ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB289

TOWARDS HYSTERESIS AWARE BAYESIAN REGRESSION AND OPTIMIZATION *

R. Roussel†, University of Chicago, Chicago, USA
A. Hanuka, SLAC National Laboratory, Menlo Park, USA

Abstract

Hysteresis effects in magnetic elements are often approximated by simple linear relationships between applied current and magnetization when conducting accelerator operations. However, with increasingly difficult operational targets for the next generation of accelerators, deviations from this approximation due to hysteresis cannot be neglected. Furthermore, experimental models used in accelerator control algorithms must also take hysteresis effects into account when predicting beam properties. Here we present initial work on incorporating hysteresis effects into Bayesian statistical models that eventually will be applied to online accelerator optimization. We treat the hysteresis cycle of accelerator elements as a latent, unobserved variable that is learned via Bayesian techniques using observations of the beam response. As a result, the hysteresis behavior of any magnetic element can be determined without directly measuring the magnetic field. The methodology used here can be extended to modeling any system that has hysteresis-like behavior, like backlash in mechanical actuators or forces exerted by elastic components.

INTRODUCTION

Hysteresis is the process by which the magnetization of magnetic elements exhibit non-linear behavior in response to external magnetic fields [1]. The magnetization depends not only the external magnetic field created by current carrying wires surrounding a magnetizable material, but also on the history of external fields. As a result, if the external magnetic field is cycled, the magnetization will not return to the original state. This complicates the process of modeling the beam response as a function of currents applied to magnetic accelerator elements as the beam behavior is now dependant not only on the current set point, but also on historical parameter states as well. Accelerator operators generally avoid these effects by staying in a limited parameter regime, where the magnetic response can be approximated by a linear fit that neglects errors due to previous parameter states. As beam quality and control targets become more challenging, errors due to hysteresis effects can no longer be ignored. Furthermore, if we wish to allow algorithms to assist in optimizing accelerator parameters, the models these algorithms use need to jointly learn the hysteresis behavior as well as the beam response.

One such method for surrogate-assisted beamline optimization is Bayesian optimization (BO) [2]. This black-box optimization algorithm uses a Bayesian statistical model of a target function (such as the transverse rms. beam size), to predicts the mean and uncertainty of any future measurement. The model is combined with an acquisition function, that predicts the value gained from any particular observation. A simple acquisition function for optimization is Upper Confidence Bound [3] which strategically chooses observations that balance the trade-off between taking advantage of previously observed extrema (exploitation) and reducing the uncertainty of the overall model (exploration). However, normal Bayesian optimization frameworks cannot account for systems that exhibit hysteresis behavior as they assume a one-to-one relationship between input parameters and output observations. A potential way to circumvent this limitation is to directly measure the magnetization of each accelerator element *in situ* via the nearby magnetic field. However, this can be prohibitively expensive to do for large numbers of elements.

Here we present a joint Bayesian model that represents both the hysteresis process as well as the beam response to the resulting magnetic fields. This is achieved without measuring the magnetization directly. Instead we measure the beam properties downstream of the magnetic element and estimate the hysteresis model parameters by maximizing the log likelihood function.

JOINT HYSTERESIS-BAYESIAN MODELS

The joint statistical model for a single magnetic element is described as follows. The first part of the model maps the external field at some time step H_t and the history of external fields $H_{0:t-1}$ to the resulting bulk magnetization M_t. The second part of the model contains a Gaussian process [4] Bayesian model than maps the magnetization to a beam property, which for simplicity, we choose to be the transverse beam size σ_x. Both of these sub-models contain free parameters that we estimate by maximizing the marginal log likelihood of the data set containing our observable measurements $\mathcal{G}_{0:t} = \{(H_0, \sigma_0), (H_1, \sigma_1), \dots, (H_t, \sigma_t)\}$.

The magnetization at a given time step is modeled via the Preisach model of hysteresis [5]. This model describes the total magnetization as a sum of many "hysterions", which can have values of either +1 or -1 as shown in Fig. 1(a), corresponding to alignment or anti-alignment with the external field respectively. Each hysterion has two associated properties, β which describes the external positive field magnitude required to flip the hysterion from the -1 state to +1 and α which describes the external negative field that is required to do the opposite. The total magnetization is given by a sum

* This work was supported by the U.S. National Science Foundation under Award No. PHY-1549132, the Center for Bright Beams.
† rroussel@uchicago.edu

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB289

of i'th elements

$$M(\mathbf{H}_{0:t}) = \sum_i m_i(\mathbf{H}_{0:t}; \beta_i, \alpha_i). \qquad (1)$$

We can extend this sum of discrete hysterions to a continuum, where the number of hysterions with a given (β, α) is defined by $\mu(\beta, \alpha)$, shown in Fig. 1(b). The hysterion density is non-zero in a triangular region bounded by the line $\alpha = \beta$ as $\alpha < \beta$ is non-physical, and two limits that define the fields required for positive $\alpha = \alpha_m$ and negative $\beta = \beta_m$ saturation of the bulk material. To calculate the magnetization as a function of external field we split the valid domain into two sub-domains $S+$ and $S-$ which corresponds to the hysterions that have a $+1$ and -1 value respectively. The equation for the magnetization thus becomes

$$M(\mathbf{H}_{0:t}) = \int_{S+} \mu(\beta, \alpha) d\beta d\alpha - \int_{S-} \mu(\beta, \alpha) d\beta d\alpha. \qquad (2)$$

An iterative procedure is used to determine the regions $S+, S-$ based on the historical record of external fields $\mathbf{H}_{0:t}$. We assume that $H_0 < \beta_m$, such that the $S-$ sub-domain is the entire valid region. If the external magnetic field is increased ($\Delta H_s = H_s - H_{s-1} > 0$, where $s = 1, 2, \dots t$) then a horizontal line is swept upwards to the final external magnetic field strength. Any regions below this horizontal line are added to the $S+$ sub-domain and removed from the $S-$ sub-domain. If the external magnetic field is then decreased (ΔH_s), a vertical line is swept from right to left to the final external field value, flipping any regions to the right of this line back to the $S-$ domain. This procedure is repeated for each intermediate step s until $s = t$, resulting in sub-domains similar to those shown in Fig. 1(c). With these sub-domains defined, the remaining task is to determine the hysterion density $\mu(\beta, \alpha)$. In this case we use a bivariate Gaussian density function with a diagonal sigma matrix $\mathbf{\Sigma} = \mathrm{diag}(l_1, l_2)$ and centered at the origin. The parameters l_1, l_2 will be estimated inside the joint model.

To describe the beam response as a function of the magnetization $f(M)$ we use a non-parametric Gaussian process model that is specified by a covariance function, $k(M, M'; \theta)$ with hyperparameters θ and a zero mean function so that we can write $f(M) \sim \mathcal{GP}(0, k(M, M'))$. In an experimental setting the observed beam response y is corrupted by noise: $y = f(M) + \epsilon$ where we assume that $\epsilon \sim \mathcal{N}(0, \sigma^2_{noise})$. Given previous measurements $\mathcal{D}_{0:t} = \{(M_0, y_0), (M_1, y_1), \dots, (M_t, y_t)\}$ the predictive probability distribution for the function value $f_{t+1} = f(M_{t+1})$ is given by

$$P(f_{t+1}|\mathcal{D}_{0:t}, M_{t+1}) = \mathcal{N}(\mu(M_{t+1}), \sigma^2(M_{t+1})), \qquad (3)$$

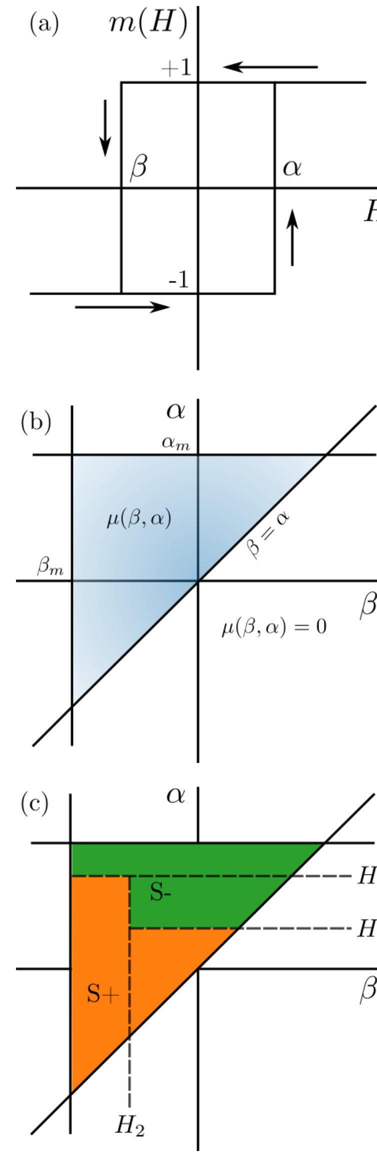

Figure 1: (Color online) Cartoon showing the elements of the Preisach hysteresis model. (a) Function of a single hysterion. (b) Hysterion density on the $\beta - \alpha$ plane. (c) $S+$ and $S-$ sub-domains after three time steps, where $H_1 > H_3 > H_2 > \beta_m$.

where

$$\mu(M_{t+1}) = \mathbf{k}^T \left[K + \sigma^2_{noise} I\right]^{-1} \mathbf{y}_{0:t} \qquad (4)$$

$$\sigma(M_{t+1}) = k(M_{t+1}, M_{t+1}) - \mathbf{k}^T \left[K + \sigma^2_{noise} I\right]^{-1} \mathbf{k} \qquad (5)$$

$$\mathbf{k} = \left[k(M_{t+1}, M_0), \dots, k(M_{t+1}, M_t)\right]^T \qquad (6)$$

$$K = \begin{bmatrix} k(M_0, M_0) & \cdots & k(M_0, M_t) \\ \vdots & \ddots & \vdots \\ k(M_t, M_0) & \cdots & k(M_t, M_t) \end{bmatrix}. \qquad (7)$$

The hyperparameters are determined by maximizing the log marginal likelihood, $\theta, \sigma_{noise} = \mathrm{argmax}_{\theta, \sigma_{noise}} \log[p(D_{0:t}; \theta, \sigma_{noise})]$, which balances model accuracy and complexity when choosing hyperparameters. However, since the magnetization $\mathbf{M}_{0:t}$ is unknown,

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB289

we also need to maximize the marginal log likelihood with respect to the hysteresis density parameters conditioned on the observable data set $\mathcal{G}_{0,t}$. As a result, our goal is to find

$$\mathrm{argmax}_{\theta,\sigma_{noise},l_1,l_2} \log[p(G_{0:t};\theta,\sigma_{noise},l_1,l_2)]. \quad (8)$$

We implemented the Preisach model in PyTorch [6] and the Gaussian process model in GPyTorch [7] which both have auto-gradient functionality, which tracks derivative information during calculations, and allows gradient descent based algorithms to be used to optimize the marginal log-likelihood.

TOY MODEL RESULTS

We evaluated the effectiveness of this modeling strategy by using it to fit a toy hysteresis model. We created a magnet model with the hysterion density $\mu(\beta,\alpha)$ given by a multivariate Gaussian centered at the origin and with a covariance matrix given by $\mathbf{\Sigma} = \mathrm{diag}((0.25, 1.5))$. The positive and negative magnetization saturation levels were set to $+1$ and -1 respectively. We specified external magnetic field strengths over $t = 20$ steps given by $H_t = -1.05\cos(2\pi t/25)$ and recorded the magnetization M_t. The beam size response to the magnetization was specified to be $f(M) = (M - 0.25)^2$ and the noise parameter was set to $\sigma_{noise} = 0.1$. We used a radial basis kernel (RBF) which contained the hyperparameters $\theta = \{A, \lambda\}$ where A is the scale factor and λ is the length scale. Results from this calculation are shown in Figs. 2(a) and (b). We then used the external field and beam size data sets to train the joint model. The Adam gradient descent algorithm was used to maximize the marginal log likelihood with respect to the hysteresis and Gaussian process parameters. The estimated hysterion density had a covariance matrix of $\mathbf{\Sigma} = \mathrm{diag}((0.29, 1.3))$ which closely matches the ground truth. The Gaussian process kernel function had a length scale $\lambda = 1.67$ and a scale factor of $A = 3.1$. Predictions from this model are shown in Fig. 2.

From these results, we see that the hybrid model shows reasonable accuracy when jointly modeling both the hysteresis process and the beam response. Comparisons to the ground truth of both the magnetization (Fig. 2(a)) and the beam response (Fig. 2(c)) as a function of the next external field show good agreement. This includes non-smooth behavior as seen in Fig. 2(c), since we explicitly account for hysteresis effects in our model. We also see that the joint model accurately predicts the shift in external field H_{t+1} that results in a minimum beam size, relative to when hysteresis effects are ignored (“hysteresis off case”).

CONCLUSION

Here we have introduced a method for simultaneously fitting the hysteresis behavior of a magnetic accelerator element and the beam response to that magnet using a joint Gaussian process based model. Next steps for this work are to experimentally demonstrate this measurement and also to integrate this method into an existing accelerator control system for optimization.

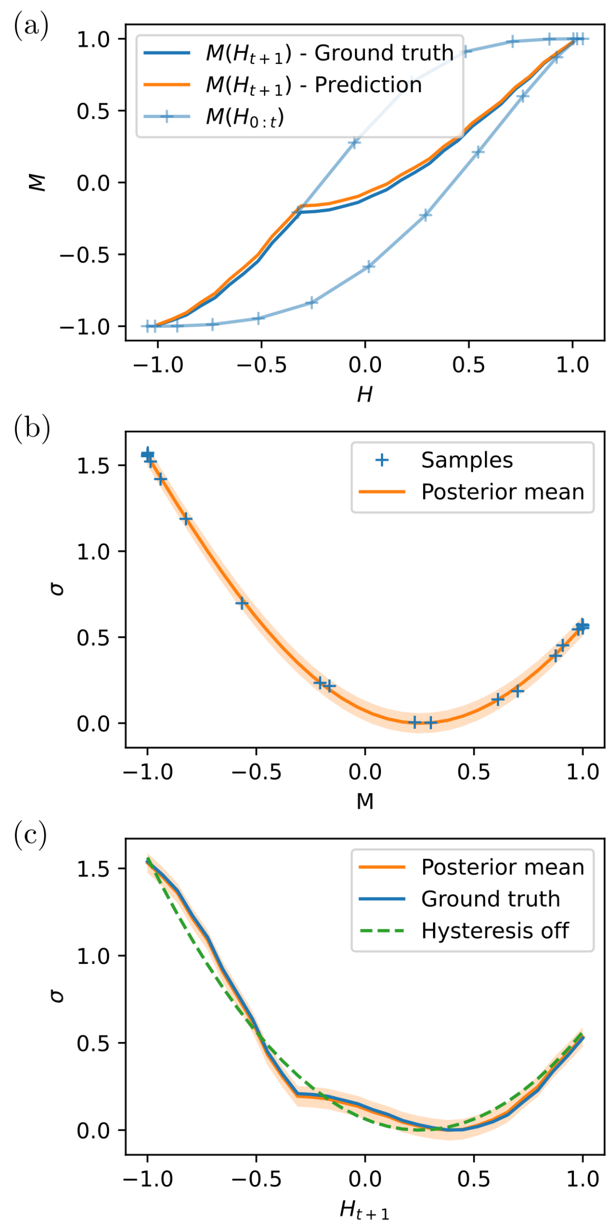

Figure 2: (Color online) Analysis of toy hysteresis model. (a) The historical hysteresis curve of the toy model over 20 steps $M(H_{0:t})$ along with the ground truth and predicted magnetization as a function of the next external field H_{t+1}. (b) Latent space representation of beam size as a function of magnetization showing historical samples and model prediction. (c) Beam size as a function of next external field, H_{t+1} comparing model predictions to the ground truth as well as the case where hysteresis effects are neglected ($M = H_{t+1}$).

ACKNOWLEDGEMENTS

We would like to thank Micol Tresoldi, Benjamin Harris and Young-Joo Yun from the University of Chicago Department of Statistics for their consulting work on this project.

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB289

REFERENCES

[1] G. Bertotti and I. D. Mayergoyz, The science of hysteresis, Boston, USA: Academic Press, 2005.

[2] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms", in *Proc. NIPS'12*, Nevada, USA, Dec. 2012, pp. 1-8.

[3] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: no regret and experimental design", in *Proc. ICML'10*, Madison, USA, Jun. 2010, pp. 1015–1022.

[4] C. E. Rasmussen and C. K. I. Williams, "Gaussian processes for machine learning", Cambridge, USA: MIT Press, 2006.

[5] F. Preisach, "Uber die magnetische Nachwirkung", *Zeitschrift fur Physik*, vol. 94, pp. 277–302, May 1935. doi:10.1007/BF01349418

[6] A. Paszke *et al.*, "PyTorch: An Imperative Style", in *Proc. NIPS'19*, Vancouver, Canada, Dec. 2019, pp. 8024–8035.

[7] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration", in *Proc. NIPS'18*, Montreal, Canada, Dec. 2018, pp. 1-11.