SciPost Phys. Core 9, 028 (2026)

On the reconstruction of kinematic distributions computed with Monte Carlo methods using orthogonal basis functions

Kirill Melnikov[1⋆], Ivan Novikov[1†] and Ivan Pedron[1,2‡]

1 Institute for Theoretical Particle Physics, Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany
2 Institute for Astroparticle Physics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

⋆ kirill.melnikov@kit.edu , † ivan.novikov@kit.edu , ‡ ivan.pedron@kit.edu

Abstract

Reconstruction of one-dimensional kinematic distributions from calculations based on high-dimensional Monte-Carlo integration is a standard problem in high-energy physics. Traditionally, this is done by collecting randomly-generated events in histograms. In this article, we explore an alternative approach, whose main idea is to approximate the target distribution by a weighted sum of orthogonal basis functions whose *coefficients* are calculated using the Monte-Carlo integration. This method has the advantage of directly yielding smooth approximations to target distributions. Furthermore, in the context of high-order perturbative calculations with local subtractions, it eliminates the so-called bin-to-bin fluctuations, which often severely affect the quality of conventional histograms. We also demonstrate that the availability of a high-quality approximation to the target distribution, for example the leading-order result in the perturbative expansion, can be exploited to construct an optimized orthonormal basis. We compare the performance of this method to conventional histograms in both toy-model and real Monte-Carlo settings, applying it to Higgs boson production in weak boson fusion as an example.

Received 2026-01-28
Accepted 2026-04-07
Published 2026-05-19
doi:10.21468/SciPostPhysCore.9.2.028

Check for updates

Contents

1 Introduction

Monte-Carlo methods play a key role in high-energy physics [1–3]. A particular application of such methods that will concern us in this paper, is the computation of distributions of observables in collider processes in the context of perturbative QCD [4].

A key step in the usage of these methods is to organize the simulated events into histograms with finite bins which, if a smooth approximation to the target distribution is desired, can be fitted with a parametric model. An interesting question is whether it is possible to construct the smooth approximation directly from a Monte-Carlo simulation, bypassing histograms as an intermediate step.

One possible approach, that we investigate in this paper, is to decompose the unknown target distribution into a complete orthonormal set of basis functions, and compute their coefficients using conventional Monte-Carlo methods. These coefficients can be thought of as *moments* of distributions, since they are given by their weighted integrals. Importantly, this procedure is as efficient as constructing a histogram, and can be used to compute a large number of observables simultaneously.

We note that this idea may not be entirely new [5]. For example, it has been used in Ref. [6] to determine the distribution of scattering angles in molecular collisions, and it is possible that there are other examples that we are not aware of. At the same time, histograms have become a method of choice in theoretical high-energy physics due to their conceptual simplicity, robustness, and ability to handle kinematic distributions spanning multiple orders of magnitude.

A complete set of basis functions is, in general, infinite-dimensional, and in any practical calculation it needs to be truncated. Although one may think that this is a challenge for accurately reproducing the target functions, we note that many kinematic distributions in high-energy physics are smooth and, roughly, have the shape of an asymmetric bell curve (transverse momenta distributions) or a smoothed rectangle (rapidity distributions). Because of this, one can expect that a limited number of basis functions would be sufficient to adequately describe them.

The method of moments has an additional key benefit when used with perturbative calculations based on local subtraction schemes [7–15]. Real and virtual corrections to differential cross sections exhibit infrared singularities, when taken separately. Local subtraction schemes add terms to the real corrections to subtract these singularities in the soft and collinear limits. From a computational Monte Carlo perspective, subtraction terms provide counter-events for each generated event, making their combination integrable. Although these counter-events are only needed in kinematic regions where additional radiation becomes indistinguishable from the no-radiation case, the counter-events often extend to the whole phase space. Since the kinematics of an event and counter-events differ, they may end up in separate histogram bins. Although this rarely happens, such events are, in some sense, catastrophic, and it takes a significant amount of CPU time to recover from them.

Moments, on the other hand, are largely immune to this issue. Indeed, since weights that are used to calculate moments are continuous functions of the kinematic observables, all events and counter-events contribute to each moment, which provides a natural way to smooth the potentially offending contributions.

An important question when using moments is the choice of basis functions. Obviously,

there is significant freedom in their choice, which can be exploited to construct the most suitable basis for describing a particular target distribution. Indeed, one can employ known sets of orthogonal functions such as harmonic functions, Legendre or Chebyshev polynomials etc. [16, 17]. However, as we will argue below, this is often not *an optimal choice*. Partially, this is related to the fact that kinematic distributions in collider physics can change by several orders of magnitude on the interval of interest. This feature is not conducive to the reconstruction of the target distribution from a relatively small number of moments.

However, if a fairly high-quality approximation to a particular target distribution is available, it can be used to construct an orthonormal basis that is optimized for describing this distribution using moments.[1] In the context of perturbative computations for collider physics, such approximations can be obtained from the leading-order distributions, which are much easier to calculate than the same distributions at higher perturbative orders. From the perspective of moments calculations, leading-order results provide high-quality approximations because QCD corrections change their shapes by tens of percent at most, whereas distributions themselves can change by orders of magnitude.

When such an optimized basis is used, the first basis function is proportional to the chosen initial approximation, and other basis functions are variations needed to describe higher-order corrections to the shape of such a distribution. In comparison to e.g. Legendre moments, an optimized basis converges faster and avoids large fluctuations in regions where the target distribution is small.

The paper is organized as follows. In Section 2 we introduce the idea of moments as an alternative to histograms, and provide the relevant mathematical details. We also discuss how to choose the highest moment to retain in the calculation, and estimate the truncation error based on the decay rate of the moments coefficients. In Section 3 we compare how the method of Legendre moments[2] performs in comparison to conventional histograms for a number of toy Monte-Carlo examples. In Section 4 we apply these ideas to a real-world example – the calculation of the next-to-next-to-leading-order (NNLO) QCD corrections to Higgs boson production in weak-boson fusion (WBF). In Section 5 we show how to use an initial approximation to the target distribution to construct an orthonormal basis that is optimized for computing it, and we apply this idea to WBF. We conclude in Section 6.

2 Legendre moments

2.1 Approximation of distributions with moments

In this section we explain how to calculate the Legendre moments and use them to reconstruct a target distribution. Suppose that we want to approximate an unknown target distribution $f(x)$ for a variable x on an interval $x \in [a, b]$. We assume that for $x \in [a, b]$, there exists a complete set of functions $e_k(x)$, which are orthonormal

$$\langle e_k, e_l \rangle = \delta_{kl} \,, \tag{1}$$

with respect to a particular scalar product

$$\langle f, g \rangle = \int_a^b w(x)\, f(x)\, g(x)\, \mathrm{d}x \,. \tag{2}$$

[1] We note that a similar idea was discussed in Ref. [18] in the context of designing a flexible parametrization for the B-meson shape function. However, we believe that it was never applied for computing higher-order QCD effects in kinematic distributions relevant for collider physics.

[2] We have used different orthogonal polynomials and found rather similar results. For this reason, we will focus on Legendre polynomials in the rest of this paper.

In Eq. (2) the weight function $w(x)$ is arbitrary except for the fact that it should be positive-definite, $w(x) > 0$, on the interval $x \in [a, b]$. As we will explain below, the choice of $w(x)$ should be driven by the features of the target distribution that we try to approximate. Making a particular choice for $w(x)$ may lead to a faster convergence of the moments-based approximation, especially in regions where $w(x)$ is large.

Equipped with the scalar product Eq. (1), we can express the target distribution as follows

$$f(x) = \sum_{k=0}^{\infty} c_k \, e_k(x) \,, \tag{3}$$

where

$$c_k = \langle f, e_k \rangle = \int_a^b f(x) \, w(x) \, e_k(x) \, \mathrm{d}x \,, \tag{4}$$

is the k-th moment of the function $f(x)$ with respect to the chosen basis.

As we mentioned in the Introduction, we will consider Legendre polynomials as a prototypical example of orthogonal polynomials. They are defined as[3]

$$P_k(x) = \frac{1}{2^k k!} \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^k (x^2 - 1)^k \,. \tag{5}$$

Legendre polynomials form a complete basis on the interval $x \in [-1, 1]$ and are orthogonal to each other

$$\int_{-1}^{1} P_k(x) P_l(x) \mathrm{d}x = \frac{2\delta_{kl}}{2k+1} \,. \tag{6}$$

To remap the Legendre polynomials to the interval $[a, b]$, we use a smooth monotonic variable transformation

$$t = t(x) \,, \qquad t(b) = 1 \,, \qquad t(a) = -1 \,. \tag{7}$$

Accounting for the normalization change, we choose the basis functions as

$$e_k(x) = \sqrt{k + 1/2} \; t'(x) P_k(t(x)) \,, \tag{8}$$

where we used the notation $t'(x) = \mathrm{d}t/\mathrm{d}x$. The functions in Eq. (8) are orthonormal with respect to the weight function $w(x) = 1/t'(x)$ on the interval $x \in [a, b]$. We note that the simplest choice for $t(x)$ is the linear mapping

$$t(x) = \frac{2x - a - b}{b - a} \,, \tag{9}$$

in which case the weight evaluates to $w(x) = (b - a)/2$.

We now discuss the computation of such moments for observables in scattering processes at colliders. Our starting point is an observation that for a given set of basis functions e_i, the relevant moments can be evaluated using the standard Monte-Carlo methods. Indeed, suppose that Φ describes the phase space of particles in a scattering process. The observable $\mathcal{O}$ is a function of a phase-space point $\phi \in \Phi$.

The distribution of the variable $\mathcal{O}$ is given by

$$\left. \frac{\mathrm{d}\sigma}{\mathrm{d}\mathcal{O}} \right|_{\mathcal{O}=x} = \int_{\Phi} \delta\left(x - \mathcal{O}(\phi)\right) \frac{\mathrm{d}\sigma}{\mathrm{d}\phi} \mathrm{d}\phi \,, \tag{10}$$

[3] Legendre polynomials can be efficiently evaluated using the Bonnet's recursion formula, and the Legendre series can be accurately evaluated using the Clenshaw algorithm [19].

where $\mathrm{d}\sigma/\mathrm{d}\phi$ is the differential cross section. Hence, the k-th moment of the $\mathcal{O}$-distribution in the interval $\mathcal{O} \in [a, b]$ reads

$$\left\langle \frac{\mathrm{d}\sigma}{\mathrm{d}\mathcal{O}}, e_k \right\rangle = \int_\Phi \Theta_a^b(\mathcal{O}(\phi)) w(\mathcal{O}(\phi)) e_k(\mathcal{O}(\phi)) \frac{\mathrm{d}\sigma}{\mathrm{d}\phi} \mathrm{d}\phi \,, \tag{11}$$

where

$$\Theta_a^b(x) = \begin{cases} 1\,, & a < x < b\,, \\ 0\,, & \text{otherwise.} \end{cases} \tag{12}$$

Integration over Φ can be performed using the standard Monte-Carlo methods. In general, phase-space points $\phi \in \Phi$ are sampled from a probability density $p(\phi)$. Therefore, the integral over Φ

$$I = \int_\Phi f(\phi)\, \mathrm{d}\phi \,, \tag{13}$$

can be interpreted as the expectation value of the function $f(\phi)/p(\phi)$, i.e.

$$I = \mathbb{E}\left[\frac{f(\phi)}{p(\phi)}\right]. \tag{14}$$

This expectation value can be estimated by drawing N random samples of the phase-space points ϕ_i, $i \in \{1, .., N\}$ from the probability distribution $p(\phi)$, and computing the average value

$$\bar{I} = \frac{1}{N} \sum_{i=1}^{N} I_i \,, \tag{15}$$

where

$$I_i = \frac{f(\phi_i)}{p(\phi_i)}. \tag{16}$$

Then, the integral I is estimated using the following formula

$$I = \bar{I} \pm \delta I \,, \tag{17}$$

where

$$\delta I = \sqrt{\frac{1}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} I_i^2 - \bar{I}^2\right)}, \tag{18}$$

is the standard deviation.

We note that many moments for the multitude of observables can be evaluated in a *single* Monte-Carlo integration, using the same set of generated phase-space samples. In fact, the construction of a histogram can be viewed as an evaluation of the integral in Eq. (11) since a histogram with the bin extents $[x_{k,\mathrm{low}}, x_{k,\mathrm{high}}]$ corresponds to rectangular functions

$$e_k(x) = \frac{\Theta_{x_{k,\mathrm{low}}}^{x_{k,\mathrm{high}}}(x)}{\sqrt{x_{k,\mathrm{high}} - x_{k,\mathrm{low}}}} \,, \tag{19}$$

with $w(x) = 1$. This set of functions is obviously orthonormal but not complete.

2.2 Truncation of moments

In any calculation that involves numerical computation of moments, the series in Eq. (3) must be truncated. This truncation causes an error in the reconstructed distribution that needs to be estimated. In addition, the numerical calculation of moments results in uncertainties in their values, which also lead to errors in the reconstruction of the target distribution.

The two sources of uncertainties have an interesting interplay, which makes the number of terms that should be retained in the series in Eq. (3), dependent on the Monte-Carlo integration uncertainty. This interplay happens because for any sufficiently smooth function $f(x)$ its moments $\langle f, e_k\rangle$ *decrease* when k increases. At the same time, the Monte-Carlo integration errors in the estimates of $\langle f, e_k\rangle$ do not decrease, and at some point the error becomes larger than the true value of the moment $\langle f, e_k\rangle$. It is then reasonable to expect that the optimal number of moments to include is determined by the moment whose "natural" value is of the order of the Monte Carlo integration error. The clarification "natural" here is important because it allows us to ignore cases where the actual value of a particular moment is accidentally close to zero, but higher moments are not small.

To this end, we need to design a procedure to find "natural" values of moments for a particular distribution. Our starting point is an observation that Legendre moments of analytic functions decrease exponentially [20]. Therefore, we expect that the k-th moment of a typical distribution can be estimated as follows

$$|\langle f, e_k\rangle| \sim c\, r^k\,, \tag{20}$$

where c and $r < 1$ are distribution-dependent constants.

We note that the sign of a moment is unknown a priori, and that individual moments may, in fact, be much smaller than what this estimate suggests. With this in mind, we model moments as independent normally-distributed random numbers with zero mean and variance $\sigma^2 = (cr^k)^2$. Hence, we write

$$\langle f, e_k\rangle \sim \mathcal{N}(0, cr^k)\,, \tag{21}$$

where

$$\mathcal{N}(x_0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}\,. \tag{22}$$

The Monte-Carlo simulation yields moments $\hat{\mu}_k = \langle f, e_k\rangle + \varepsilon_k$ with some integration error ε_k and an estimated variance $\delta\hat{\mu}_k^2$ of an integration error. Then

$$\varepsilon_k \sim \mathcal{N}(0, \delta\hat{\mu}_k)\,, \tag{23}$$

for all moments k with $0 \le k \le m-1$ for some m. Hence, we assume that calculated moments are drawn randomly from the distribution

$$\hat{\mu}_k = \langle f, e_k\rangle + \varepsilon_k \sim \mathcal{N}\left(0, \sqrt{c^2 r^{2k} + \delta\hat{\mu}_k^2}\right), \tag{24}$$

where the coefficients c, r are unknown. We then find them by fitting the values of computed moments $\hat{\mu}_k$. Technically, the fit is performed by numerically minimizing the log-likelihood function

$$\mathcal{L}(c, r) = \sum_{k=0}^{m-1} \left[\ln\left(c^2 r^{2k} + \delta\hat{\mu}_k^2\right) + \frac{\hat{\mu}_k^2}{c^2 r^{2k} + \delta\hat{\mu}_k^2} \right]. \tag{25}$$

This procedure yields an estimate of the "natural" size of moments $\langle f, e_k\rangle \sim cr^k$ that we use to compare against the Monte-Carlo integration error.

Finally, to determine the number of moments that is retained for computing the target distribution, we find the first moment $n+1$ for which the inequality $\delta\hat{\mu}_{n+1}^2 > c^2 r^{2(n+1)}$ holds.

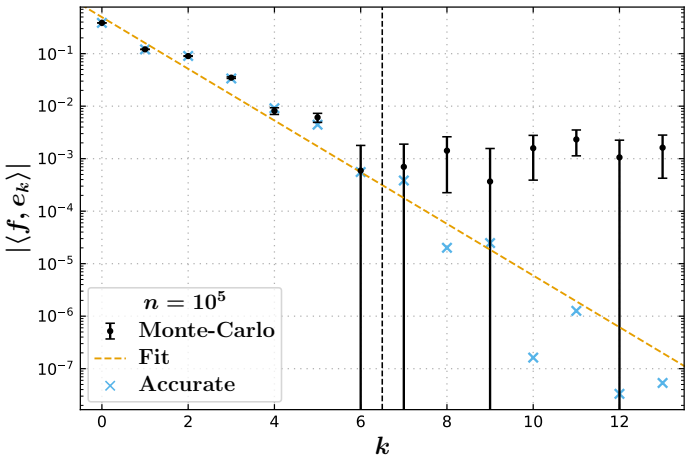

Figure 1: Legendre moments $\langle f, e_k\rangle$ of the normal distribution $f = \mathcal{N}(0,1)$ (see Eq. (22) for the definition) on the interval $[-1,2]$. The blue "Accurate" crosses are the true Legendre moments, calculated with high precision. The black data points are the moments calculated in a toy Monte-Carlo with $n = 10^5$ samples. The orange dashed line is the estimate $\langle f, e_k\rangle \sim cr^k$ of the typical magnitude of the Legendre moments, obtained from the Monte-Carlo data (black points). The vertical gray dashed line marks the truncation point, determined according to the prescription described in Sec. 2.2. The distribution itself and its reconstruction from the retained moments are shown in Figure 2.

Then, Eq. (3) is computed by including n moments in the sum on the right-hand side. We note that the number m of moments that are calculated numerically by means of the Monte-Carlo integration is chosen to be larger than the number n of moments that are used to evaluate the target distribution.

We illustrate the use of this prescription with an example in Figure 1, where we use moments to describe the function $e^{-x^2/2}/\sqrt{2\pi}$ on the interval $x \in [-1,2]$. Because the true values of moments $\langle f, e_k\rangle$ decrease with increasing k while the Monte-Carlo errors tend to a constant, there exists a moment k^* whose Monte-Carlo estimate gets dominated by statistical noise. In the current example this happens around $k^* \sim 6$.

Monte-Carlo estimates of moments can be considered reliable only to the left of this point, while to the right of it, the Monte-Carlo estimates are many orders of magnitude larger than the true values. The orange dashed line shows the trend $\langle f, e_k\rangle \sim cr^k$, with c and r estimated by minimizing the log-likelihood in Eq. (25). We remind the reader that this estimate is based on the results of the Monte-Carlo computation of the moments, but without the knowledge of the true values of $\langle f, e_k\rangle$. In spite of this, the determined trend $\langle f, e_k\rangle \sim cr^k$ provides a reasonable, albeit slightly conservative estimate of the possible magnitudes of the true values of $\langle f, e_k\rangle$. The vertical dashed line marks the truncation point determined by the prescription described above, — in this case, the first 7 moments are retained to provide an estimate of the target distribution. We will revisit this example once again in the next section.

Once the number of retained moments is fixed and their asymptotic form is established, it becomes possible to estimate the total error of the approximation. To this end, for $x \in [a,b]$ we estimate the basis functions as $e_k(x) \sim \sqrt{1/(b-a)}$, and compute the remaining truncation error. We find

$$\sum_{k=n}^{\infty} \langle f, e_k\rangle\, e_k \sim \sqrt{\sum_{k=n}^{\infty} \frac{c^2 r^{2k}}{b-a}} = \frac{cr^n}{\sqrt{(b-a)(1-r^2)}}. \tag{26}$$

Then, the total approximation error is estimated as follows

$$\left|\sum_{k=0}^{n-1}\hat{\mu}_k e_k(x) - f(x)\right| = \left|\sum_{k=0}^{n-1}\varepsilon_k e_k(x) - \sum_{k=n}^{\infty}\langle f, e_k\rangle\, e_k(x)\right| \lesssim \sqrt{\sum_{k=0}^{n-1}\delta\hat{\mu}_k^2 (e_k(x))^2 + \frac{c^2 r^{2n}}{(b-a)(1-r^2)}}\,. \tag{27}$$

In the last step we assumed that the Monte-Carlo errors in different coefficients are uncorrelated. Although this is not exactly true, we find it to be acceptable to neglect the correlations between the moments for estimating the uncertainty. The truncation criterion detailed below Eq. (25) minimizes the total approximation error.

Higher basis functions tend to be more oscillatory, and the truncation procedure described above amounts to discarding higher frequencies in the approximate distribution. This means that this truncation procedure is biased towards smoother distributions, which we actually see as a desirable form of regularization.

3 Toy examples

To illustrate the main features of the method of moments, introduced in the previous sections, we elaborate further on the toy model introduced there. To this end, we consider the normal distribution $\mathcal{N}(0,1)$ as the target distribution. We attempt to approximate it on the interval $x \in [-1,2]$. We do this by sampling points uniformly and then using these samples, weighted by the value of the function $\mathcal{N}(0,1)$, to either construct a histogram-based or a moments-based approximation to it.

Figure 2 shows the results for $n = 10^4$ and $n = 10^5$ Monte-Carlo samples. The bottom panel in each of the four plots shows the relative deviation of the two kinds of approximations from the true result. The number between parentheses in the legend for the moments is the number of retained moments after applying the truncation procedure described in Section 2.

The bands around the histograms show the Monte-Carlo integration uncertainty estimated according to Eq. (18). The uncertainty in the moments-based approximations is computed using Eq. (27); as we explained in the previous section, it accounts for both the Monte-Carlo and the truncation uncertainties. We emphasize that for each plot in Figure 2, the histograms and the moments-based approximations are constructed from the same set of generated events.

The following features of various plots in Figure 2 are worth commenting upon.

- With increased statistics, both methods converge to the true distribution.
- Higher statistics allows us to compute more moments reliably, and to incorporate them into the moments-based approximation, according to the truncation prescription that balances the Monte-Carlo and truncation errors.
- The differences between the histograms, the moments-based approximation and the true function are comparable in size *around the peak of the distribution*.
- An important advantage of moments over histograms is that they yield smooth approximations of the target distribution. This is especially obvious when we compare them to a histogram with *narrow bins*. Such a comparison is shown in the bottom-right plot of Figure 2. With narrow bins, the number of events in each bin is smaller, and the statistical fluctuations are larger. Moments, on the other hand, include information from all the generated samples, and they effectively smooth out these statistical fluctuations.

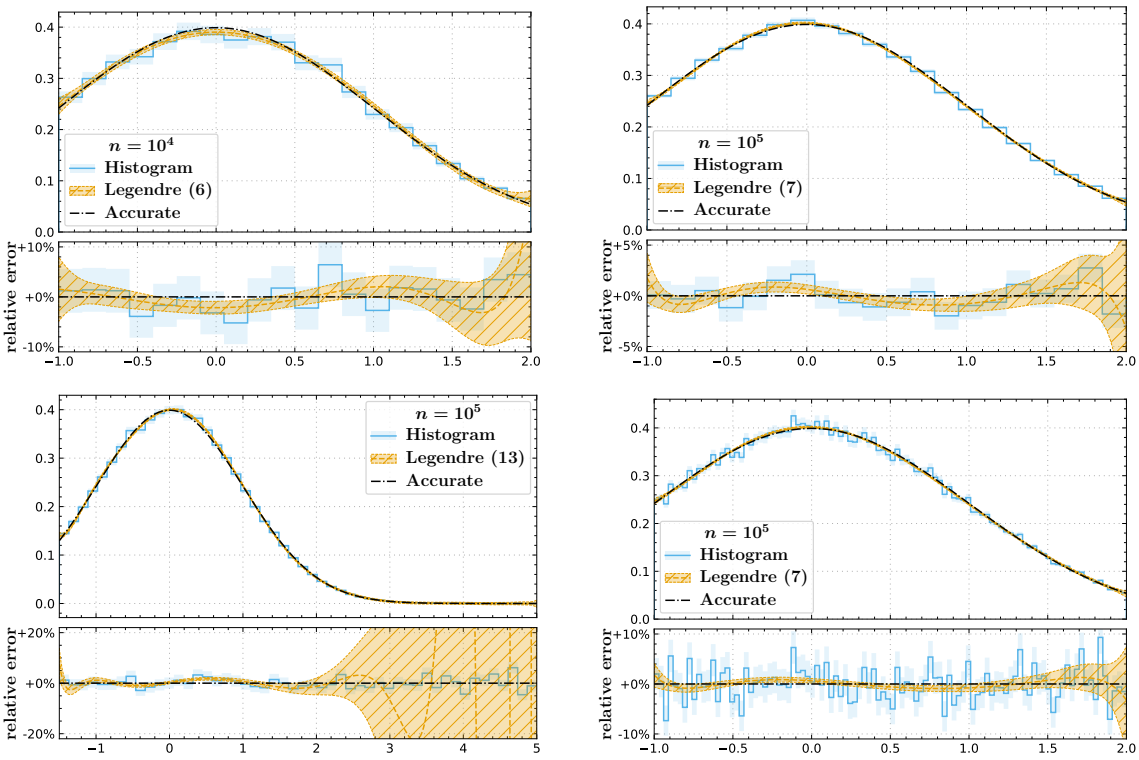

Figure 2: Normal distribution $(2\pi)^{-1/2}\exp(-x^2/2)$ reconstructed from n samples using a histogram or Legendre moments. All plots, except for the bottom-left one, compare the two approximations on the interval $[-1, 2]$. The bottom-left plot shows the approximation on the interval $[-1.5, 5]$. The plots in the top row differ in the number of Monte-Carlo samples n. The histograms in the right column differ in the bin width, while the Legendre approximation is the same. The number in parenthesis indicates the number of retained terms in the Legendre series, which is determined from the estimated Monte-Carlo uncertainties according to the truncation prescription described in Sec. 2.2. The bottom panels show the relative deviation of the reconstructed distributions from the true distribution. For each plot, the histogram and Legendre approximation are reconstructed using the same Monte-Carlo samples.

- For large values of x, the relative error of the moments-based distribution becomes very large, as illustrated in the bottom-left plot in Figure 2. This is related to the fact that the approximation using standard polynomials provides a nearly uniform approximation error across the whole interval. The exponential suppression of the target distribution at large x can be achieved by a precise point-by-point cancellation of infinitely many contributions to the sum in Eq. (3). The truncation, that we are forced to do, violates this requirement and leads to a significant deterioration in the quality of the approximation based on moments. With higher statistics, the region where this problem occurs moves to the right since it becomes possible to include more moments, but it never disappears completely. On the contrary, the relative error of a histogram is uniform, because the absolute error is proportional to the average value of the function in each bin, and we sample points uniformly.

Throughout this discussion, we assumed that the target distribution $f(x)$ is smooth and does not possess any high-frequency features. If this assumption is violated, the moments-based approximation can be expected to perform worse. For illustration, consider the left panel in Figure 3, which displays a mixture of two Gaussians: the first one has mean zero, standard

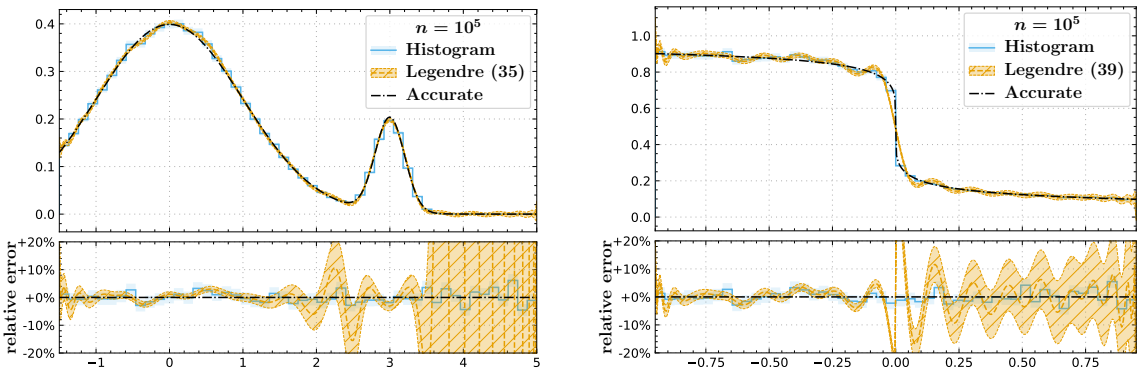

Figure 3: Examples of distributions that are more difficult to approximate using Legendre moments. The distribution on the left is a mixture of two gaussian peaks. The distribution on the right, defined in Eq. (28) has an infinite derivative at $x = 0$. The number in parenthesis indicates the number of retained terms in the Legendre series, which is determined from the estimated Monte-Carlo uncertainties according to the truncation prescription described in Sec. 2.2. The bottom panels show the relative deviation of the reconstructed distributions from the true distribution.

deviation one, and unit integral, whereas the second one is centered at $x = 3$, has a smaller standard deviation of 0.2, and the integral 0.5. The presence of a second, relatively sharp peak makes it difficult to represent the overall distribution accurately using Legendre moments. In fact, we observe that the number of moments required for a reasonable description is several times larger than in the previous examples. Moreover, the relative approximation error remains elevated even in the regions far from the second peak.

Another pathological example is shown in the right plot of Figure 3. It refers to the following target distribution

$$f(x) = \begin{cases} 1 - \frac{\alpha}{1+\sqrt{\beta|x|}}, & x < 0, \\ \frac{\alpha}{1+\sqrt{\beta|x|}}, & x > 0, \end{cases} \tag{28}$$

with parameters $\alpha = 0.4$ and $\beta = 10$. This function has an infinite derivative at $x = 0$. Again, many Legendre moments are needed to describe this function, and the approximation error is relatively large.

A possible way to deal with such problems is to split the x-interval into several sub-intervals, and construct the moments-based approximation for each sub-interval separately. We did not pursue this issue further in the current study. Instead, we continue with the discussion of how the moments-based approximation fare in practice, and how the issue of large uncertainties in tails of distributions can be addressed.

4 Moments in practice

After the toy model, we consider a realistic example. As mentioned in the introduction, our primary interest lies in perturbative QCD calculations for collider processes, and the example that we would like to discuss falls into this category. We implemented the method of moments in the NNLO Monte-Carlo integrator for Higgs boson production in weak-boson fusion (WBF) that was developed in Ref. [21]. The input parameters and the WBF event selection criteria are the same as in Ref. [21]. We will mostly focus on the distributions of the transverse momentum $p_{\perp,H}$ and the rapidity y_H of the Higgs boson, produced in the weak-boson-fusion, since these two distributions provide archetypal examples.

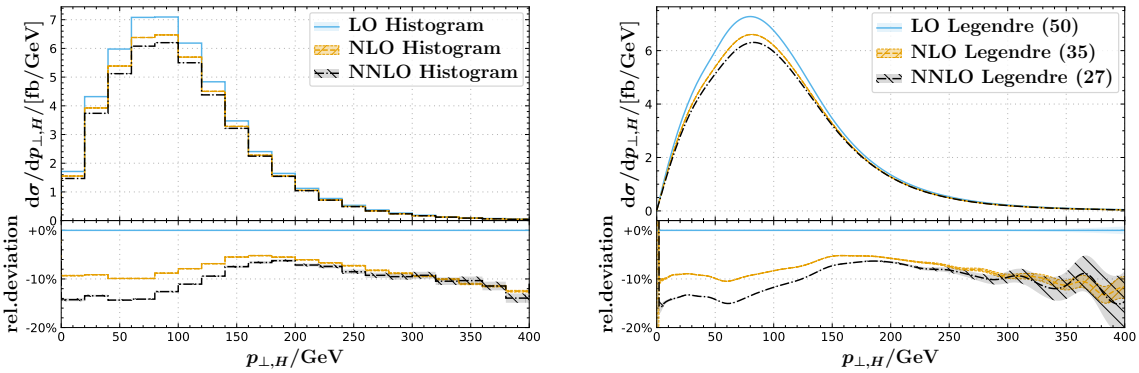

Figure 4: Distribution of transverse momentum $p_{\perp,H}$ of Higgs boson produced in WBF, reconstructed using histograms (left) or Legendre moments (right). The blue, orange, and black lines correspond to the leading-, next-to-leading, and next-to-next-to-leading order result. The bottom panels show relative deviations from the leading-order one. For Legendre moments the number in brackets indicates the number of moments retained after the truncation procedure described in Section 2.2. The uncertainty bands account for Monte-Carlo integration uncertainty, and, for Legendre moments, the truncation uncertainty.

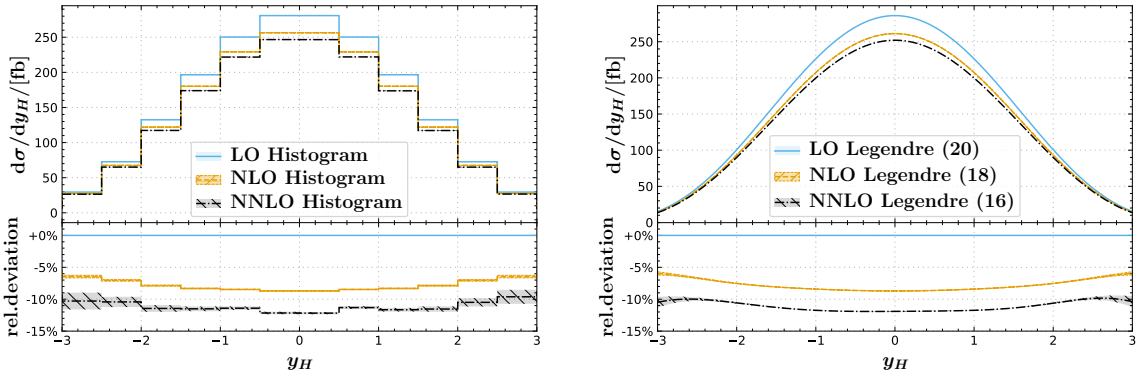

Figure 5: Distribution of rapidity y_H of Higgs boson produced in WBF, reconstructed using histograms (left) or Legendre moments (right). The blue, orange, and black lines correspond to the leading-, next-to-leading, and next-to-next-to-leading order result. The bottom panels show relative deviations from the leading-order one. For Legendre moments the number in brackets indicates the number of moments retained after the truncation procedure described in Section 2.2. The uncertainty bands account for Monte-Carlo integration uncertainty, and, for Legendre moments, the truncation uncertainty.

Figures 4 and 5 show the $p_{\perp,H}$ and y_H distributions at different perturbative orders, reconstructed using histograms and Legendre moments. The distributions, obtained with two different methods, agree within the uncertainties. As we have already discussed in the toy examples, distributions reconstructed from the Legendre series have a larger relative error in the tails of the distributions.

We can use this example to illustrate one of the key strengths of the method of moments, which is related to the fact that moments provide a natural way to smooth distributions. Indeed, as we have mentioned previously, in calculations that employ local subtractions, contributions of events with additional real partons need to be combined with corresponding counter-events to ensure cancellation of infrared singularities. Since events and counter-events

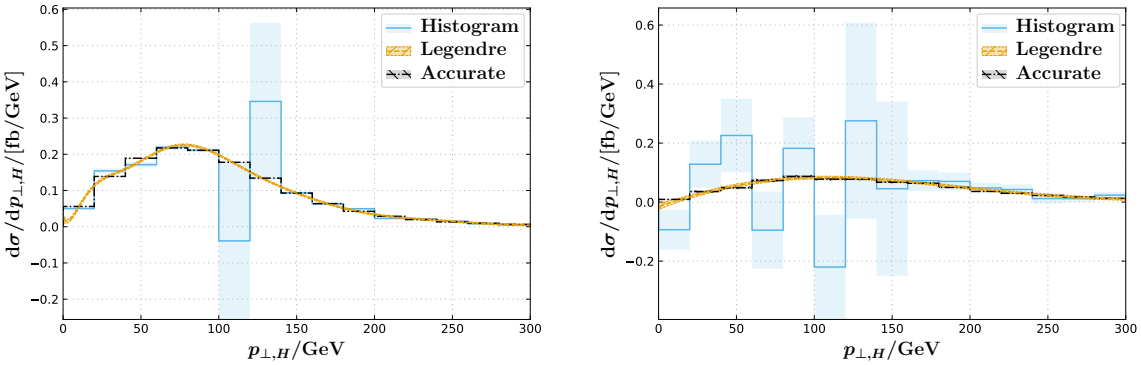

Figure 6: Examples of bin-to-bin fluctuations in some infrared-subtracted double-real NNLO corrections to WBF. The solid blue histogram ("Histogram") and the orange dashed line ("Legendre") show the correction reconstructed from the same set of generated events using a conventional histogram and using Legendre moments, respectively. The black dash-dotted histogram ("Accurate") serves as a proxy for the true distribution, as it was generated using over 100 times more statistics than in the other two approximations. The left plot shows a prominent example of a bin-to-bin fluctuation, where an event and a counterevent landed on opposite sides of the bin edge at $p_{\perp,H} \approx 120\,\text{GeV}$. The right plot shows a different infrared-subtracted double-real correction, which is dominated by such fluctuations.

can have somewhat different kinematics, they can end up in different histogram bins. In this case the necessary cancellation between events and counter-events does not happen, leading to to a dramatic fluctuation of results recorded in neighboring bins. This phenomenon is illustrated in Figure 6, with examples taken from some distributions contributing to the NNLO double-real corrections to WBF.

This problem is related to the fact that the integral over a bin of a histogram is a scalar product with a *discontinuous* function, the indicator function of the bin. As a result, at the border between the neighboring bins soft and collinear radiation can change to which bin the given event is assigned.[4] Although this may sound like a violation of the infrared safety, since this problem only occurs at the edges of histogram bins, which form a measure-zero subset of the available phase space, this phenomenon does not render the integral divergent. However, it does cause problems in numerical simulations with a finite number of samples since once such a mis-cancellation happens, significant statistics is required to overcome it, and converge to the right result.

The difference between kinematics in the event and the counter-event is related to the choice of the phase-space parameterization. It may be possible to design a phase-space parametrization that optimizes the choice of kinematics for counter-events, ensuring that the value of this observable is nearly the same in the event and all corresponding counter-events. However, even if one succeeds in finding a parametrization that has this property for one observable, it is not guaranteed to hold for the other ones. Designing a specialized phase-space parametrization for each considered observable is impossible in practice.

A possible way to mitigate bin-to-bin fluctuations is to distribute an event weight across different bins, a procedure which is usually referred to as "bin smearing". Different variants of bin smearing have been proposed, including fractional filling [22] and joint manipulation of the weights of the events and counterevents using the understanding of the singularity structure of the integrand [23]. Although these techniques are powerful, they introduce yet another smearing of original distributions, which the method of moments strives to avoid.

[4] The same issue exists at the boundaries of the acceptance regions needed for computing fiducial cross sections.

In fact, moments are largely immune to mis-cancellations between events and counter-events. Because moments are scalar products with *continuous* basis functions, a small change in the kinematic of the observable computed for the event and the counter-event leads to a similarly small change in their contributions to the moments. Correspondingly, the distributions reconstructed from Legendre moments do not show any sign of fluctuations between histogram bins, see Figure 6. We emphasize that the histograms and the Legendre series compared in each plot in Figure 6 are constructed in the same Monte-Carlo run using identical samples.

5 Basis optimization

The challenge in the moments-based reconstruction approach, that we have witnessed repeatedly in the previous sections, is that their performance deteriorates in exponentially-suppressed tails of distributions. The goal of this section is to address this issue. In particular, we will explain how a *reasonably-accurate approximation to the target distribution* can be used to construct an optimized basis for computing it. In the context of perturbative calculations, reasonably-accurate approximations to target distributions can often be inferred from the leading-order computations of the corresponding observables.

In Section 2 an orthonormal basis was constructed by mapping the target distribution from the interval $[a,b]$ to the interval $[-1,1]$ with the help of a linear variable transformation, and then using the Legendre polynomials in the transformed coordinates as the basis functions. The main idea behind the basis optimization is to trade a linear variable transformation for a non-linear one, and to choose it in such a way that a reasonably-accurate approximation to the target distribution becomes flat in the new coordinates.

Hence, if we are interested in the kinematic distribution characterized by a variable x defined on the interval $x \in [a,b]$, and if $f_0(x)$ is the approximate target distribution, we choose

$$t(x) = -1 + 2\frac{\int_a^x f_0(\tilde{x})\mathrm{d}\tilde{x}}{\int_a^b f_0(\tilde{x})\mathrm{d}\tilde{x}}\,, \tag{29}$$

as the variable transformation $x \to t$. This transformation can only be employed if the function $f_0(x)$ is sign-definite since, otherwise, the variable transformation $x \to t$ is not invertible.

We define the basis functions as follows

$$e_k(x) = \sqrt{k+1/2}\; t'(x) P_k(t(x))\,. \tag{30}$$

It is easy to see that these functions are orthonormal on the interval $x \in [a,b]$ with the weight function

$$w(x) = \big(t'(x)\big)^{-1} \propto \big(f_0(x)\big)^{-1}\,. \tag{31}$$

The completeness of this basis follows immediately from the completeness of Legendre polynomials. To distinguish the optimized basis from the standard Legendre polynomials, we call this basis "rectified".

Suppose that the kinematic distribution $f_0(x)$ differs from the true distribution $f(x)$ by a function $K(x) = f(x)/f_0(x)$. If the initial approximation $f_0(x)$ is the leading-order distribution, then K plays the role of a differential K-factor, which for many perturbative computations

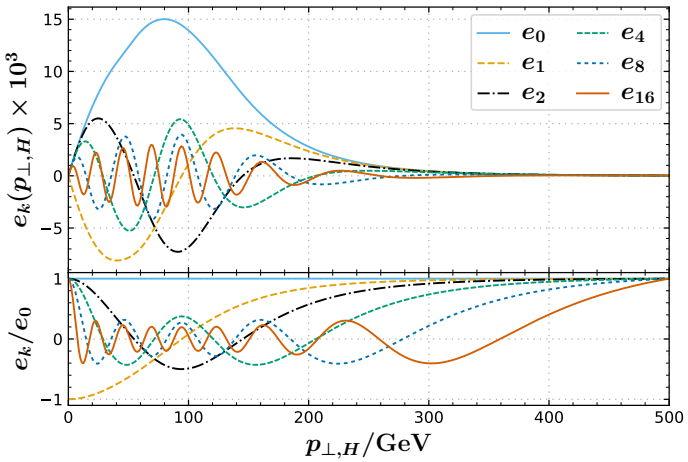

Figure 7: Basis functions constructed from an initial approximation of the distribution of Higgs boson transverse momentum in weak-boson fusion. The bottom panel illustrates the relative difference between different basis functions.

is a slowly-varying order-one function.[5] The moments of $f(x) = K(x)f_0(x)$ then evaluate to

$$c_k = \sqrt{k+1}\,\frac{\int_a^b \mathrm{d}x f_0(x)}{2}\int_{-1}^{1}\mathrm{d}t P_k(t)K(x(t))\,. \tag{32}$$

Hence, if $K(x(t))$ is flat, only a relatively small number of Legendre polynomials will significantly differ from zero.

To give an example, in Figure 7 we show the basis functions e_k for the transverse momentum distribution of the Higgs bosons produced in the weak-boson fusion. We consider the transverse momenta on the interval $0 < p_{\perp,H} < 500$ GeV. The function $f_0(x)$ is identified with the leading-order transverse momentum distribution. We approximate it by a truncated sum of Legendre polynomials calculated as in the previous sections. This can be easily done because leading-order distributions are typically inexpensive to calculate, which means that they can be accurately described, even if they are not known analytically.

It follows from Eq. (30) that the first basis function $e_0(x)$ is proportional to $f_0(x)$, and all other basis functions feature the exponential suppression of the high-$p_\perp$ region as well. Although basis functions do oscillate, as dictated by the presence of the Legendre polynomials in Eq. (30), these oscillations are mostly confined to the region around the peak of the distribution. Therefore, one can expect that this set of basis functions will describe fine features of the target distribution around its peak, and provide a smooth approximation to the target distribution in its tail.

A comparison of the calculation of the Higgs boson transverse momentum distribution based on the standard basis of Legendre polynomials, and the calculation with the optimized basis, is presented in Figure 8. We note that the number of moments that are needed to adequately describe the distribution in the two cases differs by nearly a factor three, indicating faster convergence if the optimized basis is used. Because all functions in the optimized basis are constructed to reproduce the exponentially-suppressed tail of the leading-order result, the calculated distributions at higher orders become proportional to $f_0(x)$ in the tail and do not exhibit any large fluctuations.

[5] The function $K(x)$ is "slowly-varying" when compared to possible exponential changes of target distributions $f(x)$ and $f_0(x)$.

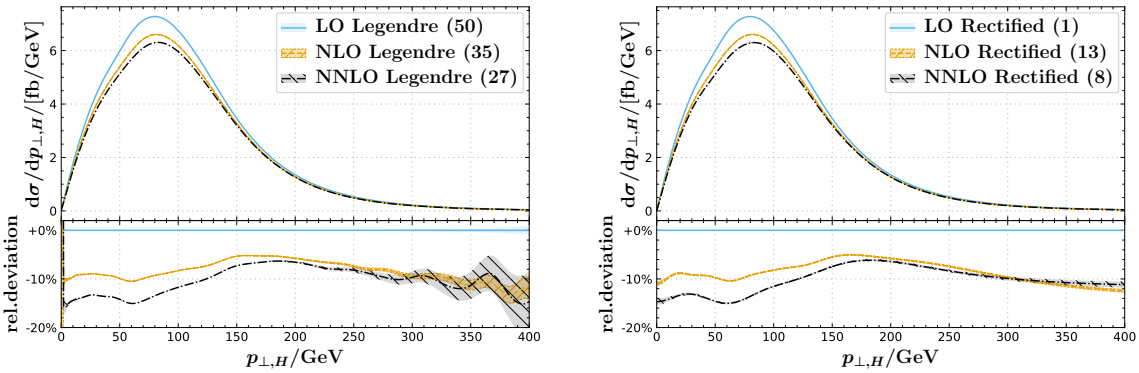

Figure 8: Distribution of transverse momentum of the Higgs boson approximated using using Legendre moments (left) and moments in a "rectified" basis constructed from the leading-order approximation (right), at different perturbative orders. The number in parenthesis indicates the number of retained terms in the Legendre series, which is determined from the estimated Monte-Carlo uncertainties according to the truncation prescription described in Sec. 2.2. The bottom panels show the relative deviation of the reconstructed distributions from the leading-order distribution. Significant improvement in the quality of the description of the high-$p_\perp$ tail of the distribution is observed when the rectified basis is used.

It should be appreciated, however, that this feature is a consequence of the way the optimized basis is constructed. This basis is not inherently more sensitive to the tail of the distribution; rather it effectively maps a significant interval in Higgs $p_\perp$ to a narrow region in $t \in [-1, 1]$ interval so that any features in the high-$p_\perp$ region are averaged over. Conceptually, this is equivalent to taking histograms with larger bins in regions where the target distribution is small, and this may lead to the distortion of the shape of the distribution at high $p_\perp$.

To illustrate this point, in the top plot of Figure 9 we compare the distribution reconstructed using the optimized basis against conventional histograms with very fine bins at leading and next-to-leading order. At NNLO we have to use coarser binning, because at this perturbative order bin-to-bin fluctuations become overwhelming if small bins are employed. Around the peak of the distribution the moments approximations are very accurate and can resolve even very fine features of the shape modification due to higher-order corrections, but the resolution deteriorates towards the tail of the distribution. The histograms and the moments approximations agree within uncertainties for $p_{\perp,H} \lesssim 350\,\text{GeV}$, but above this point some discrepancy in the slope of the reconstructed distributions is visible at NLO. The smoothing of the true distribution that arises because of the use of moments appears to be too aggressive in this region.

The origin of this problem can be seen in the bottom plot in Figure 9, which shows the same distribution in the *transformed coordinates*. We note that in these coordinates the optimized basis functions reduce to the standard Legendre polynomials. However, when written in t-variables, functions that are relatively smooth in the $p_\perp$-variable start changing very rapidly near the endpoint $t = 1$. This feature is not well-described by a few Legendre polynomials, which translates to an overly aggressive smoothing in the far-tail region of this observable. One way to solve this problem is to exclude the region where the distribution in the t-variable starts to decrease rapidly. In the current case, this corresponds to the right-most $\mathcal{O}(5\%)$ of the interval $t \in [-1, 1]$, which translates to $p_{\perp,H} \gtrsim 300\,\text{GeV}$.

We note that a similar issue is also present in other kinematic distributions that we studied, but it was not as prominent as in the example we show in Figure 9. We also stress that it is important that this problem can be easily detected by looking at a particular distribution in transformed coordinates and a "safe" region in the original coordinates can be identified.

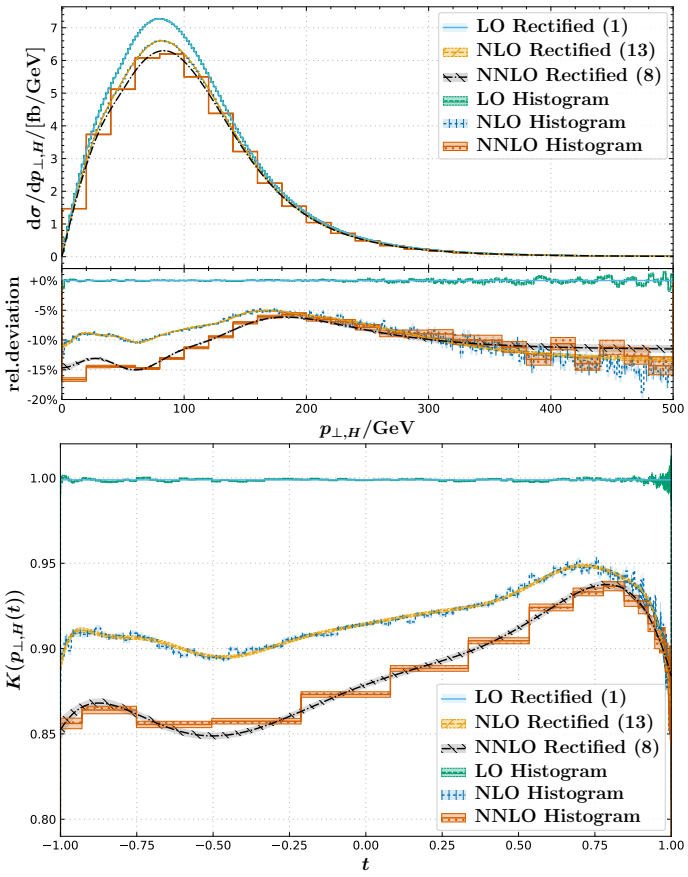

Figure 9: The top plot shows the distribution of transverse momentum of the Higgs boson in weak-boson fusion approximated using conventional histograms or moments in a "rectified" basis constructed from the leading-order approximation. The bottom plot shows the same distribution in the transformed variable $t(p_{\perp,H})$, constructed according to Eq. (29). The bin size for the histograms is 2 GeV at leading and next-to-leading orders and 20 GeV at NNLO. The number in parenthesis indicates the number of retained terms in the Legendre series, which is determined from the estimated Monte-Carlo uncertainties according to the truncation prescription described in Sec. 2.2. The bottom panel on the top plot shows the relative deviation from the leading-order distribution.

Finally, we note that the approaches discussed above are not unique. A variety of alternative reconstruction methods exist, including different functional bases or target distribution manipulations. We explored several such possibilities and obtained results of varying quality. For brevity, and because these alternatives do not qualitatively change the conclusions of our work, we do not discuss them here.

6 Conclusions

We have shown that moments-based approximations provide an attractive alternative to conventional histograms. They directly yield smooth approximations to the target distributions, and, in perturbative computations based on local subtraction schemes, they are immune to large fluctuations that arise when an event and a counter-event get accommodated into neighboring bins.

For smooth distributions with no high-frequency features, moments-based approximations converge faster than conventional histograms of comparable resolution. Furthermore, they yield compact representations of distributions that do not depend on the arbitrary choice of binning, admit efficient evaluation and integration, and are inherently differentiable, which might be useful for applications such as gradient-based optimization. Furthermore, it is straightforward to use moments to recover conventional histograms, for any desired binning.

Moreover, if an informed high-quality approximation to the target distribution is available, it can be used to construct an optimized basis of functions and moments. As we have shown, this leads to faster convergence and smooths out fluctuations in the tails of the reconstructed distributions.

While in this study we focused on one-dimensional distributions, the discussed methods generalize straightforwardly to multidimensional ones. We expect them to perform similarly well, as long as the dimension is not too high.

In conclusion, moments-based methods of reconstructing distributions from random samples can outperform conventional histograms when the underlying distributions are sufficiently smooth and when calculations are performed with local subtraction schemes. For these reasons, we advocate for further exploration and use of these methods in Monte-Carlo integrators.

Acknowledgments

We are grateful to K. Asteriadis, A. Behring and R. Röntsch for useful conversations.

Funding information This research is partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762 - TRR 257.

References

[1] F. James, *Monte Carlo theory and practice*, Rep. Prog. Phys. **43**, 1145 (1980), doi:10.1088/0034-4885/43/9/002.

[2] G. P. Lepage, *A new algorithm for adaptive multidimensional integration*, J. Comput. Phys. **27**, 192 (1978), doi:10.1016/0021-9991(78)90004-9.

[3] S. Navas et al, *Review of particle physics*, Phys. Rev. D **110**, 030001 (2024), doi:10.1103/PhysRevD.110.030001.

[4] A. Buckley et al., *General-purpose event generators for LHC physics*, Phys. Rep. **504**, 145 (2011), doi:10.1016/j.physrep.2011.03.005 [preprint doi:10.48550/arXiv.1101.2599].

[5] B. W. Silverman, *Density estimation for statistics and data analysis*, CRC Press, Boca Raton, USA, ISBN 9781315140919 (1998), doi:10.1201/9781315140919.

[6] D. G. Truhlar and N. C. Blais, *Legendre moment method for calculating differential scattering cross sections from classical trajectories with Monte Carlo initial conditions*, J. Chem. Phys. **67**, 1532 (1977), doi:10.1063/1.435057.

[7] S. Frixione, Z. Kunszt and A. Signer, *Three-jet cross sections to next-to-leading order*, Nucl. Phys. B **467**, 399 (1996), doi:10.1016/0550-3213(96)00110-1 [preprint doi:10.48550/arXiv.hep-ph/9512328].

[8] S. Catani and M. H. Seymour, *A general algorithm for calculating jet cross sections in NLO QCD*, Nucl. Phys. B **485**, 291 (1997), doi:10.1016/S0550-3213(96)00589-5 [preprint doi:10.48550/arXiv.hep-ph/9605323].

[9] A. Gehrmann-De Ridder, T. Gehrmann and E. W. N. Glover, *Antenna subtraction at NNLO*, J. High Energy Phys. **09**, 056 (2005), doi:10.1088/1126-6708/2005/09/056 [preprint doi:10.48550/arXiv.hep-ph/0505111].

[10] M. Czakon, *A novel subtraction scheme for double-real radiation at NNLO*, Phys. Lett. B **693**, 259 (2010), doi:10.1016/j.physletb.2010.08.036 [preprint doi:10.48550/arXiv.1005.0274].

[11] M. Cacciari, F. A. Dreyer, A. Karlberg, G. P. Salam and G. Zanderighi, *Fully differential vector-boson-fusion Higgs production at next-to-next-to-leading order*, Phys. Rev. Lett. **115**, 082002 (2015), doi:10.1103/PhysRevLett.115.082002 [preprint doi:10.48550/arXiv.1506.02660].

[12] V. Del Duca, C. Duhr, A. Kardos, G. Somogyi, Z. Szőr, Z. Trócsányi and Z. Tulipánt, *Jet production in the CoLoRFulNNLO method: Event shapes in electron-positron collisions*, Phys. Rev. D **94**, 074019 (2016), doi:10.1103/PhysRevD.94.074019 [preprint doi:10.48550/arXiv.1606.03453].

[13] F. Caola, K. Melnikov and R. Röntsch, *Nested soft-collinear subtractions in NNLO QCD computations*, Eur. Phys. J. C **77**, 248 (2017), doi:10.1140/epjc/s10052-017-4774-0 [preprint doi:10.48550/arXiv.1702.01352].

[14] L. Magnea, E. Maina, G. Pelliccioli, C. Signorile-Signorile, P. Torrielli and S. Uccirati, *Local analytic sector subtraction at NNLO*, J. High Energy Phys. **12**, 107 (2018), doi:10.1007/JHEP12(2018)107 [preprint doi:10.48550/arXiv.1806.09570].

[15] V. Del Duca, C. Duhr, L. Fekésházy, F. Guadagni, P. Mukherjee, G. Somogyi, F. Tramontano and S. Van Thurenhout, *NNLOCAL: Completely local subtractions*, Proc. Sci. **485**, 243 (2026), doi:10.22323/1.485.0243 [preprint doi:10.48550/arXiv.2601.02033].

[16] G. Szegö, *Orthogonal polynomials*, American Mathematical Society, New York, USA, ISBN 9780821810231 (1939).

[17] J. P. Boyd, *Chebyshev and Fourier spectral methods*, Dover, Mineola, USA, ISBN 9780486411835 (2001).

[18] Z. Ligeti, I. W. Stewart and F. J. Tackmann, *Treating the b quark distribution function with reliable uncertainties*, Phys. Rev. D **78**, 114014 (2008), doi:10.1103/PhysRevD.78.114014 [preprint doi:10.48550/arXiv.0807.1926].

[19] C. W. Clenshaw, *A note on the summation of chebyshev series*, Math. Comput. **9**, 118 (1955), doi:10.1090/S0025-5718-1955-0071856-0.

[20] H. Wang and S. Xiang, *On the convergence rates of Legendre approximation*, Math. Comput. **81**, 861 (2012), doi:10.1090/S0025-5718-2011-02549-4.

[21] K. Asteriadis, F. Caola, K. Melnikov and R. Röntsch, *NNLO QCD corrections to weak boson fusion Higgs boson production in the $H \to b\bar{b}$ and $H \to WW^* \to 4l$ decay channels*, J. High Energy Phys. **02**, 046 (2022), doi:10.1007/JHEP02(2022)046 [preprint doi:10.48550/arXiv.2110.02818].

[22] C. Bierlich et al., *Robust independent validation of experiment and theory: Rivet version 3*, SciPost Phys. **8**, 026 (2020), doi:10.21468/SciPostPhys.8.2.026 [preprint doi:10.48550/arXiv.1912.05451].

[23] E. Bothmann et al., *Event generation with Sherpa 3*, J. High Energy Phys. **12**, 156 (2024), doi:10.1007/JHEP12(2024)156 [preprint doi:10.48550/arXiv.2410.22148].