ISBN: 978-3-95450-241-7 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2022-MOP10

REMOVING NOISE IN BPM MEASUREMENTS WITH VARIATIONAL AUTOENCODERS

J. P. Edelen, M. J. Henderson, J. Einstein-Curtis, C. C. Hall, RadiaSoft LLC, CO, USA
A. L. Romanov, Fermilab, IL, USA

Abstract

Noise in beam measurements is an ever-present challenge in accelerator operations. In addition to the challenges presented by hardware and signal processing, new operational regimes, such as ultra-short bunches, create additional difficulties in routine beam measurements. Techniques in machine learning have been successfully applied in other domains to overcome challenges inherent in noisy data. Variational autoencoders (VAEs) are shown to be capable of removing significant leevels of noise. A VAE can be used as a pre-processing tool for noise removal before the de-noised data is analyzed via other methods, or the VAE can be directly used to make beam dynamics measurements. Here we present the use of VAEs as a tool for addressing noise in BPM measurements.

INTRODUCTION

In recent years machine learning (ML) has been identified as having the potential for significant impact on the modeling, operation, and control of particle accelerators (e.g. see [1, 2]). Specifically, in the diagnostics space, there have been many efforts focused on improving measurement capabilities and detecting faulty instruments. Relatively recently, ML methods have been utilized to improve optics measurements from beam position monitor data [3]. Additionally, machine learning has been used to identify and remove malfunctioning beam position monitors in the Large Hadron Collider (LHC), prior to application of standard optics correction algorithms [4]. However, noise in BPM measurements remains an issue for processing in even functioning BPMs. The ability to remove noise from these measurements would greatly improve our ability to extract meaningful information from these instruments.

Variational autoencoders (VAEs) are a well established tool for noise reduction due to the enforcement of a smoothness condition in the latent-space representation. This feature of VAEs has been applied to gravitational wave research [5, 6] and geophysical data [7], for example. Recurrent autoencoders have the added advantage of being well suited to work with data sequences. In this paper we explore the use of Variational Recurrent Autoencoders (VRAEs) to remove different power law spectra (colors) of noise from simulated BPM data in a ring. We begin with a review of our data generation model, we then analyze the noise reduction capabilities for Gaussian noise. Finally we test our method using additive noise with different power law spectra.

DATA GENERATION

To simplify exploration of this noise removal technique in this work, we only consider data generated from a simplified model of a circular (periodic) accelerator with analytic solutions [8] . Rather than composing the accelerator of discrete focusing magnets we consider a uniform focusing channel with coupled optics. This reduces the problem to that of a coupled oscillator. We are only considering motion in the transverse plane, such that the equations of motion will be:

$$\begin{aligned} \frac{d^2x}{d\theta} + \nu_x x + Cy = 0 \\ \frac{d^2y}{d\theta} + \nu_y y + Cx = 0. \end{aligned} \tag{1}$$

Where C is the coupling strength and $\theta = 2\pi f t$ is the fractional revolution period in radians. The solutions to the coupled equations of motion will then be:

$$\begin{aligned} x(\theta) = A_x \cos(\nu_+\theta) + B_x \cos(\nu_-\theta) \\ y(\theta) = A_y \cos(\nu_+\theta) + B_y \cos(\nu_-\theta), \end{aligned} \tag{2}$$

with the coupled oscillation frequency given by:

$$\nu_\pm^2 = \frac{1}{2}\left(\nu_x^2 + \nu_y^2 \pm \sqrt{\left(\nu_x^2 + \nu_y^2\right)^2 + 4C^2}\right). \tag{3}$$

The principal goal of the analysis tools developed in this paper will be to extract correct, independent frequencies from noisy, periodic measurements of x and y. These are referred to as the tunes, ν_x and ν_y, in Eq. (1). The amplitude coefficients may be uniquely determined from the initial x and y positions, but are not included here as we do not use them in analyzing performance of the methods presented. Out dataset was generated with tune values ranging from 0.05 to 1.6 and a coupling parameter of 0.1. Figure 1 shows an example of this generated data.

To create test and training data x and y, position data is sampled as if from M BPMs placed around a ring so that BPM m will have a phase offset of $\varphi = 2\pi\nu m/M$. Noise in the measurements at each turn N is sampled from a normal distribution $\mathcal{N}(0, \sigma_M)$, where the variance σ_M, representing the noisiness of each BPM, has been set by sampling from a normal distribution $\mathcal{N}(0, \sigma_{noise})$.

The Variational Recurrent Autoencoder Model

For the VAE model, since we are analyzing time series data, we use a recurrent network architecture for the encoder and decoder, which we will refer to as a VRAE. The implementation of the VRAE architecture is based on [9] and uses Long Short-Term Memory (LSTM) units for both the

ISBN: 978-3-95450-241-7 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2022-MOP10

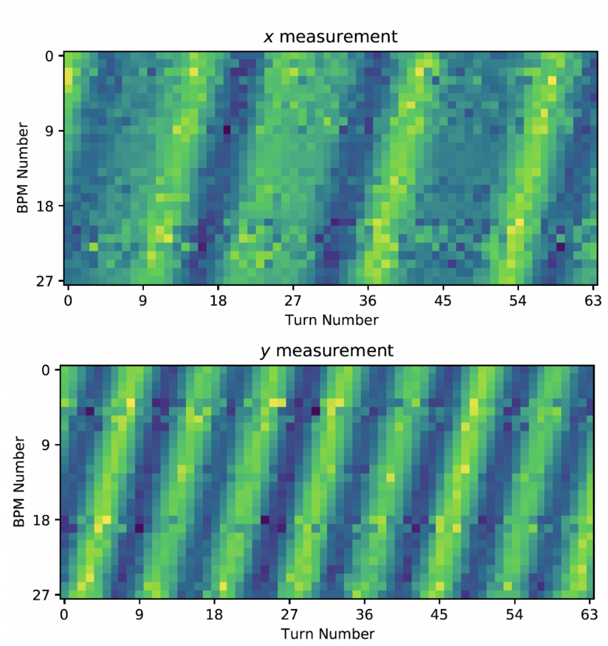

Figure 1: Example of data produced from discrete sampling of Eq. (2) to simulate BPM placement around a ring. Data was created with a Gaussian noise of $\sigma_{noise} = 0.25$ applied.

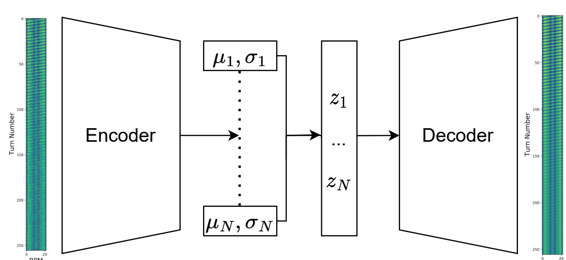

Figure 2: Structure of the VRAE showing encoder and decoder with latent layer parameterized in terms of mean μ_i and variance σ_i for the latent space distribution of dimension i.

encoder and decoder. A schematic of the VRAE is shown in Fig. 2. Both encoder and decoder contain 1 hidden layer of size 90. The latent space dimension was varied from 2 to 6. Linear layers are used to transform from the LSTM output to the distribution parameterization for sampling the latent space. The loss function is composed of two terms: the Kullback–Leibler divergence [10] — which acts as a regularization term — and the reconstruction loss. For the reconstruction loss, mean squared error between the encoder input and decoder output is used.

The VRAE was trained on generated data with a sequence length of 28 turns and with 56 features (28 BPM measurements of x and 28 measurements of y). Training data amplitude was initially normalized, however, this seemed to result in over-fitting and the model would show poor performance when applied to datasets with noise levels or periods that differed from the training set. Training with a dataset

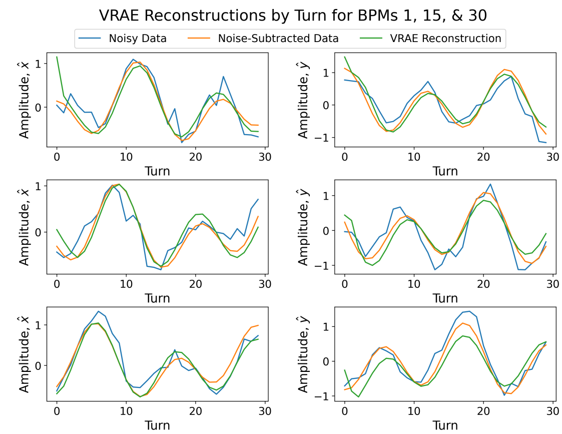

Figure 3: Comparison of autoencoder outputs and inputs without noise for several different BPMs.

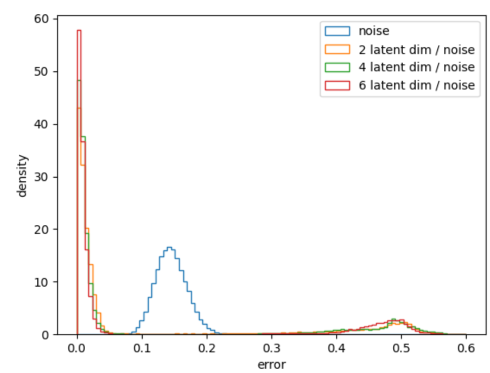

Figure 4: Histogram of the prediction error between the autoencoder output and the noise-subtracted signal for an autoencoder trained both with and without noise. Here we explore the sensitivity of the noise reduction capabilities to size of the latent space. The noise distribution is shown in blue for reference.

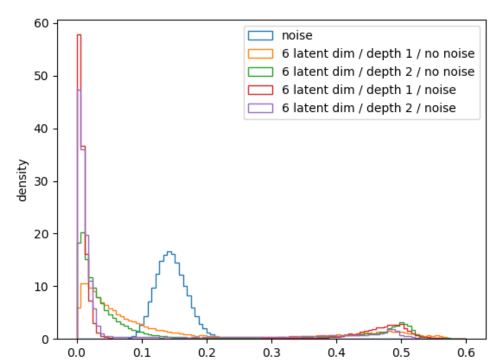

Figure 5: Histogram of the prediction error between the autoencoder output and the noise subtracted signal for an autoencoder trained both with and without noise. Here we explore the sensitivity of the noise reduction capabilities to the network depth. The noise distribution is shown in blue for reference.

ISBN: 978-3-95450-241-7 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2022-MOP10

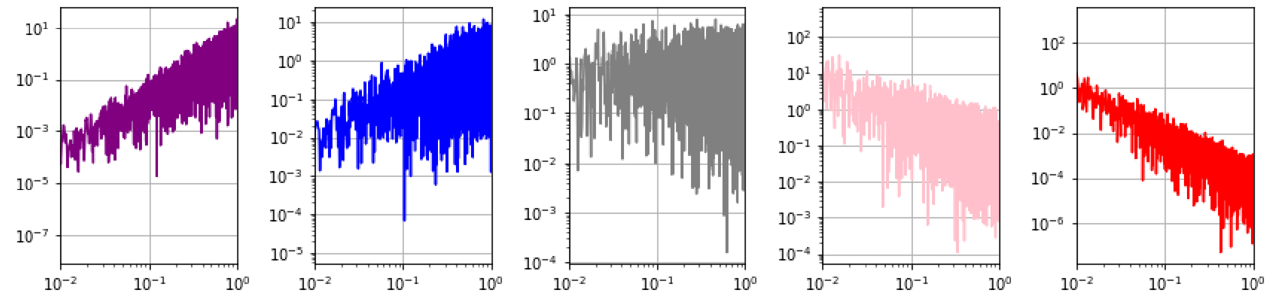

Figure 6: Power spectral densities for different colors of noise, with densities proportional to $1/f^{\beta}$. The factor β for each color of noise, from left to right, is two (violet), one (blue), zero (white), negative one (pink), and negative two (red).

of random initial amplitudes significantly improved overall performance.

Training for both types of autoencoders was performed in the RadiaSoft Sirepo computing platform on NVIDIA Tesla V100 GPUs. Datasets of 50 000 samples were generated from an analytic accelerator model with varying levels of noise, depending on testing to be run. From each dataset, 90% was used for training and 10% for validation.

GAUSSIAN NOISE REDUCTION

Before considering colored noise, we look at random Gaussian noise. Random Gaussian noise has a power spectrum identical to gray or white noise. We trained the autoencoder using data with random noise and compared the reconstruction to the input data with noise subtracted. We explored a wide range of configurations for training and testing, including different model structures. We trained on different noise levels, including on data without noise, to better understand the VRAE's noise reduction capabilities. First we considered training on 20% additive Gaussian Noise. Figure 3 shows a comparison of the autoencoder reconstruction, the noise subtracted data, and the noisy data, for the horizontal and vertical positions of three BPMs in our dataset. Here we see that the VRAE does a very good job of removing noise from the BPM data.

One way to quantify the level of noise is to compute the mean squared error for each BPM and then compute the mean squared error of the errors for all BPMs. Here the error is defined as the difference between the reconstruction and the input data with the noise subtracted. We computed the histograms of these errors over a number of trials and compared the results for several different test cases. Figure 4 shows the noise reduction capabilities of an autoencoder trained and tested on data with noise. Here we also vary the number of latent dimensions. Not only do we see a significant increase in the noise reduction capabilities of this network, but we also see an increase in the noise reduction effectiveness as we increase the number of latent dimensions. This test case is also more representative of operational data where one will not be able to remove the noise prior to training, and would instead be relying on the autoencoder to provide noise reduction.

Figure 5 compares the denoising ability of an autoencoder trained with and without noise, but while varying the number of latent layers of the network. Here the increased depth does appear to improve the noise reduction capabilities, but by far the largest effect is training with noisy data. In all cases there is a peak in the error around 0.5 — these are cases where the autoencoder does not do a good job of reconstructing the original signal.

COLORED NOISE

Our work thus far has been focused on Gaussian random noise, which has no amplitude dependence on frequency, making it white noise (gray in Fig. 6). In electronic systems operated under different environmental conditions, it is possible to have many different kinds of noise. To better understand how well this technique will generalize to real BPM systems, we explore the removal of different colors of noise. Colored noise referrers to power spectrum noise where the intensity of the noise varies with the frequency of the signal. White or gray noise has a uniform intensity with frequency. Pink noise is noise where the intensity is inversely proportional to the frequency in a $1/f$ relationship. For our study we explored denoisinig for four different types of power law noise. Figure 6 shows different spectra of power law noise with their color designations.

Figures 7 and 8 show the results for noise reduction on both Red and Violet noise, respectively, the spectral densities of which have the strongest dependence on frequency. In both cases the VRAE does a good job removing the noise and the results are fairly similar. There are subtle differences between the performance on the two different types of noise, but overall the VRAE is not sensitive to either noise type.

CONCLUSIONS

We have generated a test dataset using a linear optics model and used this dataset to evaluate the ability of VRAEs to remove different types of noise. We explored the effect of different VRAE architectures in conjunction with different noise types, thus demonstrating the efficacy of this technique under a variety of circumstances. Additional work is needed to better understand cases where the VRAE does a poor job of reconstructing signals, leading to an increase in effective

ISBN: 978-3-95450-241-7 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2022-MOP10

noise. Careful consideration for these edge cases will be important when implementing the method on operational data.

ACKNOWELDGEMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC0021699.

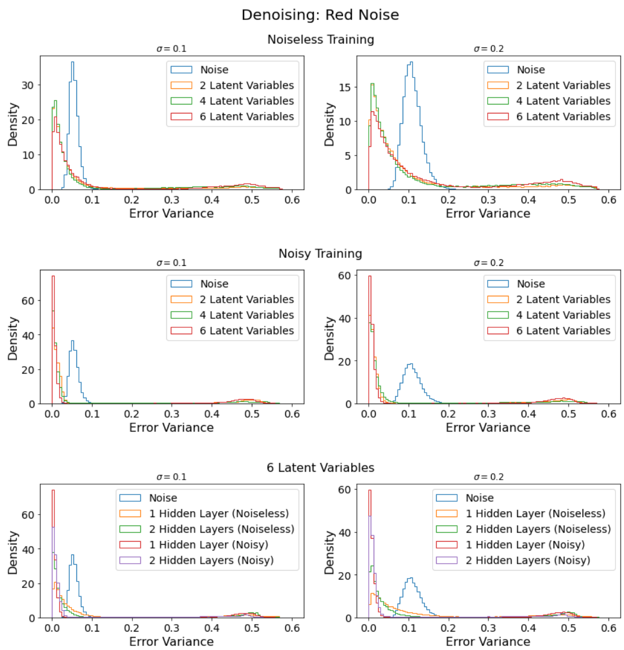

Figure 7: Noise reduction study for Red noise showing effective reduction of the noise spectrum.

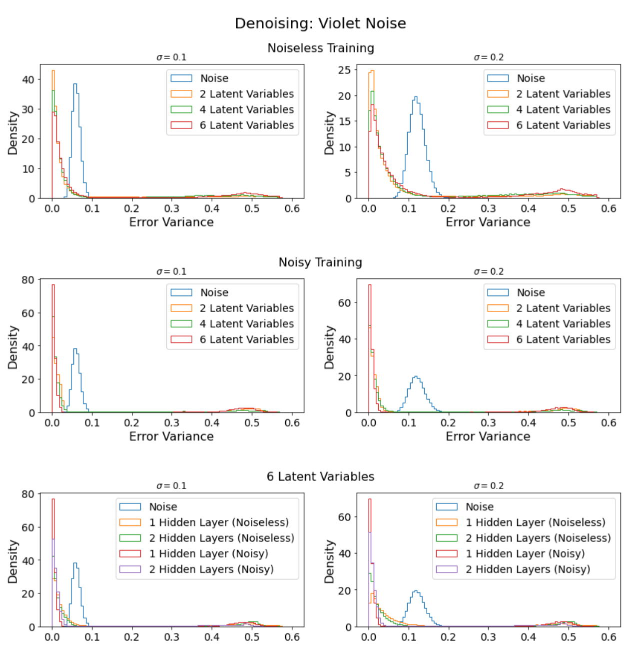

Figure 8: Noise reduction study for Violet noise showing effective reduction of the noise spectrum.

REFERENCES

[1] A. L. Edelen, S. G. Biedron, B. E. Chase, D. Edstrom, S. V. Milton, and P. Stabile, "Neural networks for modeling and control of particle accelerators," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 2, pp. 878–897, 2016. doi:10.1109/TNS.2016.2543203

[2] A. Edelen *et al.*, "Opportunities in machine learning for particle accelerators," 2018. doi:10.48550/arXiv.1811.03172

[3] E. Fol, F. S. Carlier, J. M. C. de Portugal, A. Garcia-Tabares, and R. Tomas, "Machine Learning Methods for Optics Measurements and Corrections at LHC," in *Proc. IPAC'18*, Vancouver, Canada, Apr.-May 2018, pp. 1967–1970. doi:10.18429/JACoW-IPAC2018-WEPAF062

[4] E. Fol, J. M. C. de Portugal, and R. Tomás, "Unsupervised Machine Learning for Detection of Faulty Beam Position Monitors," in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 2668–2671. doi:10.18429/JACoW-IPAC2019-WEPGW081

[5] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, "Noise reduction in gravitational-wave data via deep learning," *Phys. Rev. Res.*, vol. 2, p. 033 066, 2020. doi:10.1103/PhysRevResearch.2.033066

[6] G. Vajente *et al.*, "Machine-learning nonstationary noise out of gravitational-wave detectors," *Phys. Rev. D: Part. Fields*, vol. 101, p. 042 003, 2020. doi:10.1103/PhysRevD.101.042003

[7] D. Bhowick, D. K. Gupta, S. Maiti, and U. Shankar, "Stacked autoencoders based machine learning for noise reduction and signal reconstruction in geophysical data," 2019. doi:10.48550/arXiv.1907.03278

[8] X. Huang, S. Y. Lee, E. Prebys, and R. Tomlin, "Application of independent component analysis to fermilab booster," *Phys. Rev. ST Accel. Beams*, vol. 8, p. 064 001, 2005. doi:10.1103/PhysRevSTAB.8.064001

[9] O. Fabius and J. R. van Amersfoort, "Variational recurrent auto-encoders," 2015. doi:10.48550/arXiv.1412.6581

[10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2014. doi:10.48550/arXiv.1312.6114