The Role of Untangled Latent Spaces in Unsupervised Learning Applied to Condition-Based Maintenance

Daniel N. Wilke(✉), P. Stephan Heyns, and Stephan Schmidt

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
nico.wilke@up.ac.za
https://www.up.ac.za/centre-for-asset-integrity-management

Abstract. Advances in signal processing are complemented by advances in machine and deep learning and vice versa. In general, machine and deep learning are employed as discriminative models within a supervised setting. Progress in unsupervised generative modelling allows for generative models to be employed in discriminative (discrete classes) and deviation (continuous deviation from a baseline) tasks. This only requires the samples to be chronological. Discriminative and deviation analysis is usually based on the reconstruction loss. However, this is limited as it offers only a single scalar from which inference can be made. Generative models do, however, learn a latent representation of the data from which additional scalars can be derived. Whether these derived scalars are informative depends on the quality of the latent representations. Most learning algorithms derive a latent representation that efficiently explains the variance in the data, which can be informative when the property of interest is well explained by variance. Alternatively, a lesser known class of learning algorithms aim to learn a latent representation that aims to identify sources in the data. Hence, given the same data, an infinite number of latent representations are possible, of which only a fraction are informative. We consider three classes of latent spaces that are stochastic, entangled and untangled. Furthermore, we highlight the importance of untangled latent spaces to obtain informative signals for condition monitoring.

Keywords: Signal processing · Generative modelling · Unsupervised learning · Latent variable · Untangled latent spaces

1 The Intersection Between Signal Processing and Learning Models

Signal processing in condition-based maintenance (CBM) is primarily concerned with extracting and analysing informative features from raw time-series data that enable diagnosis and prognosis to prevent asset failure and downtime, as shown in Fig. 1. Given sensor data, $\mathbf{x}(t)$, two questions are raised:

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 38–49, 2022.
https://doi.org/10.1007/978-3-030-85584-0_5

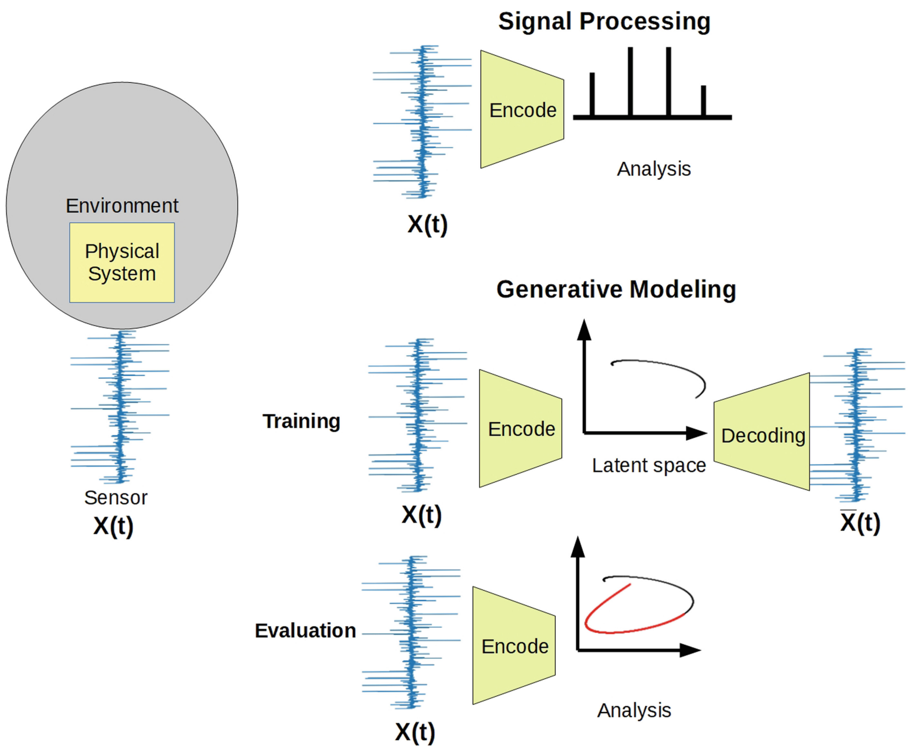


Fig. 1. Signal processing viewed as the encoding of useful or informative information from raw sensed signal data that the analyst can analyse. Signal processing is contrasted against the typical autoencoding training of generative models.

1. What is the information content of the signal, i.e. what of the physical system can be identified from the signal, and
2. Is the identifiable information in the signal captured such that the informativeness is maximised?

In signal processing, the analyst aims to isolate identifiable information from $\mathbf{x}(t)$ for analysis. The main aim is to extract useful information that informs an assets' condition. This task of separating identifiable information or extracting relevant information for the condition monitoring problem can be seen as encoding, followed by the encoded signal analysis.

The analyst's encoding task is equivalent to finding a transformation or representation that enhances specific information and attenuates extraneous signal components. This makes it possible to process the raw time-series signal into a more meaningful representation. Examples include the wavelet transform and the spectral coherence of the signal [1,2].

Alternatively, a time-series signal can be viewed as high dimensional data, i.e. as an n-dimensional vector, $\mathbf{x}(t) \in \mathbb{R}^n$, where n corresponds to the number of data points in the signal. The analyst is now tasked with finding a lower-dimensional representation or features that is/are informative. A single n-dimensional observation $\mathbf{x}(t)$ does not in itself enable the analyst to find a lower-dimensional representation. However, several strategies exist and have been employed to recast a single n-dimensional observation into multiple correlations/observations of high dimensional data. The auto-correlation matrix recasts $\mathbf{x}(t)$ into $\mathbf{X}_{AC}(t) \in \mathbb{R}^{n\times n}$ [3], while through a sliding window approach $\mathbf{X}(t) \in \mathbb{R}^{m\times l}$ recasts $\mathbf{x}(t)$ into m observations of window length l [4–6]. This approach is also used to obtain the Hankel matrix. Given some $\mathbf{X}(t)$, the aim is to find a lower-dimensional representation that enhances specific information and suppresses components unrelated to the component of interest.

In signal processing, irrespective of whether the data is being transformed or projected, the analyst is essentially conducting the following operations: finding informative transformations (e.g. [7]) and features or condition indicators (e.g. [8]), whereafter they are analysed for damage. This requires knowledge and mastering of signal processing principles, experience and extensive knowledge of the physical mechanisms that generate the measured signals.

In artificial intelligence, statistical learning [9], machine learning [10] and deep learning [11] are sub-domains that focus on developing generative models. Incidentally, generative modelling is the task of transforming high-dimensional data to a lower-dimensional representation or latent space through encoding, as shown in Fig. 1. Samples in this latent space can be resonstructed back to the high-dimensional space through decoding. This encoding-decoding process is used in training where the error between the original signal $\mathbf{x}(t)$ and the reconstructed signal $\bar{\mathbf{x}}(t)$ is minimised. Hence, the task of encoding is common between generative modelling and signal processing. An additional intersection being the analysis of the latent variables to inform on the condition of an asset.

Therefore, it is critical to find latent spaces that are informative and ensure that the latent space components correspond to distinct physical processes. In CBM under the effects of time-varying environmental and operational conditions (EOCs), this implies that time-varying components can be isolated and identified as independent latent representations [12,13]. The challenge in generative modelling is to encode informative latent spaces. Hence, by applying generative modelling with untangled latent spaces to CBM, the aim is to automate signal processing tasks that require extensive domain knowledge.

2 Process of Lower-Dimensional Representations

The process of finding lower-dimensional representations entails two steps, as shown in Fig. 2. Firstly, by finding an informative, low dimensional coordinate system, and secondly, by projecting the high dimensional data onto the low dimensional coordinate system. Depending on the generative modelling, the coordinate system can be linear (see Fig. 2), i.e. the coordinate system forms a basis that a set of vectors can describe. Some transformation function can describe a nonlinear or curvilinear coordinate system (see Fig. 2).

Statistical learning specifically aims to alleviate the analyst from finding features to merely supervising the task of finding features, a.k.a. feature engineering. In particular, statistical learning restricts itself to finding a linear coordinate system, i.e. linear basis to represent the latent space. Machine learning has similar aims, but the latent space is described by nonlinear coordinate systems to which the raw signals can be transformed. This enables the analyst to focus on analysing the projected signals instead of first finding informative features. Deep learning has more ambitious goals. It aims to automate the processing, feature engineering, and analysis of raw time-series signals, i.e., automate the entire signal processing value chain.

Hence the task of statistical, machine and deep learning is to find a coordinate system to represent the latent space onto which a time-series signal can be projected for analysis. This is achieved by defining an appropriate optimisation problem. Here, the analyst needs to decide a proper optimisation model, e.g. primal problem formulation or dual problem formulation, to solve a formulated min-max, maximisation or minimisation problem [14]. The choice of the loss function, norm (e.g. l_1-norm and l_2-norm), regulariser (e.g. l_2-norm weight norm penalty (ridge regression), or l_1-norm weight penalty (LASSO for sparsity)) and constraints (e.g. orthogonality of coordinates) allows the analyst to craft and target specific characteristics in the data that are assumed to be informative. This problem can be cast within a statistical language by assuming a distribution for the noise (e.g. Laplacian or Gaussian, which are related to the l_1-norm and l_2-norm, respectively) and prior distribution (e.g. Laplacian or Gaussian, which are associated with the l_1-norm weight norm penalty and l_2-norm weight penalty). The process of solving a formulated optimisation problem is known as training, see Fig. 1.

To summarise, data science, particularly generative modelling, requires the formulation of an optimisation that can be solved to find informative coordinate systems onto which high-dimensional data can be projected.

3 Supervised, Semi-supervised and Unsupervised Learning

The availability and type of data dictate which learning strategies are available for the analyst. In the context of anomaly or fault detection, supervised learning requires signals that were taken when a physical asset was known to be in a healthy and damaged state. This is referred to as labelled data. In turn, unsupervised learning only requires physical asset data without knowing whether it was taken while the physical asset was healthy or damaged. However, we distinguish between data where the samples are chronologically ordered for unsupervised learning, i.e. measured samples are time-stamped, or not, i.e. each sample is chronological but no time-stamps were recorded for the samples. Lastly, semi-supervised learning relates the learning with predominantly unsupervised data and access to few supervised samples. The most prevalent data in CBM is unsupervised data with chronological samples without prior knowledge of a machine's

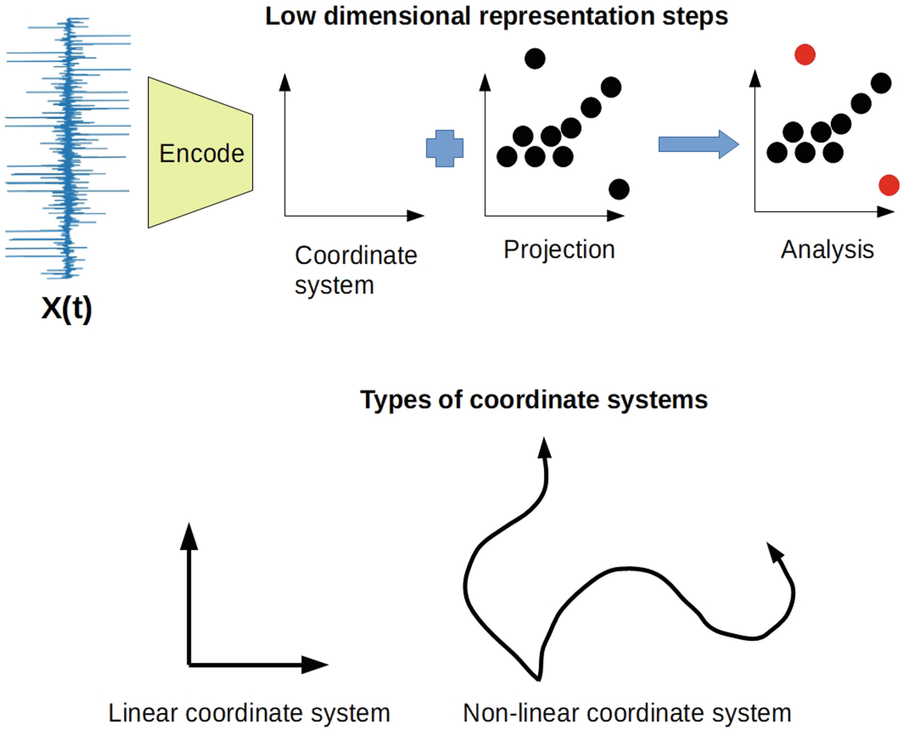


Fig. 2. Encoding a raw signal viewed as finding a coordinate system onto which the data is projected for analysis.

initial state. Data is selected to represent a reference state on which to train a model [13]. This reference state is often referred to as "healthy" when the initial data in the chronological data flow was used to train on [13]. Hence, unsupervised learning with chronologically sorted data is an ideal learning framework to train statistical, machine and deep learning models, which we will explore in more detail.

4 Unsupervised Generative Learning with Chronologically Sorted Data

Unsupervised generative learning can be achieved primarily by one of two learning principles, namely,

1. auto-associative learning [15], or
2. generative adversarial learning [16],

thereby enabling the training of generative models in an unsupervised fashion.

4.1 Auto-associative Learning

Auto-associative learning [15] or autoencoding [17] is pervasive in statistical, machine and deep learning and follows the process of encoding and decoding outlined in Fig. 1. In essence, autoencoding aims to reconstruct the input signal by encoding the signal in a lower-dimensional coordinate system. It reconstructs the signal back to the same dimensionality as the input signal through decoding. This lower-dimensional encoding is often referred to as the bottleneck, an essential mechanism to find an informative lower-dimensional signal. The aim is to learn to regenerate the samples $\mathbf{X}$ without memorizing the samples.

This is exemplified by considering m, l-dimensional samples arranged in matrix $\mathbf{X} \in \mathbb{R}^{m \times l}$. Given that we have centred the data $\mathbf{X}$, which we denote $\mathbf{C}$, we aim to find a coordinate system described by a $l \times k$ matrix $\mathbf{W}$. Here, $\mathbf{CW}$ encodes the data $\mathbf{C}$ into a k-dimensional coordinate system $\mathbf{W}$, whereas $(\mathbf{CWW}^{\mathrm{T}})$, decodes the projected data $\mathbf{CW}$ back to the original l-dimensional space.

Given $k = l$ with $\mathbf{W} = \mathbf{I}$ then encoding results in $\mathbf{CI} = \mathbf{C}$, and decoding gives $\mathbf{CI}^{\mathrm{T}} = \mathbf{C}$ perfect reconstruction. However, using $\mathbf{W} = \mathbf{I}$ is equivalent to memorising the data instead of learning useful information about the data, again highlighting the importance of a lower-dimensional latent space $k \ll l$.

Learning algorithms can be constructed to optimally reconstruct signals. For this case, the aim is to maximise the variance explained in a dataset by finding an optimal lower-dimensional coordinate system to reconstruct signals. Strategies that focus on reconstruction include principal component analysis (PCA), autoencoders (AE) and variational autoencoders (VAE) [15,18,19]. Alternatively, learning algorithms can be constructed to find an informative lower-dimensional coordinate system [20–24], or an informative latent space. Latent focussed approaches include independent component analysis (ICA) and beta-variational autoencoders (β-VAEs) [5,25].

Autoencoding with a Reconstruction Focus: Principal component analysis (PCA) [19] and singular value decomposition (SVD) [26] encapsulate the foundation of autoencoding that aims to maximise the variance explained or aimed at reconstructing the signal as efficiently as possible.

Given $\mathbf{C}$, we aim to find a new orthogonal coordinate system described by a $l \times k$ matrix $\mathbf{W}$. With $k \ll l$ to describe the data, we can achieve it by minimising the following constrained optimisation problem:

$$\mathbf{W}^{*} = \arg \min_{\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}} \left((\mathbf{CW})\mathbf{W}^{\mathrm{T}} - \mathbf{C}\right) : \left((\mathbf{CW})\mathbf{W}^{\mathrm{T}} - \mathbf{C}\right). \tag{1}$$

$\mathbf{CW}$ encodes the data $\mathbf{C}$ into a k-dimensional coordinate system $\mathbf{W}$, whereas $\mathbf{W}^{\mathrm{T}}(\mathbf{CW})\mathbf{W}^{\mathrm{T}}$ decodes the projected data $\mathbf{CW}$ back to the original l-dimensional space. Since $\mathbf{C} \in \mathbb{R}^{m \times l}$, we require a double contraction, denoted: to reduce the expression to a scalar. Autoencoding (AE) [15], Variational Autoencoding (VAE) [27] and Singular Spectrum Analysis (SSA) [28] aim to find a coordinate system that is efficient at explaining the variance in the signal, which in some cases can be informative if the aspect of interest manifests as the variance. It is untangled if it is the sole aspect to do so, however, this is seldom the case.

Autoencoding with a Latent Space Focus: Independent component analysis (ICA) [20,29,30] complements autoencoding that aims to have an informative latent space or aims to find an informative decomposition. ICA terminology often refers to the components of such latent spaces as the sources [31]. The application of ICA in prognostic maintenance of renewable energy systems is well known, but the connection to untangling is not that well established [32].

Machine learning and deep learning offer improved abilities to untangle latent spaces over linear models [33], but remain largely unexplored for CBM applications. Most time-series studies are focussed on human speech and animal acoustics [34].

Given $\mathbf{C}$, we aim to find an informative new coordinate system $l \times k$ matrix $\mathbf{W}$, such that the projected data

$$\mathbf{S} \approx \mathbf{C}\mathbf{W} \tag{2}$$

is maximally statistically independent and non-Gaussian, according to some measure of non-Gaussianity. The measures of non-Gaussianity define the characteristics of the expected sources, such as kurtosis and negentropy [20,30,35,36]. Damaged signals often manifest as non-Gaussian sources in the measured data [21,24,37], making ICA useful for fault diagnosis. This is also aligned with the developments of the signal processing community, where different measures of non-Gaussianity are used to identify damaged machine components [21,24,37,38].

Note, the reconstruction of the given data is given by

$$\mathbf{S}\mathbf{W}^{-1} \approx \mathbf{C}, \tag{3}$$

but was obtained by finding statistically independent latent variables, which implies that $\mathbf{S}$ is uncorrelated

$$\mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{I}. \tag{4}$$

Hence, ICA aims to find a statistically independent latent space during training in contrast to the reconstruction-based focused models. Each latent component represents an isolated physical process or source contributing to the measured signal. This directly implies a latent space that is untangled and informative as the measured signal is decomposed into independent components that can be readily analysed and interpreted.

In machine learning and deep learning, an extension of variational autoencoders (VAE) towards an untangled latent space is β–VAE [25].

4.2 Generative Adversarial Learning

The premise of adversarial training is to transform the unsupervised generative modelling problem into a supervised classification problem. This is achieved by introducing two sub-models in the adversarial learning framework: a generator and a discriminator model. The generator model generates new samples *a priori* from a chosen latent space, while the discriminator classifies between actual

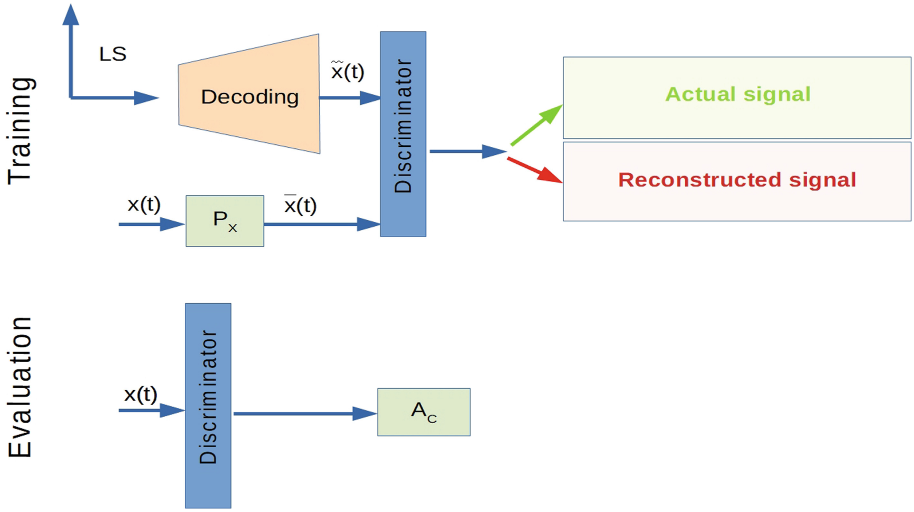


Fig. 3. Adversarial training turns an unsupervised learning problem into supervised learning by introducing a discriminator between actual $\mathbf{x}(t)$ and the reconstructed $\tilde{\mathbf{x}}(t)$ signals. Training constructs a decoding network and discriminator network. Evaluation then uses the discriminator to evaluate whether a signal $\mathbf{x}(t)$ is from the training set data or not by analysing, A_c, the discriminator signal.

signals and generated signals. In an adversarial framework, the two models are trained together.

In a basic adversarial framework, the latent space is chosen to be stochastic without any structure enforced. The implication is that the latent space is uninformative. However, the adversarial framework makes an additional measure available in the form of the discriminator.

A generative adversarial network (GAN) discriminator can be informative in condition monitoring applications [13]. However, an informative latent space would supplement adversarial learning approaches. Ref. [12] extended on the work of Ref. [13] with latent space conditioning semi-supervised learning for CBM applications. Improvements to construct informative latent spaces for GANs include adversarial latent autoencoders [39].

5 Latent Representations

Recall, the latent representations given the same data are not all equivalent. Let us consider concrete manifestations of latent representations using a foundational example. Since faults or environmental and operating conditions can manifest in several variations from a nominal signal, e.g. variations in amplitude, frequency, phase, and offset, we consider a simple sine wave signal with amplitude magnitude variation (between 1 (black) and 5 (white)) as shown in Fig. 4. A low-frequency signal over 10 s is purposefully constructed for clarity.

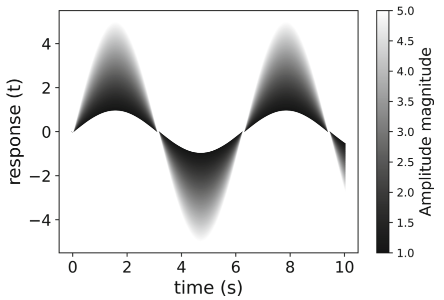
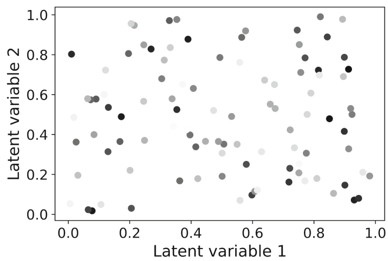
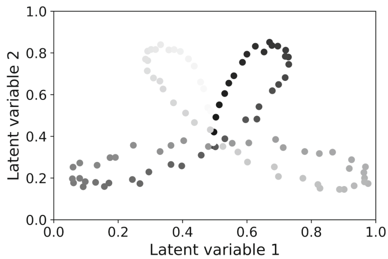


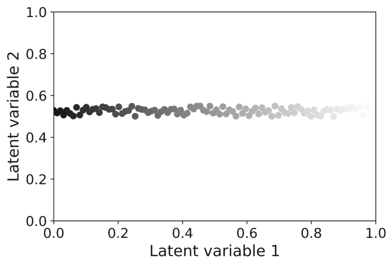
Fig. 4. Foundational example of amplitude magnitude change in a sine wave signal from 1 (black) to 5 (white).



(a)



(b)



(c)

Fig. 5. Latent representations for the amplitude variation from 1 (black) to 5 (white) of the sine wave given as (a) stochastic, (b) entangled and (c) untangled.

Let us focus on finding latent representations highlighting the variation in amplitude magnitude (given by the colour variation of the data points from black (amplitude magnitude of 1) to white (amplitude magnitude of 5)) for the sine wave. We depict a two-dimensional latent space representation for a stochastic latent space in Fig. 5(a), entangled latent space in Fig. 5(b) and untangled latent space in Fig. 5(c).

Given, we obtain a stochastic latent space signal, as the amplitude varies in the sine wave. It is clear that the stochastic latent space signal does not inform the amplitude variation, i.e. the latent space is uninformative. Suppose we would instead obtain an entangled latent space signal, as the amplitude varies in the sine wave. The entangled latent space signal does contain some information of the amplitude variation. In this case, some analysis may be required to interpret the amplitude variation from this entangled latent signal. Lastly, the untangled latent space signal informs the amplitude variation independent of any other variation in the latent signal, resulting in an informative latent signal requiring no additional or minimalistic processing for analysis and interpretation.

Given a condition monitoring problem, where damage is variance or amplitude driven (e.g. modulation due to bearing impacts), we may find similar performance between reconstruction focussed (e.g. SSA, AE or VAE) and latent focussed autoencoding (e.g. ICA or β–VAE) as the variance is a good independence measure for damage in this example. Here, we will obtain an entangled or untangled latent space depending on the additional variations in the signal. However, should damage manifest weakly in the signal's variance, we may find that untangled latent spaces are restricted to latent focussed autoencoding. An untangled latent space is also critical in CBM under the effects of time-varying environmental and operational conditions (EOCs). It allows for informative time-varying components to be identified and isolated [12,13], or uninformative components to be identified and suppressed.

6 Conclusions

This study explored the role of untangled latent spaces obtained with semi-supervised or unsupervised learning. An infinite number of latent spaces exist, of which only a fraction are informative. Significant effort is required to obtain an untangled and informative latent space. However, the additional latent signals available for analysis given an untangled and informative latent space make this endeavour all worthwhile. In some cases, the variance may be a good proxy for statistical independence, given that the fault of interest manifests in the variance of a sensed signal. In these cases, a reconstruction-focused learning strategy may result in a partially untangled latent space that is informative. However, should the fault of interest not manifest in the signal variance or under time-varying operating conditions, then a latent focussed learning strategy is imperative to obtain an untangled and informative latent space.

References

1. Kruczek, P., Zimroz, R., Antoni, J., Wyłomańska, A.: Generalized spectral coherence for cyclostationary signals with α-stable distribution. Mech. Syst. Sig. Process. **159**, 107737 (2021)
2. Al-Badour, F., Sunar, M., Cheded, L.: Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques. Mech. Syst. Sig. Process. **25**(6), 2083–2101 (2011)
3. Cadzow, J.A., Baseghi, B., Hsu, T.: Singular-value decomposition approach to time series modelling. IEE Proc. F Commun. Radar Sig. Process. **130**(3), 202–210 (1983)
4. Bozzo, E., Carniel, R., Fasino, D.: Relationship between Singular Spectrum Analysis and Fourier analysis: theory and application to the monitoring of volcanic activity. Comput. Math. Appl. **60**(3), 812–820 (2010)
5. He, Q., Feng, Z., Kong, F.: Detection of signal transients using independent component analysis and its application in gearbox condition monitoring. Mech. Syst. Sig. Process. **21**(5), 2056–2071 (2007)
6. Bekiroglu, K., Tekeoglu, A., Andriamanalimanana, B., Sengupta, S., Chiang, C.-F., Novillo, J.: Hankel-based unsupervised anomaly detection. In: 2020 American Control Conference (2020)
7. Buzzoni, M., Antoni, J., d'Elia, G.: Blind deconvolution based on cyclostationarity maximization and its application to fault identification. J. Sound Vibr. **432**, 569–601 (2018)
8. Antoni, J., Borghesani, P.: A statistical methodology for the design of condition indicators. Mech. Syst. Sig. Process. **114**, 290–327 (2019)
9. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics, Springer, New York (2001)
10. Aggarwal, C.C.: Outlier Analysis. Springer, New York (2013)
11. Goodfellow, I.J., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
12. Baggeröhr, S., Booyse, W., Heyns, P., Wilke, D.: Novel bearing fault detection using generative adversarial networks. In: Condition Monitoring and Diagnostic Engineering Management (2018)
13. Booyse, W., Wilke, D.N., Heyns, S.: Deep digital twins for detection, diagnostics and prognostics. Mech. Syst. Sig. Process. **140**, 106612 (2020)
14. Snyman, J.A., Wilke, D.N.: Practical Mathematical Optimization. Springer, Heidelberg (2018)
15. Kramer, M.A.: Nonlinear principal component analysis using autoassociative neural networks. AIChE J. **37**(2), 233–243 (1991)
16. Goodfellow, I.J., et al.: Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS 2014, vol. 2, pp. 2672–2680, MIT Press, Cambridge (2014)
17. Makhzani, A., Frey, B.: k-sparse autoencoders. In: International Conference on Learning Representations (ICLR) (2014)
18. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3/4), 321–377 (1936)
19. Golyandina, N.: Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing (2020)
20. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Netw. **13**(4–5), 411–430 (2000)

21. Hou, S., Wentzell, P.D.: Fast and simple methods for the optimization of kurtosis used as a projection pursuit index. Anal. Chim. Acta **704**(1–2), 1–15 (2011)
22. He, Q., Du, R., Kong, F.: Phase space feature based on independent component analysis for machine health diagnosis. J. Vibr. Acoust. Trans. ASME **134**(2), 1–11 (2012)
23. Debals, O., Lathauwer, L.D.: Stochastic and deterministic tensorization for blind signal separation. In: Lecture Notes in Computer Science, vol. 9237, no. 1, pp. 3–13 (2015)
24. Qian, Y., Yan, R.: Gearbox fault diagnosis in a wind turbine using single sensor based blind source separation. J. Sens. **2016**, 1–14 (2016)
25. Burgess, C.P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., Lerchner, A.: Understanding disentangling in β-VAE. arXiv (2018)
26. Jiang, Y., Tang, B., Qin, Y., Liu, W.: Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD. Renew. Energy **36**(8), 2146–2153 (2011)
27. Kingma, D.P., Welling, M.: An introduction to variational autoencoders. Found. Trends® Mach. Learn. **12**(4), 307–392 (2019)
28. Sulandari, W., Subanar, Lee, M.H., Rodrigues, P.C.: Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. Energy **190**, 116408 (2020)
29. Herault, J., Jutten, C., Ans, B.: Detection de grandeurs primitives dans un message composite par une architeture de calcul neuromimetique en apprentissage non supervise. In: Colloque sur le traitement du signal et des images, pp. 1017–1022 (1985)
30. Herault, J., Jutten, C.: Space or time adaptive signal processing by neural network models. In: AIP Conference Proceedings, vol. 151, no. 1, pp. 206–211 (1986)
31. Li, Z., Yan, X., Wang, X., Peng, Z.: Detection of gear cracks in a complex gearbox of wind turbines using supervised bounded component analysis of vibration signals collected from multi-channel sensors. J. Sound Vibr. **371**, 406–433 (2016)
32. Afridi, Y.S., Ahmad, K., Hassan, L.: Artificial intelligence based prognostic maintenance of renewable energy systems: a review of techniques, challenges, and future research directions (2021)
33. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
34. Sainburg, T., Thielk, M., Gentner, T.Q.: Latent space visualization, characterization, and generation of diverse vocal communication signals. bioRxiv (2019)
35. Hyvrinen, A., Oja, E.: Independent component analysis: a tutorial. Technical report, TUH (1999)
36. Hyvärinen, A.: Independent component analysis: recent advances. Philos. Trans. Roy. Soc. A Mathe. Phys. Eng. Sci. **371**(1984), 20110534 (2013)
37. Cheng, W., Jia, Z., Chen, X., Gao, L.: Convolutive blind source separation in frequency domain with kurtosis maximization by modified conjugate gradient. Mech. Syst. Sig. Process. **134**, 106331 (2019)
38. Peña, D., Prieto, F.J., Viladomat, J.: Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure. J. Multivar. Anal. **101**(9), 1995–2007 (2010)
39. Pidhorskyi, S., Adjeroh, D.A., Doretto, G.: Adversarial latent autoencoders. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) (2020)