Proximal Biasing for Bayesian Optimization and Characterization of Physical Systems

Ryan Roussel
SLAC National Accelerator Laboratory
Menlo Park, CA 94025
rroussel@slac.stanford.edu

Auralee Edelen
SLAC National Accelerator Laboratory
Menlo Park, CA 94025
edelen@slac.stanford.edu

Abstract

Bayesian techniques have been shown to be extremely efficient in optimizing expensive to evaluate black box functions, in both computational (offline design) and physical (online experimental control) contexts. Optimizing physical systems often comes with extra challenges due to costs associated with changing parameters in real life experimentation, such as measurement location in physical space or mechanical/electrical actuation in scientific equipment. In these cases, the cost of changing a given input parameter is often proportional to the magnitude of the change. To minimize these costs, optimization algorithms can simply limit the maximum distance travelled in input space during each step. However, hard restrictions on travel distances inhibits global exploration advantages normally afforded by Bayesian optimization algorithms. In this work, we describe a proximal weighting term that can bias acquisition functions towards localized exploration, while still allowing for large travel distances, if far away points are predicted to be valuable for observation. We describe a use case where this weighting is used to minimize the uncertainty of a particle accelerator Bayesian model in a smooth manner, which in turn, minimizes temporal costs associated with changing input parameters.

1 Introduction

The efficient global optimization of black box problems is a challenge faced by many who work in the physical sciences [1–3]. Bayesian optimization (BO) [4, 5] has been shown to perform well in solving physical [6–9] optimization problems for which gradient information is unavailable, while naturally treating statistical noise inherent in most physical scenarios. In these cases, a statistical model, normally a Gaussian process (GP) [10], is combined with an acquisition function that determines the best set of input parameters to observe next. There are a wide variety of acquisition functions which can serve several purposes, including global optimization [11, 12] and characterization through uncertainty quantification [13].

However, most if not all acquisition functions ignore costs associated with varying input parameters in physical systems. For example, if we want to determine the location of maximum concentration of chemical contaminants in a wide area [14], we must expend resources in terms of time or effort to sample separate locations in input space. Widely used acquisition functions such as Expected Improvement (EI) [15] or Upper Confidence Bound (UCB) [16] do not take these costs into account, resulting in large parameter changes during each optimization step to maximize expected value from future observations of the target function. This degrades performance when the cost of traversing input space is significant, relative to the cost of making observations.

Furthermore, rapid changes in input parameters are unacceptable when operating complex experimental equipment during experiments. For example, particle accelerators that are used to study

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

chemical [17], physical [18], or biological [19] systems are extremely complex machines that rely on the interconnection between many subsystems. Individual accelerator subsystems are often autonomously controlled by internal feedback loops that maintain performance or prevent damage to components in response to external changes. If external changes swing widely from one extreme to another, these internal feedback loops can take a long time to regain stability or in the worst case, can fail completely. As a result, it is critical that controlled parameters are varied in a adiabatic manner.

Naive methods for encouraging this behavior in Bayesian optimization removes the exploration aspect of BO necessary for global optimization. For example, we could choose the exploration-exploitation trade-off parameter of EI or UCB [20] such that it heavily weights exploitation. Alternatively, we could simply restrict the maximum travel distance that the optimizer can travel in each step by zeroing the acquisition function outside a given radius from the current location. These methods are not ideal, as they remove the exploration aspect of BO, potentially causing the algorithm to get stuck in local extrema.

To solve this problem, we propose including a squared exponential biasing factor in acquisition functions used for Bayesian optimization, which minimizes the costs of traversing input space while still allowing for exploration. We provide a description of our biasing function along with an instructive synthetic example. Then we describe how we used this modification to efficiently characterize a physical accelerator during online operations where changing parameters incurs a temporal cost similar in magnitude to the measurement cost.

2 Proximal Biasing

We bias the acquisition function of a Bayesian optimization loop by simply multiplying the original acquisition function $\alpha(\mathbf{x})$ by a biasing term $\Psi(\mathbf{x}, \mathbf{x_0}; \Sigma)$. We choose the biasing term to be a squared exponential function giving the biased acquisition function

$$\tilde{\alpha}(\mathbf{x}) = \alpha(\mathbf{x})\Psi(\mathbf{x}, \mathbf{x_0}; \Sigma) \tag{1}$$

$$\Psi(\mathbf{x}, \mathbf{x_0}; \Sigma) = \exp\Big(-\frac{1}{2}(\mathbf{x} - \mathbf{x_0})^T \Sigma^{-1} (\mathbf{x} - \mathbf{x_0})\Big) \tag{2}$$

where $\mathbf{x_0}$ is the most recently observed point and $\Sigma = \mathrm{diag}(\lambda)$ is a diagonal covariance matrix chosen prior to optimization. Each element of the covariance matrix describes how strongly biased the algorithm is towards nearby points. In the general framework of cost-aware Bayesian optimization [21] this term can be interpreted as an inverse cost function $c(\mathbf{x}, \mathbf{x}') = 1/\Psi(\mathbf{x}, \mathbf{x_0}; \Sigma)$. Through this framework we can apply this cost function to any previously defined acquisition function used in serialized optimization, including those expressed analytically or calculated numerically using Monte Carlo estimation.

A demonstration of the biasing term's effect on optimization is shown in Figure 1. Our goal is to characterize the first objective function used in the constrained TNK multi-objective optimization test problem [22]. To achieve this, we use the following active learning-based acquisition function

$$\alpha(\mathbf{x}) = \sigma(\mathbf{x}) \prod_{i=1}^{N} P_i[g_i(\mathbf{x}) \geq h_i] \tag{3}$$

where $P_i[g_i(\mathbf{x}) \geq h_i]$ represents the probability of satisfying the i'th problem constraint given by $g_i(\mathbf{x}) \geq h_i$. We run optimization trials with and without proximal biasing, starting with a single initialized point and iterating for 25 steps, retraining the GP model hyperparameters at each step using marginal log likelihood maximization.

Figure 1(a) shows the path taken by the optimizer without proximal biasing in input space during optimization. The optimizer chooses points that aggressively reduce the uncertainty in the GP model while attempting to remain in the valid region (red shading), resulting in a chaotic sampling pattern. We also observe that the optimizer makes several large excursions outside of the valid region to explore limits of the input domain.

Figure 1(b) shows optimization of the same problem with the addition of an isotropic proximal biasing term with $\lambda = 0.01$. In this case we clearly see a smoothing of the path through input space and a reduction of total travel length over the optimization path. We repeated each optimization trial in Figure 1 with the same initial point 25 times and calculated the average total distance traveled

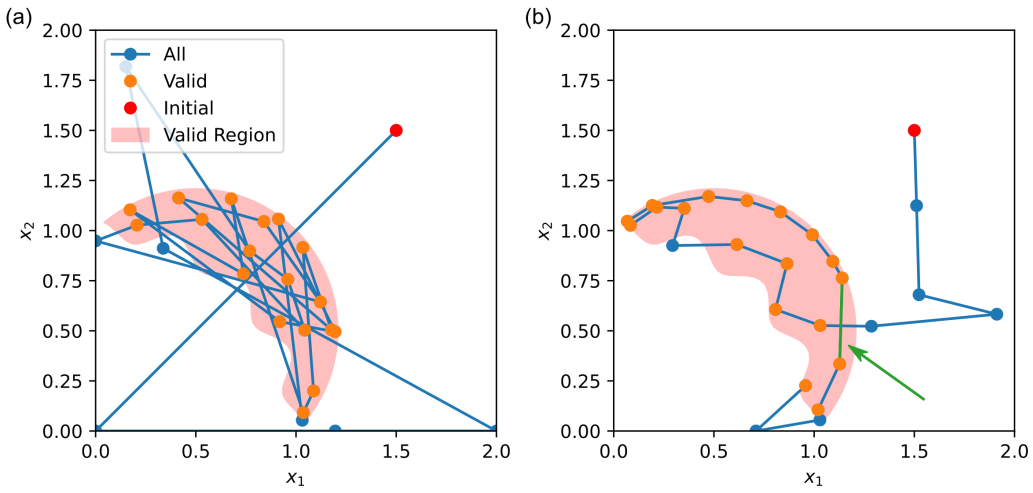

Figure 1: Plots showing Bayesian optimization sampling patterns with and without proximal biasing for the TNK test problem. Optimization begins with a single sample (red) and travels through 2D input space. Points that satisfy the TNK problem constraints (red shading) are labeled valid (orange). (a) Optimization path without proximal biasing. (b) Optimization path with proximal biasing where $\Sigma = 0.01\mathbf{I}$. The green arrow denotes an optimization step where the optimizer "jumps" over previous measurements.

by the optimizer for each. The unbiased optimizer travelled a distance of 20.1 ± 1.3 and the biased optimizer travelled a distance of 5.9 ± 0.5. Of particular interest is the optimization step denoted by the green line segment in Figure 1(b), where the optimizer "jumps" over previously observed points where the uncertainty is expected to be low. By biasing the acquisition function using a function that is non-zero everywhere, we enable the optimizer to make large jumps in input space when necessary, while reducing the travel distance during optimization.

3 Application to Characterizing Particle Accelerators

We used this technique to characterize the quality of beams produced by the Argonne Wakefield Accelerator facility at Argonne National Laboratory [23]. The figure of merit we use to describe beam quality is the beam's "emittance" [24]. Our goal is to measure the beam emittance as a function of four tunable parameters in the accelerator using the acquisition function defined in Eq. 3.

When conducting experiments at a real accelerator, it is important to reduce the amplitude of adjustments to accelerator parameters. Many electrical and mechanical subsystems in the accelerator are controlled by elements which trade adjustment speed for precision. As a result, in these subsystems the temporal cost of making substantial changes in control parameters is significant. Furthermore, since most subsystems in the accelerator are interconnected and rely on internal feedback algorithms to maintain performance, rapid changes in accelerator parameters can introduce instability issues into these systems [25], which can result in long periods of time to reach stability.

We explored the beam response with respect to four accelerator parameters. Two parameters were the currents applied to two solenoid magnets (referred to here as "focusing" and "matching") near the electron source. The other two parameters are the currents applied to two focusing magnets further down the beamline (referred to here as "DQ4" and "DQ5"). The measurement of beam emittance was done using a specialized diagnostic that has a small dynamic range [26]. As a result some regions of input space did not produce valid measurements. The validity of each measurement was used as a constraint function during characterization.

Figure 2(a) shows results from an characterization run with no proximal biasing. Each input parameter varies widely from step to step, accumulating cost associated with changing each parameter. We also observe that the fraction of valid measurements converges to approximately 0.55. Figure 2(b)

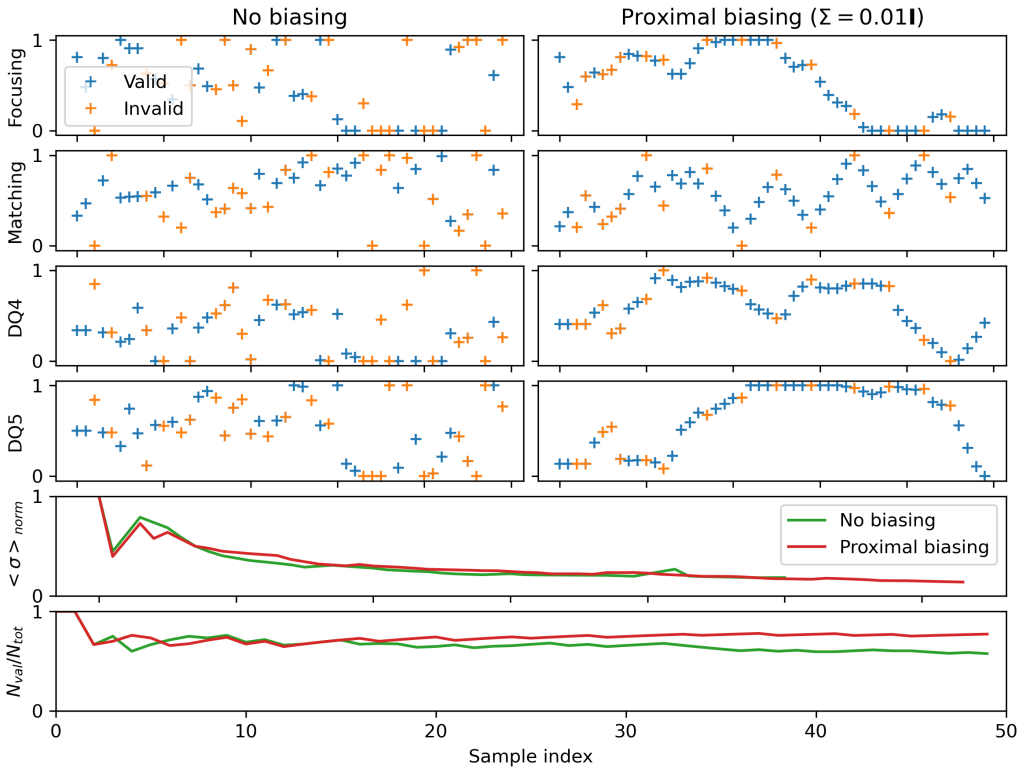

Figure 2: Experimental characterization progress as a function of sample index with and without proximal biasing. Top four rows trace normalized free parameter values. Fifth row plots the fractional mean uncertainty of the GP model relative to the initial uncertainty. Last row plots the fraction of valid measurements.

on the other hand shows that the proximal term ($\Sigma = 0.01\mathbf{I}$) significantly reduces the magnitude of adjustments to each input parameter. The travel distance in normalized parameter space with biasing (16.8 u) was significantly less than the travel distance without biasing (59.6 u), representing a 3.5 fold reduction.

We saw similar performance in minimizing our objective function for each trial, ie. the total model uncertainty, despite the smaller steps in parameter space when biasing is included. It is certainly possible that performance can be negatively impacted in other situations [9] so balancing the potential impacts of decreased convergence speed with reductions in observation costs is a requirement for ideal optimization.

Finally, we also observe that biasing increases the number of valid measurements during optimization, where the fraction of valid measurements converges to 0.77. We believe that this is due to the improved model confidence regarding constraints in a local region around previous measurements where the proximal biasing term is largest.

4 Conclusion

Here we have introduced and demonstrated a modification to classical Bayesian optimization that biases optimizers towards making small steps in input space, while preserving exploration advantages afforded by Bayesian optimization. By choosing a biasing function that never reaches zero, we allow the optimizer to choose large travel distances if the measurement value is large. Our modification to standard acquisition functions is especially important for optimization of physical systems like particle accelerators, since minimizing the costs of changing parameters is critical towards minimizing the overall cost of optimization. Code and data for this work can be found in the public repository `https://github.com/roussel-ryan/neurips_proximal`.

5 Acknowledgements

This work was supported by the U.S. Department of Energy, under DOE Contract No. DE-AC02-76SF00515 and the Office of Science, Office of Basic Energy Sciences

References

[1] Alexander Scheinker. Adaptive Machine Learning for Robust Diagnostics and Control of Time-Varying Particle Accelerator Components and Beams. *Information*, 12(4):161, April 2021. doi: 10.3390/info12040161. URL https://www.mdpi.com/2078-2489/12/4/161. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[2] E. A. Baltz, E. Trask, M. Binderbauer, M. Dikovsky, H. Gota, R. Mendoza, J. C. Platt, and P. F. Riley. Achievement of Sustained Net Plasma Heating in a Fusion Experiment with the Optometrist Algorithm. *Scientific Reports*, 7(1):6425, July 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-06645-7. URL https://www.nature.com/articles/s41598-017-06645-7. Number: 1 Publisher: Nature Publishing Group.

[3] Marcus M. Noack, Kevin G. Yager, Masafumi Fukuto, Gregory S. Doerk, Ruipeng Li, and James A. Sethian. A Kriging-Based Approach to Autonomous Experimentation with Applications to X-Ray Scattering. *Scientific Reports*, 9(1):11809, August 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-48114-3. URL https://www.nature.com/articles/s41598-019-48114-3. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Applied mathematics;Computational science;Design, synthesis and processing;Scientific data;Software Subject_term_id: applied-mathematics;computational-science;design-synthesis-and-processing;scientific-data;software.

[4] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1):148–175, January 2016. ISSN 1558-2256. doi: 10.1109/JPROC.2015.2494218.

[5] Stewart Greenhill, Santu Rana, Sunil Gupta, Pratibha Vellanki, and Svetha Venkatesh. Bayesian Optimization for Adaptive Experimental Design: A Review. *IEEE Access*, 8:13937–13948, 2020. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.2966228. Conference Name: IEEE Access.

[6] Mitchell McIntire, Tyler Cope, Stefano Ermon, Daniel Ratner, and others. Bayesian optimization of FEL performance at LCLS. In *Proceedings of the 7th International Particle Accelerator Conference*. IOP Publishing, 2016.

[7] Adi Hanuka, J Duris, J Shtalenkova, D Kennedy, A Edelen, D Ratner, and X Huang. Online tuning and light source control using a physics-informed Gaussian process. page 5, 2019.

[8] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, and D. Ratner. Bayesian Optimization of a Free-Electron Laser. *Physical Review Letters*, 124(12), 2020. ISSN 0031-9007. doi: 10.1103/physrevlett.124.124801. URL http://arxiv.org/abs/1909.05963.

[9] Ryan Roussel, Adi Hanuka, and Auralee Edelen. Multiobjective Bayesian optimization for online accelerator tuning. *Physical Review Accelerators and Beams*, 24(6):062801, June 2021. doi: 10.1103/PhysRevAccelBeams.24.062801. URL https://link.aps.org/doi/10.1103/PhysRevAccelBeams.24.062801. Publisher: American Physical Society.

[10] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006. ISBN 978-0-262-18253-9. OCLC: ocm61285753.

[11] Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. The application of Bayesian methods for seeking the extremum. *Towards global optimization*, 2(117-129):2, 1978.

[12] Harold J Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. 1964.

[13] David J. C. MacKay. Information-Based Objective Functions for Active Data Selection. *Neural Computation*, 4(4):590–604, July 1992. ISSN 0899-7667. doi: 10.1162/neco.1992.4.4.590. URL https://doi.org/10.1162/neco.1992.4.4.590.

[14] Guillaume Pirot, Tipaluck Krityakierne, David Ginsbourger, and Philippe Renard. Contaminant source localization via Bayesian global optimization. *Hydrology and Earth System Sciences*, 23 (1):351–369, January 2019. ISSN 1027-5606. doi: 10.5194/hess-23-351-2019. URL https://hess.copernicus.org/articles/23/351/2019/. Publisher: Copernicus GmbH.

[15] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, December 1998. ISSN 1573-2916. doi: 10.1023/A:1008306431147. URL https://doi.org/10.1023/A:1008306431147.

[16] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 1015–1022, Madison, WI, USA, June 2010. Omnipress. ISBN 978-1-60558-907-7.

[17] Jacques-Philippe Colletier, Michael R. Sawaya, Mari Gingery, Jose A. Rodriguez, Duilio Cascio, Aaron S. Brewster, Tara Michels-Clark, Robert H. Hice, Nicolas Coquelle, Sébastien Boutet, Garth J. Williams, Marc Messerschmidt, Daniel P. DePonte, Raymond G. Sierra, Hartawan Laksmono, Jason E. Koglin, Mark S. Hunter, Hyun-Woo Park, Monarin Uervirojnangkoorn, Dennis K. Bideshi, Axel T. Brunger, Brian A. Federici, Nicholas K. Sauter, and David S. Eisenberg. De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature*, 539(7627):43–47, 2016. doi: 10.1038/nature19825. URL https://doi.org/10.1038/nature19825. ISBN: 1476-4687.

[18] M. P. Jiang, M. Trigo, I. Savić, S. Fahy, É. D. Murray, C. Bray, J. Clark, T. Henighan, M. Kozina, M. Chollet, J. M. Glownia, M. C. Hoffmann, D. Zhu, O. Delaire, A. F. May, B. C. Sales, A. M. Lindenberg, P. Zalden, T. Sato, R. Merlin, and D. A. Reis. The origin of incipient ferroelectricity in lead telluride. *Nature Communications*, 7(1):12291, 2016. doi: 10.1038/ncomms12291. URL https://doi.org/10.1038/ncomms12291. ISBN: 2041-1723.

[19] Iris D. et al. Young. Structure of photosystem II and substrate binding at room temperature. *Nature*, 540(7633):453–457, 2016. doi: 10.1038/nature20161. URL https://doi.org/10.1038/nature20161. ISBN: 1476-4687.

[20] Eric Brochu, Vlad M. Cora, and Nando de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *arXiv:1012.2599 [cs]*, December 2010. URL http://arxiv.org/abs/1012.2599. arXiv: 1012.2599.

[21] Eric Hans Lee, Valerio Perrone, Cedric Archambeau, and Matthias Seeger. Cost-aware Bayesian Optimization. *arXiv:2003.10870 [cs, stat]*, March 2020. URL http://arxiv.org/abs/2003.10870. arXiv: 2003.10870.

[22] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. ISSN 1941-0026. doi: 10.1109/4235.996017. Conference Name: IEEE Transactions on Evolutionary Computation.

[23] Manoel Conde, Sergey Antipov, Darrell Doran, Wei Gai, Qiang Gao, Gwanghui Ha, Chunguang Jing, Wanming Liu, Nicole Neveu, John Power, and others. Research program and recent results at the Argonne Wakefield Accelerator Facility (AWA). *Proc. IPAC'17*, pages 2885–2887, 2017.

[24] Peter W Hawkes and Erwin Kasper. *Principles of electron optics*, volume 3. Academic press, 1996.

[25] 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 Transactions on Nuclear Science*, 63(2):878–897, April 2016. ISSN 1558-1578. doi: 10.1109/TNS.2016.2543203. Conference Name: IEEE Transactions on Nuclear Science.

[26] Min Zhang. Emittance formula for slits and pepper-pot measurement. Technical Report FNAL-TM–1988, Fermi National Accelerator Lab., Batavia, IL (United States), 1996.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]