ISBN: 978-3-95450-230-1 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2021-THOB03

ADAPTIVE CONTROL AND MACHINE LEARNING FOR PARTICLE ACCELERATOR BEAM CONTROL AND DIAGNOSTICS*

A. Scheinker†, Los Alamos National Laboratory, Los Alamos, USA

Abstract

In this tutorial, we start by reviewing some topics in control theory, including adaptive and model-independent feedback control algorithms that are robust to uncertain and time-varying systems, and provide some examples of their application for particle accelerator beams at both hadron and electron machines. We then discuss recent developments in machine learning (ML) and show some examples of how ML methods are being developed for accelerator controls and diagnostics, such as online surrogate models that act as virtual observers of beam properties. Then we give an overview of adaptive machine learning (AML) in which adaptive model-independent methods are combined with ML-based methods so that they are robust for and applicable to time-varying systems. Finally, we present some recent applications of AML for accelerator controls and diagnostics. In particular we present recently developed adaptive latent space tuning methods and show how they can be used as virtual adaptive predictors of an accelerator beam's longitudinal phase space as well as all of the other 2D projections of a beam's 6D phase space. Throughout the tutorial we will present recent results of various algorithms which have been applied at the LANSCE ion accelerator, the EuXFEL and LCLS FELs, the FACET plasma wakefield accelerator facility, the NDCXII ion accelerator, and the HiRES compact UED.

INTRODUCTION

The control of charged particle beams in particle accelerator facilities is a very challenging task due to the time variation and complexity of the beams and of the machines. Accelerators are typically composed of hundreds-thousands of coupled components which include radio frequency (RF) resonant cavities used for acceleration as well as magnets for focusing of the beams. The performance of large RF systems is known to drift with time due to external disturbances such as vibrations in the case of superconducting cavities and temperature fluctuations for normal conducting cavities which perturb their resonant frequencies. On slower time scales environmental temperature changes result in slight variation of RF cables or analog RF components such as mixers or local oscillators which also introduces phase and amplitude shifts in the highly sensitive high frequency RF systems. The performance of magnets is also perturbed and uncertain due to issues such as power source ripple, hysteresis, and misalignments.

Charged particle beams themselves are also highly complex and time-varying objects which live in a 6 dimensional phase space (x, y, z, p_x, p_y, p_z) which is impossible to measure directly or quickly. While newer electron machines are able to measure 2D longitudinal phase space projections (z, E) hadron machines are many times limited to scalar beam position or current monitor-based measurements online. More detailed measurements are possible but typically rely on slow emittance scans or wire scanners, which cannot be performed online in real-time without disrupting operations. Furthermore, both hadron and electron machines suffer from time-varying initial phase space distributions at their sources and undergo complex collective effects such as space charge forces. In the case of highly relativistic intense electron beams collective effects such as coherent synchrotron radiation are also an issue.

Because of all of the complexities and uncertainties described above, advanced controls and diagnostics are of great importance in the accelerator community. Control theory methods, including adaptive and model-independent feedback control algorithms exist which are robust to uncertain and time-varying systems and we provide some examples of their application for particle accelerator beams at both hadron and electron machines. We also discuss recent developments in machine learning (ML) and show some examples of how ML methods are being developed for accelerator controls and diagnostics, such as online surrogate models that act as virtual observers of beam properties. Then we give an overview of adaptive machine learning (AML) in which adaptive model-independent methods are combined with ML-based methods which are used to train surrogate models directly from raw data.

ADAPTIVE CONTROL

Model-independent feedback methods have been developed by the control theory community with an emphasis of robustness to un-modeled disturbances and changes to system dynamics. One classic adaptive control result is given for a scalar linear system of the form

$$\dot{x}(t) = ax(t) + bu(t), \tag{1}$$

where the values of a and b are unknown. Such a system cannot be stabilized with simple proportional integral derivative (PID)-type feedback, but if the sign of b is known, for example if $b > 0$, a stable equilibrium of (1) can be established at $x = 0$ by the following nonlinear controller

$$u(t) = \theta(t)x(t), \quad \dot{\theta}(t) = -kx^2(t), \quad k > 0. \tag{2}$$

This approach does not depend on a detailed knowledge of system dynamics, but has major limitations which are: 1). The sign of the unknown term b must be known and cannot be time-varying. 2). The presence of an arbitrarily small

* Work supported by Los Alamos National Laboratory
† ascheink@lanl.gov

ISBN: 978-3-95450-230-1 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2021-THOB03

un-modeled disturbance in the dynamics (1) can destabilize the closed loop nonlinear system [1–3].

For a long time, the main limitation of nonlinear and adaptive control approaches was an inability to handle a sign-changing time varying coefficient $b(t)$ in system (1) which multiplies the control input $u(t)$, such as $b(t) = \cos(2\pi f t)$ which changes sign repeatedly thereby changing the effect of control input $u(t)$. For particle accelerators such variation comes from the fact that the beam at a certain location is influenced by many upstream components, such as magnet settings as well as the initial phase space distribution of the beam entering into the particle accelerator. Changes in input beams and accelerator components upstream have an influence on the response of quantities such as beam loss relative to downstream components. For example, consider a state $x(t)$ which describes beam loss in a particle accelerator, whose minimization is desired, which is influenced by a large collection of quadrupole magnets $\mathbf{u} = (u_1, \ldots, u_m)$. The effect of a single magnet, u_m, depends on the initial beam's phase space as it enters the accelerator from the source and also on the settings of all of the other quadrupole magnets that are upstream, $u_{i<m}$ and changes with time as the upstream magnets are adjusted and as the initial beam conditions change. One day decreasing x may require decreasing the current of magnet u_m and another day it might have to be increased.

The control and stabilization of time-varying systems is notoriously difficult, even simple linear time-varying systems are difficult to analyze in general because standard eigenvalue techniques break down and stability can only be proven by using Lyapunov theory [1]. Recently, a nonlinear extremum seeking (ES) feedback control method was developed which could stabilize and minimize the analytically unknown outputs of a wide range of dynamics systems, scalar and vector-valued which can be time-varying, nonlinear and open loop unstable with unknown control directions [4–6]. The ES method is applicable to a wide range of nonlinear and time-varying systems of the form

$$\begin{aligned} \dot{x} &= a(t)x(t) + b(t)u(\hat{y}(t)), \\ \dot{\mathbf{x}} &= A(t)\mathbf{x}(t) + B(t)\mathbf{u}(\hat{y}(t)), \\ \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(\hat{y}(t)), t), \\ \hat{y}(\mathbf{x}, t) &= y(\mathbf{x}, t) + n(t), \end{aligned} \tag{3}$$

which include scalar time-varying linear systems, vector-valued time-varying linear systems, and vector-valued nonlinear time-varying systems, where in each case the feedback control u is based only on a noise-corrupted measurement $\hat{y}(t)$ of an analytically unknown cost function $y(\mathbf{x}, t)$. For example, a measurable but analytically unknown cost function an be the sum of beam loss along a many kilometer long particle accelerator, which depends on all accelerator parameters and on the initial 6D phase space of the beam being accelerated.

For accelerator applications, the ES method can tune groups of parameters, $\mathbf{p} = (p_1, \ldots, p_m)$. For example, tuned parameters might include RF cavity amplitude and phase set points as well as magnet power supply voltages or currents. The adaptive ES algorithm dynamically tunes parameters according to

$$\dot{p}_j = \psi_j \left(\omega_j t + k \hat{y}(\mathbf{x}, t) \right), \tag{4}$$

where ω_i are distinct dithering frequencies defined as $\omega_i = \omega r_i$ with $r_i \neq r_j$ for $i \neq j$, k is a feedback gain. The ψ_j may be chosen from a large class of functions which may be non-differentiable and not even continuous, such as square waves which are easily implemented in digital systems [6]. The only requirements on the ψ_j are that for a given time interval $[0, t]$ they are measurable with respect to the L^2 norm and that they are mutually orthogonal in Hilbert space in the weak sense relative to all measurable functions $f(t) \in L^2[0, t]$ in the limit as $\omega \to \infty$, which can be written as

$$\begin{aligned} \lim_{\omega\to\infty} \int_0^t \psi_i(\tau)\psi_j(\tau) d\tau &= 0, \ \forall i \neq j, \\ \lim_{\omega\to\infty} \int_0^t \psi_i(\tau) f(\tau) d\tau &= 0, \ \forall i, \ \forall f(t) \in L^2[0, t], \\ \lim_{\omega\to\infty} \int_0^t \psi_i^2(\tau) f(\tau) d\tau &= \int_0^t c_i f(\tau) d\tau, \end{aligned}$$

$\forall i, \forall f(t) \in L^2[0, t], \ c_i > 0.$

One particular implementation of the ES method is especially convenient for particle accelerator applications because the tuning functions ψ_i have analytically guaranteed bounds despite acting on analytically unknown and noisy functions, which guaranteed known update rates and limits on all tuned parameters [5]:

$$u_i = \sqrt{\alpha_i \omega_i} \cos \left(\omega_i t + k \hat{y}(\mathbf{x}, t) \right). \tag{5}$$

The utility of this approach is clearly demonstrated by considering a system of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(\hat{y}(t)), t), \quad \dot{p}_i = \sqrt{\alpha \omega_i} \cos \left(\omega_i t + k \hat{y}(\mathbf{x}, t) \right), \tag{6}$$

which results in average dynamics that minimize the noise-corrupted unknown function $y(\mathbf{x}, t)$:

$$\dot{\bar{p}}_i = -\frac{k\alpha}{2} \frac{\partial y(\mathbf{x}, t)}{\partial p_i}. \tag{7}$$

This method has been utilized for various particle accelerator applications including real-time betatron oscillation minimization in a time-varying magnetic lattice at the SPEAR3 synchrotron [7], for maximization of the output power of the Linac Coherent Light Source (LCLS) free electron laser (FEL) and of the European X-ray FEL [8], for real-time multi-objective optimization for simultaneous trajectory control and emittance growth minimization at the AWAKE plasma wakefield acceleration facility at CERN [9], and for beam loss minimization by automatically tuning the amplitude and phase set points of multiple RF cavities at the Los Alamos Neutron Science Center (LANSCE) linear ion accelerator. One limitation of adaptive methods such as the ES approach is that they are local feedback-based methods and it is possible for them to get stuck in a local minimum when optimizing for an analytically unknown output function.

ISBN: 978-3-95450-230-1
ISSN: 2673-5350
doi:10.18429/JACoW-IBIC2021-THOB03

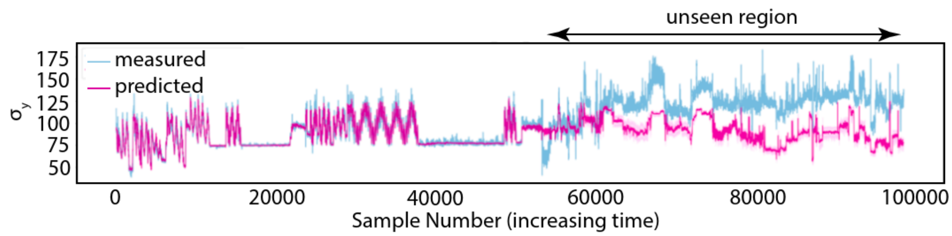

Figure 1: The accuracy of the σ_y prediction quickly degrades once the system has had time to evolve and leaves the span of the collected training data set. Such an approach would have to be continuously and repeatedly re-trained to maintain accuracy, which is infeasible and defeats the purpose of the ML-based diagnostic [16].

LIMITATIONS OF ML FOR TIME-VARYING SYSTEMS

Machine learning (ML) tools are being developed that can learn representations of complex accelerator dynamics directly from data. ML methods have been utilized to develop surrogate models to act as virtual diagnostics [10], powerful polynomial chaos expansion-based surrogate models have been used for uncertainty quantification [11], convolutional neural networks have been used for time-series classification and forecasting in accelerators [12], Bayesian Gaussian processes utilize learned correlations in data/physics-informed kernels [13], surrogate models can help speed up simulation-based optimization [14], and various ML methods have been used for beam dynamics studies at CERN [15].

A major limitation of ML methods, and an active area of research in the ML community, is the problem of time-varying systems, known as distribution shift. If a system changes with time then the data that was used to train an ML-based tool will no longer provide an accurate representation of the system of interest, and the accuracy of the ML tool will degrade. Distribution drift is a challenge for all ML methods including neural networks for surrogate models, the use of neural networks to represent cost functions or optimal policies in reinforcement learning, and even for methods such as Gaussian processes which utilize learned correlations in their kernels. Incorporating methods to deal with distribution shift is a major need for the accelerator community because accelerators and their beams change unpredictably with time. This challenge is illustrated by Fig. 1 which demonstrates that an ML-based prediction quickly degrades in accuracy as the system changes over time, which in this case is the σ_y beam prediction for the LCLS [16]. Such an approach would have to be continuously and repeatedly re-trained to maintain accuracy, which is infeasible and defeats the purpose of the ML-based diagnostic.

ADAPTIVE ML FOR TIME-VARYING SYSTEMS

Efforts have begun to combine the robustness of adaptive feedback with the global representations that can be learned with ML methods to develop adaptive machine learning (AML) for time-varying systems. The first such result combined neural networks and model-independent feedback to provide fast, robust, and automatic control over the energy vs time phase space of electron bunches in the LCLS [17]. Recently, AML methods have been studied in more generality for adaptive tuning of the inputs and outputs of ML tools such as neural networks for time-varying systems [18].

Some of the most powerful ML tools are encoder-decoder generative convolutional neural networks (CNN) which can be used to find highly efficient nonlinear functions that can project incredibly high dimensional input spaces, which may be combinations of images and vectors, down to a very low dimensional latent space, before generating back up to a high dimensional representation [19, 20]. Encoder-decoder networks have been used for anomaly detection [21], time-series data [22], and for optimization of deep generative models [23].

A novel approach to AML for time-varying systems is now being developed which utilizes such generative CNN-based encoder-decoders to adaptively tune directly the low-dimensional latent space representation (as small as 2 dimensions), for incredibly high dimensional systems (hundreds of thousands - millions of parameters) [24–26]. The setup of such an encoder-decoder generative CNN is shown in Figure 2. The network takes inputs that are 2D images of beam phase space distributions together with vectors of accelerator parameter settings such as magnets and RF systems. The high dimensional inputs are squeezed down to a low-dimensional latent space representation from which a collection of distributions is then generated, as shown in Figure 3 for a 2-dimensional latent space representation.

The method works by first performing a supervised learning-based training in which we have access to input-output pairs of the form $(\mathbf{x}_{\text{in}}, \mathbf{X}_{\text{in}}, \hat{\mathbf{Y}}_{\text{out}})$ where $\mathbf{x}_{\text{in}}$ are vectors of accelerator parameter inputs, $\mathbf{X}_{\text{in}}$ are stacks of 2D phase space image inputs. The generative half of the encoder-decoder CNN builds back up to a high dimensional output $\hat{\mathbf{Y}}_{\text{out}}$ which is a $752,640 = 224 \times 224 \times 15$ dimensional output with the 15 channels representing the 15 2D projections of the 6D phase space: (x, y), (x, z), (x, x'), (x, y'), (x, E), (x', y), (x', z), (x', y'), (x', E), (y, z), (y, y'), (y, E), (y', z), (y', E), (z, E) in the HiRES UED as shown in Figure 2 and in a similar setup the output is a $1,228,800 = 128 \times 128 \times 75$ dimensional object with the 75 channels representing the 15 2D projections of the 6D phase at 5 different locations in FACET-II, shown in 3. By

ISBN: 978-3-95450-230-1 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2021-THOB03

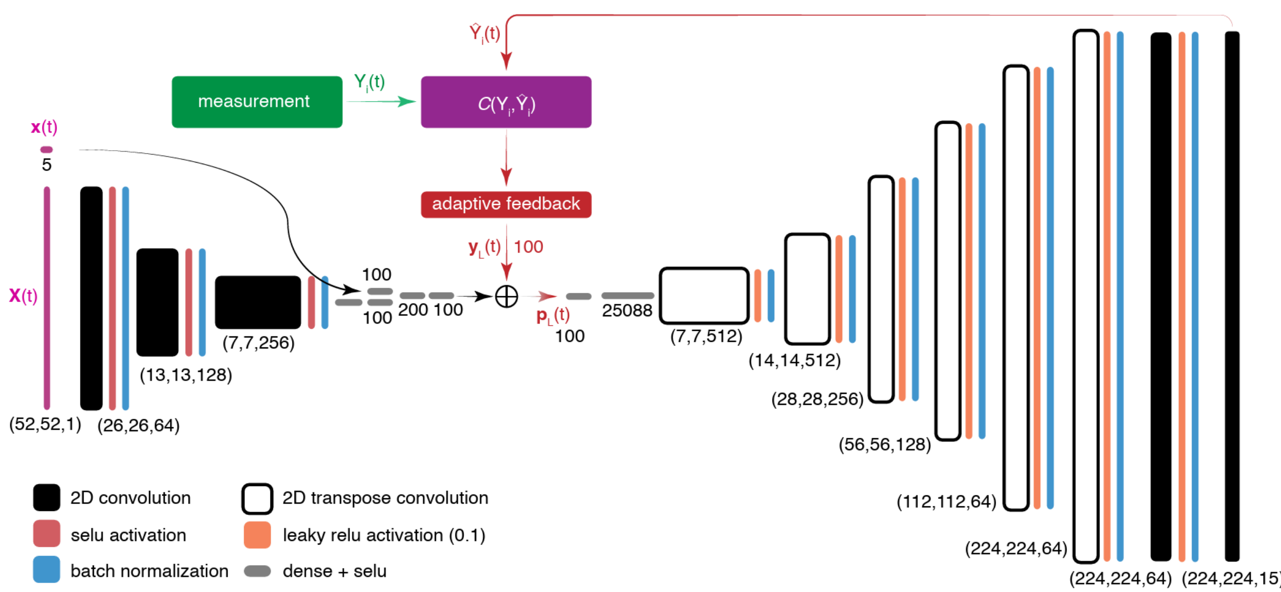

Figure 2: The HiRES encoder-decoder CNN structure for the AML setup is shown with layer sizes such as (224, 224, 15) representing an output of 15 filters of image size 214 × 214 each. The dense layer widths shown as single numbers.

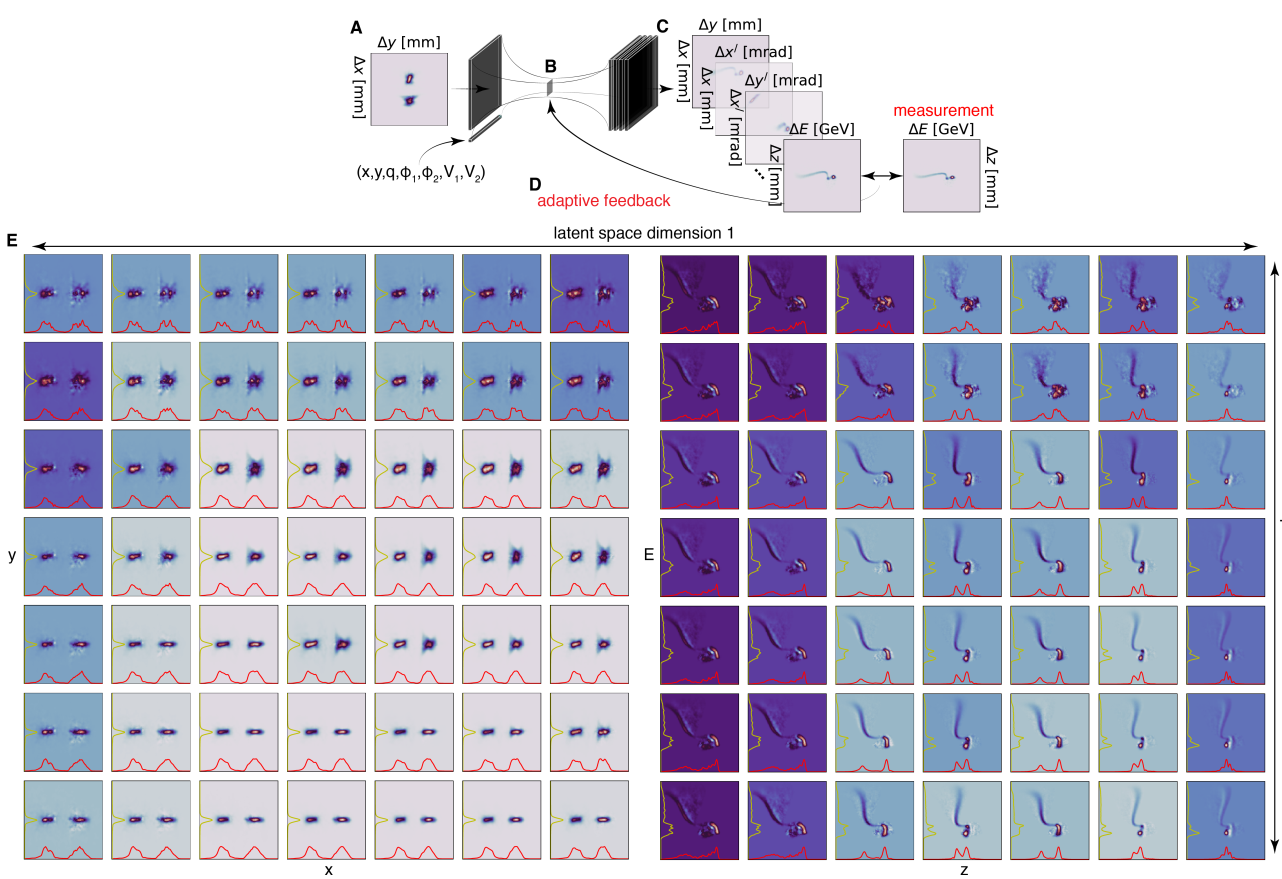

Figure 3: An encoder-decoder convolutional neural network setup is shown which takes an image of an electron beam's (x, y) phase space distribution as an input together with a vector of accelerator parameters (A). The high dimensional inputs are squeezed down to a 2 dimensional latent space (B), from which 75 2D distributions are then generated which are all 15 2D projections of the beam's 6D phase space at 5 different particle accelerator locations (C). Some of the projections, such as the (z, E) longitudinal phase space distributions can be compared to TCAV-based measurements to guide adaptive feedback which takes place in the low dimension latent space to compensate for unknown changes in both the accelerator parameters and in the initial beam distribution (D). The variation of the (x', y') and (z, E) 2D phase space projections is shown as one moves through the 2D latent space learned by the network and adaptively tuned (E) [25].

ISBN: 978-3-95450-230-1 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2021-THOB03

forcing the generative half of the CNN to predict such high dimensional outputs which contain all of the projections of the beam's 6D phase space simultaneously, we are forcing the CNN to learn the relationships between various phase space projections as well as their correlations and physics constraints within the particle accelerator system for which the network is being trained.

In both the HiRES and the FACET-II setup we are considering a mapping of inputs to outputs of the form

$$\hat{\mathbf{Y}}_{\text{out}}(t) = \mathbf{F}\left(\mathbf{x}_{\text{in}}(t), \mathbf{X}_{\text{in}}(t)\right), \qquad (8)$$

where both the accelerator parameters $\mathbf{x}_{\text{in}}(t)$ and the input beam $\mathbf{X}_{\text{in}}(t)$ are expected to change unpredictably with time and furthermore we assume that we will not have access to non-invasive and accurate measurements of these changes. Furthermore, once the accelerator is operational, we lose access to most of the true measurements of the beam's phase space $\mathbf{Y}_{\text{out}}(t)$ which could be compared to their predictions from the generative CNN $\hat{\mathbf{Y}}_{\text{out}}(t)$. However, most advanced accelerators do have access to non-invasive measurements of some subset of the beam's phase space, for example transverse deflective cavities together with dipole magnets can be used to measure the beam's longitudinal phase space (LPS) 2D (z, E) distribution as is routinely done at the LCLS.

In order to accurately predict $\hat{\mathbf{Y}}_{\text{out}}(t)$ without knowledge of the time-varying accelerator beam and component measurements $(\mathbf{x}_{\text{in}}(t), \mathbf{X}_{\text{in}}(t))$, we rely on the fact that the generative CNN has learned the correlations within the system and respects the physics constraints in the data and therefore we use just the available measurements, such as the LPS distribution or energy spread spectrum measurements, which we denote as $\hat{\mathbf{Y}}_i(t) \in \hat{\mathbf{Y}}_{\text{out}}(t)$.

We compare just these predictions to their measurements and operate the trained generative CNN in a un-supervised adaptive manner in which we apply feedback directly on the low-dimensional latent space representation in order to track the time-varying measurements by actively minimizing a cost function in real time, of the form:

$$C(\mathbf{Y}_i(t), \hat{\mathbf{Y}}_i(t)) = \iint \left|\mathbf{Y}_i(t) - \hat{\mathbf{Y}}_i(t)\right| dY_i, \qquad (9)$$

which is minimized by adaptively tuning the latent space parameters $\mathbf{y}_L = (y_1, \ldots, y_n)$, according to the model-independent ES algorithm described above, according to:

$$\frac{dy_j(t)}{dt} = \sqrt{\alpha_i \omega_i} \cos\left(\omega_j t + k_j C(\mathbf{Y}_i(t), \hat{\mathbf{Y}}_i(t))\right), \qquad (10)$$

as shown in Figure 2.

Note that with this implementation, the relationship in Equation (8) is now being approximated by

$$\hat{\mathbf{Y}}_{\text{out}}(t) \approx \hat{\mathbf{F}}\left(\mathbf{y}_L(t)\right), \qquad (11)$$

where $\hat{\mathbf{F}}$ is the generative half of the CNN and $\hat{\mathbf{Y}}_{\text{out}}(t)$ is now parameterized by the low dimensional latent space vector $\mathbf{y}_L(t)$ without needing access to measurements of $(\mathbf{x}_{\text{in}}(t), \mathbf{X}_{\text{in}}(t))$.

One example of such convergence for the FACET-II setup with a 7-dimensional latent space is shown in Figure 4, which shows the trajectory taken by ES in the latent space from a starting point very far from the correct input distribution and accelerator parameters $(\mathbf{x}_{\text{in}}(t), \mathbf{X}_{\text{in}}(t))$ as it converges to the global minimum, with the components (y_1, y_n) for $n \in \{2, 3, 4, 5, 6, 7\}$ shown overlaid on top of the cost function surface. Figure 5 shows the results of the convergence

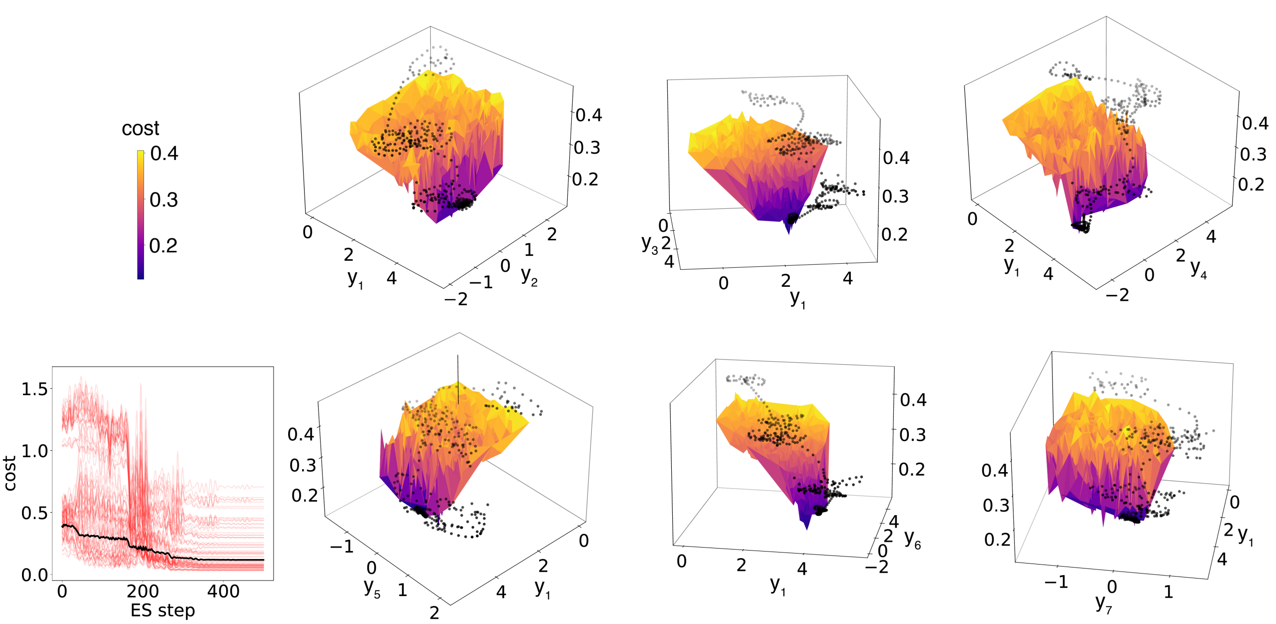

Figure 4: Several 3D projections (y_1, y_n) for $n \in \{2, 3, 4, 5, 6, 7\}$ of convergence within the 7D latent space are shown with the adaptively tuned trajectory shown as black dots lifted slightly above the surface of the cost function. The cost convergence is also shown and seen to take approximately 400 steps to converge.

ISBN: 978-3-95450-230-1 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2021-THOB03

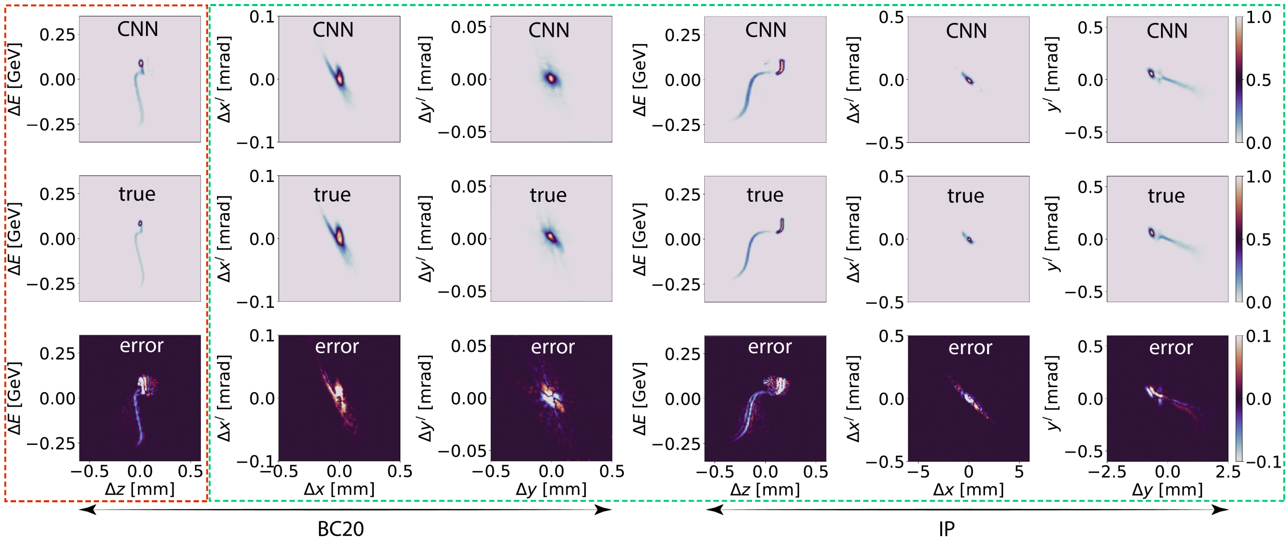

Figure 5: Predictions of the 7D latent space model of the phase space at bunch compressor 20 of FACAT-II (BC20) and at the interaction point (IP). The red dashed box shows a LPS diagnostic that was used as part of the cost function while the other 2D phase space projections in the green dashed box were unseen by the CNN which is correctly predicting projections of the beam's 2D phase space not only at BC20, but also at the unseen IP location.

which gives a close match of various 2D phase space projections throughout the accelerator despite feedback acting only based on a single LPS measurement.

CONCLUSION

This work demonstrates an adaptive ML approach to high dimensional time-varying systems in general and in particular for particle accelerator applications in which both the accelerator components and the input beams change unpredictably with time due to various external disturbances. By training a deep convolutional encoder-decoder style generative neural network and forcing it to predict all 2D projections of the beam's 6D phase space simultaneously this physics-informed approach gives accurate predictions for unseen phase space projections by adaptively matching only a measurable distribution.

REFERENCES

[1] H. K. Khalil, *Nonlinear Systems*, Prentice Hall, 2002.

[2] P. Ioannou and B. Fidan, *Adaptive control tutorial,* Society for Industrial and Applied Mathematics, 2006.

[3] P. A. Ioannou and J. Sun, *Robust adaptive control,* Courier Corporation, 2012.

[4] A. Scheinker, "Model Independent Beam Tuning", in *Proc. IPAC'13*, Shanghai, China, May 2013, paper TUPWA068, pp. 1862–1864.

[5] A. Scheinker, "Simultaneous stabilization and optimization of unknown, time-varying systems," 2013 American Control Conference, IEEE, 2013. `https://ieeexplore.ieee.org/document/6580232`

[6] A. Scheinker and D. Scheinker, "Bounded extremum seeking with discontinuous dithers," *Automatica*, vol. 69, pp. 250-257, 2016. `doi:10.1016/j.automatica.2016.02.023`

[7] A. Scheinker, X. Huang, and J. Wu, "Minimization of betatron oscillations of electron beam injected into a time-varying lattice via extremum seeking," *IEEE Transactions on Control Systems Technology*, vol 26, no .1, pp. 336-343, 2017. `doi:10.1109/TCST.2017.2664728`

[8] A. Scheinker *et al.*, "Model-independent tuning for maximizing free electron laser pulse energy," *Physical Review Accelerators and Beams*, vol. 22, no .8, p. 082802, 2019. `doi:10.1103/PhysRevAccelBeams.22.082802`

[9] A. Scheinker *et al.*, "Online multi-objective particle accelerator optimization of the AWAKE electron beam line for simultaneous emittance and orbit control," *AIP Advances*, vol. 10, no .5, p. 055320, 2020. `doi:10.1063/5.0003423`

[10] C. Emma *et al.*, "Machine learning-based longitudinal phase space prediction of particle accelerators," *Physical Review Accelerators and Beams*, vol. 21, no .11, p. 112802, 2018. `doi:10.1103/PhysRevAccelBeams.21.112802`

[11] A. Adelmann, "On nonintrusive uncertainty quantification and surrogate model construction in particle accelerator modeling," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 7, pp.383–416, 2019. `doi:10.1137/16M1061928`

[12] S. Li *et al.*, "A Novel Approach for Classification and Forecasting of Time Series in Particle Accelerators," *Information*, vol. 12, no .3, p. 121, 2021. `doi:10.3390/info12030121`

[13] J. Duris *et al.*, "Bayesian optimization of a free-electron laser," *Physical review letters*, vol. 124, no. 12, p. 124801, 2020. `doi:10.1103/PhysRevLett.124.124801`

[14] A. Edelen *et al.*, "Machine learning for orders of magnitude speed up in multiobjective optimization of particle accelerator systems," *Physical Review Accelerators and Beams*, vol. 23, p .4, p. 044601, 2020. `doi:10.1103/PhysRevAccelBeams.23.044601`

[15] E. Fol *et al.*, "Supervised learning-based reconstruction of magnet errors in circular accelerators," *The European Phys-*

ISBN: 978-3-95450-230-1 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2021-THOB03

ical Journal Plus, vol. 136, no .4, pp. 1-19, 2021. doi:10.1140/epjp/s13360-021-01348-5

[16] Figure shared by collaborators at SLAC National Accelerator Laboratory.

[17] A. Scheinker *et al.*, “Demonstration of model-independent control of the longitudinal phase space of electron beams in the linac-coherent light source with femtosecond resolution,” *Physical review letters*, vol. 121, no .4, p. 044801, 2018. doi:10.1103/PhysRevLett.121.044801

[18] A. Scheinker, “Adaptive Machine Learning for Robust Diagnostics and Control of Time-Varying Particle Accelerator Components and Beams,” *Information*, vol. 12, no .4, p. 161, 2021. doi:10.3390/info12040161

[19] A. L. Caterini, A. Doucet, and D. Sejdinovic. “Hamiltonian variational auto-encoder." arXivpreprint arXiv:1805.11328, 2018. https://arxiv.org/abs/1805.11328

[20] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, C. A. Stevens, and L. Carin, “Variational autoencoder for deep learning of images, labels and captions," *Advances in neural information processing systems*, vol. 29, pp. 2352-2360, 2016. https://proceedings.neurips.cc/paper/2016/file/eb86d510361fc23b59f18c1bc9802cc6-Paper.pdf

[21] J. An and S. Cho, “Variational autoencoder based anomaly detection using reconstruction probability,” Special lecture on IE., vol. 2, no .1, pp. 1-18, 2015. http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf

[22] R. Maulik, A. Mohan, B. Lusch, S. Madireddy, P. Balaprakash, and D. Livescu, “Time-series learning of latent-space dynamics for reduced-order model closure,” *Physica D: Nonlinear Phenomena*, vol. 405, p. 132368, 2020. doi:10.1016/j.physd.2020.132368

[23] A. Tripp, E. Daxberger, and J. M. Hernández-Lobato. “Sample-efficient optimization in the latent space of deep generative models via weighted retraining,” *Advances in Neural Information Processing Systems*, 33, 2020. https://proceedings.neurips.cc/paper/2020/file/81e3225c6ad49623167a4309eb4b2e75-Paper.pdf

[24] A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, “Adaptive deep learning for time-varying systems with hidden parameters: Predicting changing input beam distributions of compact particle accelerators,” arXiv preprint arXiv:2102.10510, 2021. https://arxiv.org/abs/2102.10510

[25] A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, “Adaptive Latent Space Tuning for Non-Stationary Distributions,” arXiv preprint arXiv:2105.03584, 2021. https://arxiv.org/abs/2105.03584

[26] A. Scheinker, “Adaptive Machine Learning for Time-Varying Systems: Low Dimensional Latent Space Tuning,” arXiv preprint arXiv:2107.06207, 2021. https://arxiv.org/abs/2107.06207