ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-MOPOPT034

SURROGATE-BASED BAYESIAN INFERENCE OF TRANSVERSE BEAM DISTRIBUTION FOR NON-STATIONARY ACCELERATOR SYSTEMS

H. Fujii*, N. Fukunishi, RIKEN Nishina Center, Wako, Japan
M. Yamakita, Tokyo Institute of Technology, Meguro-ku, Japan

Abstract

Constraints on the beam diagnostics available in real-time and time-varying beam source conditions make it difficult to provide users with high-quality beams for long periods without interrupting experiments. Although surrogate model-based inference is useful for inferring the unmeasurable, the system states can be incorrectly inferred due to manufacturing errors and neglected higher-order effects when creating the surrogate model. In this paper, we propose to adaptively assimilate the surrogate model for reconstructing the transverse beam distribution with uncertainty and underspecification using a sequential Monte Carlo from the measurements of quadrant beam loss monitors. The proposed method enables sample-efficient and training-free inference and control of the time-varying transverse beam distribution.

MATHEMATICAL BACKGROUND

Accelerator Systems with Unknown Drift and Parameter Uncertainties

The dynamical system of the accelerator with measurement error can be written as

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}, \boldsymbol{u}) \tag{1}$$
$$\boldsymbol{y} = h(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{x} \equiv \boldsymbol{x}(t) \in \mathbb{R}^n$ is the vector of the time-varying hidden states, $\boldsymbol{u} \in \mathbb{R}^m$ is the control input state, t is the time. $\boldsymbol{y} \in \mathbb{R}^d$ is the measurement, and $\boldsymbol{\epsilon} = [\epsilon_1, ..., \epsilon_d]^T$ is the independent measurement noise.

To produce a quality beam of interest by inferring hidden states $\boldsymbol{x}$, we introduce a surrogate model as

$$\dot{\hat{\boldsymbol{x}}} = \hat{f}(\hat{\boldsymbol{x}}, \boldsymbol{u}) \tag{3}$$
$$\hat{\boldsymbol{y}} = \hat{h}(\hat{\boldsymbol{x}}, \boldsymbol{u}), \tag{4}$$

where $\hat{\boldsymbol{x}} \equiv \hat{\boldsymbol{x}}(t) \in \mathbb{R}^n$, and $\hat{\boldsymbol{y}} \in \mathbb{R}^d$ are the counterpart of the real system Eq. (1)(2) built based on the parameterized system model. One may infer the hidden state by simply minimizing the loss function

$$C(\boldsymbol{y}) = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2. \tag{5}$$

However, the accurate state $\boldsymbol{x}$ cannot be necessarily obtained due to the model uncertainties and possible multiple hidden states when the measurement is sparse. The goal of the following Approximate Bayesian Computation (ABC) framework [1] is to sequentially filter and track the time-varying hidden state estimates $\hat{\boldsymbol{x}}$ by minimizing the expectation of the cost function in Eq. (5) throughout the operation of the accelerator complex.

* hiroki.fujii@riken.jp

Markov-Chain Monte-Carlo (MCMC) for the Initial Posterior Inference

Given an observation $\boldsymbol{y}$, the initial posterior distribution of interest for a given fixed control input $\boldsymbol{u}$ can be given by

$$p(\hat{\boldsymbol{x}}|\hat{\boldsymbol{y}}) = \frac{p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})p(\hat{\boldsymbol{x}})}{p(\hat{\boldsymbol{y}})}, \tag{6}$$

here subscripts for denoting the time steps are omitted in this section, e.g., $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_{k=0}$. Since $p(\hat{\boldsymbol{x}}|\hat{\boldsymbol{y}})$ is a probability distribution, we can rewrite this equation as

$$p(\hat{\boldsymbol{y}}) = \int p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})p(\hat{\boldsymbol{x}})d\hat{\boldsymbol{x}}. \tag{7}$$

The probability density function $p(\hat{\boldsymbol{y}})$ given a measurement $\boldsymbol{y}$ with normally distributed sensor noise with standard deviation $\boldsymbol{\sigma}_n = \{\sigma_{n1}, ..., \sigma_{nd}\}$ is expressed as

$$p(\hat{\boldsymbol{y}}) = \prod_j^d \frac{1}{\sqrt{2\pi}\sigma_{nj}} \exp\left(-\frac{|\hat{y}_j - y_j|^2}{2\sigma_{nj}^2}\right) \tag{8}$$

and $p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}})$ drawn by surrogate model equation Eq. (4). Then, the proposal $p(\hat{\boldsymbol{x}})$ of a general nonlinear system can be sampled using MCMC [2]. In this work, the parallel tempering MCMC (PT-MCMC) algorithm is adopted to efficiently find the optimal and promote mixing across the state spaces. Details on PT-MCMC can be found in [3].

Particle Filtering for Tracking Hidden States

The sequential importance resampling is a class of particle filter algorithm. The sequence is initialized with a set of N particles representing $p(\hat{\boldsymbol{x}})$, and assign the normalized weights $W_{k=1}^s = 1/N$, for all $s = 1, ..., N$. Then particles at a discrete time step $k-1$ are sequentially propagated for a new time step k to update prior $p(\hat{\boldsymbol{x}}_k^s|\hat{\boldsymbol{x}}_{k-1}^s, \boldsymbol{u}_{k-1})$. The weight update of s-th particle is performed based on the measurement and cosine similarity

$$p(\hat{\boldsymbol{y}}_k^s|\hat{\boldsymbol{x}}_k^s, \boldsymbol{u}_k, \boldsymbol{u}_{k-1}) \propto \mathrm{sim}_k(\Delta_k\hat{\boldsymbol{y}}^s, \Delta_k\boldsymbol{y}) \prod_j^d \exp\left(-\frac{|\hat{y}_j^s - y_j|^2}{2\sigma_{nj}^2}\right). \tag{9}$$

Here, $\mathrm{sim}_k(\Delta_k^s\hat{\boldsymbol{y}}, \Delta_k\boldsymbol{y})$ is a similarity measure between the responses of the actual system $\Delta_k\boldsymbol{y} \equiv \boldsymbol{y}_k - \boldsymbol{y}_{k-1}$ and the surrogate model $\Delta_k^s\hat{\boldsymbol{y}} \equiv \hat{\boldsymbol{y}}_k^s - \hat{\boldsymbol{y}}_{k-1}^s$ expressed by

$$\mathrm{sim}_k(\Delta_k^s\hat{\boldsymbol{y}}, \Delta_k\boldsymbol{y}) = \begin{cases} \frac{1+\tau(\Delta_k^s\hat{\boldsymbol{y}}, \Delta_k\boldsymbol{y})}{2} & (\|\Delta_k^s\hat{\boldsymbol{y}}\| > 2\sigma_n) \\ 1 & (\|\Delta_k^s\hat{\boldsymbol{y}}\| \le 2\sigma_n) \end{cases} \tag{10}$$

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-MOPOPT034

and $\tau(\boldsymbol{a},\boldsymbol{b}) \equiv \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}$ is the cosine similarity. The inequality conditions in Eq. (10) are introduced such that the noise level does not adversely affect the weight updates. Then the sample particles are reweighed according to Eq. (6) as

$$W_k^s = \frac{p(\hat{\boldsymbol{y}}_k^s|\hat{\boldsymbol{x}}_k^s, \boldsymbol{u}_k, \boldsymbol{u}_{k-1})W_{k-1}^s}{\sum_s p(\hat{\boldsymbol{y}}_k^s|\hat{\boldsymbol{x}}_k^s, \boldsymbol{u}_k, \boldsymbol{u}_{k-1})W_{k-1}^s}. \tag{11}$$

After each update of states and weights, the particles are resampled by systematic resampling to obtain new equally-weighted particles to prevent degeneracy when the effective sample size $ESS = \left(\sum_{s=1}^N (W_k^s)^2\right)^{-1}$ is below the user-defined threshold [4].

NUMERICAL EXPERIMENTS

CNN Modeling of the Transverse Beam Dynamics with Parameterized Beam Source Condition

In the numerical experiment, the system Eq. (1) is defined to mimic the beam transport line after the AVF cyclotron as in Fig. 1(a) of Ref. [5]. Control inputs $\boldsymbol{u} \in \mathbb{R}^7$ are assigned to applied excitation current for a quadrupole triplet (u_1, u_2, u_3), horizontal and vertical steerers (u_4, u_5), and a subsequent quadrupole doublet (u_6, u_7).

This study's incident particle distribution containing $N = 40,000$ particles is characterized by skewed bi-Gaussian transverse beam distribution with skewness parameters $\alpha_{\mathrm{x,y}}$ [6, 7], while fixing the beam width. Here, the skewness in x-axis α_{x} modifies the standard normal distribution $\phi(\mathrm{x}) = \frac{1}{\sqrt{2\pi}}\exp(-\mathrm{x}^2/2)$ to be $\phi_\alpha(\mathrm{x}) = 2\phi(\mathrm{x})\Phi(\alpha\mathrm{x})$, where $\Phi(\mathrm{x})$ is the cumulative distribution function of the normal distribution $\Phi(\mathrm{x}) = \int_{-\infty}^{\mathrm{x}} \phi(t)dt = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\mathrm{x}}{\sqrt{2}}\right)\right]$. The same applies to the y-axis. Note that the skewness is a practical problem for the beam extracted by the AVF Cyclotron of Nishina Center as in Fig. 5 of Ref. [5].

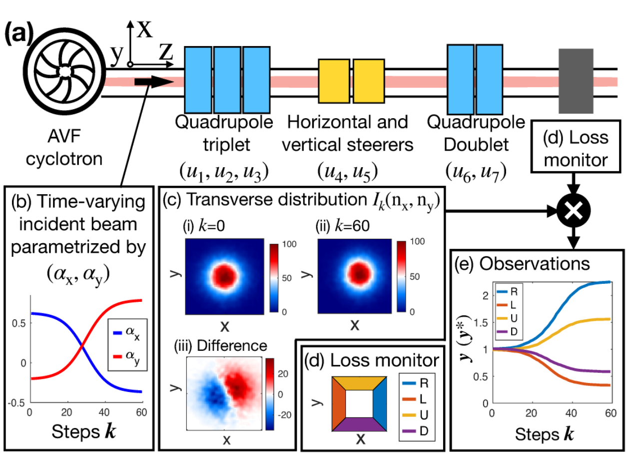

Figure 1: (a) Beam transport system's overview. (b) Time variation of hidden states $\alpha_{\mathrm{x,y}}$ over discrete time steps k. (c) Transverse distribution $I(n_{\mathrm{x}}, n_{\mathrm{y}})$ at the position of the beam loss monitors (d). (e) The number of incoming particles for each loss monitor relative to the initial observation $\boldsymbol{y}^*$.

The observation function $h(\boldsymbol{x},\boldsymbol{u}) \in \mathbb{R}^4$ is defined to sum up the transverse two-dimensional histogram $I(n_{\mathrm{x}}, n_{\mathrm{y}}) \in \mathbb{R}^{52\times52}$ as in Fig. 1(c) overlapped by the beam loss monitors on each side (R: Right, L: Left, U: Up, D: Down) as in Fig. 1(d), where $n_{\mathrm{x}} \in \mathbb{N}$ and $n_{\mathrm{y}} \in \mathbb{N}$ are indexes of pixels which evenly discretize the field of view at the beam loss monitor location.

As summarized in Fig. 1, we assume the skewness parameters of the incident beam before the transport slowly vary like a sigmoid function over 60 discrete time steps as in Fig. 1(b). The transverse beam distributions at $k = 0$ (Fig. 1(c-i)) and at $k = 60$ (Fig. 1(c-ii)) differ as shown in Fig. 1(c-iii) as a result of these time-varying skewness parameter. The subset Fig. 1(e) describes the time evolution of the observations by loss monitor Fig. 1(d) at each side relative to the optimized value y_j^* at $k = 0$.

Following the similar procedure in [8][9], the corresponding simulation-based surrogate model is trained using Convolutional Neural Network to output transverse distribution estimate $\hat{I}(n_{\mathrm{x}}, n_{\mathrm{y}}) \in \mathbb{R}^{52\times52}$ as a function of the excitation currents and hidden parameters of the incident beam distribution $\alpha_{\mathrm{x,y}} \in [-3, 3]$. The average reconstruction error of the predicted transverse distribution $\hat{I}(n_{\mathrm{x}}, n_{\mathrm{y}})$ defined by the equation below resulted in around 7 % for the test data.

$$Error = \sum_{n_{\mathrm{x}},n_{\mathrm{y}}} |\hat{I}(n_{\mathrm{x}}, n_{\mathrm{y}}) - I(n_{\mathrm{x}}, n_{\mathrm{y}})| / \sum_{n_{\mathrm{x}},n_{\mathrm{y}}} |I(n_{\mathrm{x}}, n_{\mathrm{y}})|$$

We take into account the input bias in surrogate modeling, e.g., i-th component of the input is assumed to be

$$\hat{u}_i = (1 + \delta_i)u_i, \tag{12}$$

where δ_i is the hidden state representing the scaling error.

In the later sections, the nine-dimensional vector $\hat{\boldsymbol{x}} = [\delta_1,...,\delta_m,\ \alpha_{\mathrm{x}},\ \alpha_{\mathrm{y}}]^T$ is considered as the hidden states to be updated by sequential Monte-Carlo (SMC). By computing the likely hidden states $\hat{\boldsymbol{x}}$ from measurements $\boldsymbol{y}$ with the ABC framework, surrogate-based reconstruction of the transverse beam distribution can be made even if invasive diagnostic devices such as wire scanners or scintillators are not insertable remotely due to insufficient room for installation or because of their damage threshold.

Process Variance for Hidden States Tracking

Changes in the hidden parameters are updated by conditioning the probability distribution of states, which is spread by artificial process noise, with the observation at each step k. For simplicity of calculation, a Gaussian uncertainty with a standard deviation σ_{vi} up to 5 % of the maximum i-th training excitation current $\boldsymbol{U}$ is considered for δ_i in Eq. (12); $\sigma_{vi} = 0.05 \max U_i$. In the sequential filtering process, $\Delta\hat{\boldsymbol{u}} \sim \mathcal{N}(\Delta\boldsymbol{u}, \gamma\boldsymbol{\Sigma}_v)$ is added to the surrogate model input for each move $\Delta\boldsymbol{u}$ on the actual system. Here, γ is the squared average of the control input norm for each move relative to $\max U_i$ expressed as $\gamma \equiv \frac{1}{m}\sum_i^m \frac{\Delta u_i^2}{\max U_i^2}$, and $\boldsymbol{\Sigma}_v = \mathrm{diag}\,\boldsymbol{\sigma}_v^2$ where $\boldsymbol{\sigma}_v^2 = [\sigma_{v1}^2, ..., \sigma_{vm}^2]$.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-MOPOPT034

On the other hand, the process variances of the skewness parameters $\alpha_{\mathrm{x,y}}$ are determined from the variances of the resultant MCMC particle distribution. The MCMC is initialized first by finding $\hat{x}$ which minimizes the cost Eq. (5) while the initial guesses $\alpha_{\mathrm{x,y}} = 0$ are used for the skewness parameters, and then subsequent proposal moves are accepted for the criterion conditioned by Eq. (8).

Measurement Based Control while Tracking the Hidden State Space for the Underspecified Model

In this section, we aim at matching the observation $\boldsymbol{y}$ to the target observation $\boldsymbol{y}^* \in \mathbb{R}^d$ while tracking the hidden states with SMC. For the weighted sum of the observations on the surrogate model $\hat{\boldsymbol{y}}_k = \sum_s W_k^s \hat{\boldsymbol{y}}_k^s$, the optimal input at time $k+1$ can be determined by the following two steps

$$\boldsymbol{u}^* = \arg\min_{u} \|(\boldsymbol{y}_k - \boldsymbol{y}^*) - (\hat{\boldsymbol{y}}_k - \hat{\boldsymbol{y}}_{k+1})\|_2^2 \quad (13)$$

$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \lambda(\boldsymbol{u}^* - \boldsymbol{u}_k). \quad (14)$$

Please note that if $\hat{\boldsymbol{y}}_k$ is equal to $\boldsymbol{y}_k$, $\hat{\boldsymbol{y}}_{k+1}$ is made close to $\boldsymbol{y}^*$. Here, the exponential convergence factor λ $(0 < \lambda \leq 1)$ is introduced to regulate the input move within the proximity of the input state $\boldsymbol{u}_k$. In this study, the optimization problem Eq. (13) is solved using the simplex method [10] for the measurement estimates $\hat{\boldsymbol{y}}$ inferred by the surrogate model with tracked hidden states $\hat{\boldsymbol{x}}_k = \sum_s W_k^s \hat{\boldsymbol{x}}_k^s$ with $s = 5,000$. Here, the standard deviation of the Gaussian noise ϵ_j is set to 5 % relative to y_j^*, while the measurement uncertainty σ_j is designed to be 10 % of the desired values for each output y_j^*. To demonstrate the robust behavior of the proposed scheme to underspecified model, the beam current is gradually enhanced by 10 % with a factor which varies like a sigmoid function similar to $\alpha_{\mathrm{x,y}}$. In Fig. 2, the control strategy Eq. (14) for a randomly picked input state $\boldsymbol{u}_0$ is demonstrated for (a) $\lambda = 1/3$ and (b) $\lambda = 1/2$. Figure 2(a) demonstrates the successful stabilization of the time-varying system only with a few samples around the optimal $\boldsymbol{y}^*$ for the noisy sensor measurement, while a slightly delayed response is observed for the higher rate of change in hidden states $|\alpha_{k+1} - \alpha_k|$. Figure 2(b) suggests that large λ can lead to unstable response because too large input move toward optimal $\boldsymbol{u}^*$ can promote quick degeneracy of sample particles.

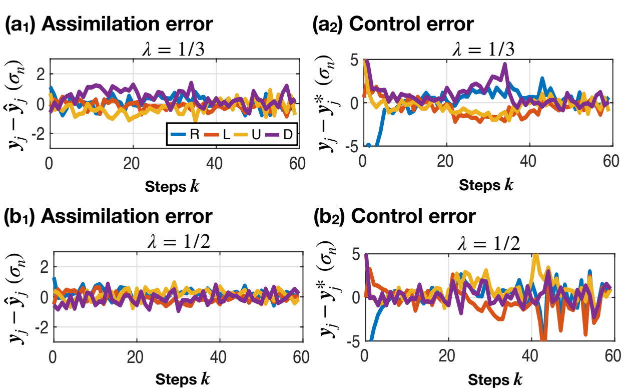

Figure 2: Results of control attempts toward y_j^* $(j \in R, L, U, D)$ by Eq. (13) for (a) $\lambda = 1/3$ and (b) $\lambda = 1/2$.

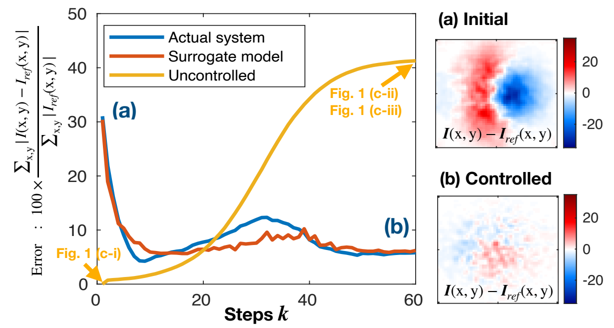

Figure 3: Left: The blue line shows the relative error of $I(n_{\mathrm{x}}, n_{\mathrm{y}})$ when the model's state is controlled based on the distribution $\hat{I}(n_{\mathrm{x}}, n_{\mathrm{y}})$ represented by the red line toward $I_{ref}(n_{\mathrm{x}}, n_{\mathrm{y}})$. The yellow line is the uncontrolled reference when the state evolves as in Fig. 1. Right: Error in the transverse distribution $I(n_{\mathrm{x}}, n_{\mathrm{y}})$ at (a) $k = 0$ and (b) $k = 60$.

Inference Based Control Strategy for the Transverse Beam Distribution

Instead of the control strategy based on observable $\boldsymbol{y}$, we design the control law as to minimize $\sum_{n_{\mathrm{x}},n_{\mathrm{y}}} |\hat{I}_{k+1}(n_{\mathrm{x}}, n_{\mathrm{y}}) - I_{ref}(n_{\mathrm{x}}, n_{\mathrm{y}})|^2$ by replacing the cost function in Eq. (13), where the reference distribution $I_{ref}(n_{\mathrm{x}}, n_{\mathrm{y}})$ is defined to be the one shown in Fig. 1(c-i). The left window in Fig. 3 illustrates the responses of the actual system and surrogate model to this control strategy beginning at a randomly picked control state $\boldsymbol{u}_0$ under the time-varying hidden state as in Fig. 1 (b). In this scenario, the error of the actual system (blue) and surrogate model (red) are close throughout the state space tracking by the particle filter, and the distribution error is minimized as shown in the right panes of Fig. 3. As a result, the error at the final state (at $k = 60$) is much less than the uncontrolled case (yellow) whose distribution began by $I_{ref}(n_{\mathrm{x}}, n_{\mathrm{y}})$.

CONCLUSION

We checked the efficacy of the similarity-based hidden state tracking and control using SMC on a simulation-based surrogate model. The method has proven its sample-efficient and robust behavior for uncertainties and under-specifications. The inference-based optimization is demonstrated for the beam distribution estimate of the assimilated surrogate model. Future directions would be the high-dimensional extension of the SMC algorithms potentially with Ref. [11] and [12], error compensation of the surrogate model itself, and global tracking of multiple plausible clusters of possible states. The theoretical bound on the exponential convergence factor λ should be also investigated.

ACKNOWLEDGMENTS

The authors gratefully acknowledge valuable discussions with Hiroyasu Tsukamoto.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-MOPOPT034

REFERENCES

[1] M. A. Beaumont, "Approximate bayesian computation," *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 379–403, 2019. doi:10.1146/annurev-statistics-030718-105212

[2] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré, "Markov chain Monte Carlo without likelihoods," *Proceedings of the National Academy of Sciences*, vol. 100, no. 26, pp. 15 324–15 328, Dec. 2003. doi:10.1073/pnas.0306899100

[3] M. Baragatti, A. Grimaud, and D. Pommeret, "Likelihood-free parallel tempering," *Statistics and Computing*, vol. 23, no. 4, pp. 535–549, Jul. 2013. doi:10.1007/s11222-012-9328-6. 1108.3423

[4] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of Nonlinear Filtering*, pp. 4–6, Dec. 2009.

[5] H. Fujii, A. Scheinker, A. Uchiyama, S. J. Gessner, O. Kamigaito, and N. Fukunishi, "Extremum Seeking Control For The Optimization Of Heavy Ion Beam Transportation," *Proceedings of the 18th Annual Meeting of Particle Accelerator Society of Japan*, vol. WEOB01, pp. 146–150, Aug. 2021.

[6] A. O'Hagan and T. Leonard, "Bayes estimation subject to uncertainty about parameter constraints," *Biometrika*, vol. 63, no. 1, pp. 201–203, 1976. doi:10.2307/2335105

[7] G. S. Mudholkar and A. D. Hutson, "The epsilon–skew–normal distribution for analyzing near-normal data," *Journal of Statistical Planning and Inference*, vol. 83, no. 2, pp. 291–309, 2000. doi:10.1016/S0378-3758(99)00096-8

[8] A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, "Adaptive Latent Space Tuning for Non-Stationary Distributions," May 2021. arXiv: 2105.03584. doi:10.48550/arXiv.2105.03584

[9] A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, "An adaptive approach to machine learning for compact particle accelerators.," *Scientific reports*, vol. 11, no. 1, p. 19 187, Sep. 2021, ISSN: 2045-2322. doi:10.1038/s41598-021-98785-0

[10] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, Jan. 1965. doi:10.1093/comjnl/7.4.308

[11] J. Poterjoy, "A localized particle filter for high-dimensional nonlinear systems," *Monthly Weather Review*, vol. 144, no. 1, pp. 59–76, Jan. 2016. doi:10.1175/MWR-D-15-0163.1

[12] F. Septier and G. W. Peters, "Langevin and Hamiltonian Based Sequential MCMC for Efficient Bayesian Filtering in High-Dimensional Spaces," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 312–327, Mar. 2016. doi:10.1109/JSTSP.2015.2497211