Machine Learning Methods for Cross Section Measurements

by

Krish Desai

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Benjamin Nachman, Co-chair
Professor Uros Seljak, Co-chair
Professor Joshua Bloom
Professor Saul Perlmutter

Spring 2025

Machine Learning Methods for Cross Section Measurements

To my family, mentors, and friends

for their unwavering support and encouragement

Contents

List of Figures

List of Tables

Chapter I

Moment Unfolding: direct deconvolution of distribution moments

I.A   Why moments: physics context and QCD calculations

Statistical moments of probability distributions serve as powerful tools in physics, providing a concise and theoretically meaningful way to characterise complex phenomena. In the context of particle physics, and particularly in quantum chromodynamics (QCD), working at the level of moments rather than distributions offers significant advantages for both theoretical calculations and experimental measurements. This section explores the fundamental importance of moments in physics, their theoretical foundations, and their specific applications in QCD.

I.A.1   The Theoretical Significance of Moments in Physics

Statistical moments represent a systematic way to characterise probability distributions, with each successive moment providing additional information about the shape and properties of the distribution. The kth moment of a probability density $p(z)$ is defined as

$$\langle Z^k \rangle = \int_{-\infty}^{\infty} z^k \, p(z) \, \mathrm{d}z \tag{I.1}$$

While the complete probability distribution contains all available information, working with moments is often much more tractable. In many physical theories, moments can

be calculated analytically even when full distributions cannot. Each moment has a direct physical interpretation—the first moment corresponds to the mean, the second moment to the variance, and higher moments characterize asymmetry (skewness) and peakedness (kurtosis) and so on. Moments therefore provide much more concrete insight into the properties of an observable than distributions.

In many physical theories, including QCD, there exist relatively simple evolution equations for how moments change as a function of energy scale. If certain moments exhibit universal behaviour across different physical systems, this too reveals universality in the fundamental underlying principles. For many physical systems, including those governed by QCD, moments provide a natural language for connecting theoretical predictions with experimental measurements, particularly when examining how observables scale with energy.

I.A.2   Moments in QCD Calculations

In quantum chromodynamics, moments play a particularly crucial role in understanding the scaling behaviour of parton distributions and fragmentation functions. The DGLAP equations form the theoretical framework for parton distribution function evolution.

I.A.2.i   DGLAP Evolution Equations

The Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution equations [**Altarelli:1977zs**,

dokshitzerCalculationStructureFunctions1977, **gribovDeepInelasticEp1972**,

1] govern how parton distribution functions evolve with energy scale. While these

equations are integro–differential equations in their most general form, they transform

into ordinary differential equations when expressed in terms of moments, significantly

simplifying their solution [2–12].

For a parton distribution function $f(x, Q^2)$, where x is the momentum fraction

and Q^2 is the energy scale, the $n-$th moment is defined as

$$M_n(Q^2) = \int_0^1 x^{n-1} \, f(x, Q^2)\mathrm{d}x. \tag{I.2}$$

The DGLAP equations for these moments take the form

$$\frac{\mathrm{d}M_n(Q^2)}{\mathrm{d}\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \, P_n \, M_n(Q^2). \tag{I.3}$$

where P_n are the moments of the splitting functions [13–17]. This formulation converts

the complex integro–differential evolution equations into a set of simple ordinary

differential equations, one for each moment independently.

I.A.2.ii   Operator Product Expansion

The Operator Product Expansion (OPE) in QCD naturally expresses deep inelastic

scattering structure functions in terms of moments [18]. In this formalism, the

moments of structure functions are related to matrix elements of local operators, which have clear physical interpretations and scaling properties. For example, the moments of the structure function $F_2(x, Q^2)$ can be expressed as

$$M_n(Q^2) = \int_0^1 x^{n-2} \, F_2(x, Q^2) \, \mathrm{d}x = \sum_i C_n^i(Q^2) \, \langle O_n^i \rangle. \tag{I.4}$$

where C_n^i are Wilson coefficients calculable in perturbation theory, and $\langle O_n^i \rangle$ are matrix elements of local operators [19].

I.A.2.iii   Event Shape Moments

Event shape observables, which characterise the geometrical properties of energy flow in collisions, are often studied through their moments. Moments of event shapes like thrust, broadening, and jet mass distributions provide sensitive tests of perturbative QCD and allow precise extractions of the strong coupling constant α_s [20, 21]. The moments of these event shapes can be calculated as

$$\langle e^n \rangle = \int_0^{e_{\max}} e^n \, \frac{1}{\sigma} \frac{\mathrm{d}\sigma}{\mathrm{d}e} \, \mathrm{d}e, \tag{I.5}$$

where e is the event shape variable, and $\frac{1}{\sigma} \frac{\mathrm{d}\sigma}{\mathrm{d}e}$ is its normalized differential cross section. These moments can be predicted in perturbative QCD, typically expressed as power series in α_s:

$$\langle e^n \rangle = A_n \left(\frac{\alpha_s}{2\pi} \right) + B_n \left(\frac{\alpha_s}{2\pi} \right)^2 + \mathcal{O}(\alpha_s^3), \tag{I.6}$$

plus non perturbative corrections that scale as inverse powers of the centre of mass energy.

I.A.3   Experimental Significance of Moments in QCD

From an experimental perspective, moments offer several advantages that make them particularly valuable for QCD studies.

Some of the most precise determinations of the strong coupling constant α_s come from measurements of moments. For example, moments of event shapes measured at LEP have provided determinations of α_s with uncertainties at the few-percent level [cite –KD]. Studies by AMY, TASSO, OPAL, DELPHI, L3, and ALEPH collaborations have all used moment–based analyses to extract α_s values that contribute significantly to the world average [cite –KD].

The scaling behavior of moments with energy provides a direct and robust window into the running of QCD couplings. By measuring how moments of distributions change with energy scale, experiments can observe the logarithmic scaling violations that are a hallmark of asymptotic freedom in QCD [cite –KD].

Many theoretical QCD calculations are more readily available (and more precise) for moments than for full distributions. For certain observables, perturbative calcula-

tions may exist to next–to–next–to–leading order (NNLO) or beyond for moments, while full distribution calculations may only be available at lower orders [cite –KD].

In practice, measuring moments directly can reduce certain experimental uncertainties, particularly those related to detector resolution and acceptance effects in regions of phase space with sparse statistics.

Traditionally, there was no straightforward method to directly unfold moments. Instead, moments were computed after first measuring and unfolding the full distribution. This typically involved binning the data, unfolding the binned distribution, and then computing moments from the unfolded histogram. The limitation of this approach though, is the binning itself, not the moments qua moments.

The use of discrete bins introduces a bias in moment calculations. For a variable z binned into n_{bins} with centers $z_{\mathrm{bin},i}$ and counts N_i, the binned approximation of the $k-$th moment is

$$\langle z^k \rangle_{\mathrm{bin}} = \frac{1}{N} \sum_{i=1}^{n_{\mathrm{bins}}} N_i z_{\mathrm{bin},i}^k \tag{I.7}$$

This approximation introduces a bias relative to the true moment, which grows with the order of the moment computed, and is particularly large in regions where the distribution varies rapidly within bins [cite –KD].

These limitations motivate the development of unbinned approaches that can directly unfold moments without first discretizing the data, which is precisely the focus of the Moment Unfolding method described in this chapter.

I.A.4   Applications in Jet Physics

In the context of jet physics specifically—a central arena for testing QCD—moments of jet substructure observables provide particularly powerful probes of QCD dynamics **[cite –KD]**>. Several notable applications deserve special mention.

The energy scale dependence of jet properties is best studied through their moments. The moments of jet substructure observables such as jet mass, width, and multiplicity exhibit characteristic scaling with jet energy that can be predicted by perturbative QCD. For example, the first moment of the jet mass scales approximately as **[cite –KD]**>

$$\langle m^2 \rangle \propto \alpha_s(p_T) \, p_T^2 \, \ln(R/R_0) \tag{I.8}$$

where p_T is the jet transverse momentum, R is the jet radius, and R_0 is the reference scale. Higher moments have different scaling behaviours, providing multiple handles on the underlying QCD dynamics.

The moments of jet substructure variables are also used in the classification of jets. This is because jet moments differ characteristically between quark–initiated

and gluon–initiated jets, making them valuable for jet flavour tagging. For instance, gluon jets typically have higher multiplicity and broader mass distributions than quark jets, reflected in the moments of these distributions **[cite –KD]**.

Jet substructure moments also provide a window to study non–perturbative effects and hadronization. The interplay between perturbative and non-perturbative effects in QCD is particularly evident in moments of jet observables. Lower moments often capture the perturbative, large–angle radiation physics, while higher moments become increasingly sensitive to non–perturbative effects like hadronization **[cite –KD]**.

In the context of Soft-Collinear Effective Theory (SCET), moments of jet substructure observables provide direct constraints on the parameters of the effective theory, offering a bridge between first–principles QCD and phenomenological models of jet formation **[cite –KD]**.

I.A.4.i   Moments in Physics Beyond the Standard Model

While moments are crucial for precision QCD studies, their utility extends to searches for BSM physics. Certain moments of distributions can exhibit enhanced sensitivity to new physics scales. For example, higher moments of mass distributions can be more sensitive to heavy particle contributions than the full distribution average **[cite –KD]**.

Moments can therefore provide a model–independent way to parametrize deviations from Standard Model predictions, similar to the role of effective field theory coefficients. By measuring a series of moments, experiments can place constraints on a wide class of possible new physics scenarios without committing to specific models [cite –KD]. Hence, changes in the moment structure of distributions can serve as an early indicator of anomalous physics, even before the full nature of the anomaly is understood <kdcite>.

I.A.5   The Case for Direct Moment Unfolding

Given the theoretical significance of moments in QCD and their experimental advantages, there is a strong motivation for developing methods that can directly unfold moments from detector–level data without first unfolding the entire distribution. Such methods would eliminate binning biases inherent in traditional approaches, while potentially improving precision and reducing computational cost by focusing directly on the quantities of interest. They would enable moment measurements in higher–dimensional phase spaces where binned approaches become impractical to provide more direct comparisons with theoretical predictions that are formulated at the level of moments.

The remainder of this chapter presents the Moment Unfolding method, a novel approach that addresses precisely this need by directly deconvolving moments of distributions using machine learning techniques inspired by the Boltzmann distribution and implemented through a GAN–like architecture. This method offers a significant advance in precision moment measurements, enabling more stringent tests of QCD predictions and potentially uncovering subtle effects that might be obscured in traditional analyses.

I.B   A GAN–like Method to Unfold Moments

The direct extraction of distribution moments without first reconstructing the entire distribution presents a unique challenge in unfolding methodology. Traditional approaches typically reconstruct the full differential cross–section before computing moments, introducing unnecessary computational complexity. The Moment Unfolding technique presented in this chapter takes an entirely different approach by directly targeting the moments themselves.

The basic idea underlying Moment Unfolding is the connection between statistical moments and the Boltzmann distribution from statistical mechanics. In thermodynamics, the Boltzmann distribution represents the probability distribution that maximizes entropy while satisfying certain constraints on average quantities (internal energy). This principle can be adapted for our unfolding problem, where we seek to find a reweighting of simulated events that accurately reproduces the moments of the true particle–level distribution.

I.B.1   Boltzmann inspired reweighting

The central mathematical construction of Moment Unfolding is a reweighting function inspired by the Boltzmann factor,

$$g(z) = \frac{1}{P(\boldsymbol{\beta})} \exp\left(-\sum_{a=1}^{n} \beta_a \, z^a\right) \tag{I.9}$$

where z is the particle–level observable whose moments we wish to unfold, β_a are parameters to be determined (analogous to Lagrange multipliers in statistical mechanics), P is a normalization factor similar to the partition function, and n is the number of moments we aim to unfold simultaneously.

This form is particularly powerful because it directly connects to the maximum entropy principle in statistical mechanics. Just as the Boltzmann distribution maximizes entropy subject to constraints on average energy, this reweighting function maximizes the binary cross entropy loss between the discriminator's classification output between the reweighted distribution and the prior distribution while enforcing constraints on the moments. When applied to a simulated particle–level distribution with probability density $q_{\text{Gen.}}(z)$, this reweighting function produces a modified distribution

$$\widetilde{q}(z) = g(z) \cdot q_{\text{Gen.}}(z) \tag{I.10}$$

The training objective then is to determine the optimal values of the parameters β_a such that the moments of $\widetilde{q}(z)$ match those of the true underlying distribution $p(z)$.

I.B.2   Adversarial Training Framework

Given the theoretical foundation provided by the Boltzmann form, we now need a practical method to determine the optimal parameters. The task can very naturally be framed as a two–level optimization problem:

1. The parameters β_a control the reweighting at the particle level

2. The quality of the reweighting must be assessed at the detector level

This task is highly non–trivial because the detector response function—which maps particle-level to detector–level observables—-is typically stochastic and known only implicitly through simulation.

To address this, Moment Unfolding employs an adversarial training architecture inspired by Generative Adversarial Networks (GANs), but with significant modifications tailored to the unfolding problem. Unlike traditional GANs that generate new samples, our "generator" $g(z)$ reweights existing particle-level simulated events according to the Boltzmann form. The discriminator is a neural network classifier $d(x)$ that attempts to distinguish between real detector–level data and the reweighted simulation at detector level.

The system is then trained adversarially, with the generator attempting to fool the discriminator while the discriminator attempts to correctly classify events. This setup is illustrated in Figure **[fig:moment-unfolding-architecture –KD]**, showing

the flow of information and the adversarial training process. In this framework, the optimal parameters β_a are those that make the reweighted detector–level simulation indistinguishable from the real detector–level data. Once these parameters are found, the moments of the particle–level reweighting are the unfolded moments.

I.B.3   Mathematical Formalism

We can formalize the Moment Unfolding approach as follows. Let $p(x)$ be the detector–level data distribution and $fq(z,x)$ be the joint distribution of particle–level and detector–level variables in the MC, Generation and Simulation. Then $q(z)$ and $q(x)$ are the marginal distributions of Generation and Simulation respectively. Let $r(x|z)$ be the detector response function. The assumption of a universal detector response can be formalized as

$$p(x|z) = q(x|z) = r(x|z). \tag{I.11}$$

The goal is to find parameters β_a such that the moments of the reweighted distribution $g(z) \cdot q(z)$ match the true moments of the underlying particle–level distribution, $p(z)$.

The detector–level distribution after reweighting is

$$\widetilde{q}(x) = \int g(z) \cdot q(z,x) \, \mathrm{d}z. \tag{I.12}$$

The adversarial training process optimizes a weighted binary cross entropy loss function,

$$\mathcal{L}[g, d] = -\mathbb{E}_{X \sim p(x)}[\log d(X)] - \mathbb{E}_{X \sim \widetilde{q}(X)}[(1 - d(x))] \tag{I.13}$$

This loss function encourages the discriminator to output high values for real data and low values for reweighted simulation, while the generator tries to make the reweighted simulation indistinguishable from real data.

Once the optimal parameters β_a are determined, the unfolded moments can be computed directly from the reweighted Generation

$$\langle Z^k \rangle_{\text{unfolded}} = \frac{\sum_{z \in \text{Gen.}} g(z) \cdot z^k}{\sum_{z \in \text{Gen.}} g(z)} \tag{I.14}$$

This approach differs fundamentally from traditional unfolding methods that first reconstruct the full distribution and then compute moments. By directly targeting the moments themselves, we avoid both the binning artifacts and dimensionality challenges associated with binned methods, and the computational complexity associated with unbinned full distribution unfolding.

I.B.4   Theoretical Properties

The Moment Unfolding method possesses several important theoretical properties that justify its application to physics problems. Under certain conditions (detailed

in Appendix[ref –KD]), the Moment Unfolding method provably converges to the correct moments.

The parameter n (the number of moments being unfolded) serves as an implicit regularization parameter. By limiting the number of moments, we constrain the complexity of the reweighting function, helping to stabilize the unfolding process against statistical fluctuations. This creates a natural bias-variance tradeoff: including more moments allows for a more flexible reweighting function but increases sensitivity to statistical fluctuations. In practice, the optimal choice of n depends on the physics application and available statistics.

I.B.4.i   Connection to the Maximum Entropy Principle

The Boltzmann form of the reweighting function has a deep connection to the maximum entropy principle in statistical mechanics. The reweighting function maximizes the entropy of the reweighted distribution relative to the prior (simulation) distribution, subject to constraints on the moments. This provides a principled physical and informational theoretic basis for the method: among all possible reweighting functions that reproduce the target moments, the Boltzmann form is the one that introduces the least additional information beyond what is required by the moment constraints.

I.B.5 Comparison with Traditional GAN Architectures

While Moment Unfolding draws inspiration from GANs, it differs from traditional GAN architectures, in that the so–called "generator" does not generate data; it generates weights. Traditional GANs generate new samples by mapping random noise through a neural network. In contrast, Moment Unfolding reweights existing simulated events, preserving the physics correlations already present in the simulation and focusing only on correcting the distribution's moments.

Furthermore, unlike conventional GANs where the generator can represent any function within the capacity of the neural network, our generator has an extremely specific parametric form constrained by the Boltzmann expression. This constraint massively reduces the effective degrees of freedom in the optimization process and encodes our prior knowledge about the problem structure. The parameters β_a in the Moment Unfolding approach have clear physical interpretations as Lagrange multipliers enforcing moment constraints. This interpretability contrasts with the typically opaque nature of neural network weights in traditional GANs.

Finally, traditional GANs operate at a single level, with the generator and discriminator acting on the same "type" of data. Moment Unfolding operates across two levels, particle–level and detector–level, which might not even have the same

dimensionality, which is the source of the additional complexity but is essential for addressing the detector response effect.

I.B.6   Practical Considerations

While theoretically elegant, the Moment Unfolding method has practical considerations and limitations that must be understood in order to apply the method correctly and effectively.

Like all reweighting–based methods, Moment Unfolding requires that the Generation distribution has overlapping support with the Truth distribution. Regions of phase space not covered by the MC cannot be properly unfolded. With finite statistics, extremely large weights can arise, in regions of phase space with sparse MC coverage as the generator tries and fails to reweight a zero density into a non–zero density. Various regularization techniques must be employed to mitigate these effects if they arise in a given analysis.

The optimal number of moments to unfold depends on the physics application, available statistics, and the complexity of the true distribution. Too few moments may miss important features of the distribution, while too many may lead to overfitting and numerical instabilities.

I.B.7   Extension to Differential Measurements

An extension of the Moment Unfolding approach is to study moments as a function of another observable, such as transverse momentum. This is achieved by making the parameters β_a functions of the conditional variable,

$$g(z; p_T) = \frac{1}{P(\boldsymbol{\beta}(p_T))} \exp\left(-\sum_{a=1}^{n} \beta_a(p_T)\, z^a \right) \qquad (\text{I.15})$$

This allows the study of how moments evolve with energy scale or other experimental conditions, providing direct comparisons with theoretical predictions of scaling behaviors in QCD.

The parameters $\beta_a(p_T)$ can be modeled as polynomials, splines, or neural networks, depending on the expected complexity of their dependence on p_T. This extension significantly enhances the physics applications of the method, enabling studies of how observable moments scale with energy.

I.B.8   Theoretical Foundations in Statistical Mechanics

The mathematical form at the heart of Moment Unfolding has deep connections to statistical mechanics and information theory, providing a solid theoretical foundation for the method. In statistical mechanics, the Boltzmann distribution arises as the solution to the maximum entropy problem. The maximum entropy problem is the problem of finding the probability distribution that maximizes entropy subject to

constraints on expectation values (typically energy). The solution takes the form

$$p(s) = \frac{1}{Z} e^{-\beta E(s)} \tag{I.16}$$

where $E(s)$ is the energy of state s, β is the inverse temperature, and Z is the partition function.

The Moment Unfolding algorithm borrows this idea to include multiple constraints on different moments, with each β_a serving as a Lagrange multiplier for the constraint on the a-th moment. This connection to maximum entropy principles provides a theoretical justification for the specific form of the reweighting function and suggests that it is, in some sense, the most natural choice for the moment unfolding problem. The success of this approach in practical applications, as will be demonstrated in subsequent sections, confirms the value of this theoretical grounding and highlights the power of cross-disciplinary approaches in developing new data analysis techniques for particle physics.

I.C   Machine Learning Implementation

The theoretical framework introduced in the previous section provides the foundation for Moment Unfolding, but implementing this approach in practice requires careful architectural design and optimization strategies. This section details the machine learning implementation that enables robust moment extraction through our GAN–like approach.

The generator in Moment Unfolding differs fundamentally from traditional GAN generators. Instead of producing new samples through a complex neural network, our generator implements the Boltzmann–inspired weighting function, Eq. I.9. This function is applied to existing simulated particle–level events to reweight them, producing a weighted distribution that matches the moments of the true distribution. The parameters β_a are the trainable parameters of the generator, and P is a normalization constant estimated at the batch level,

$$\hat{P} = \sum_{z \in \text{batch}} \exp\left(-\sum_{a=1}^{n} \beta_a z^a\right). \tag{I.17}$$

For differential moment measurements, where we study the dependence of moments on another variable like transverse momentum (p_T), the β_a parameters become functions of p_T, as described in Eq. I.15. In our implementation, we parameterize $\beta_a(p_T)$ as linear functions,

$$\beta_a(p_T) = \beta_a^{(0)} + \beta_a^{(1)} p_T \tag{I.18}$$

This linear parameterization is motivated by empirical observations that the ratios of spectra between different simulations often exhibit approximately linear dependence on kinematic variables **[cite –KD]**.

The discriminator network $d(x)$ takes detector–level features as input and outputs a probability that a given event came from the real data rather than the reweighted simulation. For our implementation, we use a dense neural network with three fully–connected hidden layers with 50 nodes per layer. ReLU activation functions are applied to intermediate layers, and a sigmoid activation function is applied to the output layer.

This relatively simple architecture provides sufficient capacity to distinguish between data and simulation distributions.

The training objective is formulated as a weighted version of the Maximum Likelihood Classifier (MLC) loss **[cite –KD]**,

$$L[g, d] = - \sum_{x \in \text{Data}} \log d(x) - \sum_{(z,x) \in (\text{Gen., Sim.})} g(z)(1 - d(x)) \tag{I.19}$$

where the first sum is over detector-level events from real data, and the second sum is over matched pairs of particle–level and detector–level events from simulation, weighted by the generator function $g(z)$.

This loss function differs from the standard Binary Cross Entropy (BCE) loss often used in GANs, providing smoother gradients and better convergence properties

for our specific application. While the BCE loss function could also be used (and indeed gives similar empirical performance), the MLC loss function provides certain theoretical guarantees about the convergence of the method to the correct moments [cite –KD].

I.C.1   Training Procedure

The training procedure for Moment Unfolding follows the standard adversarial approach, but with specific adaptations for the reweighting framework.

The generator parameters β_a are initialized to small random values near zero, corresponding to minimal reweighting (i.e., starting close to the original simulation). The discriminator weights are initialized using standard neural network initialization techniques[cite HeNormal initialization –KD].

During each training step, we sample batches of events from Data, Generation, and Simulation. The training proceeds by alternating between optimizing the discriminator and the generator. First, with the generator parameters fixed, the discriminator weights are updated to minimize the loss function. Then with the discriminator weights fixed, the generator parameters are updated to maximize the loss function. This alternating optimization continues for a fixed number of epochs, or until convergence criteria are met.

We employ the ADAM optimizer **[cite –KD]** with an initial learning rate of 5×10^{-4} for both the discriminator and generator The learning rate is reduced by a factor of 0.5 every 20 epochs to ensure stable convergence.

To ensure stable training and prevent overfitting, a few regularization techniques are employed. Gradient values are clipped to the range $[-1, 1]$ to prevent large updates that could destabilize training, particularly important for the generator parameters which can be sensitive to statistical fluctuations. Batch normalization **[cite –KD]** is applied after each hidden layer in the discriminator to stabilize training and reduce sensitivity to weight initialization and learning rates.

I.C.2 Gradient Updates

For the discriminator, gradient computation follows standard backpropagation through the neural network. For the generator, gradients are computed with respect to the parameters β_a by differentiating the loss function.

$$\frac{\partial L}{\partial \beta_a} = -\sum_{(z,x) \in (\text{Gen., Sim.})} \frac{\partial g(z)}{\partial \beta_a}(1 - d(x)), \tag{I.20}$$

where

$$\frac{\partial g(z)}{\partial \beta_a} = -g(z)\left(z^a - \frac{1}{P}\frac{\partial P}{\partial \beta_a}\right) = -g(z)(z^a - \langle z^a \rangle_g) \tag{I.21}$$

Here, $\langle z^a \rangle_g$ represents the $a-$th moment computed from the reweighted distribution. This formulation provides a direct connection between the gradient updates and the moment matching objective of the method.

I.C.3   Implementation Details

he method is implemented using KERAS [cite –KD] with the TENSORFLOW2 backend [cite –KD], leveraging their automatic differentiation capabilities for efficient gradient computation. NUMPY [cite –KD] is used for data manipulation, and MATPLOTLIB [cite –KD] is employed for visualization.

The computational requirements of Moment Unfolding are remarkably light compared to many deep learning applications. Training typically takes less than five minutes per observable on an NVIDIA RTX6000 GPU for the case studies presented in this dissertation. Even the potentially computationally intensive calculation of the partition function P, which requires summation over all events in a batch is implemented efficiently using vectorized operations in TENSORFLOW.

Tab. I.1 summarizes the key hyperparameters used in our implementation. These hyperparameter values were determined through systematic experimentation and provide a good balance between training stability and model performance across a range of physics scenarios.

Hyperparameter	Value	Description
Discriminator layers	3	Number of hidden layers in the discriminator
Discriminator nodes	50	Number of nodes per hidden layer
Activation function	ReLU	Activation function for hidden layers
Output activation	Sigmoid	Activation function for the output layer
Optimizer	ADAM	Optimization algorithm used for training
Learning rate	5×10^{-4}	Initial learning rate for the optimizer
Batch size	10,000	Number of training events in each batch
Training epochs	50–100	Number of full passes through the training dataset
Gradient clip value	1.0	Maximum allowed gradient norm for clipping

Table I.1: Summary of training hyperparameters used in the model. These values control the architecture, optimization behaviour, and regularization of the training process.

I.C.4   Extensions to Multiple Observables

While the basic formulation of Moment Unfolding applies to moments of a single observable, the method can be extended to handle multiple observables simultaneously. For multiple observables $z_1, z_2, ..., z_d$, we can define joint moments as $\langle z_1^{k_1} z_2^{k_2} \cdots z_d^{k_d} \rangle$. The reweighting function then takes the form,

$$g(z_1, z_2, ..., z_d) = \frac{1}{P} \exp\left(- \sum_{k_1, k_2, ..., k_d} \beta_{k_1, k_2, ..., k_d} z_1^{k_1} z_2^{k_2} \cdots z_d^{k_d} \right) \qquad (I.22)$$

In practice, including all possible joint moments can quickly become intractable as the number of observables increases. To address this, one can employ strategies such as

- Limiting the total order of moments (e.g., $k_1 + k_2 + ... + k_d \leq K_{max}$),

- Including only specific joint moments known to be physically relevant, and

- Using factorized forms that capture the most important correlations while maintaining computational feasibility.

I.C.5   Uncertainty Estimation

Accurate uncertainty estimation is crucial for meaningful physics measurements. In Moment Unfolding, uncertainties on the extracted moments are obtained through bootstrap resampling. Statistical uncertainties are estimated by generating N bootstrap replicas of the data by resampling with replacement, applying Moment Unfolding to each replica, and computing the standard deviation of the resulting moment values across all replicas. This procedure accounts for statistical uncertainties in both the data and the simulation.

Systematic uncertainties related to theoretical modeling, detector effects, and other sources are evaluated by varying the underlying simulation according to systematic uncertainty sources, applying Moment Unfolding to each variation, and taking the envelope of the resulting moment values as the systematic uncertainty.

This approach allows for comprehensive uncertainty quantification that accounts for all relevant sources of uncertainty in the measurement.

I.C.6   Validation Procedures

Several validation procedures are employed to ensure the reliability of the Moment Unfolding method. Closure tests verify that the method can correctly recover known moments from simulated data. The method's internal consistency is verified by checking that the moments computed from the reweighted Generation match the target moments, the reweighted Simulation reproduces the features of the Data, and that the discriminator is unable to distinguish between Simulation and Data.

Results from Moment Unfolding are also compared with those from traditional unfolding methods like Iterative Bayesian Unfolding followed by moment calculation. Agreement between different methods provides additional confidence in the results, while disagreement may indicate methodological issues that require further investigation.

I.C.7   Code Availability and Reproducibility

The implementation of Moment Unfolding is available as open–source Python code, facilitating reproducibility and adoption by the scientific community. This open–source approach promotes transparency, enables independent verification of results, and facilitates improvements and extensions to the method by the broader scientific community. The full code repository can be found at [GitHub URL] [cite

–KD], and the specific datasets used for the studies in this dissertation are archived on Zenodo [cite –KD].

By combining a theoretically motivated approach with practical machine learning implementation details, Moment Unfolding provides a robust and efficient method for extracting moments directly from data. The next sections will demonstrate the application of this method to jet substructure measurements, showcasing its performance on realistic physics problems.

I.D   Case studies

I.D.1   Gaussian experiments

Before applying Moment Unfolding to jet physics scenarios, it is essential to validate the method in a controlled environment where the ground truth is precisely known. To this end, the approach is first demonstrated using a one–dimensional Gaussian example, providing a clear illustration of the method's capabilities while establishing a benchmark for its performance.

I.D.1.i   Experimental Setup

For this controlled study, we generate datasets from Gaussian distributions with known parameters. The Truth dataset comprises $10,000$ events drawn from a Gaussian distribution with mean $\mu_{\text{Truth}} = 0.0$ and variance $\sigma^2_{\text{Truth}} = 1$. The Generation dataset comprises $100,000$ events drawn from a Gaussian distribution with mean $\mu_{\text{Gen.}} = -0.5$ and variance $\sigma^2_{\text{Gen.}} = 1.0$. Both distributions are then subjected to detector effects modelled as additive Gaussian noise with resolution parameter $\sigma_{\text{det}} = 0.5$. This setup mimics the typical scenario in particle physics where the MC differs from the true distribution, and both are observed through an imperfect detector.

Figure [fig:gaussian-setup –KD] illustrates the particle–level and detector–level distributions, showing how the detector resolution affects both the truth and simulation datasets.

I.D.1.ii   Results

Figure [fig:gaussian-results –KD] shows the results of applying Moment Unfolding to the Gaussian example. The left panel displays the particle–level distributions, the Truth distribution (blue), the Generation (orange), and the reweighted Generation (black dashed line). Even visually, the close overlap between the Truth histogram and the reweighted Generation demonstrates the success of the Moment Unfolding procedure.

The right panel of Figure [fig:gaussian-results –KD] provides a more quantitative perspective by showing the discriminator–optimized loss landscape as a function of the Boltzmann parameters β_1 and β_2. This landscape visualization allows us to verify that the maximum of the loss function (red star) aligns with the parameter values that produce the correct moments, providing empirical confirmation of the method's theoretical properties.

These results confirm that Moment Unfolding successfully recovers the true moments of normal distributions. Since normal distributions are fully characterized by their first two moments, the reweighted distribution closely matches the entire

truth distribution in this Gaussian case even though the method explicitly targets only the first two moments. We should not expect this to be the case for more complex distributions with non–trivial higher moments.

This controlled study confirms that Moment Unfolding performs as expected in a well–understood scenario, providing confidence to extend the method to more complex physics environments.

I.D.2   Jet Substructure in Collider Physics

Having validated Moment Unfolding in a controlled environment, we now apply the method to a realistic high–energy physics scenario: measuring moments of jet substructure observables in proton–proton collisions at the Large Hadron Collider (LHC).

I.D.2.i   Datasets

The simulated samples used for this study follow the setup from Andreassen et al. [cite –KD]. Proton–proton collisions at $\sqrt{s} = 14$ TeV are simulated using PYTHIA 8.243 with Tune 26 [cite –KD] (standing in for MC) and HERWIG 7.1.5 [cite –KD] (standing in for nature). The DELPHES 3.4.2 [cite –KD] fast simulation of the CMS detector is used to model detector effects, configured with particle flow reconstruction [cite –KD]. Jets are clustered using the anti-k_T algorithm [cite –KD] with radius

parameter $R = 0.4$, implemented in FASTJET 3.3.2[cite –KD]. To reduce acceptance effects, only leading jets in events with a Z boson having transverse momentum $p_T^Z > 200$ GeV are studied. This setup allows us to test Moment Unfolding in a realistic environment where the "Truth" (HERWIG particle–level) and "Generation" (PYTHIA particle–level) are different generators with distinct physics models, and both are observed through an imperfect detector to produce "Data", HERWIG + DELPHES and "Simulation", PYTHIA + DELPHES.

I.D.2.ii   Observables

Four key jet substructure observables are unfolded, each providing different insights into the internal structure of jets.

1. **Jet Mass** (m): The invariant mass of the jet, calculated as

$$m = \sqrt{\sum_k E_k^2 - \sum_k \mathbf{p}_k^2},$$

(I.23)

where the sum runs over all constituents of the jet, and E_k and $\mathbf{p}_k$ are their energies and three–momenta.

2. **Jet Charge** (q): An infrared–safe but not collinear–safe observable that measures the jet's electric charge [cite –KD].

$$q = \frac{1}{\sum_k \sqrt{p_{T,k}}} \sum_k q_k \sqrt{p_{T,k}},$$

(I.24)

where q_k is the electric charge of constituent k.

3. **Jet Width** (w): A measure of the radial distribution of radiation within the jet.

$$w = \frac{1}{p_{T,\text{jet}}} \sum_k p_{T,k} \Delta R_k, \tag{I.25}$$

where ΔR_k is the angular distance between constituent k and the jet axis.

4. **Groomed Momentum Fraction** (z_g): The momentum sharing between subjets after Soft Drop grooming **[cite –KD]** with $z_{\text{cut}} = 0.1$ and $\beta = 0$.

$$z_g = \frac{p_{T,\text{subleading}}}{p_{T,\text{leading}} + p_{T,\text{subleading}}} \tag{I.26}$$

These observables span a diverse range of jet properties, including both infrared–and–collinear safe observables (m and w), a Sudakov safe observable, z_g**[cite –KD]**, and q, which is only infrared safe.

I.D.2.iii Results

Table **??** presents the first and second moments of each jet observable extracted using Moment Unfolding, compared with the true values (from HERWIG) and those from the original generation (PYTHIA). The uncertainties in the Truth and Generation columns are computed by bootstrapping the respective datasets. For the Moment Unfolding column, the uncertainty includes the bootstrap uncertainty from

Table I.2: Moments of jet observables at particle level. First and second moments of m, q, w and z_g are shown for truth (HERWIG), generation (PYTHIA) and Moment Unfolding. Uncertainties in the truth and generation columns are estimated via bootstrap resampling; uncertainties in the unfolding column combine in quadrature the generation bootstrap uncertainty with the empirical 1σ spread from repeated unfolding on the same dataset.

	Truth (HERWIG)	Generation (PYTHIA)	Moment Unfolding
$\langle m \rangle$	$(2.182 \pm 0.030) \times 10^1$	$(2.064 \pm 0.043) \times 10^1$	$(2.173 \pm 0.047) \times 10^1$
$\langle m^2 \rangle$	$(6.049 \pm 0.222) \times 10^2$	$(5.360 \pm 0.350) \times 10^2$	$(6.115 \pm 0.364) \times 10^2$
$\langle q \rangle$	$(1.006 \pm 0.037) \times 10^{-2}$	$(1.582 \pm 0.038) \times 10^{-2}$	$(1.090 \pm 0.040) \times 10^{-2}$
$\langle q^2 \rangle$	$(1.216 \pm 0.082) \times 10^{-2}$	$(1.508 \pm 0.074) \times 10^{-2}$	$(1.207 \pm 0.074) \times 10^{-2}$
$\langle w \rangle$	$(1.498 \pm 0.025) \times 10^{-1}$	$(1.231 \pm 0.029) \times 10^{-1}$	$(1.499 \pm 0.029) \times 10^{-1}$
$\langle w^2 \rangle$	$(3.370 \pm 0.113) \times 10^{-2}$	$(2.421 \pm 0.128) \times 10^{-2}$	$(3.374 \pm 0.128) \times 10^{-2}$
$\langle z_g \rangle$	$(2.334 \pm 0.029) \times 10^{-1}$	$(2.457 \pm 0.030) \times 10^{-1}$	$(2.353 \pm 0.059) \times 10^{-1}$
$\langle z_g^2 \rangle$	$(6.789 \pm 0.166) \times 10^{-2}$	$(7.425 \pm 0.165) \times 10^{-2}$	$(6.767 \pm 0.330) \times 10^{-2}$

the Generation and the empirical uncertainty from the unfolding process added in quadrature.

These results demonstrate that Moment Unfolding successfully recovers the true moments from the data for all observables within statistical uncertainties. The method effectively corrects for both the differences between the PYTHIA and HERWIG physics models and the detector distortions introduced by DELPHES.

Figure **[fig:jet-substructure-distributions –KD]** shows the particle–level distributions for each observable, comparing the truth, generation, and reweighted generation after applying Moment Unfolding While the method explicitly targets only the first two moments, it produces reasonable agreement across the entire distribution,

particularly for the regions that contribute most significantly to the moments. Figure

[fig:jet-loss-landscapes –KD] displays the discriminator–optimized loss landscapes

for each observable, confirming that the maximum of the loss function aligns with the

parameter values that produce the correct moments. The blue ellipses represent the

1σ confidence intervals for the Moment Unfolding parameters, and they all contain

the true parameter values (red dots), indicating proper uncertainty estimation.

I.D.3   Momentum–Dependent Unfolding

One of the most powerful features of Moment Unfolding is its ability to extract

moments as a function of another observable, such as jet transverse momentum.

When implemented with momentum–dependent parameters in the Boltzmann weight

function,

$$g(z; p_T) = \frac{1}{P(p_T)} \exp\left(-\sum_{a=1}^{n} \beta_a(p_T) z^a\right), \tag{I.27}$$

the generator unfolds moments conditional on the jet p_T.

This ability to unfold moments as functions of p_T is a necessary feature for any

modern unfolding method. Many QCD phenomena exhibit strong scale dependence,

including parton shower evolution, hadronization effects, and detector response.

Understanding the momentum–dependence is necessary to account for these variations.

Jets at different p_T values populate different regions of the substructure phase space.

A momentum–dependent reweighting is therefore important for better coverage of the phase space. As we will see, the momentum–dependent approach effectively leverages the statistical power of the entire dataset while allowing for local adaptations in different p_T regions, resulting in more precise measurements overall.

The implementation of momentum–dependent parameters, $\beta_a(p_T)$ introduces additional complexity that must be carefully managed. We found that linear parametrization of $\beta_a(p_T)$ provides a good balance between flexibility and stability for the jet substructure observables studied.[1] More complex parametrizations might be warranted for observables with a strongly non–linear p_T dependence. However, increasing the number of parameters in the momentum–dependent approach must be approached with caution, since that precisely undoes the regularisation imposed by the Moment Unfolding method. If the training remains unstable despite appropriately constraining the functional form of $\beta_a(p_T)$ regularization in the neural network training may become necessary.

I.D.3.i Inclusive Distributions

Figure **[fig –KD]** presents the inclusive distributions of jet mass, jet charge, jet width, and groomed momentum fraction after applying momentum–dependent Moment Unfolding. Compared to the non-momentum-dependent results shown in

---

[1]Moment Unfolding can be run as a binned method by parametrizing $\beta_a(p_T)$ as a piecewise constant function.

Figure [fig –KD], these distributions demonstrate markedly better agreement with the truth. This improvement occurs because the momentum-dependent approach effectively implements a more refined correction, allowing the reweighting function to adapt to the changing detector response and physics differences across the p_T) spectrum. By parametrising β_a as a function of p_T, we account for the fact that both detector effects and physics modelling discrepancies can vary significantly with jet energy.

The inclusive distributions obtained through momentum–conditioned unfolding also reveal subtle physical features that might be obscured in the non–conditioned approach. For example, in the jet mass distribution, the shape of the low–mass region shows better agreement with truth when using momentum conditioning, correctly capturing the interplay between soft radiation and resolution effects that varies with jet p_T.

I.D.3.ii   Differential analysis

When the momentum–dependent unfolding approach is used for a differential analysis of moments as a function of p_T, it provides a comprehensive understanding of scaling behaviours. This dual perspective offers particularly powerful constraints on theoretical models, as it simultaneously tests the overall distribution shapes and

their evolution with energy scale. Figure **[fig:pt-dependent-moments –KD]** shows the first and second moments of each observable as a function of p_T.

The results demonstrate that Moment Unfolding successfully recovers the p_T–dependent moments of all the observables, capturing the variations in these quantities with jet p_T. This capability is particularly valuable for testing theoretical predictions of how jet properties scale with energy, a fundamental aspect of QCD phenomenology.

These momentum–dependent results allow us to analyze the scaling behaviour of each observable. The jet mass shows expected direct scaling with p_T, consistent with theoretical expectations from perturbative QCD **[cite –KD]**, the mean jet charge decreases slightly with increasing p_T, reflecting the increased role of gluon radiation at higher energies **[cite –KD]**, the jet width decreases with p_T, consistent with the collimation of high–energy jets due to the Lorentz boost **[cite –KD]**, and the groomed momentum fraction shows minimal p_T dependence, a feature expected from the scale–invariant nature of the splitting functions in QCD **[cite –KD]**.

These results not only validate the Moment Unfolding method but also demonstrate its utility for extracting physically meaningful insights about jet formation and evolution. These scaling behaviours, when applied to real collider data, serve as stringent tests of perturbative QCD and provide constraints on Monte Carlo tuning parameters.

I.E   Comparison with Alternative Methods

To assess the performance of Moment Unfolding relative to existing approaches, we compare our results with those obtained from two alternative methods,

1. **OmniFold**: A state–of–the–art unbinned unfolding method that reconstructs the full distribution [cite –KD], and

2. **Iterative Bayesian Unfolding (IBU)**: A traditional binned unfolding method followed by moment calculation [cite –KD].

For a comprehensive comparison, we also include a variation on IBU that applies bin–by–bin corrections to account for binning effects in the moment calculation. Figure [fig:method-comparison –KD] shows the moments of each observable as a function of jet p_T for all methods, comparing them to the true values. The lower panels show the ratio to truth, highlighting differences between methods.

This comparison highlights some of Moment Unfolding's most important properties. The Moment Unfolding algorithm achieves comparable or better precision than OmniFold for most observables, despite, or perhaps due to, its more targeted approach. This demonstrates the advantage of directly targeting the moments without having to unfold the entire distribution. Both unbinned methods (Moment Unfolding and OmniFold) significantly outperform the binned approaches (IBU and

IBU+Correction) in precision, highlighting the benefits of avoiding binning artifacts. The bin–by–bin correction improves the performance of IBU but still does not match the precision of the unbinned methods, indicating that the fundamental limitations of binning cannot be fully mitigated by post–hoc corrections.

However, Moment Unfolding requires significantly less computational resources than OMNIFOLD due to its non–iterative nature. On average, Moment Unfolding can be run in about 1% of the time required for OMNIFOLD, while achieving comparable or better precision. IBU is about 10^4 times faster than Moment Unfolding. This comparison highlights the practical advantages of Moment Unfolding in scenarios where computational resources are limited or where multiple unfolding tasks need to be performed, such as in systematic uncertainty studies. That said, in analyses where computation is at a high premium, and the binning artifacts introduced by IBU are acceptable, IBU continues to offer a low–tech, light unfolding procedure.

These results establish Moment Unfolding as a powerful tool for precision measurements of statistical moments in high-energy physics, offering advantages in both statistical precision and computational efficiency compared to existing methods.

I.F   Conclusion

Moment Unfolding represents a significant advancement in the statistical toolkit for particle physics measurements, providing a novel approach to directly extract moments of distributions without the intermediate step of reconstructing the full spectrum. This methodology addresses several challenges in traditional unfolding techniques while offering unique advantages for theoretical comparison and interpretation. An intrinsically unbinned method, Moment Unfolding eliminates binning artifacts and dimensionality limitations that affect traditional methods. By using a physically–motivated reweighting function based on the Boltzmann distribution, it provides a principled approach to the unfolding problem, with clear connections to the maximum entropy principle from statistical mechanics.

The adversarial training framework effectively addresses the two–level nature of the unfolding problem, allowing particle–level corrections to be optimized based on detector–level comparisons. By extending this basic architecture to unfold moments conditional on momentum, Moment Unfolding can be used for detailed studies of how moments scale with energy, providing powerful constraints on theoretical models.

As a non-iterative method that is designed specifically to unfold moments rather than full spectra, Moment Unfolding offers significant computational advantages over iterative approaches, particularly for systematic uncertainty studies that require mul-

tiple unfolding runs. The application to jet substructure measurements in Sec. I.D.2 demonstrates the practical utility of Moment Unfolding for real physics analyses. The method successfully recovers both inclusive and differential moments of diverse jet observables, providing enhanced precision compared to traditional binned approaches and comparable or better accuracy than full-distribution unbinned methods. These capabilities make Moment Unfolding particularly valuable for precision tests of QCD, where theoretical predictions are often most precise at the level of moments rather than full distributions.

Although Moment Unfolding excels at extracting moments, it does not guarantee optimal reconstruction of the full distribution, particularly when higher moments are significant. The choice of how many moments to unfold represents a regularization parameter that must be carefully selected based on the physics application and available statistics. The basic formulation discussed above is designed for moments of a single observable, though extensions to joint moments are possible.

I.F.1   Towards Unfolding Distributions

While Moment Unfolding offers significant advantages for moment–specific measurements, many analyses require the full unfolded distribution rather than just its moments. This raises the question: can the Boltzmann–inspired approach be extended

to unfold entire distributions? In a sense, this should be relatively straightforward. If the degree of the polynomial in the exponent of the Boltzmann factor governs the number of moments unfolded, by simply replacing the polynomial with a Taylor series,

$$g(z) = \frac{1}{P} \exp\left(-\sum_{a=1}^{\infty} \beta_a z^a\right), \tag{I.28}$$

we could, in principle, constrain all moments simultaneously, thereby unfolding the entire distribution. However, since the degree of the polynomial was the regularizing parameter for the inverse problem, by setting it to infinity this approach removes the regularisation central to the algorithm. The resulting ill-posed inverse problem requires significant regularization.

These challenges motivate the development of the Reweighting Adversarial Network (RAN) method, which will be introduced in the next chapter. RAN builds upon the foundations laid by Moment Unfolding, extending the approach to full distribution unfolding while maintaining its unbinned, non–iterative, adversarial nature. By incorporating more flexible neural network parameterizations and Wasserstein metrics for distribution comparison, RAN addresses the limitations of Moment Unfolding while preserving its core strengths. This evolution from moment–specific to full–distribution unfolding represents a natural progression in developing a comprehensive framework for unbinned cross-section measurements in particle physics.

Bibliography

[1]   M. Thomson. *Modern particle physics*. eng. Oclc: 840462909. Cambridge, United Kingdom: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6. URL: http://digitale-objekte.hbz-nrw.de/storage/2013/11/04/file_5/5371837.pdf (visited on 07/18/2025).

[2]   D. Kotlorz and S. V. Mikhailov. "Cut Moments and a Generalization of DGLAP Equations". In: *Journal of High Energy Physics* 06.6 (2014), p. 065. ISSN: 1029-8479. DOI: 10.1007/jhep06(2014)065. arXiv: 1404.5172 [hep-th]. (Visited on 07/18/2025).

[3]   D. Kotlorz and A. Kotlorz. "Evolution Equations for Truncated Moments of the Parton Distributions". In: *Physics Letters B* 644.4 (2007), pp. 284–287. ISSN: 03702693. DOI: 10.1016/j.physletb.2006.11.054. arXiv: hep-ph/0610282. (Visited on 07/18/2025).

[4]   P. Skands. "Introduction to QCD". In: *Searching for New Physics at Small and Large Scales.* 2013, pp. 341–420. DOI: `10.1142/9789814525220_0008`. arXiv: `1207.2389 [hep-ph]`. (Visited on 07/18/2025).

[5]   G. Dissertori, I. G. Knowles, and a. M. Schmelling. *Quantum Chromodynamics: High Energy Experiments and Theory.* International Series of Monographs on Physics. Oxford, New York: Oxford University Press, Apr. 2003. ISBN: 978-0-19-850572-3.

[6]   G. Altarelli. "Partons in Quantum Chromodynamics". In: *Physics Reports* 81.1 (1982), p. 1. ISSN: 0370-1573. DOI: `10.1016/0370-1573(82)90127-2`. (Visited on 07/18/2025).

[7]   G. Altarelli. *Elementary Particles.* 1st ed. Vol. 4. 1. Berlin ; Heidelberg: Springer, 2008. ISBN: 3-540-74202-6.

[8]   D. J. Gross and F. Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Physical Review Letters* 30.26 (June 1973), pp. 1343–1346. DOI: `10.1103/PhysRevLett.30.1343`. (Visited on 07/18/2025).

[9]   D. J. Gross and F. Wilczek. "Asymptotically Free Gauge Theories. I". In: *Physical Review D* 8.10 (Nov. 1973), pp. 3633–3652. DOI: `10.1103/PhysRevD.8.3633`. (Visited on 07/18/2025).

[10] H. D. Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Physical Review Letters* 30.26 (1973), pp. 1346–1349. DOI: `10.1103/PhysRevLett.30.1346`. (Visited on 07/18/2025).

[11] S. Moch, J. A. M. Vermaseren, and A. Vogt. "The Three-Loop Splitting Functions in QCD: The Non-Singlet Case". In: *Nuclear Physics B* 688.1-2 (2004), pp. 101–134. ISSN: 05503213. DOI: `10.1016/j.nuclphysb.2004.03.030`. arXiv: `hep-ph/0403192`. (Visited on 07/18/2025).

[12] A. Vogt, S.-O. Moch, B. Ruijl, T. Ueda, and J. A. M. Vermaseren. "Four-Loop Results on Anomalous Dimensions and Splitting Functions in QCD". In: *PoS* Radcor2017 (Jan. 2018). Ed. by A. Hoang and C. Schneider, p. 046. DOI: `10.22323/1.290.0046`. arXiv: `1801.06085 [hep-ph]`.

[13] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. 8. Cambridge: Cambridge University Press, Feb. 2011. ISBN: 978-0-521-54589-1. DOI: `10.1017/cbo9780511628788`. (Visited on 07/18/2025).

[14] G. P. Salam. "Elements of QCD for hadron colliders". In: (Jan. 2011). arXiv:1011.5131 [hep-ph]. DOI: `10.48550/arXiv.1011.5131`. URL: `http://arxiv.org/abs/1011.5131` (visited on 07/18/2025).

[15] L. Schoeffel. "An Elegant and Fast Method to Solve QCD Evolution Equations. Application to the Determination of the Gluon Content of the Pomeron". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 423.2 (1999), pp. 439–445. ISSN: 0168-9002. DOI: `10.1016/s0168-9002(98)01316-3`. (Visited on 07/18/2025).

[16] D. Strózik-Kotlorz and A. Kotlorz. "DGLAP Evolution of the Truncated Moments of PDFs in Spin Physics". In: *Physics of Particles and Nuclei* 45.1 (2014), pp. 52–53. ISSN: 1531-8559. DOI: `10.1134/s1063779614010985`. (Visited on 07/18/2025).

[17] A. Simonelli. "Analytic Solutions of the DGLAP Evolution and Theoretical Uncertainties". In: 84.8 (Aug. 2024), p. 867. DOI: `10.1140/epjc/s10052-024-13169-8`. arXiv: `2401.13663 [hep-ph]`. (Visited on 07/18/2025).

[18] M. Göckeler, R. Horsley, H. Perlt, P. E. L. Rakow, G. Schierholz, and A. Schiller. "Operator Product Expansion on the Lattice: Analytic Wilson Coefficients". In: Lat2006 (2006), p. 119. DOI: `10.22323/1.032.0119`. arXiv: `hep-lat/0610064`. (Visited on 07/18/2025).

[19] W. A. Bardeen, A. J. Buras, D. W. Duke, and T. Muta. "Deep-Inelastic Scattering beyond the Leading Order in Asymptotically Free Gauge Theories".

In: *Physical Review D* 18.11 (1978), p. 3998. DOI: `10.1103/PhysRevD.18.3998`. (Visited on 07/18/2025).

[20] T. Gehrmann, G. Luisoni, and P. F. Monni. "Resummation of jet rates and event-shape distributions in eˆ+ eˆ -". In: *Parton radiation and fragmentation from LHC to FCC-ee.* 2017, pp. 97–102.

[21] S. Catani, L. Trentadue, G. Turnock, and B. R. Webber. "Resummation of Large Logarithms in e+ E- Event Shape Distributions". In: *Nucl. Phys. B* 407 (1993), pp. 3–42. DOI: `10.1016/0550-3213(93)90271-p`.