Sustainable Matrix Element Method through Deep Learning

HSF-CWP-018

Mark Neubauer[1]

[1]University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

August 18, 2017

1 Introduction

The Matrix Element (ME) Method [1–4] is a powerful technique which can be utilzed for measurements of physical model parameters and direct searches for new phenomena. It has been used extensively by collider experiments at the Tevatron for SM measurements and Higgs boson searches [5–10] and at the LHC for measurements in the Higgs and top quark sectors of the SM [11–17]. The ME method is based on *ab initio* calculation of the probabilty density function $\mathcal{P}$ of an event with observed final-state particle momenta $\mathbf{x}$ to be due to a physics process ξ with theory parameters $\boldsymbol{\alpha}$. One can compute $\mathcal{P}_\xi(\mathbf{x}|\boldsymbol{\alpha})$ by means of the factorization theorem from the corresponding partonic cross-sections of the hard scattering process involving parton momenta $\mathbf{y}$ and is given by

$$\mathcal{P}_\xi(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{\sigma_\xi^{\text{fiducial}}(\boldsymbol{\alpha})} \int d\Phi(\mathbf{y}_{\text{final}})\, dx_1\, dx_2\, \frac{f(x_1)f(x_2)}{2sx_1x_2}\, |\mathcal{M}_\xi(\mathbf{y}|\boldsymbol{\alpha})|^2\, \delta^4(\mathbf{y}_{\text{initial}} - \mathbf{y}_{\text{final}})\, W(\mathbf{x},\mathbf{y}) \tag{1}$$

where and x_i and $\mathbf{y}_{\text{initial}}$ are related by $y_{\text{initial},i} \equiv \frac{\sqrt{s}}{2}(x_i, 0, 0, \pm x_i)$, $f(x_i)$ are the parton distribution functions, $\sqrt{s}$ is the collider center-of-mass energy, $\sigma_\xi^{\text{fiducial}}(\boldsymbol{\alpha})$ is the total cross section for the process ξ (with $\boldsymbol{\alpha}$) times the detector acceptance, $d\Phi(\mathbf{y})$ is the phase space density factor, $\mathcal{M}_\xi(\mathbf{y}|\boldsymbol{\alpha})$ is the matrix element (typically at leading-order (LO)), and $W(\mathbf{x},\mathbf{y})$ is the probability density (aka "transfer function") that a selected event $\mathbf{y}$ ends up as a measured event $\mathbf{x}$.

One can use calculations of Eqn. 1 in a number of ways to search for new phenomena at particle colliders. For measurement of model parameters $\boldsymbol{\alpha}$, one would maximize the likelihood function for observed events $\mathcal{L}(\boldsymbol{\alpha})$ given by

$$\mathcal{L}(\boldsymbol{\alpha}) = \prod_i \sum_k f_k \mathcal{P}_{\xi_k}(\mathbf{x}_i|\boldsymbol{\alpha}) \tag{2}$$

where f_k are the fractions of (non-interfering) processes contributing to the data. For new particle searchres, one can (using Bayes' Theorem [18]) compute for a hypthosized signal S the probability $P(S|\mathbf{x})$ given by

$$P(S|\mathbf{x}) = \frac{\sum_i \beta_{S_i} \mathcal{P}_{S_i}(\mathbf{x}|\boldsymbol{\alpha}_{S_i})}{\sum_i \beta_{S_i} \mathcal{P}(\mathbf{x}|\boldsymbol{\alpha}_{S_i}) + \sum_j \beta_{B_j} \mathcal{P}(\mathbf{x}|\boldsymbol{\alpha}_{B_j})} \tag{3}$$

where, S_i and B_j, denote all signal and background processes relevant to the considered phase space and β are the *a priori* expected process fractions. According to the Neyman-Pearson Lemma [19], Eqn. 3 is the optimal discriminant function for S in the presence of B and can be used to extract a signal fraction in the data.

2 Advantages over Training-based Methods

As a multivariate analysis approach, the ME method brings in several unique and desirable features, most notably it (1) does not require training data being an *ab initio* calculation of event probabilities, (2) incorporates all available kinematic information of a hypothesized process, including all correlations, and (3) has a clear physical meaning in terms transition probabilities within the framework of quantum field theory.

3 Limitations of ME Method with Current Techniques

One drawback to the ME Method is that it has traditionally relied on LO matrix elements, although nothing in principle limits the ME method to LO calculations. Techniques that accomodate initial-state QCD radiation within the LO ME framework using transverse boosting and dedicated transfer functions to integrate over the transverse momentum of initial-state partons have been developed [20]. Another challenge is development of the transfer functions which rely on tediously hand-crafted fits to full simulated Monte-Carlo events.

The most serious difficulty in the ME method, and the one which has limited its applicability to searches for beyond-the-SM physics and precision measurements at collider experiments, is that it is very *computationally intensive*. If this limitation could be overcome, then it would enable more widespread use of ME methods for analysis of LHC data. This could be particularly important for extending the new physics reach of the HL-LHC which will be dominated by increases in integrated luminosity rather than center-of-mass collision energy.

Accurate evaluation of Eqn. 1 is computationally challenging primarily for two reasons: (1) it involves high-dimensional integration over a large number of events, signal and background hypotheses, and systematic variations and (2) it involves sharply-peaked integrands[1] over a large domain in phase space. In reference to point (1), the matrix element $\mathcal{M}_\xi(\mathbf{y}|\boldsymbol{\alpha})$ in the method involves all partons in the $n \to m$ process, so when the 4-momentum of particles are not completely measured experimentally (e.g. neutrinos), one must integrate over the missing information which increases the dimensionality of the integration. In reference to point (2), a clever technique to re-map the phase space in order to reduce the sharpness of integrate in that space in an automated way (MADWEIGHT [21]) is often used in conjunction with a matrix element calculation package (MADGRAPH_aMCNLO [22]). In practice, evaluation of definite integrals by the ME approach invokes techniques such as importance sampling (see VEGAS [23, 24] and FOAM [25]) or recursive stratified sampling (see MISER [26]) Monte Carlo integration. Acceleration of some of these techniques on modern computing architectures has been achieved, for example concurrent phase space sampling in VEGAS on GPUs.

4 Sustainable ME Method using Deep Learning

Despite the attractive features of the ME method and promise of further optimization and parallelization of the evaluation of Eqn. 1, the computational burden of the ME technique will continue to limit is range of applicability for practical data analysis without new and innovative approaches. This is especially true when one considers the process of producing a physics publication which involves many selection, sample and systematic iterations for which ME calculations are required. The primary idea put forward in this Section is to utilize modern *machine learning techniques to dramatically speed up the numerical evaluation of Eqn. 1* and therefore broaden the applicability of the ME method to the benefit of the HL-LHC physics program.

Applying neural networks to numerical integration problems is plausible but not new (see [27–29], for example). The technical challenge is to design a network which is sufficiently rich to encode the complexity of the ME calculation for a given process over the phase space relevant to the signal process. Deep Neural Networks (DNNs) are stong candidates for networks with sufficient complexity to achieve good approximation of Eqn. 1, possibly in conjunction with smart phase-space mapping such as described in [21]. Promising demonstration of the power of Boosted Decision Trees [30, 31] and Generative Adversarial Neural Networks [32] for improved Monte Carlo integration can be found in [33]. Once a set of DNNs representing of definite integrals of the form of Eqn. 1 to good approximation are generated, evaluation of the ME method calculations via the DNNs will be very fast. These DNNs can be throught of as preserving the essence of ME calculations in a way that allows for fast forward execution. The net result is that the DNNs can enable the ME method to be both *nimble* and *sustainable*, neither of which is true today.

5 Example Analysis Flow

The overall strategy is to do the expensive full ME calculations as infrequently as possible, ideally once for DNN training and once more for a final pass before publication, with the DNNs utilized as a good approximation in between. A future analysis flow using the ME method with DNNs might look something like the following: One performs a large number of ME calculations using a traditional numerical integration technique like VEGAS or FOAM on a large CPU resource like an HPC, Cloud or the Grid, ideally exploiting acceleration on many-core devices like GPUs or even FPGAs. The DNN training data is generated from the phase space sampling in performing the full integration in this initial pass, and DNNs are trained either *in situ* or *a posteriori*. The accuracy of the DDN-based ME calculation can be assessed through this procedure. As the analysis develops and progresses through selection and/or sample changes, systematic treatment, etc., the DNN-based ME calculations are used in place of the time-consuming, full ME calculations to make the analysis nimble and to preserve the ME calculations. Before a result using the ME method is published, a final pass using full ME calculation would likely be performed both to maximize the numerical precision or sensitivity of the results and to validate the analysis evolution via the DNN-based approximations.

[1] a consequence of imposing energy/momentum conservation in the processes

6 Roadmap

There are several activities which are proposed to further develop the idea of a Sustainable Matrix Element Method. The first is to establish a cross-experiment group interested in developing the ideas presented in this Section, along with a common software project for ME calculations, for example in the spirit of [34]. Given the nature of the challenges for a sustainable ME method, this is area which is very well-suited for impactful collaboration with computer scientists and those working in machine learning, so effort should be placed in establishing those connections. Using a few test cases (e.g. $t\bar{t}$ or $t\bar{t}h$ production), evaluation of DDN choices and configurations, developing methods for DNN training from full ME calculations and direct comparisons of the integration accuracy between Monte Carlo and DNN-based calculations should be undertaken. More effort should also be placed in developing compelling applications of the ME method for HL-LHC physics. In the longer term and after successfully demonstrating the value of the methods on a few test cases, we propose exploring the possibilty of Sustainable-Matrix-Element-Method-as-a-Service (SMEMaaS) where shared software and infrastructure could be used through a common API.

References

[1] K. Kondo. "Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models." *J. Phys. Soc. Jap.* 57 (1988), pp. 4126–4140. DOI: 10.1143/JPSJ.57.4126.

[2] F. Fiedler et al. "The Matrix Element Method and its Application in Measurements of the Top Quark Mass." *Nucl. Instrum. Meth.* A624 (2010), pp. 203–218. DOI: 10.1016/j.nima.2010.09.024. arXiv: 1003.1316 [hep-ex].

[3] I. Volobouev. "Matrix Element Method in HEP: Transfer Functions, Efficiencies, and Likelihood Normalization." *ArXiv e-prints* (Jan. 2011). arXiv: 1101.2259 [physics.data-an].

[4] F. Elahi and A. Martin. "Using the modified matrix element method to constrain $L_\mu - L_\tau$ interactions." *Phys. Rev.* D96.1 (2017), p. 015021. DOI: 10.1103/PhysRevD.96.015021. arXiv: 1705.02563 [hep-ph].

[5] V. M. Abazov et al. "A precision measurement of the mass of the top quark." *Nature* 429 (2004), pp. 638–642. DOI: 10.1038/nature02589. arXiv: hep-ex/0406031 [hep-ex].

[6] A. Abulencia et al. "Precision measurement of the top quark mass from dilepton events at CDF II." *Phys. Rev.* D75 (2007), p. 031105. DOI: 10.1103/PhysRevD.75.031105. arXiv: hep-ex/0612060 [hep-ex].

[7] T. Aaltonen et al. "First Measurement of ZZ Production in panti-p Collisions at $\sqrt{s} = 1.96$-TeV." *Phys. Rev. Lett.* 100 (2008), p. 201801. DOI: 10.1103/PhysRevLett.100.201801. arXiv: 0801.4806 [hep-ex].

[8] T. Aaltonen et al. "Inclusive Search for Standard Model Higgs Boson Production in the WW Decay Channel using the CDF II Detector." *Phys. Rev. Lett.* 104 (2010), p. 061803. DOI: 10.1103/PhysRevLett.104.061803. arXiv: 1001.4468 [hep-ex].

[9] V. M. Abazov et al. "Observation of Single Top Quark Production." *Phys. Rev. Lett.* 103 (2009), p. 092001. DOI: 10.1103/PhysRevLett.103.092001. arXiv: 0903.0850 [hep-ex].

[10] T. Aaltonen et al. "First Observation of Electroweak Single Top Quark Production." *Phys. Rev. Lett.* 103 (2009), p. 092002. DOI: 10.1103/PhysRevLett.103.092002. arXiv: 0903.0885 [hep-ex].

[11] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[12] S. Chatrchyan et al. "Measurement of the properties of a Higgs boson in the four-lepton final state." *Phys. Rev.* D89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[13] G. Aad et al. "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector." *Phys. Rev.* D91.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.91.012006. arXiv: 1408.5191 [hep-ex].

[14] V. Khachatryan et al. "Measurement of spin correlations in $t\bar{t}$ production using the matrix element method in the muon+jets final state in pp collisions at $\sqrt{s} = 8$ TeV." *Phys. Lett.* B758 (2016), pp. 321–346. DOI: 10.1016/j.physletb.2016.05.005. arXiv: 1511.06170 [hep-ex].

[15] V. Khachatryan et al. "Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method." *Eur. Phys. J.* C75.6 (2015), p. 251. DOI: 10.1140/epjc/s10052-015-3454-1. arXiv: 1502.02485 [hep-ex].

[16] G. Aad et al. "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector." *Eur. Phys. J.* C75.7 (2015), p. 349. DOI: 10.1140/epjc/s10052-015-3543-1. arXiv: 1503.05066 [hep-ex].

[17] G. Aad et al. "Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s}$ =8 TeV with the ATLAS detector using the Matrix Element Method." *Phys. Lett.* B756 (2016), pp. 228–246. DOI: 10.1016/j.physletb.2016.03.017. arXiv: 1511.05980 [hep-ex].

[18] M. Bayes and M. Price. "An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S." *Philosophical Transactions* 53 (1763), pp. 370–418. DOI: 10.1098/rstl.1763.0053. eprint: http://rstl.royalsocietypublishing.org/content/53/370.full.pdf+html. URL: http://rstl.royalsocietypublishing.org/content/53/370.short.

[19] J. Neyman and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 231.694-706 (1933), pp. 289–337. ISSN: 0264-3952. DOI: 10.1098/rsta.1933.0009. eprint: http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf. URL: http://rsta.royalsocietypublishing.org/content/231/694-706/289.

[20] J. Alwall, A. Freitas, and O. Mattelaer. "The Matrix Element Method and QCD Radiation." *Phys. Rev.* D83 (2011), p. 074010. DOI: 10.1103/PhysRevD.83.074010. arXiv: 1010.2263 [hep-ph].

[21] P. Artoisenet et al. "Automation of the matrix element reweighting method." *JHEP* 12 (2010), p. 068. DOI: 10.1007/JHEP12(2010)068. arXiv: 1007.3300 [hep-ph].

[22] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[23] G. P. Lepage. "A new algorithm for adaptive multidimensional integration." *Journal of Computational Physics* 27.2 (1978), pp. 192–203. ISSN: 0021-9991. DOI: http://dx.doi.org/10.1016/0021-9991(78)90004-9. URL: http://www.sciencedirect.com/science/article/pii/0021999178900049.

[24] T. Ohl. "Vegas revisited: Adaptive Monte Carlo integration beyond factorization." *Comput. Phys. Commun.* 120 (1999), pp. 13–19. DOI: 10.1016/S0010-4655(99)00209-X. arXiv: hep-ph/9806432 [hep-ph].

[25] S. Jadach. "Foam: A general-purpose cellular Monte Carlo event generator." *Computer Physics Communications* 152.1 (2003), pp. 55–100. ISSN: 0010-4655. DOI: http://dx.doi.org/10.1016/S0010-4655(02)00755-5. URL: http://www.sciencedirect.com/science/article/pii/S0010465502007555.

[26] W. H. Press and G. R. Farrar. "RECURSIVE STRATIFIED SAMPLING FOR MULTIDIMENSIONAL MONTE CARLO INTEGRATION." *Submitted to: Comp.in Phys.* (1989).

[27] Z. Zhe-Zhao, W. Yao-Nan, and W. Hui. "Numerical integration based on a neural network algorithm." 8 (Aug. 2006), pp. 42–48.

[28] L. y. Xu and L. j. Li. "The New Numerical Integration Algorithm Based on Neural Network." *Third International Conference on Natural Computation (ICNC 2007)*. Vol. 1. Aug. 2007, pp. 325–328. DOI: 10.1109/ICNC.2007.730.

[29] L. Yan, J. Di, and K. Wang. "Spline Basis Neural Network Algorithm for Numerical Integration." *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering* 7.3 (2013), pp. 458–461. ISSN: PISSN:2010-376X, EISSN:2010-3778. URL: http://waset.org/Publications?p=75.

[30] J. Friedman, T. Hastie, and R. Tibshirani. "Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)." *Ann. Statist.* 28.2 (Apr. 2000), pp. 337–407. DOI: 10.1214/aos/1016218223. URL: http://dx.doi.org/10.1214/aos/1016218223.

[31] J. H. Friedman. "Greedy function approximation: A gradient boosting machine." *Ann. Statist.* 29.5 (Oct. 2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: http://dx.doi.org/10.1214/aos/1013203451.

[32] I. J. Goodfellow et al. "Generative Adversarial Networks." *ArXiv e-prints* (June 2014). arXiv: 1406.2661 [stat.ML].

[33] J. Bendavid. "Efficient Monte Carlo Integration Using Boosted Decision Trees and Generative Deep Neural Networks" (2017). arXiv: 1707.00028 [hep-ph].

[34] *MeMEMta: Modular Matrix Element Implementation.* URL: https://github.com/MoMEMta.