.2 $\pm$ 2.1	0.2 $\pm$ 0.1
Homogeneous	-1.9 $\pm$ 0.3	0.1 $\pm$ 0.0	1825.9 $\pm$ 11.9	26.9 $\pm$ 2.2	1207.0 $\pm$ 151.3	5.3 $\pm$ 0.7	0.1 $\pm$ 0.1
Homogeneous	2.3 $\pm$ 0.4	0.0 $\pm$ 0.0	1824.6 $\pm$ 10.1	10.5 $\pm$ 1.1	1243 $\pm$ 133.4	2.0 $\pm$ 0.3	0.1 $\pm$ 0.1
Homogeneous	∞	0.0 $\pm$ 0.0	1155.1 $\pm$ 96.9	0.2 $\pm$ 0.0	1443.9 $\pm$ 160.8	0.1 $\pm$ 0.0	0.1 $\pm$ 0.0
Non-homogeneous	-13.1 $\pm$ 0.6	7.6 $\pm$ 1.1	1917.6 $\pm$ 12.4	252.9 $\pm$ 27.0	1403.0 $\pm$ 181.8	94.8 $\pm$ 60.7	3.9 $\pm$ 0.6
Non-homogeneous	-9.4 $\pm$ 0.7	3.0 $\pm$ 0.4	1920.1 $\pm$ 6.4	133.4 $\pm$ 17.5	1414.4 $\pm$ 189.5	48.1 $\pm$ 48.9	3.4 $\pm$ 0.5
Non-homogeneous	-3.4 $\pm$ 0.7	0.8 $\pm$ 0.1	1922.3 $\pm$ 7.2	38.6 $\pm$ 5.6	1390.9 $\pm$ 139.0	12.5 $\pm$ 1.9	3.1 $\pm$ 0.4
Non-homogeneous	2.7 $\pm$ 0.6	0.2 $\pm$ 0.0	1922.9 $\pm$ 5.0	10.4 $\pm$ 1.8	1381.2 $\pm$ 137.6	12.6 $\pm$ 16.2	3.1 $\pm$ 0.5
Non-homogeneous	∞	0.0 $\pm$ 0.0	1956.6 $\pm$ 67.1	0.0 $\pm$ 0.0	1257.6 $\pm$ 162.3	25.1 $\pm$ 34.0	3.0 $\pm$ 0.4
Travelling Wave	-11.7 $\pm$ 0.6	320.1 $\pm$ 106.2	1913.0 $\pm$ 13.5	1170.2 $\pm$ 57.8	1415.6 $\pm$ 129.8	334.2 $\pm$ 82.3	1129.0 $\pm$ 235.4
Travelling Wave	-5.9 $\pm$ 0.5	225.5 $\pm$ 78.9	1927.6 $\pm$ 9.6	1108.1 $\pm$ 32.7	1346.6 $\pm$ 115.3	212.2 $\pm$ 89.4	40.2 $\pm$ 1.5
Travelling Wave	-1.8 $\pm$ 0.6	199.9 $\pm$ 50.0	1930.4 $\pm$ 13.7	1084.1 $\pm$ 38.3	1267.8 $\pm$ 110.2	241.6 $\pm$ 67.7	38.3 $\pm$ 0.6
Travelling Wave	1.8 $\pm$ 0.8	235.9 $\pm$ 52.0	1932.8 $\pm$ 18.8	1094.9 $\pm$ 29.4	1390.5 $\pm$ 163.5	280.8 $\pm$ 62.8	36.7 $\pm$ 0.4
Travelling Wave	∞	409.5 $\pm$ 13.5	1189.0 $\pm$ 43.3	1115.9 $\pm$ 1.3	1548.1 $\pm$ 75.4	342.5 $\pm$ 5.6	35.8 $\pm$ 0.3
Segmented	-11.6 $\pm$ 0.0	1589.0 $\pm$ 38.2	1792.1 $\pm$ 25.2	1515.1 $\pm$ 84.4	1264.2 $\pm$ 116.8	1526.5 $\pm$ 77.4	891.8 $\pm$ 117.3
Segmented	-9.1 $\pm$ 0.0	1594.1 $\pm$ 36.0	1794.9 $\pm$ 29.3	1532.5 $\pm$ 78.8	1261.0 $\pm$ 131.1	1491.6 $\pm$ 60.2	913.2 $\pm$ 117.5
Segmented	-3.1 $\pm$ 0.0	1638.1 $\pm$ 16.3	1792.5 $\pm$ 27.3	1520.5 $\pm$ 68.8	1191.5 $\pm$ 127.6	1480.8 $\pm$ 35.0	767.3 $\pm$ 20.3
Segmented	5.5 $\pm$ 0.0	1664.0 $\pm$ 46.2	1786.2 $\pm$ 28.7	1483.3 $\pm$ 25.0	1158.0 $\pm$ 134.0	1470.5 $\pm$ 40.8	616.3 $\pm$ 4.7
Segmented	∞	889.9 $\pm$ 0.0	1840 $\pm$ 12.8	1006.1 $\pm$ 0.0	1019.1 $\pm$ 0.0	1516.1 $\pm$ 0.0	852.0 $\pm$ 0.1

TABLE III: Mean squared errors (times 10^3) between absolute value of Fourier transform of actual basis generating the data for each case (averaged over 20 Monte-Carlo experiments) and the matrix $\mathbf{D}$ obtained from the algorithms (**WIMF**), (**WIKSVD**), (**ICA**), (**DMD**), (**EMD**) and (**PCA**). Highlighted values show the best performance.

Data	SNR [dB]	WIKSVD	ICA	DMD	EMD	PCA	WIMF
Homogeneous	-11.8 $\pm$ 0.4	0.5 $\pm$ 0.1	1732.0 $\pm$ 23.9	229.5 $\pm$ 18.2	891.3 $\pm$ 151.2	59.9 $\pm$ 7.9	0.2 $\pm$ 0.1
Homogeneous	-5.9 $\pm$ 0.6	0.1 $\pm$ 0.1	1719.3 $\pm$ 21.3	62.5 $\pm$ 7.8	838.2 $\pm$ 150.2	12.6 $\pm$ 2.1	0.1 $\pm$ 0.0
Homogeneous	-1.9 $\pm$ 0.3	0.0 $\pm$ 0.0	1725.7 $\pm$ 24.9	26.1 $\pm$ 2.2	733.0 $\pm$ 104.9	4.9 $\pm$ 0.6	0.0 $\pm$ 0.0
Homogeneous	2.3 $\pm$ 0.4	0.0 $\pm$ 0.0	1723.9 $\pm$ 25.3	10.1 $\pm$ 1.0	768.8 $\pm$ 86.2	1.8 $\pm$ 0.2	0.0 $\pm$ 0.0
Homogeneous	∞	0.0 $\pm$ 0.0	782.7 $\pm$ 110.2	0.2 $\pm$ 0.0	1020.5 $\pm$ 111.3	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0
Non-homogeneous	-13.1 $\pm$ 0.6	4.3 $\pm$ 0.6	1767.4 $\pm$ 20.5	238.7 $\pm$ 25.0	958.8 $\pm$ 122.0	78.9 $\pm$ 52.5	1.8 $\pm$ 0.4
Non-homogeneous	-9.4 $\pm$ 0.7	1.5 $\pm$ 0.2	1772.0 $\pm$ 15.6	126.0 $\pm$ 16.5	944.4 $\pm$ 135.7	38.4 $\pm$ 40.0	1.5 $\pm$ 0.3
Non-homogeneous	-3.4 $\pm$ 0.7	0.3 $\pm$ 0.0	1779.4 $\pm$ 15.4	36.0 $\pm$ 5.4	918.6 $\pm$ 158.6	8.7 $\pm$ 14.2	1.3 $\pm$ 0.3
Non-homogeneous	2.7 $\pm$ 0.6	0.1 $\pm$ 0.0	1771.7 $\pm$ 15.3	9.5 $\pm$ 1.7	869.4 $\pm$ 151.3	8.4 $\pm$ 12.7	1.3 $\pm$ 0.3
Non-homogeneous	∞	0.0 $\pm$ 0.0	1753.3 $\pm$ 79.8	0.0 $\pm$ 0.0	948.3 $\pm$ 116.4	1.3 $\pm$ 0.3	18.9 $\pm$ 28.3
Travelling Wave	-11.7 $\pm$ 0.6	102.6 $\pm$ 62.5	1878.9 $\pm$ 14.6	668.1 $\pm$ 23.1	1113.0 $\pm$ 74.4	147.8 $\pm$ 38.9	628.3 $\pm$ 187.2
Travelling Wave	-5.9 $\pm$ 0.5	8.9 $\pm$ 1.4	1893.0 $\pm$ 8.3	597.0 $\pm$ 2.7	1040.7 $\pm$ 64.7	26.0 $\pm$ 4.5	34.0 $\pm$ 0.8
Travelling Wave	-1.8 $\pm$ 0.6	5.1 $\pm$ 0.8	1899.5 $\pm$ 12.0	588.0 $\pm$ 1.6	970.5 $\pm$ 63.8	10.1 $\pm$ 2.3	32.9 $\pm$ 0.6
Travelling Wave	1.8 $\pm$ 0.8	4.0 $\pm$ 1.1	1902.7 $\pm$ 16.4	585.5 $\pm$ 0.5	1088.8 $\pm$ 150.5	4.1 $\pm$ 1.0	31.6 $\pm$ 0.3
Travelling Wave	∞	1.5 $\pm$ 0.2	1081.0 $\pm$ 40.1	585 $\pm$ 0	1446.5 $\pm$ 60.8	0.1 $\pm$ 0.1	31.3 $\pm$ 0.3
Segmented	-11.6 $\pm$ 0.0	1037.5 $\pm$ 36.0	1373.8 $\pm$ 24.3	949.2 $\pm$ 33.3	724.3 $\pm$ 80.7	957.9 $\pm$ 35.3	536.7 $\pm$ 71.4
Segmented	-9.1 $\pm$ 0.0	1067.2 $\pm$ 47.9	1383.5 $\pm$ 25.9	907.1 $\pm$ 27.8	626.9 $\pm$ 88.2	938.5 $\pm$ 34.9	502.7 $\pm$ 98.8
Segmented	-3.1 $\pm$ 0.0	1249.5 $\pm$ 79.9	1382.9 $\pm$ 25.1	863.7 $\pm$ 45.3	496.2 $\pm$ 39.5	923.7 $\pm$ 26.5	397.7 $\pm$ 128.4
Segmented	5.5 $\pm$ 0.0	1175.7 $\pm$ 18.7	1372.1 $\pm$ 24.5	973.5 $\pm$ 38.0	448.5 $\pm$ 63.8	908.8 $\pm$ 21.7	428.5 $\pm$ 6.6
Segmented	∞	623.5 $\pm$ 0.0	1337.8 $\pm$ 25.0	348.7 $\pm$ 0.0	426.6 $\pm$ 0.0	887.2 $\pm$ 0.0	357.9 $\pm$ 0.1

PCA and EMD achieve decent localization of the modes but fail to extract the correct frequency content, as evident in Table III. WIMF, EMD, and PCA all demonstrate some degree of separation and localization, but are not tightly bound to their spatial region. In particular, WIMF does not achieve stronger localization because the wave equation constraints are placed on each mode rather than the complete solution, and the discontinuity between segments does not satisfy the wave equation for each individual mode.

F. General discussion on wave-informed matrix factorization

In the performance analysis of this method, we observe that *wave-informed matrix factorization* successfully recovers pure sinusoids with zero phase in the homogeneous vibrations case in Section IV-B and recovers decaying sinusoids in the inhomogeneous vibrations case in Section IV-C. In Section IV-D, it

recovers approximate cosines that start at zero and eventually converge to the sinusoid shape, when the true result has a phase shift. For the multi-segment case in Section IV-E, it recovers sinusoidal vibrations that locally to each region.

We would like to emphasize here that hard constraining the wave equation, i.e. solving an optimization similar to

$$\arg \min_{\mathbf{D}, \mathbf{X}, N, \mathbf{k}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}^T\|_F^2 + \frac{\lambda}{2} (\|\mathbf{D}\|_F^2 + \|\mathbf{X}\|_F^2) \\ \text{s.t. } \mathbf{L}\mathbf{D}_i = -k_i^2 \mathbf{D}_i, \forall i \in [N]$$

would be equivalent to solving, for $\mathbf{D}$, the eigenvectors of $\mathbf{L}$ that closely span (in the ℓ_2 sense) the column space of $\mathbf{Y}$. These are pure sinusoids with zero phase. In contrast to this, we observe that the columns of $\mathbf{D}$ are dictated by both data and the soft constraint introduced by the wave equation. In particular, we note that the columns extracted in most of the

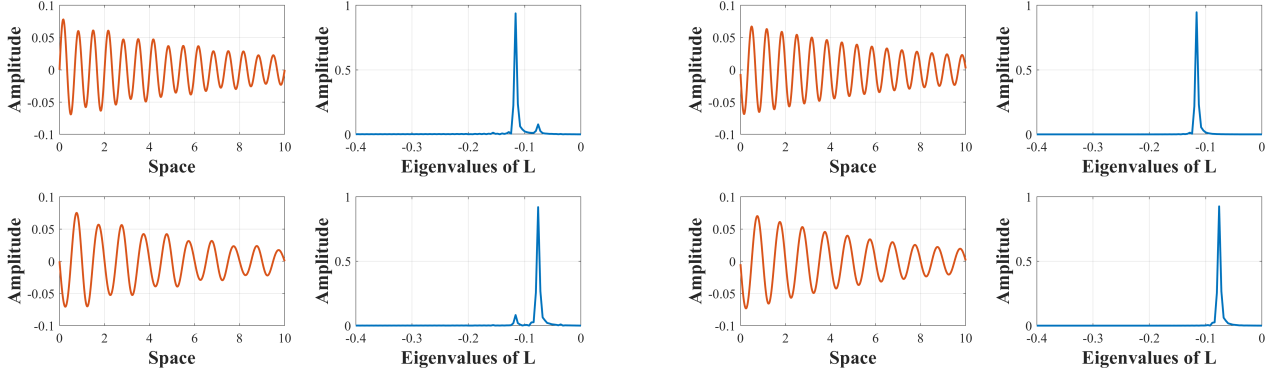


Fig. 5: (a) Two columns from the basis obtained from PCA and the absolute value in the transformed domain (transformed to the eigenbasis of L) (b) Two columns from the basis obtained from wave-informed matrix factorization and the absolute value in the transformed domain

cases only approximately satisfy the wave-constraint and do not satisfy the wave equation. This property distinguishes our algorithm from traditional factorization methods that are either data dependent or are entirely parameterized by the assumed physics. We attribute this to the recovery of components such as exponentially decaying sinusoids (which do not satisfy the wave equation).

V. CONCLUSIONS

We have introduced wave-informed matrix factorization and developed a framework and an algorithm with provable, global optimality guarantees for the same. More generally, this work introduced a methodology to enforce linear homogeneous partial differential equation to influence a matrix factorization algorithm, and demonstrated this for the case of the time-independent version of the Helmholtz wave equation. The output from the algorithm was compared with that of state-of-the-art algorithms for modal and component analysis. We demonstrated that the wave-informed approach learns representations that are more physically relevant and practical for the purpose of modal analysis.

Future work will include generalizing this approach to a variety of linear PDEs beyond the wave equation (especially to adapt the works [49], [52] to our framework) as well as wave propagation along more than one dimension and extension to applications in baseline-free anomaly detection for structural health monitoring [27], [29].

VI. ACKNOWLEDGEMENTS

This work is partially supported by NSF EECS-1839704, NSF CISE-1747783, NIH NIA 1R01AG067396, ARO MURI W911NF-17-1-0304, and NSF-Simons MoDL 2031985.

REFERENCES

- [1] S. Mehrkanoon, "Deep shared representation learning for weather elements forecasting," *Knowledge-Based Systems*, vol. 179, pp. 120–128, 2019.
- [2] M. Ringsquandl, S. Lamparter, R. Lepratti, and P. Kröger, "Knowledge fusion of manufacturing operations data using representation learning," in *Proc. of the IFIP International Conference on Advances in Production Management Systems*. Springer, 2017, pp. 302–310.
- [3] N. Zobeiry, J. Reiner, and R. Vaziri, "Theory-guided machine learning for damage characterization of composites," *Composite Structures*, vol. 246, p. 112407, 2020.
- [4] A. Mohamed, G. Hinton, and G. Penn, "Understanding how deep belief networks perform acoustic modelling," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 4273–4276.
- [5] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [6] A. Stanbridge and D. Ewins, "Modal testing using a scanning laser doppler vibrometer," *Mechanical systems and signal processing*, vol. 13, no. 2, pp. 255–270, 1999.
- [7] Y. Yang, C. Dorn, T. Mancini, Z. Talken, S. Nagarajaiah, G. Kenyon, C. Farrar, and D. Mascareñas, "Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-nyquist), video measurements," *Journal of Sound and Vibration*, vol. 390, pp. 232–256, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022460X16306897>
- [8] D. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, pp. 251 – 259, 06 1991.
- [9] A. Cunha and E. Caetano, "Experimental modal analysis of civil engineering structures," *Sound and Vibration*, vol. 40, 06 2006.
- [10] K. D. Marshall, "Modal analysis of a violin," *Journal of the Acoustical Society of America*, vol. 77, pp. 695–709, 1985.
- [11] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [12] M. Wathelet, B. Guillier, P. Roux, C. Cornou, and M. Ohnberger, "Rayleigh wave three-component beamforming: signed ellipticity assessment from high-resolution frequency-wavenumber processing of ambient vibration arrays," *Geophysical Journal International*, vol. 215, no. 1, pp. 507–523, 07 2018. [Online]. Available: <https://doi.org/10.1093/gji/ggy286>
- [13] L. Yu and Z. Tian, "Guided wave phased array beamforming and imaging in composite plates," *Ultrasonics*, vol. 68, pp. 43–53, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0041624X16000160>
- [14] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. 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] N. Benjamin Erichson, M. Muehlebach, and M. W. Mahoney, "Physics-informed autoencoders for lyapunov-stable fluid flow prediction," *arXiv [physics.comp-ph]*, p. 14, 2019.
- [16] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.
- [17] M. A. Nabian and H. Meidani, "Physics-driven regularization of deep neural networks for enhanced engineering design and analysis," *Journal*

- of *Computing and Information Science in Engineering*, vol. 20, no. 1, 2019.
- [18] A. M. Tartakovsky, C. O. Marrero, P. Perdikaris, G. D. Tartakovsky, and D. Barajas-Solano, "Learning parameters and constitutive relationships with physics informed deep neural networks," *arXiv [math.AP]*, p. 22, 2018.
 - [19] Y. Long, X. She, and S. Mukhopadhyay, "Hybridnet: Integrating model-based and data-driven learning to predict evolution of dynamical systems," in *Proc. of the Conference on Robot Learning*, 2018, p. 10.
 - [20] K. Shukla, P. C. Di Leoni, J. Blackshire, D. Sparkman, and G. E. Karniadakis, "Physics-informed neural network for ultrasound nondestructive quantification of surface breaking cracks," *arXiv [cs.LG]*, p. 19, 2020.
 - [21] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (PGNN): An application in lake temperature modeling," *arXiv [cs.LG]*, vol. 1710.11431, p. 11, 2017.
 - [22] D. Lei, X. Chen, and J. Zhao, "Opening the black box of deep learning," *arXiv [cs.LG]*, vol. 1805.08355, p. 27, 2018.
 - [23] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "State estimation for unobservable distribution systems via deep neural networks," in *Proc. of the IEEE Power & Energy Society General Meeting*. Portland: IEEE, 2018, pp. 1–5.
 - [24] T.-j. Luo, Y. Fan, L. Chen, G. Guo, and C. Zhou, "EEG signal reconstruction using a generative adversarial network with Wasserstein distance and temporal-spatial-frequency loss," *Frontiers in Neuroinformatics*, vol. 14, p. 15, 2020.
 - [25] L. Mosser, O. Dubrule, and M. J. Blunt, "Stochastic seismic waveform inversion using generative adversarial networks as a geological prior," *Mathematical Geosciences*, vol. 52, pp. 53–79, 2020.
 - [26] J. B. Harley and J. M. 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.
 - [27] K. S. Alguri, J. Melville, and J. 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.
 - [28] K. S. Alguri, C. C. Chia, and J. B. Harley, "Model-driven, wavefield baseline subtraction for damage visualization using dictionary learning," vol. 2, 2017, pp. 2276–2283.
 - [29] —, "Sim-to-real: Employing ultrasonic guided wave digital surrogates and transfer learning for damage visualization," *Ultrasonics*, vol. 111, p. 106338, 2021.
 - [30] K. S. Alguri and J. B. Harley, "Transfer learning of ultrasonic guided waves using autoencoders: A preliminary study," *AIP Conf. Proc.*, vol. 2102, no. 1, p. 50013, 2019.
 - [31] H. V. Tetali, K. S. Alguri, and J. B. Harley, "Wave Physics Informed Dictionary Learning In One Dimension," in *Proc. of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2019, pp. 1–6.
 - [32] H. V. Tetali, J. B. Harley, and B. D. Haeffele, "Wave-informed matrix factorization with global optimality guarantees," *arXiv preprint arXiv:2107.09144*, 2021.
 - [33] S. Bhojanapalli, B. Neyshabur, and N. Srebro, "Global optimality of local search for low rank matrix recovery," in *Advances in Neural Information Processing Systems*, 2016, pp. 3873–3881.
 - [34] R. Ge, J. D. Lee, and T. Ma, "Matrix completion has no spurious local minimum," in *Advances in Neural Information Processing Systems*, 2016, pp. 2973–2981.
 - [35] D. Park, A. Kyrillidis, C. Caramanis, and S. Sanghavi, "Non-square matrix sensing without spurious local minima via the burer-monteiro approach," *arXiv preprint arXiv:1609.03240*, p. 14, 2016.
 - [36] R. Ge, C. Jin, and Y. Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis," in *Proc. of the International Conference on Machine Learning*, vol. 70. JMLR. org, 2017, pp. 1233–1242.
 - [37] B. Haeffele, E. Young, and R. Vidal, "Structured low-rank matrix factorization: Optimality, algorithm, and applications to image processing," in *Proc. of the International Conference on Machine Learning*, 2014, pp. 2007–2015.
 - [38] B. D. Haeffele and R. Vidal, "Structured low-rank matrix factorization: Global optimality, algorithms, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 6, 2019.
 - [39] F. Bach, "Convex relaxations of structured matrix factorizations," *arXiv preprint arXiv:1309.3117*, p. 33, 2013.
 - [40] E. Figueiredo and J. Brownjohn, "Three decades of statistical pattern recognition paradigm for shm of bridges," *Structural Health Monitoring*, vol. 0, no. 0, p. 14759217221075241, 0. [Online]. Available: <https://doi.org/10.1177/14759217221075241>
 - [41] G. Strang, *Computational Science and Engineering*. Wellesley-Cambridge Press, 2007.
 - [42] G. Golub and C. Van Loan, *Matrix Computations*, ser. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.
 - [43] F. Chung and S.-T. Yau, "Discrete green's functions," *Journal of Combinatorial Theory, Series A*, vol. 91, no. 1-2, pp. 191–214, 2000.
 - [44] N. Srebro and A. Shraibman, "Rank, trace-norm and max-norm," in *Proc. of the International Conference on Computational Learning Theory*. Springer, 2005, pp. 545–560.
 - [45] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.
 - [46] C. Malherbe and N. Vayatis, "Global optimization of lipschitz functions," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2314–2323.
 - [47] A. V. Oppenheim and R. W. Schaffer, *Digital signal processing*, ser. Prentice-Hall international editions. Prentice-Hall, 1975. [Online]. Available: <https://www.worldcat.org/oclc/01031348>
 - [48] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Rev.*, vol. 31, no. 4, pp. 614–627, 1989. [Online]. Available: <https://doi.org/10.1137/1031128>
 - [49] C. Dorn, S. Dasari, Y. Yang, C. Farrar, G. Kenyon, P. Welch, and D. Mascareñas, "Efficient full-field vibration measurements and operational modal analysis using neuromorphic event-based imaging," *Journal of Engineering Mechanics*, vol. 144, no. 7, 7 2018.
 - [50] Y. Yang, C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, and D. Mascareñas, "Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification," *Mechanical Systems and Signal Processing*, vol. 85, pp. 567–590, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327016303272>
 - [51] H. Ma, Y.-Z. Lin, and Z. Nie, "Physical interpretation of principal component analysis for structural dynamics through string vibration," *International Journal of Structural Stability and Dynamics*, 2019.
 - [52] Z. Lai, I. Alzugaray, M. Chli, and E. Chatzi, "Full-field structural monitoring using event cameras and physics-informed sparse identification," *Mechanical Systems and Signal Processing*, vol. 145, p. 106905, 2020.
 - [53] S. Haykin, "Radar signal processing," *IEEE ASSP Magazine*, vol. 2, no. 2, pp. 2–18, 1985.
 - [54] J. A. Jensen, *Estimation of blood velocities using ultrasound: a signal processing approach*. Cambridge University Press, 1996.
 - [55] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO signal processing*. Springer Science & Business Media, 2006.
 - [56] D. S. Jones, "Introduction to wave phenomena. by a. hirose and k. e. lonngren. wiley, 1985. 355 pp. £53." *Journal of Fluid Mechanics*, vol. 162, p. 503–503, 1986.
 - [57] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.
 - [58] Z. Liao, L. Song, P. Chen, Z. Guan, Z. Fang, and K. Li, "An effective singular value selection and bearing fault signal filtering diagnosis method based on false nearest neighbors and statistical information criteria," *Sensors (Basel, Switzerland)*, vol. 18, 2018.
 - [59] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994, higher Order Statistics. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0165168494900299>
 - [60] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," *Journal of Computational Dynamics*, vol. 1, no. 2, pp. 391–421, 2014.
 - [61] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.
 - [62] D. Kim, M. Park, and H.-S. Oh, "Bidimensional statistical empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 191–194, 2012.
 - [63] C. Damerval, S. Meignen, and V. Perrier, "A fast algorithm for bidimensional emd," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 701–704, 2005.
 - [64] R. J. Allemang, A. W. Phillips, and M. R. Allemang, "Application of principal component analysis methods to experimental structural dynamics," in *Structural Dynamics, Volume 3*. Springer, 2011, pp. 477–498.

- [65] G. Rilling and P. Flandrin, "One or two frequencies? the empirical mode decomposition answers," *IEEE transactions on signal processing*, vol. 56, no. 1, pp. 85–95, 2007.
- [66] B. R. Noack, W. Stankiewicz, M. Morzyński, and P. J. Schmid, "Recursive dynamic mode decomposition of transient and post-transient wake flows," *Journal of Fluid Mechanics*, vol. 809, pp. 843–872, 2016.



Harsha Vardhan Tetali has a PhD in Electrical and Computer Engineering from University of Florida, Gainesville. Prior to this, he received his B.Tech in Electronics and Communication Engineering from SVNIT, Surat and M.Tech in Electrical Engineering from IIT Gandhinagar. His research spans signal processing, machine learning, and optimization theory. He currently works as a staff engineer at Marvell Technology.



Joel B. Harley (S'05-M'14) received his B.S. degree in Electrical Engineering from Tufts University in Medford, MA, USA. He received his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Carnegie Mellon University in Pittsburgh, PA, USA in 2011 and 2014, respectively. In 2018, he joined the University of Florida, where he is currently an associate professor in the Department of Electrical and Computer Engineering. Previously, he was an assistant professor in the Department of Electrical and Computer Engineering at the University

of Utah. His research interests include integrating novel signal processing, machine learning, and data science methods for the analysis of waves and time-series data. Dr. Harley's awards and honors include the 2021 Achenbach Medal from the International Workshop on Structural Health Monitoring, a 2020 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society Star Ambassador Award, a 2020 and 2018 Air Force Summer Faculty Fellowship, a 2017 Air Force Young Investigator Award, a 2014 Carnegie Mellon A. G. Jordan Award (for academic excellence and exceptional service to the community). He has published more than 90 technical journal and conference papers, including four best student papers. He is a member of the Ultrasonics, Ferroelectrics, and Frequency Control Society, a member of the IEEE Signal Processing Society, and a member of the Acoustical Society of America.



Benjamin D. Haeffele is an Associate Research Scientist in the Mathematical Institute for Data Science and the Center for Imaging Science at Johns Hopkins University. His research interests involve developing theory and algorithms for processing high-dimensional data at the intersection of machine learning, optimization, and computer vision. In addition to basic research in data science he also works on a variety of applications in medicine, microscopy, and computational imaging. He received his Ph.D. in Biomedical Engineering at Johns Hopkins University in 2015 and his B.S. in Electrical Engineering from the Georgia Institute of Technology in 2006.

This paper has supplementary downloadable material available at <https://arxiv.org/pdf/2312.13584.pdf> provided by the author. The material includes some additional proofs which have been left out in the main body of the paper. This material is 370 KB in size.