Reply to Report 2 for 2308.00027v1

Title: "Returning CP-observables to the Frames They Belong"

Authors: Jona Ackerschott, Rahool Kumar Barman, Dorival Gonçalves, Theo Heimel, Tilman Plehn

We would like to thank the referee for carefully reading our manuscript and providing detailed and valuable comments. The version of the manuscript that we resubmit addresses the aspects that the report brought to our consideration. Please find below the comments from the referee and our answers.

1. **Referee's comment**: *Section 2.1. The cINN is the salient feature of your method, yet it is barely introduced in this section. From reading the many references, it is possible to understand what you mean in eq (1) and (2), notaly the conditional part of the INN, but otherwise the paper cannot be understood alone. It is obvious that this paper builds up on your previous work, and the very short description of cINNs is made to shorten an already lengthy paper. However, for an average user that would not have read the references, this part is very much unclear and would benefit from a few more paragraphs (describing how the loss function is built from Bayes' theorem, and how the "s" and "t" networks are used in a coupling block for example). It does make a bit of redundancy for a reader that followed your previous work, but will improve the "standalone" readability of the paper*

   Author response: We concur with the referee's comments and suggestions regarding the introduction and motivation for cINNs. Taking these into account, we have revised the relevant discussion in the Sec. 2.1 of the updated manuscript. The revised version aims to improve its standalone readability and motivate the application of cINNs more pedagogically, emphasizing the connection between cINNs and Bayes theorem in particular.

2. **Referee's comment**: *Section 2.2. The justification for the periodic spline is a bit lacking, and the link with Sec 2.3 is absolutely not clear. Similarly, the choice of a uniform latent distribution (instead of the usual Gaussian) is not justified, unless it has a link with the periodicity, which emphasizes the need for more explanation.*

   Author response:

   Certain phase space directions exhibit periodic behavior. Without the use of periodic splines, such distributions are poorly reconstructed in the boundary regions, which we referred to as "undesirable boundary effects" in the previous version of our manuscript.

In the updated version, we have clarified this aspect in the begining of Sec. 2.2: "As we will discuss in Sec. 2.3, some phase space directions are periodic and the spline transformations lead to a poor reconstruction of the respective distributions in the boundary regions". We also indicate the periodic parameters in the parametrization adopted for our analysis, as discussed in Sec. 2.3.

Regarding the uniform latent distributions used for periodic dimensions, we agree that this discussion warrants further clarification. The uniform latent distribution is inherently more suitable for the bounded observables encountered in our analysis. This choice of distribution aligns with the periodic nature of the observables, making it more efficient in mitigating the undesirable boundary effects. In the revised version of the paper, we have underlined this reasoning explicitly in the third-last paragraph of Sec. 2.2: "As a last detail, we factorize the base distribution $q(z)$ (compare Eq. (3)) into a Gaussian part for non-periodic and a uniform part for periodic dimensions. The latter is a more natural choice for a bounded feature like an azimuthal angle. Effectively, this amounts to a factor of one for each uniform dimension, ignoring constant terms to the loss. Naturally, we still have to consider the Jacobian for all dimensions." We hope that these modifications address your concerns and provide a clearer understanding of our methodology.

3. **Referee's comment**: *Section 2.3. Could you shortly elaborate on the calculations that lead to 9 dof for a single top? This section seems out of context compared to the rest of the paper. You list a set of variables that fully determine the system, yet afterwards only select a few for your unfolding (even some like b4 that were not in the list). Unless you apply a MMD lose to all these distributions? The MMD is also lacking some details (the reader can follow the reference of course, but it makes the reading again a bit uneasy).*

   Author response: Let's consider the leptonic decay mode of the top quark: $t \to (W \to \ell\nu)b$ where the top quark and W boson are off-shell, and b, ℓ and ν are on-shell and massless. The top quark can be parametrized by four degrees of freedom (dof): invariant mass m_t, transverse momentum $p_{T,t}$, pseudorapidity η_t, and azimuthal angle ϕ_t, in the jet coordinate system. Next, we boost into the rest frame of the top quark: $\tilde{p}_t^t = (m_t, 0)$, $\tilde{p}_W^t = (E_W^t, \vec{p}_W^t)$, and $\tilde{p}_b^t = (m_t - E_W^b, -\vec{p}_W^t)$, where the superscript represent the rest frame, and tilde represents the four-vectors. Due to the on-shell condition for the b, we require only three dofs to determine the four-momentum of the W boson. Next, we boost to the rest frame of the W boson, where only two more dofs are required to parametrize the four-momentum of the ℓ and ν, owing to the two on-shell conditions for ν and ℓ.

   Regarding the selection of observables to parametrize the $t\bar{t}$ system in the unfolding model (Eq. 7), we choose observables that would most effectively capture the relevant new physics phenomena, which in this case is the CP-structure in the Higgs-top interaction. Therefore, we go beyond the basic ones typically used in cartesian or jet coordinate system. For example, we include variables like θ_{CS} and $\Delta\Phi_{\ell d}$ in the parameterization, since they are particularly sensitive to the CP-structure. We choose the 18 observables in Eq. 7 in such a way that it is always possible to fully determine the four-momentum of the $t\bar{t}$ system. It also must be noted that alternative parametrizations are also viable.

We would like to clarify that the phase space parametrization chosen to represent the $t\bar{t}$ system in Eq. 7 does not include b_4. Despite not being included in the list of variables, the generated parton-level distribution for b_4 matches very well with the parton truth distribution, which displays the ability of our unfolding model to fully reconstruct the $t\bar{t}$ phase space at the parton truth-level. We also do not use MMD in our analysis. We elaborate on this choice in the first paragraph of Sec. 2.3 in the updated version of our manuscript.

4. **Referee's comment**: *Overall, the section 2 does not allow for easy reproducibility of the method, considering it is one of the acceptance criteria of the journal. This concerns the technical aspects of the cINN (see comments 1 and 2), but also of the training itself.*

   Author response: As discussed earlier in our answers to referee comments 1, 2, and 3, we have performed a thorough revision of the discussion in Section 2 to improve the readability and structure of the cINN considered in the analysis. These detailed descriptions will enable the reproducibility of our method.

   Regarding the training, we documented our loss function (Eq. (3)) as well as hyperparameters and the used gradient descent algorithm. These should suffice to reproduce our results. Furthermore, note that we use standard methods, which are further documented in our references.

   Overall, the only non-standard method that we use is the periodic splines, which we introduced in section 2.2. Therefore, we are confident that reproducibility is ensured.

5. **Referee's comment**: *Section 3 intro (SMEFT). The rest of the paper being entirely about α, this discussion about SMEFT seems unnecessary. Only the last sentence could justify you can have a CP-observable of dim-4 (thereby motivating the whole use of the method on ttH), but it is not really obvious from the phrasing.*

   Author response: The discussion on SMEFT was included to illustrate the connection between the $\{\kappa_t, \alpha\}$ parametrization, predominantly focused in the paper, and the conventionally utilized dimension-6 SMEFT framework. We realize that the phrasing in

Section 3 might not have made this connection immediately apparent. We have revised the wording at the beginning of Sec. 3 to explicitly highlight this connection: "In this paper, we predominantly focus on the $\{\kappa_t, \alpha\}$ parametrization in Eq. (12), which can be linked to the standard SMEFT framework used for general LHC analyses at mass dimension six."

6. **Referee's comment**: *Section 3.2. Could you elaborate on the fact that background events can be included later in the analysis? Unless mistaken, all the plots are about signal events that are unfolded. How does this compare when the backgrounds are dominant, does it matter, and how does it affect the sensitivity ?*

   Author response: The referee is correct in noting that presented results are from unfolding signal events only. Our focus on signal events was mainly to demonstrate the promising potential of the cINN-based unfolding method in reconstructing the full parton-level phase-space where new physics effects are maximal.

   The inclusion of backgrounds in our analysis is the crucial next step that would impact the sensitivity of our model. We acknowledge that this is an important aspect, and we aim to address it in future work. In Sec. 3.2, we discuss the use of a separate classifier to distinguish the signal from the background. This would allow the unfolding model to maintain its focus on the correct reconstruction of the new physics effects in the signal events at the parton-level, while the background effects can be tackled by the classifier. This would be an important extension of this paper, and we plan to explore these aspects in future work, which we specifically mentioned in the new draft in the first paragraph of Sec. 3.2: "For our simple study, we will sketch how we can avoid modeling this step, in principle. A more thorough investigation of how the inclusion of background affects our method, will be left to future work."

7. **Referee's comment**: *Section 3.3 (Jet combinatorics). You talk about variable number of jets, but nowhere in the text do you explain how you deal with this. Is it the same method as Ref. [60]? If so, this would be worth explaining a bit more, especially considering combinatorics are a major issue in most data analyses.*

   Author response: The method is generally similar, with some distinctions. For example, we do not use the number of jets as an additional input, as done in Ref. [60]. Rather, we consider an input vector of fixed length to parametrize the detector-level information, which includes the four-momentum of up to six light jets, where the missing jets are zero-padded. The network generates the kinematics for the two hard-scattered light jets at the parton level among other candidates, unfolding the detector-level information containing up to six light jets. This explanation has been added to Sec. 3.3

(Results - Jet combinatorics) of the updated version: 'We train one network on SM events without ISR and one network on events with up to six light-flavored jets. In the latter case, we ensure that the input vector to the network has a fixed length by zero-padding the missing jets. We note that our approach to tackle a variable number of jets differs from that in Ref. [60] where the numbers of jets is incorporated as an observable in the training dataset.'

8. **Referee's comment**: *Section 3.3 (Model dependence). Outside the iterative method, how could this bias be dealt with in an analysis? As a systematic uncertainty? Does the fact that the cINN does not extrapolate to non-zero values of α when trained on SM mean that for each α that experimentalists want to try, a new cINN must be trained? Would this not be a major shortcoming?*

   Author response: If one wants to just detect deviations from the SM, it is enough to train the cINN once on SM data. Here, one could use simulated data to check which deviation from $\alpha = 0$ would lead to a significant deviation from the SM in a realistic scenario. If this deviation is not detected by the experiment, this places a bound on α without the need to train the cINN more than once. This bound would be conservative due to the bias towards the SM.

   We added a discussion on this in the Sec. 3.3 of the new draft: "Naturally, any remaining effects of such a bias would lead to a systematic uncertainty on α, which, as mentioned above, is likely to be reducible by the iterative method proposed in Ref. [62]. Luckily the bias here will always be towards $\alpha = 0$, such that deviations from the SM will always infer new physics. Conversely, checking which value of α in our data leads to new physics in simulation, allows us to place a bound on α for data. Any model bias only affects how conservative this bound will be."

   In theory the mentioned issue could only become a serious shortcoming, if one wants to go further and place a bound on a previously detected non-SM α value. However, one would have to see if an iterative approach, as mentioned in the paper (see Ref. [1]) is not sufficient to remove the bias in so far as to make it negligible.

   Note that some inherent model dependence is always present due to missing information on detector level. This is the case for any kind of unfolding technique, including all classical methods, since they are either fitted to SM data or based on Bayes theorem with an SM prior. Hence, part of the model dependence cannot be framed as a shortcoming of our specific method.

9. **Referee's comment**: *Section 4. The assertion that the iterative method could remove the model bias is a bit strong. It is likely, but given that it has not been attempted here, better*

be a bit more conservative.

Author response: We agree with the referee's comments and have revised the text in Section 3 to clarify it: "On the other hand, this bias is significantly larger than the uncertainty band, which suggests that there might be potential for reducing the model dependence through the iterative method proposed in Ref. [62]." Likewise, we have also updated the discussion in Section 4 to clarify it further: "For the former, we have found that there exists a small but significant model dependence, which could be potentially reduced through Bayesian iterative improvement, though this would require a further detailed analysis beyond the scope of the current paper".

10. **Referee's comment**: *General comment. This method looks very promising and would be worth applying to either CMS or ATLAS analyses. Your paper is of course a proof of concept, so it does not address much the challenges met during a real life analysis (exp and th uncertainties, jet energy corrections, possibly datadriven methods, combinatorics, particles outside acceptance, etc), which you can of course not be blamed for. Still, when novel methods like this one are presented in a paper, it is always on a somewhat idealized scenario and for the case that works well (though granted, you consider complications like jet combinatorics with ISR). By experience, "no plan survives first contact with data", and there is always more complication when using novel methods in a real life data analysis. It would therefore be interesting to know your opinion or experience with the challenges such analyses could face that you did not have, when applied to data. It may probably fall outside your scope of the paper, you could therefore disregard this comment, but some considerations specifically targeted at experimentalists that would want to use your technique could potentially enrich your paper and make it more easily advertisable to that audience.*

    Author response: We agree with the referee's comments that, being a proof of concept, it does not fully explore the various challenges encountered in real-life data analysis. A detailed investigation of these challenges is beyond the scope of the present analysis, however, we recognize the importance of addressing them for applications to real data. We plan to explore these challenges more comprehensively in future work to improve the applicability and usage of our method to a broader audience.

We have fully addressed all the referee comments, and we hope that with these clarifications and associated changes made to the manuscript, the paper can be accepted for publication in SciPost.

References

[1] M. Backes, A. Butter, M. Dunford, and B. Malaescu, *An unfolding method based on conditional Invertible Neural Networks (cINN) using iterative training*, arXiv:2212.08674 [hep-ph].