Redatuming physical systems using symmetric autoencoders

Pawan Bharadwaj[*]

Indian Institute of Science, CV Raman Rd., Bengaluru, Karnataka 560012, India

Matthew Li and Laurent Demanet

Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

This paper considers physical systems described by hidden states and indirectly observed through repeated measurements corrupted by unmodeled nuisance parameters. A network-based representation learns to disentangle the coherent information (relative to the state) from the incoherent nuisance information (relative to the sensing). Instead of physical models, the representation uses symmetry and stochastic regularization to inform an autoencoder architecture called SymAE. It enables redatuming, i.e., creating virtual data instances where the nuisances are uniformized across measurements.

I. INTRODUCTION

In contemporary sciences, there is increasing reliance on experimental designs involving measurements that are corrupted by unmodeled and uncontrollable *nuisance variations*. For example, in geophysics, specifically passive time-lapse seismic monitoring, the recorded seismic data are generated by uncontrollable sources related to tectonic stress changes in the subsurface [1] or ocean-wave activity [2]. Similarly, in astronomy, the fluorescent emissions which characterize the lunar surface's chemical composition fluctuate depending on unmodeled solar flares [3,4]. Such designs introduce ambiguity when determining whether changes in data represent *coherent information*, i.e., a signal indicating changes related to the underlying physical state, or conversely represent *nuisance information*, i.e., incidental variations due to noise inherent to the data acquisition. Nevertheless, these uncontrollable experiments (along with others listed in Table I) remain the only feasible avenue to measure and study certain physical phenomena. This motivates the development of algorithmic tools to reliably disentangle the nuisance noise from coherent signals in these settings.

This paper focuses on a subclass of such experiments where variations in the measurements occur across two different scales: one in which the physical processes affecting the coherent information occur, and the other scale corresponding to noise processes that induce variations in the nuisance information. We assume there exists a strong scale separation in which the noise processes occur at a significantly faster rate than the former. This assumption is crucial as it enables us to

neglect variations in the coherent information within a collection of closely spaced measurements. We refer to these closely spaced (in scale) or repeated measurements as *instances*, and groups of instances altogether describe a *(physical) state*; the dichotomy in scaling and its relation to instances and states is illustrated in Fig. 1.

We further assume that the experiments we consider abundantly produce measurements describing the same physical state, albeit with different nuisance variations. Examples of experiments that satisfy these conditions are listed in Table I. Critically, in principle, having access to a sufficiently dissimilar collection of instances enables disentanglement of coherent information from nuisance information without any reference to the underlying physical model [5].

Our proposed approach to disentanglement decomposes each measurement into separate *latent codes* which are correlated, but typically not equivalent, to the corresponding (unknown) parametric representations underlying each source of information. These coherent and nuisance latent codes are determined from an autoencoding architecture with an encoder, which maps into the latent space, and a decoder, which reconstructs the data in a near-lossless fashion. Notably, this avoids explicit modeling of both the physics and the nuisances by instead relying on data to inform these properties. This framework can be seen as a vast generalization of multichannel blind deconvolution [6], where neural networks replace the convolutional signal model, the coherent information replaces the unknown source, and the nuisance variations correspond to the unknown filters.

Disentanglement into coherent and incoherent latent variables enables a reliable comparison of coherent information between instances collected from different states. However, this relegates the analysis into a latent space that is abstractly related to the physical system. To that end, we propose an additional mechanism, called *redatuming*, which converts from latent coordinates back into the nominal data space representation with specifically chosen properties. This involves combining coherent information from one instance with the nuisance information from a reference instance in order to

---

[*]Also at Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

TABLE I. A representative list of experiments, where numerous instances are measured at each state that describes the physical phenomenon of interest. The instances exhibit dissimilarities due to the nuisance variations while sharing coherent information about the state.

Experiment; goal is to characterize the variation of . . .	Measured instances in each state comprise . . .	Components of information in each instance	
		Coherent information describing the state	Nuisance information
Seismic time-lapse imaging; subsurface with time [7,8]	Waves propagating through a given subsurface region, but generated by different uncontrollable sources	Mechanical properties of the subsurface region	Source signature, location, and mechanism
Seismology; source mechanism among different earthquakes [9,10]	Waves from an earthquake measured at different receivers	Earthquake's spectrum, geology near epicenter region, etc.	Multipathing in the subsurface, Doppler effects from rupture propagation, etc.
Lunar x-ray fluorescence spectroscopy; lunar geology with location [4,11]	Fluorescence from regions that share a given rock type under different solar conditions	Elemental composition of the rock type	Solar-flare information
Asteroseismology [12,13]; pulsating mechanism among rapidly oscillating Ap [14] or Delta Scuti [15] stars	Brightness of a given star measured in different temporal windows	Physics that is symmetrical under time translation, e.g., star's internal structure, kappa opacity mechanism, etc.	Short-lived excitation mechanisms, e.g., surface convection, variable radial velocity, etc.

synthesize a *virtual* instance that is not originally measured (i.e., not present in the dataset). The relevance of such virtual data instances is that they can be engineered to share their nuisance information with another (measured) data instance so that any remaining discrepancy can be solely explained from differences in the underlying coherent physical states. This entire process can be alternatively viewed as "swapping the physics" between states. We conjecture that this type of re-datuming can help rethink how to approach inverse problems with significant uncertainties in the forward model.

We illustrate redatuming using the example of time-lapse geophysical subsurface monitoring cited in Table I. Here, seismic surveys are conducted to measure the subsurface properties (the coherent information) indirectly by recording reflected and transmitted elastic waves generated by un-controlled or unreliably modeled mechanisms (the nuisance information). In this setting, it is reasonable to assume that changes to the complex heterogeneous subsurface occur on a significantly slower timescale (e.g., on the order of months) than the variations in the uncontrollable seismic sources (on the order of hours or days). As such, the variations in the subsurface mechanical properties, as functions of the lateral

scale dependent on distance x and depth z, can be safely neglected within each state. The goal is to detect and characterize changes in subsurface between the states, e.g., to distinguish between "state #1" and "state #2" along the abscissa of Fig. 2 [16]. In each state, the measured instances constitute the time-dependent (indexed using t) wave field recorded at a given set of receivers (indexed using r) in the medium [plotted in Figs. 2(a) and 2(c)].

In effect, the seismic sources are realized with randomized signatures and locations (cf. the nuisance variations visualized along the ordinate of Fig. 2) . This confounds direct visual comparison between the measured instances Figs. 2(a) and 2(c) as it is unclear whether the localized changes (indicated by the blue arrow) are to be attributed to changes in the medium or the source. Redatuming overcomes this ambiguity by generating a *virtual* instance, plotted in Fig. 2(d). This virtual instance is engineered by replacing the subsurface information of Fig. 2(a) with that of Fig. 2(c) while retaining its source/nuisance information. As a result, the virtual instance can be subtracted from the reference instance [here Fig. 2(a)] to qualify or quantify potential subsurface changes via standard imaging techniques. We emphasize that redatuming enables domain experts to perform data analysis using traditional tools without any reference to the implicit latent space.

These ideas are inspired by recent machine learning literature where redatuming is instead referred to as *styling*, or *deep-fakes* (see, e.g., [17–19]), and the reliance on multiple instances is referred to as *weak supervision* [5,20]. However, we note that these communities primarily apply these tools to images with significant visual structure wherein nuisance information relates to the "image style," and the coherent information relates to the "image content." This paper instead introduces the idea of redatuming to scientific signals, enabling us to quantify virtual-instance accuracy against explicit synthetic models rigorously.
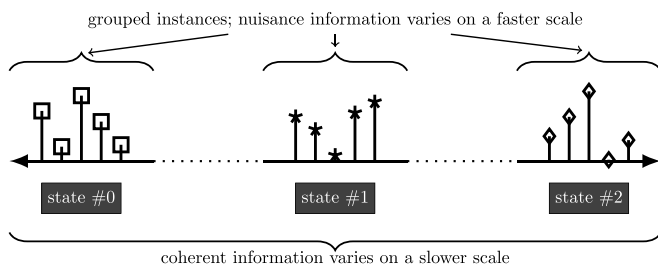

FIG. 1. Illustration of separation between two scales. The coherent information (marker shape) related to the state varies on a much slower scale compared to the nuisance information (marker distance from the baseline). Here, each marker represents a measurement.
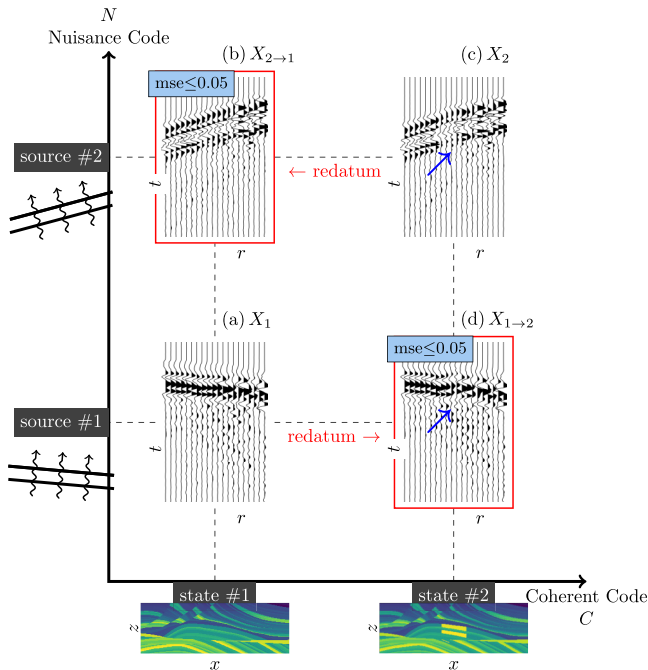
FIG. 2. Redatuming is equivalent to swapping the coherent and nuisance information in SymAE's latent space. Here, as SymAE learns to represent the information on medium (coherent) and incoming plane-wave sources (nuisance) separately, the recorded wave field [(a) and (c)] in a seismic experiment (see Table I) can be redatumed to generate virtual measurements [(b) and (d)]. A low error confirms that the redatuming operator captures salient features of the wave-equation modeling despite a multipath propagation due to the complex medium inhomogeneities.

II. OUR CONTRIBUTIONS

To achieve redatuming, we propose an unsupervised deep-learning architecture called symmetric autoencoder (SymAE). Achieving the requisite disentangled latent representation with SymAE requires two deliberate architectural design choices: (1) The encoder for the coherent latent variables is constrained to be symmetric with respect to the ordering of the instances indexed by nuisance variations. (2) The remaining latent-code dimensions are encouraged to encode independent information by stochastic regularization that promotes dissimilarity among the instances [21]. Therefore, these remaining latent components are designed to *not* represent the coherent information and correspond only to the nuisance variations.

Once the coherent and nuisance information are disentangled in the latent space, redatuming is equivalent to decoding a hybrid latent code, specifically, a hybridization of the coherent code from one state and the nuisance code of an instance from another state. We provide numerical evidence that SymAE's redatuming preserves and captures the salient features of the underlying physical modeling operator, thus enabling the use of virtual data points for subsequent downstream tasks such as parameter estimation. We numerically validate that the virtual instances generated without reference to the physics satisfy the governing wave equation up to a low relative mean-squared error. This indicates that SymAE redatuming is consistent with, or preserves, the physics of wave propagation.

The concept of redatuming appears in the context of traditional seismic inversion [22–26]. The major differences with our current generalized approach, however, are (1) the seismic-specific redatuming is limited to swapping sources or receivers from one state to another—in contrast, SymAE aims to swap any information that is coherent across the instances; and (2) seismic redatuming either requires prior knowledge about the subsurface or uses physics-derived relations with convolutions or cross-correlations—in contrast, SymAE derives the redatuming operators from the recorded data in an unsupervised manner, unlocking processing for far more general situations than cross-correlations allow. We refer the reader to [27–29] for examples of analytical-based redatuming applied to specific geophysical settings.

SymAE heavily relies on imposing symmetries in the encoder to separate the latent code. This idea of using symmetry, or equivalently physical priors, to promote structure in the neural networks has been proposed in various works; for instance, [30] embedded even/odd symmetry of a function and energy conservation into a neural network by adding special *hub* layers; [31] propose gauge equivariant CNN layers to capture rotational symmetry; and [32] structures their networks following a Hamiltonian in order to learn physically conserved quantities and symmetries. The choice of symmetry is bespoke to each application, and the identification of valid symmetries in our physical prior is one of the contributions and insights of SymAE.

III. DATA POINTS AND NOTATION

In this section, we describe the training set $\{X_i\}_{i=1,...,n_X}$ that SymAE encodes to produce a compressed and disentangled representation. We reiterate that we presume the scale separation illustrated in Fig. 1 in our dataset. As such, each data point X_i contains multiple instances that repeatedly capture the same physical state ϵ_i, but each instance may differ on account of nuisance variations. We uniformly sample from 1 to n_ϵ to generate the state labels $\{\epsilon_i\}_{i=1,...,n_X}$ for our synthetic experiments; in practice, the experimental conditions determine this sampling distribution. We emphasize that knowledge of the state labels is *not* necessary for either training or testing since our framework is purely unsupervised. We index the instances in data point as $X_i[\tau]$ for $\tau = 1, \ldots, n_\tau$ such that $X_i = [X_i[1]; \ldots; X_i[n_\tau]]$. Each instance $X_i[\tau]$ is represented as k-dimensional vectors, and the determination of k is specific to each experiment. For the seismic experiment depicted in Fig. 2(a), each instance is a source gather, where the dimension k is the product of the number of receivers and the length of the time series. Each X_i comprises several sources that illuminate the same subsurface region. In our notation, $[A; B]$ denotes a vertical concatenation of two vectors A and B. Again, the collection of instances $\{X_i[\tau]\}_{\tau=1,...,n_\tau}$ for a fixed index i shares the same coherent information to the state ϵ_i but varies by τ-specific nuisance variations.

IV. ARCHITECTURE

We refer the reader to [33] for an accessible tutorial on autoencoders [16]. Functionally, autoencoders are comprised of two components: an encoder Enc that maps each data point X_i into latent code $H_i = \text{Enc}(X_i)$, and a decoder Dec that attempts to reconstruct to X_i from the code. Traditionally,
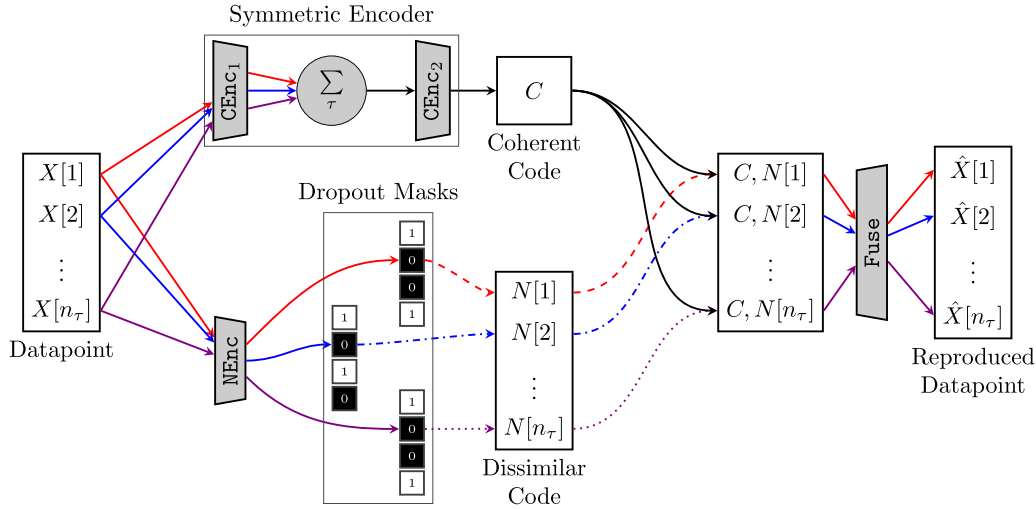
FIG. 3. Architecture of symmetric autoencoder [37]. The information that is coherent across the instances of a data point can only propagate through the network via solid arrows—notice the stochastic regularization employed to prevent its propagation. We used colored arrows to indicate the propagation of the remaining instance-specific nuisance information—notice that a symmetric function, i.e., symmetric with respect to the order of the instances, prevents its propagation. As a result, the autoencoder disentangles the coherent information from the nuisance variations in the latent space. We omitted the subscript i for X, C, and N.

both functions Enc and Dec are determined by minimizing the reconstruction loss

$$\text{Enc}, \text{Dec} = \underset{\text{Enc}, \text{Dec}}{\arg\min} \sum_i \|X_i - \text{Dec}[\text{Enc}(X_i)]\|^2 \quad (1)$$

over the training dataset. When nonlinear parametrizations are used for both Enc and Dec, the latent representation no longer describes the geometry of the datasets using linear subspaces [35]. However, this representation can efficiently compress the information [36].

SymAE builds on nonlinear autoencoders but requires additional modifications as a direct application of traditional autoencoding ideas will not ensure that the coherent and nuisance information are encoded into separate components (dimensions) in the latent space. To achieve this separation, SymAE relies on the unique encoder structure as depicted in Fig. 3 [21]. The encoder structure can be mathematically described by

$$\text{Enc}(X_i) = [\text{CEnc}(X_i); \text{NEnc}(X_i[1]); \dots; \text{NEnc}(X_i[n_\tau])]. \quad (2)$$

This output corresponds to a latent code H_i which is partitioned into interpretable components. Specifically, each data point $X_i = [X_i[1]; \dots; X_i[n_\tau]]$ is represented as a structured latent code $H_i = [C_i; N_i[1]; \dots; N_i[n_\tau]]$ in which the subcomponents $C_i = \text{CEnc}(X_i)$ contain coherent information in X_i while the remaining subcomponents $N_i[\tau] = \text{NEnc}(X_i[\tau])$ encode the complementary instance-specific nuisance information. Note the dimensions l and m of the latent codes $C_i \in \mathbb{R}^l$ and $N_i[\cdot] \in \mathbb{R}^m$ are user-specified hyperparameters which need not coincide.

Subsequently, SymAE's decoder Fuse nonlinearly combines code C_i with each instance-specific code $N_i[\cdot]$ to reconstruct the original data point, instance by instance, viz.,

$$\hat{X}_i = \text{Dec}(H_i) = \text{Dec}([C_i; N_i[1]; \dots; N_i[n_\tau]])$$
$$= [\text{Fuse}([C_i; N_i[1]]); \dots; \text{Fuse}([C_i; N_i[n_\tau]])]. \quad (3)$$

We do not enforce any constraints on Fuse in our experiments and parametrize it with standard deep-learning building blocks [34].

We ensure that CEnc, the *coherent encoder*, encodes at most the coherency or similarity among the instances in X_i by enforcing invariance with respect to permutations of the instances within the data point. Mathematically, this condition requires that

$$C_i = \text{CEnc}(X_i) = \text{CEnc}(X_i[\Pi(1{:}n_\tau)]) \quad (4)$$

for all permutations Π along the instance dimension of the output. This symmetry invokes the dichotomy of scale assumed in the data; since only nuisance variations are assumed to vary along the "fast scale" τ and that any coherent changes are negligible, this permutation invariance ensures that nuisance information cannot be encoded using CEnc without significant loss of information. It follows from the pigeonhole principle that only coherent information can remain in C_i if a low autoencoding loss is achieved.

SymAE's coherent encoder explicitly achieves the invariance mentioned above using permutation-invariant network architectures following [39] which provide universal approximation guarantees for symmetric functions. These architectures use pooling functions such as the mean or the max across the instances to ensure permutation invariance. We refer to [40] for a review of alternative pooling functions, including attention-based pooling. In our experiments, the data due to each source instance are transformed using CEnc$_1$ and summed along the instance dimension. This output is then processed by CEnc$_2$ resulting in

$$C_i = \text{CEnc}_2\left(\frac{1}{n_\tau}\sum_{\tau=1}^{n_\tau}\text{CEnc}_1(X_i[\tau])\right), \quad (5)$$

yielding the network architecture of CEnc. Intuitively, CEnc$_1$ extracts coherent information from each of the instances while

the summation encourages them to be aligned. This information is further compressed using $\mathtt{CEnc}_2$. The functions $\mathtt{CEnc}_1$ and $\mathtt{CEnc}_2$ are parametrized by compositions of fully connected layers and convolutional layers. We emphasize that the key observation in Eq. (5) is that the summation of the *transformed instances* $\mathtt{CEnc}_1(X_i[\tau])$ is symmetric with respect to the ordering of instances. This ensures that the desired symmetry [Eq. (4)] is achieved.

In contrast, the purpose of $\mathtt{NEnc}$, the *nuisance encoder*, is to capture the nuisance information specific to each instance of a data point. Critically, we do not want the decoder $\mathtt{Fuse}$ to ignore the C_i component in favor of using purely $N_i[\cdot]$ information for reconstruction. We desire disentanglement of the latent codes. Whereas $\mathtt{CEnc}$ achieves this via symmetry, for the nuisance encoder this separation is encouraged through the use of *stochastic regularization*, viz.,

$$N_i[\tau] = \mathtt{NEnc}(X_i[\tau]) + \text{``strong noise''}. \qquad (6)$$

Intuitively, this idea hinges on the assumption that coherent information does not vary with the "fast scale" τ indexing each instance. As such, obfuscating each element N_i via noise introduces artificial dissimilarities along this scale; this, therefore, encourages the decoder to instead rely on the coherent code (held constant for each instance [cf. Eq. (3)] to reconstruct the coherent information. Similarly, as before, it follows from the pigeonhole principle that the nuisance codes N_i must contain at most information relevant to nuisance information if a low autoencoding loss is achieved.

In our experiments, we implement this noise using either Bernoulli dropout regularization [41] with probability p or Gaussian dropout with unit mean and $p(1 - p)$ variance [42,43]. In either case, the strength of the noise is proportional to p, which is a hyperparameter the user must tune [38]. Critically, however, each N_i must still be expressive enough to encode nuisance-specific information. The balance between regularization strength p and the dimension (i.e., expressivity) of the latent codes is user determined on an external validation set. The SymAE components $\mathtt{NEnc}$, $\mathtt{CEnc}$, and $\mathtt{Fuse}$ are trained concurrently by minimizing Eq. (1) with the regularization mechanism just described. At test time, the entirety of the N_i code is sent unaltered and unobfuscated into the decoder. We emphasize that the stochastic regularization for $\mathtt{NEnc}$ is not employed to reduce overfitting and improve generalization error in the conventional sense (see, e.g., [44] for a survey on stochastic techniques used in neural network training). Instead, the intention of this regularization is to promote learning dissimilar representations across nuisance codes. Our numerical results did not observe the need for an additional regularization mechanism to prevent overfitting. The low dimensionality and symmetry of the coherent latent code is in itself a form of regularization against overfitting. However, if necessary, stochastic regularization can still be employed for $\mathtt{CEnc}$, provided the symmetry constraint remains intact during training.

Finally, note that we only constrained the encoders to avoid "cross talk" while disentangling the coherent and nuisance information. Implicitly the success of SymAE, therefore, requires a sufficiently large number of instances with *dissimilar* nuisance variations in order to achieve the desired structure of the latent space. We leave an examination of characterizations of physical models which are amenable to disentanglement to future work.

V. REDATUMING INTO VIRTUAL INSTANCES

A trained SymAE learns a representation with disentangled coherent and nuisance information. Redatuming data becomes equivalent to manipulations in the latent space, as illustrated in Fig. 2, where virtual instances are generated by swapping latent coordinates. In general, the coherent information in the τth instance of a data point X_i can be swapped with that of another data point X_j using

$$\hat{X}_{i \to j}[\tau] = \mathtt{Fuse}([\mathtt{CEnc}(X_j); \mathtt{NEnc}(X_i[\tau])]). \qquad (7)$$

Here, X_j is an observation of a different state compared to X_i. Notice that the nuisance information in the virtual data point $\hat{X}_{i \to j}$ is identical to that of the original data point X_i. Consequently, we attribute the difference between X_i and $\hat{X}_{i \to j}$ to the changes between the physical states. As a demonstration, the observed and virtual instances from the seismic experiment are embedded into the SymAE's latent space in Fig. 2.

VI. EXPERIMENTS

We now detail the application of SymAE towards experiments that monitor subsurface changes using seismic waves. As noted earlier, the measurements vary on two different (time) scales. (1) The slower timescale is associated with the subsurface changes that typically occur in the order of months. As such, the goal is to detect or determine variations in the coherent (subsurface) information between seismic surveys (e.g., baseline and monitor). (2) The faster time is usually on the order of the duration of the seismic survey, i.e., either hours or days; the variation in the coherent information is negligible on this scale. During each survey, waves from numerous uncontrollable sources, here taken to be the nuisance information, are recorded as instances. For our synthetic experiments, an instance is modeled as the pressure wave field $u(\mathbf{x}, t)$ from a finite-difference solver with absorbing boundary conditions for the acoustic wave equation:

$$\frac{1}{c^2(\mathbf{x})} \frac{\partial^2 u}{\partial t^2} - \nabla \cdot (\nabla u) = \delta(\mathbf{x} - \mathbf{x}_s) w(t). \qquad (8)$$

Here, $\mathbf{x} = [x, z]$ denotes the Cartesian coordinate vector and t denotes time. The medium is parametrized using the wave velocity $c(\mathbf{x})$. During the forward modeling, we vary c at a slower rate compared to source parameters, i.e., position $\mathbf{x}_s$ and signature $w(t)$ that determine the nuisance variation in each modeled instance.

We justify that SymAE captures the salient features of the physics of wave propagation using a simple illustration. Consider the wave velocity of a medium that varies in a 2 km × 2 km region, shown in Figs. 4(a)–4(c), across three states with $\epsilon_i \in \{1, 2, 3\}$ as described in Table II. Point sources at $\mathbf{x}_s = [R \cos(\theta), R \sin(\theta)]$ with signature $w(t)$ are used for modeling instances. The random variables R and θ are uniformly distributed on [0.8, 1.0] km and [0, 2π], respectively. The random source wavelet $w(t)$ has a duration of 0.13 s, is sampled from a standard normal distribution, and is convolved with a 25 Hz high-cut filter. After solving
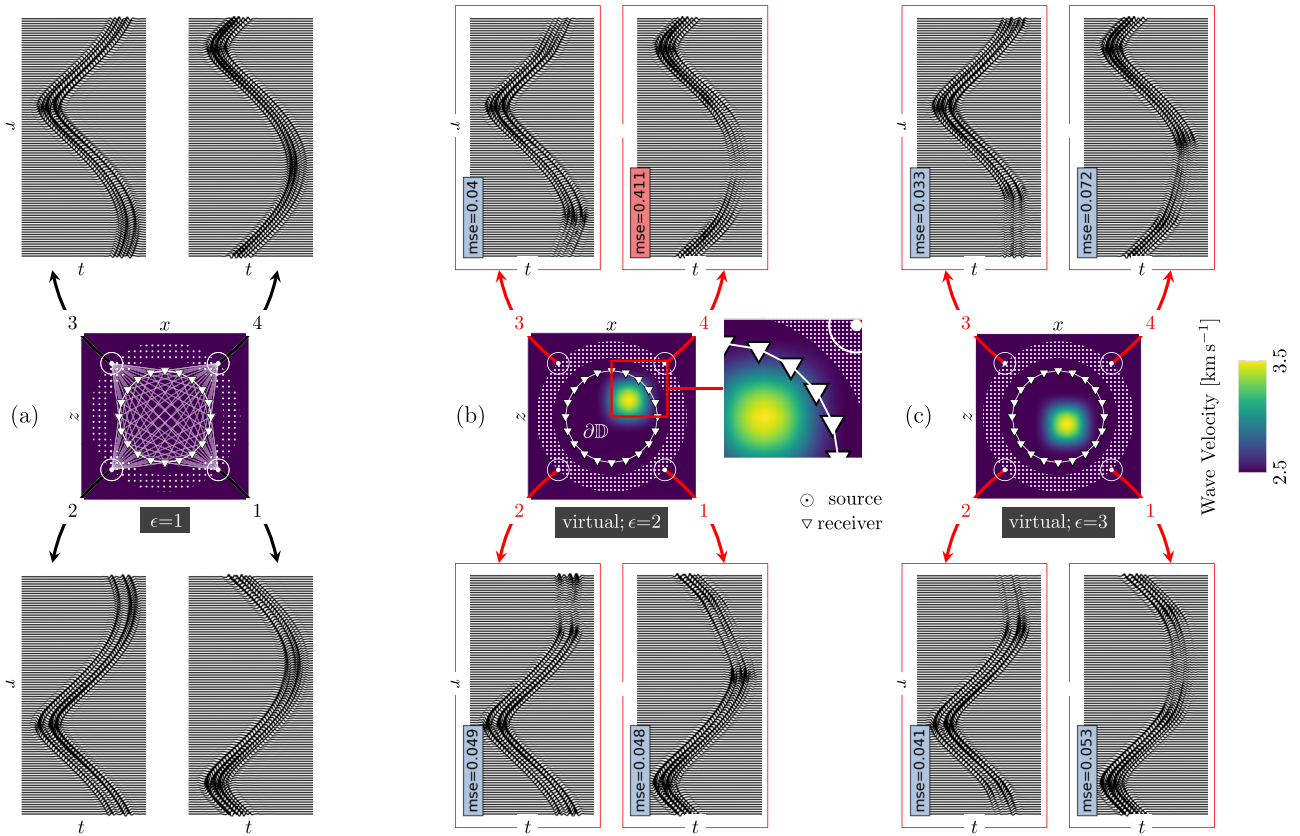
FIG. 4. Deep redatuming of waves recorded on $\partial\mathbb{D}$ due to point sources in the dotted region. This experiment illustrates that SymAE can isolate (coherent) information on the medium (Gaussian) perturbation in its representation; however, the perturbation has to lie within $\partial\mathbb{D}$. (a) $X_1[1:4]$; original instances from the first state with a homogeneous medium. (b) $\hat{X}_{1\rightarrow2}[1:4]$; virtual instances generated after swapping code C_1 of the first state with code C_2 of the second state. A high MSE in $\hat{X}_{1\rightarrow2}[4]$ indicates that C_2 fails to represent the effects of the Gaussian perturbation entirely. Note that this perturbation extends beyond $\partial\mathbb{D}$. (c) $\hat{X}_{1\rightarrow3}[1:4]$; same as (b), except a lower MSE, means that C_3 satisfactorily represents the medium when the perturbations lie within $\partial\mathbb{D}$.

Eq. (8), the acoustic wave field is sampled at 160 time steps and 100 evenly distributed receiver locations on a circle $\partial\mathbb{D}$ to form an instance of dimension $k = 16\,000$. We generated 8000 instances per state and considered a total of 6000 data points (with $n_\tau = 20$) for training and testing. It is important to note that the distribution of the forcing term in Eq. (8) is independent of the state ϵ to facilitate disentanglement.

We have invariably used the same medium parameters during the forward modeling of the instances in each state. Therefore, we hypothesize that (1) the medium parameters $c(\mathbf{x})$ characterize the coherent information represented by the

TABLE II. A time-lapse seismic experiment that justifies the SymAE's capture of the physics of wave propagation.

State ϵ	Medium perturbation[a]	MSE[b]	Virtual MSE[c]
1	None; homogeneous	< 0.01	
2	Not entirely inside $\partial\mathbb{D}$[d]	< 0.01	High (> 0.3)
3	Inside $\partial\mathbb{D}$[d]	< 0.01	Low (< 0.1)

[a]Gaussian perturbation.

[b]Normalized mean-squared error between the true X and reconstructed $\hat{X}$ data points for both training and testing.

[c]Between virtual and synthetic instances after redatuming.

[d]Receiver circle with center $[0, 0]$ and radius 600 m.

code C_i, i.e., CEnc encodes the information related to the entire medium in each state; and (2) the forcing term $w(t)$ and the source position $\mathbf{x}_s$ characterize the nuisance information, represented by N_i, of a given instance. We test these hypotheses numerically and show that CEnc does not encode the entire medium but only a portion that is coherently illuminated by all the sources. After redatuming, we compute relative mean squared error (MSE) between the virtual instances (generated by deep redatuming) and synthetic instances; a low MSE signifies that the virtual instance satisfies the governing wave equation in Eq. (8) with appropriate medium and source parameters. We now redatum four sources, $X_1[1:4]$ as in Fig. 4(a), picked from the first state ($\epsilon = 1$). First, we swapped C_1 of these measurements with C_2 to include the physics of wave propagation related to the Gaussian perturbation in $\epsilon = 2$. The virtual instances $\hat{X}_{1\rightarrow2}$ are plotted in Fig. 4(b); it can be observed that source information (position and signature) remained intact during redatuming, confirming that C_2 does not represent any of the source effects. Furthermore, notice that most of the virtual instances have low MSE, for example, $\hat{X}_{1\rightarrow2}[1:3]$, indicating that C_2 captured a significant portion of the Gaussian perturbation. However, the virtual instances with source locations close to that of $\hat{X}_{1\rightarrow2}[4]$, plotted in Fig. 4(b), have high MSE. What is unique about these sources? It is evident from the ray paths in Fig. 4(a) that these high-MSE
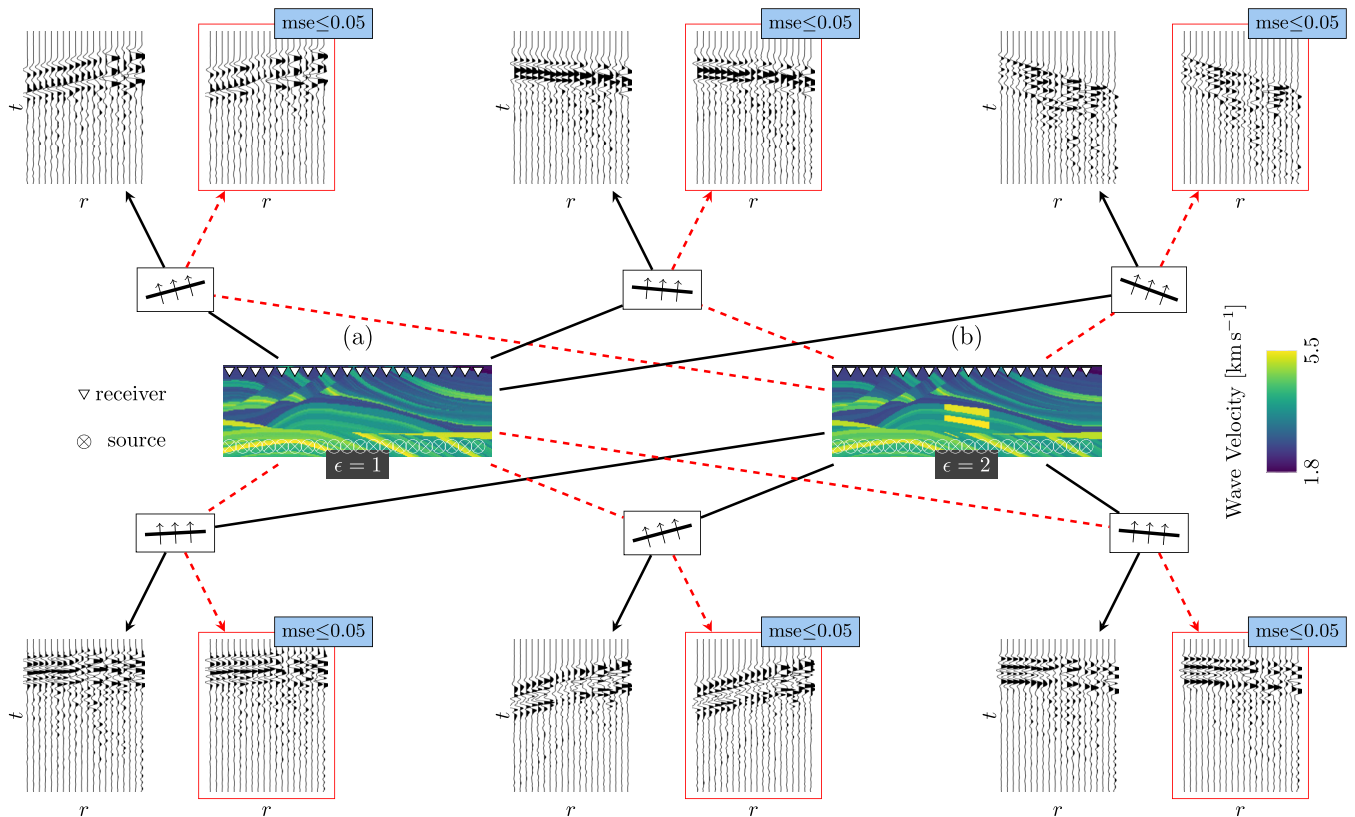
FIG. 5. Seismic Marmousi experiment of Fig. 2 where incident plane waves on the bottom of the medium undergo multipath propagation due to the complex inhomogeneities; three original instances (solid arrows) from each state are plotted and reveal the complex wave field. Each instance results from a plane wave with a unique source wavelet and angle of arrival. Virtual instances (dashed arrows) generated after swapping the coherent code from (a) to (b) and vice versa have a low error, showcasing the success of deep redatuming with SymAE.

source locations illuminate the portion of the Gaussian perturbation outside $\partial \mathbb{D}$. On the other hand, the region inside $\partial \mathbb{D}$ is coherently illuminated irrespective of the source position. We infer that SymAE's coherent code only represents the propagation effects of inhomogeneities inside $\partial \mathbb{D}$. In order to further confirm this inference, we then generated virtual instances corresponding to the state $\epsilon = 3$, where the Gaussian perturbation is entirely inside $\partial \mathbb{D}$ as depicted in Fig. 4(c). We notice that all the virtual instances have low MSE. Therefore, we conclude that SymAE learned to differentiate the coherently illuminated portion of the medium by the waves without the need for physics. This means, for seismic monitoring experiments, all the sources must coherently illuminate the time-lapse medium changes of interest.

The experiment in Fig. 2 involves seismic-wave propagation in a complex two-dimensional structural model, which is commonly known as the Marmousi model [45] in exploration seismology. The structural complexities will lead to multipath propagation. The *P*-wave velocity plots of this model for $\epsilon = 1$ and $\epsilon = 2$, with source-reciever geometry, are in Figs. 5(a) and 5(b), respectively. The forcing term represents a random plane-wave source input at the bottom of the model. The source wavelet $w(t; \tau)$ is generated by convolving a Ricker wavelet whose dominant frequency is sampled from $\{10, -12.5, \ldots, 20\}$ Hz, with a random time series (of 0.4 s duration) sampled from a standard normal distribution. In this case, the change in the medium parameters from $\epsilon = 1$ to

$\epsilon = 2$ is coherently illuminated by all the plane-wave sources, similar to the $\epsilon = 3$ Gaussian perturbation of the previous example. Therefore, the virtual measurements after redatuming are expected to have low MSE, as confirmed by the results in Fig. 5.

VII. CONCLUSIONS

We propose an autoencoder architecture for poorly controlled scientific experiments that produce an abundance of incompletely modeled measurements of a physical system. The autoencoder learns a data representation that disentangles the coherent information inherent to the physical state, from the nuisance modifications inherent to the experimental configuration, in a model-free fashion. Two ideas are critical: (1) leveraging symmetry under reordering of the data instance in order to represent the coherent information in a first encoder, and (2) stochastic regularization in order to prevent coherent information from being represented by the second encoder. As a result, the architecture can perform redatuming, i.e., the swapping of physics in order to create virtual measurements.

Following the double-blind review process, a Github repository containing reproducible code is available at [46].

---

[1] K. Aki and P. G. Richards, *Quantitative Seismology* (University Science Book, 2002).

[2] J.-P. Montagner, A. Mangeney, and E. Stutzmann, *Seismology and Environment*, edited by H. Gupta, Encyclopedia of Solid Earth Geophysics: Encyclopedia of Earth Sciences Series (Springer, Cham, 2020).

[3] E. Tandberg-Hanssen and A. G. Emslie, *The Physics of Solar Flares* (Cambridge University Press, Cambridge, UK, 1988), Vol. 14.

[4] S. Narendranath, P. Athiray, P. Sreekumar, B. Kellett, L. Alha, C. Howe, K. Joy, M. Grande, J. Huovelin, I. Crawford *et al.*, Lunar x-ray fluorescence observations by the Chandrayaan-1 x-ray spectrometer (C1XS): Results from the nearside southern highlands, Icarus **214**, 53 (2011).

[5] F. Locatello, S. Bauer, M. Lucic, G. Raetsch, S. Gelly, B. Schölkopf, and O. Bachem, Challenging common assumptions in the unsupervised learning of disentangled representations, in *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research Vol. 97, edited by K. Chaudhuri and R. Salakhutdinov (PMLR, 2019), pp. 4115–4124.

[6] G. Xu, H. Liu, L. Tong, and T. Kailath, A least-squares approach to blind channel identification, IEEE Trans. Signal Process. **43**, 2982 (1995).

[7] J. P. Verdon, J.-M. Kendall, D. J. White, D. A. Angus, Q. J. Fisher, and T. Urbancic, Passive seismic monitoring of carbon dioxide storage at Weyburn, Leading Edge **29**, 200 (2010).

[8] R. Kamei and D. Lumley, Passive seismic imaging and velocity inversion using full wavefield methods, SEG Technical Program Expanded Abstracts **2014**, 2273 (2014).

[9] K. Aki, Scaling law of earthquake source time-function, Geophys. J. Int. **31**, 3 (1972).

[10] P. M. Shearer, G. A. Prieto, and E. Hauksson, Comprehensive analysis of earthquake source spectra in Southern California, J. Geophys. Res. B **111**, B06303 (2006).

[11] P. Clark and I. Adler, Utilization of independent solar flux measurements to eliminate nongeochemical variation in x-ray fluorescence data, in *Lunar and Planetary Science Conference Proceedings* (Pergamon Press, New York, 1978), Vol. 3, pp. 3029–3036.

[12] C. Aerts, J. Christensen-Dalsgaard, and D. W. Kurtz, *Asteroseismology* (Springer Science & Business Media, New York, 2010).

[13] G. Handler, Asteroseismology, arXiv:1205.6407 (2012).

[14] N. N. Samus and O. V. Durlevich, VizieR online data catalog: Combined general catalogue of variable stars (Samus + 2004), VizieR Online Data Catalog, II (2004).

[15] W. W. Campbell and W. Wright, A list of nine stars whose velocities in the line of sight are variable, Astrophys. J. **12** (1900).

[16] Note that the structural complexities of the media result in multipath wave propagation, thereby complicating the inverse problem.

[17] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, CT-GAN: Malicious tampering of 3D medical imagery using deep learning, in *28th USENIX Security Symposium* (USENIX Security 19) (USENIX Association, Santa Clara, CA, 2019), pp. 461–478.

[18] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, Synthesizing Obama: Learning lip sync from audio, ACM Trans. Graph. **36**, 2017.

[19] C. Bregler, M. Covell, and M. Slaney, Video rewrite: Driving visual speech with audio, in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, *SIGGRAPH '97* (ACM Press/Addison-Wesley, New York, 1997), pp. 353–360.

[20] F. Locatello, B. Poole, G. Raetsch, B. Schölkopf, O. Bachem, and M. Tschannen, Weakly-supervised disentanglement without compromises, in *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research Vol. 119, edited by H. D. III and A. Singh (PMLR, 2020), pp. 6348–6359.

[21] In previous work, we used focusing constraints [47] to maximize this dissimilarity and regularize blind deconvolution.

[22] K. Wapenaar, Retrieving the Elastodynamic Green's Function of an Arbitrary Inhomogeneous Medium by Cross Correlation, Phys. Rev. Lett. **93**, 254301 (2004).

[23] G. T. Schuster and M. Zhou, A theoretical overview of model-based and correlation-based redatuming methods, Geophysics **71**, SI103 (2006).

[24] G. Schuster, in *Seismic Interferometry* (Cambridge University Press, Cambridge, 2009), Vol. 9780521871, pp. 1–260.

[25] W. A. Mulder, Rigorous redatuming, Geophys. J. Int. **161**, 401 (2005).

[26] K. Wapenaar, J. Thorbecke, J. Van Der Neut, F. Broggini, E. Slob, and R. Snieder, Marchenko imaging, Geophysics **79**, WA39 (2014).

[27] A. Mordret, N. M. Shapiro, and S. Singh, Seismic noise-based time-lapse monitoring of the Valhall overburden, Geophys. Res. Lett. **41**, 4945 (2014).

[28] S. A. De Ridder, B. L. Biondi, and R. G. Clapp, Time-lapse seismic noise correlation tomography at Valhall, Geophys. Res. Lett. **41**, 6116 (2014).

[29] J. van der Neut and K. Wapenaar, Adaptive overburden elimination with the multidimensional Marchenko equation, Geophysics **81**, T265 (2016).

[30] M. Mattheakis, P. Protopapas, D. Sondak, M. Di Giovanni, and E. Kaxiras, Physical symmetries embedded in neural networks, arXiv:1904.08991 (2019).

[31] T. S. Cohen, M. Weiler, B. Kicanaoglu, and M. Welling, Gauge equivariant convolutional networks and the icosahedral CNN, Proc. Mach. Learn. Res. **97**, 1321 (2019).

[32] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks, arXiv:1906.01563 (2019).

[33] C. Doersch, Tutorial on variational autoencoders, arXiv:1606.05908 (2016).

[34] We choose to use a deterministic autoencoding strategy for simplicity. It is possible to formalize the ideas in this paper using the variational autoencoding framework.

[35] J. Klys, J. Snell, and R. Zemel, Learning latent subspaces in variational autoencoders, arXiv:1812.06190.

[36] B. Dai and D. Wipf, Diagnosing and enhancing VAE models, arXiv:1903.05789 (2019).

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.023118 for Tensorflow-style [48] algorithms that detail the implementation of SymAE.

[38] We provided more details on hyperparameter tuning in the Supplemental Material.

[39] M. Zaheer, S. Kottur, S. Ravanbhakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola, Deep sets, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates, Red Hook, NY, 2017), pp. 3394–3404.

[40] M. Ilse, J. M. Tomczak, and M. Welling, Attention-based deep multiple instance learning, arXiv:1802.04712 (2018).

[41] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: a simple way to prevent neu-

ral networks from overfitting, J. Mach. Learn. Res. **15**, 1929 (2014).

[42] S. Wang and C. Manning, Fast dropout training, in *Proceedings of the 30th International Conference on Machine Learning*, edited by S. Dasgupta and D. McAllester (PMLR, Atlanta, Georgia, 2013), Vol. 28, pp. 118–126.

[43] D. P. Kingma, T. Salimans, and M. Welling, Variational dropout and the local reparameterization trick, Adv. Neural Inf. Process. Syst. **28**, 2575 (2015).

[44] A. Labach, H. Salehinejad, and S. Valaee, Survey of dropout methods for deep neural networks, arXiv:1904.13310.

[45] G. S. Martin, R. Wiley, and K. J. Marfurt, Marmousi2: An elastic upgrade for Marmousi, The Leading Eage **25**, 156 (2006).

[46] https://github.com/pawbz/SymAE.

[47] P. Bharadwaj, L. Demanet, and A. Fournier, Focused blind deconvolution, IEEE Trans. Signal Process. **67**, 3168 (2019).

[48] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, TensorFlow: large-scale machine learning on heterogeneous distributed systems, arXiv:1603.04467 (2016).