Stronger symbolic summary statistics for the LHC

Nathalie Soybelman* **Anja Butter** **Tilman Plehn**
Institut für Theoretische Physik
Universität Heidelberg
Germany

Johann Brehmer
Center for Data Science
New York University
United States

Abstract

Analyzing the high-dimensional data collected at the Large Hadron Collider experiments often requires a balance between maximizing sensitivity and maintaining interpretability by domain experts. We propose a new algorithm to construct powerful summary statistics for LHC processes in the form of simple symbolic expressions. First, we extract latent information from a chain of simulators; through symbolic regression on this data we then learn approximately sufficient statistics. Observables constructed in this way can be used as plug-in replacements for established summary statistics, potentially improving the precision of scientific results without adding any overhead. In Higgs production in weak boson fusion, our algorithm rediscovers well-known heuristics and proposes new, moderately complex formulas that rival the new physics reach of neural networks.

1 Introduction

The experiments at the Large Hadron Collider (LHC) collect vast amounts of high-dimensional data, which need to be turned into precise measurements of the fundamental properties of nature. While particle physicists have developed a chain of high-fidelity simulators for particle collisions, the likelihood function that they implicitly define is intractable. This is a major challenge for inference and a number of simulation-based (or likelihood-free) inference methods have been developed to overcome it [1–6].

Particle physicists typically tackle this problem by not working with the high-dimensional raw data directly, but reducing its dimensionality by extracting a summary statistic. Once the data is one-dimensional, the likelihood function can be estimated through univariate density estimation methods like histograms, which makes both frequentist and Bayesian inference possible. However, designing a summary statistic is problem-specific and difficult. Traditionally, simple kinematic variables are often chosen, but these are known to lose information, leading to less precise measurements [7–9]. Recently, neural summary statistics have been proposed that are approximately sufficient [10–12]. While they can be more powerful, they are less easily interpretable, sometimes perceived by more conservative domain scientists as black boxes. In addition, integrating them in established data analysis pipelines may require some effort. These (real and perceived) drawbacks have limited their adaptation so far.

Here we introduce a new kind of summary statistics for LHC processes that is both approximately optimal and highly interpretable. Our algorithm first extracts additional latent information from LHC simulators and augments a training dataset with it. Next, we use symbolic regression on the augmented data to learn the approximately sufficient statistics of the problem. The output of this procedure is a simple analytical expression like "$p_{T1}\, p_{T2}\, \sin \Delta\phi$" that is self-explanatory to domain scientists, can be "stored" on a piece of paper, and can be directly plugged into a typical particle

*nathalie@soybelman.de

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

physics analysis without adding overhead. We demonstrate the potential of this approach in an analysis of the properties of the Higgs boson.

2 Related work

Simulation-based inference Many phenomena in science are modeled with simulators that do not have a tractable likelihood function. Parameter inference in these models requires simulation-based inference methods such as Approximate Bayesian Computation [2, 3] or techniques based on neural surrogates of the likelihood [13–16], likelihood ratio [10, 17–21], or posterior [22–28]; see Ref. [4] for a recent review and Refs. [6, 11, 12] for applications to particle physics.

Learned summary statistics Several methods have been proposed to learn powerful summary statistics. This work is closely related to the SALLY algorithm [10–12], in which a neural network (or boosted tree [29]) learns the *score*, the gradient of the log likelihood with respect to the parameters of interest. Neural summaries can also be learned by mutual information maximization [30] or based on an exponential family approximation [31, 32]. Finally, it is possible to make summaries robust to nuisance parameters [33, 34].

Even before these neural network–based methods, particle physicists have developed domain-specific techniques for optimal summary statistics known as "optimal observables" [35–38]. Unlike our method, these algorithms require approximations on the underlying physics process, for instance the detector interactions.

Interpretable machine learning for particle physics In the context of particle physics, interpretable and explainable machine learning are still under-explored topics. References [39, 40] proposed a method to extract a set of the most relevant high-level observables from a trained neural network. Symbolic regression has previously been used to learn scientific laws from data [41–44], but also for interpretable classifiers in particle physics [45].

3 Optimal observables from symbolic regression

Setup We consider an LHC process in which we want to infer some parameters of interest θ from observed data $\{x_i\}$. These data consist of multiple i. i. d. events, each characterized by a vector of basic particle properties such as reconstructed energies and masses. The relation between parameters and collected data is modeled by a simulator, which allows us to sample $x \sim p(x|\theta)$, but the likelihood function $p(x|\theta)$ is intractable. Our goal is to find a summary statistic $\hat{t}(x)$ such that inference on a summary dataset $\{\hat{t}(x_i)\}$ allows a precise measurement of θ.

We focus on the case where the parameters θ are in the neighborhood of some reference parameter point θ_0. This is common in precision measurements of the LHC, for instance when θ are the Wilson coefficients of the Standard Model (SM) effective field theory (EFT) and the reference point is the SM, $\theta_0 = 0$.

Sufficient statistics In the local approximation $|\theta - \theta_0| \ll 1$, the likelihood function is in the exponential family and the components of the score

$$t(x|\theta_0) = \nabla_\theta \log p(x|\theta)\Big|_{\theta_0} \tag{1}$$

are its sufficient statistics [10, 31]. In other words, as long as we consider parameter points close to the reference point θ_0, the score represents the perfect summary statistic or optimal observables; reducing a dataset to the values of the score for each event would not lose any information on θ. Unfortunately, the score is in general intractable.

Learning the score Our goal is therefore to learn an approximation of the score. First, we leverage the fact that particle physics simulators are not entirely black boxes, but let us access some information that characterizes the latent process in the simulator [11, 12]. More precisely, for each simulated event x we can also access the ground-truth parton-level phase space point z as well as the squared matrix element $|\mathcal{M}(z|\theta)|^2$. With these quantities, we can compute the *joint score* in the space of latents and observables

$$t(x, z|\theta_0) = \frac{\nabla_\theta |\mathcal{M}(z|\theta_0)|^2}{|\mathcal{M}(z|\theta_0)|^2} - \frac{\nabla_\theta \sigma_{\text{tot}}(\theta_0)}{\sigma_{\text{tot}}(\theta_0)} \tag{2}$$

for each simulated event. Here $\sigma_{\text{tot}}(\theta)$ is the total cross section of the process.

The usefulness of the joint score is not immediately obvious, as it explicitly depends on the latent z, which is only available in simulations. However, Refs. [10–12] have shown that the score $t(x|\theta_0)$ can be linked to the joint score $t(x, z|\theta_0)$ through the minimum of the mean-squared-error (MSE) functional:

$$t(x|\theta_0) = \arg\min_{g(x)} \mathbb{E}_{x,z\sim p(x,z|\theta_0)} \left|g(x) - t(x, z|\theta_0)\right|^2 . \tag{3}$$

We implement this equation in a practical algorithm with symbolic regression. Instead of the abstract variational family $g(x)$, we consider the function family of analytical expressions of a maximum length, consisting of kinematic variables, mathematical operations, and numerical constants. Instead of the exact minimization in Eq. (3), we use a genetic algorithm that performs random mutations on a set of symbolic expressions and keeps solutions with a higher probability if they lower the MSE. Hyperparameters allow us to choose the trade-off between complexity of the symbolic expression and the value of the MSE in Eq. (3), which we expect to be related to the quality of the summary statistic.

All in all, our algorithm thus consists of two phases: first, we simulate events; for each event we save the observable data x together with the joint score $t(x, z)$. Second, we run a symbolic regression algorithm on the joint score $t(x, z)$ to learn an analytical expression $\hat{t}(x)$ that approximates the score.

The output of our algorithm is a simple symbolic expression like $p_{T1}\, p_{T2}\, \sin \Delta\phi$. As an approximation of the score, summary statistics constructed in this way will be approximately equal to the sufficient statistics. They are therefore likely to improve the precision of the analysis results compared to established heuristics. At the same time, they are as simple to handle as conventional kinematic variables: these observables can simply be "stored" by writing their definition on a piece of paper. Using it in a data analysis pipeline is equally straightforward and does not require, for instance, the evaluation of a neural network. Finally, the score can easily be interpreted by domain experts.

Implementation We simulate events with a standard simulator chain consisting of MADGRAPH5 2.6.7 [46], PYTHIA 8.3 [47], and DELPHES 3.4.2 [48]. MADMINER 0.7.5 [49] is used for book keeping and to compute the joint score. We use a modified version of PYSR [50] to learn the functional form of the score and add a parameter-fit afterburner to finetune the constants in the analytical expression.

4 Experiments

We demonstrate the potential of our symbolically regressed observables in two precision measurements of the properties of the Higgs boson at the LHC. Signs of new physics are expected to manifest themselves in such measurements as deviations from the Standard Model predictions.

First, we verify our algorithm in a toy problem and consider ZH production without decays, shower, and detector effects. This simple process has only two particles in the final state and is fully described by only two degrees of freedom. We consider the problem of measuring the Wilson coefficient of the dimension-six operator $\mathcal{O}_B$ [51], using the SM $\theta_0 = 0$ as reference point. As summary statistic, our algorithm learns a polynomial in the transverse momentum p_T and the difference of the pseudorapidities[2] $|\eta_Z - \eta_H|$. This is in agreement with our expectations about the structure of the process.

We then focus on a realistic scenario: Higgs production in the weak boson fusion mode, including a decay of the Higgs boson into muons and detector effects. This process has substantially more complex kinematics and its final state involves two jets, collimated sprays of charged particles, in addition to the muons that the Higgs decays into. Our goal is to measure the Wilson coefficient of the CP-violating dimension-six operator $\mathcal{O}_{W\widetilde{W}}$ [51]. Again, we use the SM as reference point, $\theta_0 = 0$.

[2]The pseudorapidity measures the angle between the direction of a particle and the beam axis.

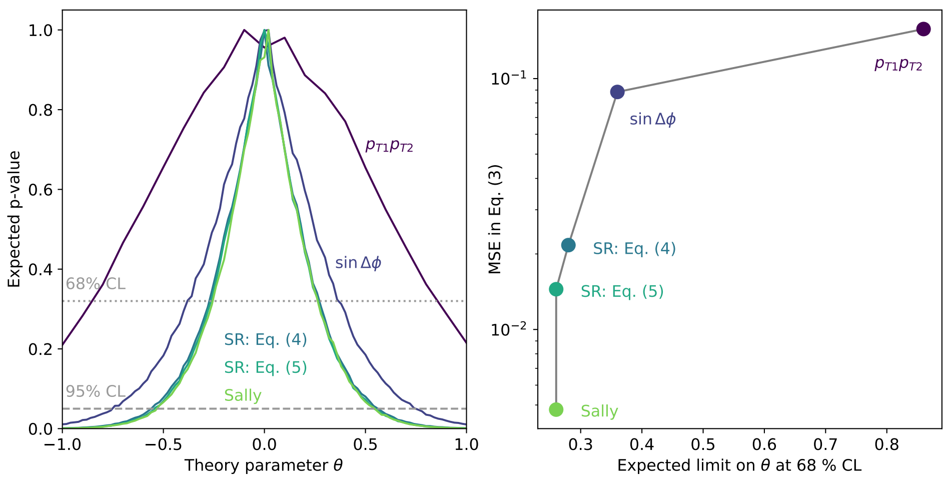

Figure 1: Left: Expected p-values on θ based on different summary statistics, using a synthetic dataset generated with $\theta_{\text{true}} = 0$. We compare simple heuristics, more complex symbolic summary statistics found by our symbolic regression (SR) algorithm, and a neural network trained with the SALLY algorithm. Right: Relation between the expected upper bounds with the MSE in Eq. (3).

Our symbolic regression algorithm finds a number of summary statistics of increasing complexity. The simplest one is $\hat{t}_1(x) = \sin\Delta\phi$, where $\Delta\phi$ is the difference between the azimuthal angles of the two jets. This is in fact an established heuristic [52]: When set to focus on simple expression, our algorithm thus rediscovers a quantity well-known to (and hand-engineered by) domain scientists.

At higher complexity, our algorithm learns

$$\hat{t}_2(x|\theta_0) = p_{T1}\, p_{T2}\, \sin\Delta\phi \tag{4}$$

and ultimately

$$\hat{t}_3(x|\theta_0) = -p_{T1}\,(p_{T2} + c)(a - b\Delta\eta)\sin(\Delta\phi + d)\,, \tag{5}$$

where p_{Ti} are the transverse momentum of the jets, $\Delta\eta$ is the difference in pseudorapidities between the two jets and a, b, c, d are numerical constants.

We then study how powerful these summary statistics are by computing expected confidence regions on the parameter of interest θ, using a synthetic "observed" dataset generated for $\theta_{\text{true}} = 0$. The left panel of Fig. 1 shows the expected p-values. The more complex summary statistics found by our symbolic regression algorithm lead to stronger exclusion bounds on the parameter of interest than the simple heuristics $p_{T1}\, p_{T2}$ and $\sin\Delta\phi$. In fact, our symbolic observables lead to confidence limits virtually indistinguishable from those based on neural networks trained with the SALLY algorithm.

Finally, in the right panel of Fig. 1 we investigate how predictive the MSE in Eq. (3) is of the performance on the downstream task of measuring θ. Indeed, summary statistics that have a lower MSE with respect to the joint score later enable more powerful measurements, visible as lower upper bounds on θ. However, the returns are diminishing: while more and more complex summary statistics can further improve the fit to the joint score, this does not translate to substantial improvements in the parameter measurement. Moderately complex symbolic expressions provide a good compromise of near-perfect measurements and straightforward interpretability.

5 Outlook

Machine learning has a huge potential to improve the sensitivity of LHC measurements. However, the lack of interpretability of many models has been a limiting factor in the widespread adaptation of new methods. We propose a new class of summary statistics based on symbolic regression. They approximate the sufficient statistics and are therefore nearly optimal for a given problem. At the same time, they are as easy to interpret, store, and implement in analyses as traditional kinematic heuristics.

For the theoretically interesting case of Higgs production in weak boson fusion, our algorithm reinvents the well-known sine of the azimuthal angle between the tagging jets as a powerful summary statistics. It also discovers new, slightly more complex expressions that allow us to measure fundamental properties of nature with higher precision, with almost no loss of performance compared to neural network–based summary statistics.

Acknowledgments

We are very grateful to Miles Cranmer for help on PySR, Kyle Cranmer for his insights, and Markus Schumacher for enlightening and fun discussion on optimal observables and how to use them in Higgs physics. The research of AB and TP is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762 – TRR 257 *Particle Physics Phenomenology after the Higgs Discovery*.

References

[1] Peter J. Diggle and Richard J. Gratton. Monte Carlo Methods of Inference for Implicit Statistical Models. In *Journal of the Royal Statistical Society: Series B (Methodological)*, volume 46, pages 193–212, 1984. doi: 10.1111/j.2517-6161.1984.tb01290.x.

[2] Donald B. Rubin. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics*, 12(4):1151–1172, 1984. ISSN 0090-5364. doi: 10.1214/aos/1176346785. URL https://doi.org/10.1214/aos/1176346785.

[3] Mark A. Beaumont, Wenyang Zhang, and David J. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002. ISSN 00166731. doi: 10.1111/j.1937-2817.2010.tb01236.x.

[4] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[5] Johann Brehmer, Kyle Cranmer, Irina Espejo, Felix Kling, Gilles Louppe, and Juan Pavez. Effective LHC measurements with matrix elements and machine learning. In *19th International Workshop on Advanced Computing and Analysis Techniques in Physics Research: Empowering the revolution: Bringing Machine Learning to High Performance Computing (ACAT 2019) Saas-Fee, Switzerland, March 11-15, 2019*, 2019. URL http://arxiv.org/abs/1906.01578.

[6] Johann Brehmer and Kyle Cranmer. Simulation-based inference methods for particle physics. In Paolo Calafiura, David Rousseau, and Kazuhiro Terao, editors, *Artificial Intelligence for High Energy Physics*. World Scientific, 2021. doi: 10.1142/12200.

[7] Johann Brehmer, Kyle Cranmer, Felix Kling, and Tilman Plehn. Better Higgs boson measurements through information geometry. *Phys. Rev. D*, 95(7):073002, 2017. doi: 10.1103/PhysRevD.95.073002.

[8] Johann Brehmer, Felix Kling, Tilman Plehn, and Tim M. P. Tait. Better Higgs-CP Tests Through Information Geometry. *Phys. Rev. D*, 97(9):095017, 2018. doi: 10.1103/PhysRevD.97.095017.

[9] Johann Brehmer, Sally Dawson, Samuel Homiller, Felix Kling, and Tilman Plehn. Benchmarking simplified template cross sections in WH production. *JHEP*, 11:034, 2019. doi: 10.1007/JHEP11(2019)034.

[10] Johann Brehmer, Gilles Louppe, Juan Pavez, and Kyle Cranmer. Mining gold from implicit models to improve likelihood-free inference. *Proceedings of the National Academy of Sciences of the United States of America*, 117(10):5242–5249, mar 2020. ISSN 10916490. doi: 10.1073/pnas.1915980117. URL http://arxiv.org/abs/1805.12244{%}0Ahttp://dx.doi.org/10.1073/pnas.1915980117http://www.ncbi.nlm.nih.gov/pubmed/32079725.

[11] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. A guide to constraining effective field theories with machine learning. *Physical Review D*, 98(5):052004, 2018. ISSN 24700029. doi: 10.1103/PhysRevD.98.052004.

[12] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. Constraining Effective Field Theories with Machine Learning. *Physical Review Letters*, 121(11):111801, 2018. ISSN 10797114. doi: 10.1103/PhysRevLett.121.111801.

[13] George Papamakarios, David C. Sterratt, and Iain Murray. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. *International Conference on Artificial Intelligence and Statistics*, 2018. URL `http://arxiv.org/abs/1805.07226`.

[14] Conor Durkan, George Papamakarios, and Iain Murray. Sequential Neural Methods for Likelihood-free Inference. *arXiv:1811.08723*, nov 2018. URL `http://arxiv.org/abs/1811.08723`.

[15] Jan-Matthis Lueckmann, Giacomo Bassetto, Theofanis Karaletsos, and Jakob H. Macke. Likelihood-free inference with emulator networks. In Francisco Ruiz, Cheng Zhang, Dawen Liang, and Thang Bui, editors, *Proceedings of The 1st Symposium on Advances in Approximate Bayesian Inference*, volume 96 of *Proceedings of Machine Learning Research*, pages 32–53. PMLR, 2018. URL `http://arxiv.org/abs/1805.09294`.

[16] Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440–4458, sep 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1960.

[17] Kyle Cranmer, Juan Pavez, and Gilles Louppe. Approximating Likelihood Ratios with Calibrated Discriminative Classifiers. *arXiv:1506.02169*, 2015. URL `http://arxiv.org/abs/1506.02169`.

[18] Markus Stoye, Johann Brehmer, Gilles Louppe, Juan Pavez, and Kyle Cranmer. Likelihood-free inference with an improved cross-entropy estimator. *arXiv:1808.00973*, 2018. URL `http://arxiv.org/abs/1808.00973`.

[19] Traiko Dinev and Michael U. Gutmann. Dynamic Likelihood-free Inference via Ratio Estimation (DIRE). *arXiv:1810.09899*, oct 2018. URL `http://arxiv.org/abs/1810.09899`.

[20] Joeri Hermans, Volodimir Begy, and Gilles Louppe. Likelihood-free MCMC with Amortized Approximate Ratio Estimators. *arXiv:1903.04057*, 2019. URL `http://arxiv.org/abs/1903.04057`.

[21] Benjamin Kurt Miller, Alex Cole, Patrick Forré, Gilles Louppe, and Christoph Weniger. Truncated Marginal Neural Ratio Estimation. 7 2021. doi: 10.5281/zenodo.5043707.

[22] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *32nd International Conference on Machine Learning, ICML 2015*, 2:1530–1538, may 2015.

[23] George Papamakarios and Iain Murray. Fast e-free inference of simulation models with Bayesian conditional density estimation. In *Advances in Neural Information Processing Systems*, pages 1036–1044, 2016.

[24] Jan Matthis Lueckmann, Pedro J. Gonçalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H. Mackey. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in Neural Information Processing Systems*, 2017-December:1290–1300, nov 2017. ISSN 10495258.

[25] Rafael Izbicki, Ann B. Lee, and Taylor Pospisil. ABC–CDE: Toward Approximate Bayesian Computation With Complex High-Dimensional Data and Limited Simulations. *Journal of Computational and Graphical Statistics*, 28(3):481–492, 2019. ISSN 15372715. doi: 10.1080/10618600.2018.1546594.

[26] Brooks Paige and Frank Wood. Inference networks for sequential monte carlo in graphical models. *33rd International Conference on Machine Learning, ICML 2016*, 6:4434–4444, feb 2016.

[27] Dustin Tran, Rajesh Ranganath, and David M. Blei. Hierarchical implicit models and likelihood-free variational inference. In I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, editors, *Advances in Neural Information Processing Systems*, volume 2017-December, pages 5524–5534. 2017.

[28] Pedro LC Rodrigues, Thomas Moreau, Gilles Louppe, and Alexandre Gramfort. Leveraging global parameters for flow-based neural posterior estimation. *arXiv preprint arXiv:2102.06477*, 2021.

[29] Suman Chatterjee, Nikolaus Frohner, Lukas Lechner, Robert Schöfbeck, and Dennis Schwarz. Tree boosting for learning EFT parameters. 7 2021.

[30] Yanzhi Chen, Dinghuai Zhang, Michael Gutmann, Aaron Courville, and Zhanxing Zhu. Neural approximate sufficient statistics for implicit models. *arXiv preprint arXiv:2010.10079*, 2020.

[31] Justin Alsing and Benjamin Wandelt. Generalized massive optimal data compression. *Monthly Notices of the Royal Astronomical Society: Letters*, 476(1):L60–L64, 2018. ISSN 1745-3925. doi: 10.1093/mnrasl/sly029.

[32] Justin Alsing, Benjamin Wandelt, and Stephen Feeney. Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology. *Monthly Notices of the Royal Astronomical Society*, 477(3):2874–2885, 2018. ISSN 13652966. doi: 10.1093/mnras/sty819.

[33] Pablo de Castro and Tommaso Dorigo. INFERNO: Inference-Aware Neural Optimisation. *Computer Physics Communications*, 244:170–179, 2019. ISSN 00104655. doi: 10.1016/j.cpc.2019.06.007.

[34] Justin Alsing and Benjamin Wandelt. Nuisance hardened data compression for fast likelihood-free inference. *Monthly Notices of the Royal Astronomical Society*, 488(4):5093–5103, 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1900.

[35] D. Atwood and A. Soni. Analysis for magnetic moment and electric dipole moment form factors of the top quark via e+e-'tt». *Physical Review D*, 45(7):2405–2413, 1992. ISSN 05562821. doi: 10.1103/PhysRevD.45.2405.

[36] M. Davier, L. Duflot, F. Le Diberder, and A. Rougé. The optimal method for the measurement of tau polarization. *Physics Letters B*, 306(3-4):411–417, 1993. ISSN 03702693. doi: 10.1016/0370-2693(93)90101-M.

[37] M. Diehl and O. Nachtmann. Optimal observables for the measurement of three gauge boson couplings in $e^+e^- \to W^+W^-$. *Zeitschrift für Physik C Particles and Fields*, 62(3):397–411, 1994. ISSN 01709739. doi: 10.1007/BF01555899.

[38] V. M. Abazov, B. Abbott, B. S. Acharya, M. Adams, T. Adams, J. P. Agnew, G. D. Alexeev, G. Alkhazov, A. Alton, A. Askew, and et al. Measurement of the top quark mass using the matrix element technique in dilepton final states. *Physical Review D*, 94(3), Aug 2016. ISSN 2470-0029. doi: 10.1103/physrevd.94.032004. URL `http://dx.doi.org/10.1103/PhysRevD.94.032004`.

[39] Taylor Faucett, Jesse Thaler, and Daniel Whiteson. Mapping Machine-Learned Physics into a Human-Readable Space. *Phys. Rev. D*, 103(3):036020, 2021. doi: 10.1103/PhysRevD.103.036020.

[40] Julian Collado, Jessica N. Howard, Taylor Faucett, Tony Tong, Pierre Baldi, and Daniel Whiteson. Learning to identify electrons. *Phys. Rev. D*, 103(11):116028, 2021. doi: 10.1103/PhysRevD.103.116028.

[41] Michael D. Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324:81 – 85, 2009.

[42] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: a Physics-Inspired Method for Symbolic Regression. *Sci. Adv.*, 6(16):eaay2631, 2020. doi: 10.1126/sciadv.aay2631.

[43] Miles D Cranmer, Rui Xu, Peter Battaglia, and Shirley Ho. Learning symbolic physics with graph networks. *arXiv preprint arXiv:1909.05862*, 2019.

[44] Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. *arXiv preprint arXiv:2006.11287*, 2020.

[45] Kyle Cranmer and R Sean Bowman. Physicsgp: A genetic programming approach to event selection. *Computer Physics Communications*, 167(3):165–176, 2005.

[46] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[47] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.

[48] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014. doi: 10.1007/JHEP02(2014)057.

[49] Johann Brehmer, Felix Kling, Irina Espejo, and Kyle Cranmer. MadMiner: Machine Learning-Based Inference for Particle Physics. *Computing and Software for Big Science*, 4(1), 2020. ISSN 2510-2036. doi: 10.1007/s41781-020-0035-2.

[50] Miles Cranmer. Pysr: Fast & parallelized symbolic regression in python/julia, September 2020. URL `http://doi.org/10.5281/zenodo.4041459`.

[51] Kaoru Hagiwara, Sumio Ishihara, R Szalapski, and D Zeppenfeld. Low energy effects of new interactions in the electroweak boson sector. *Physical Review D*, 48(5):2182, 1993.

[52] V. Hankele, G. Klämke, D. Zeppenfeld, and T. Figy. Anomalous higgs boson couplings in vector boson fusion at the cern lhc. *Physical Review D*, 74(9), Nov 2006. ISSN 1550-2368. doi: 10.1103/physrevd.74.095001. URL `http://dx.doi.org/10.1103/PhysRevD.74.095001`.

A Checklist

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] We describe the approximations that our approach is based on.
   (c) Did you discuss any potential negative societal impacts of your work? [No] Our work is most relevant to research into the fundamental properties of nature, which we deem a positive societal impact in itself. Due to their extremely general nature, it is however possible that simulation-based inference methods can be applied to more nefarious purposes.
   (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] Our work does not have new theoretical results. We cite the theoretical results we are building on.
   (b) Did you include complete proofs of all theoretical results? [N/A] Our work does not have new theoretical results. We cite the theoretical results we are building on.
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)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [No] Not in this workshop paper, though we will make a longer version with all relevant practical details available.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (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)? [No]
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? [Yes]
   (b) Did you mention the license of the assets? [No] The licenses can be found through the citations.
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (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]