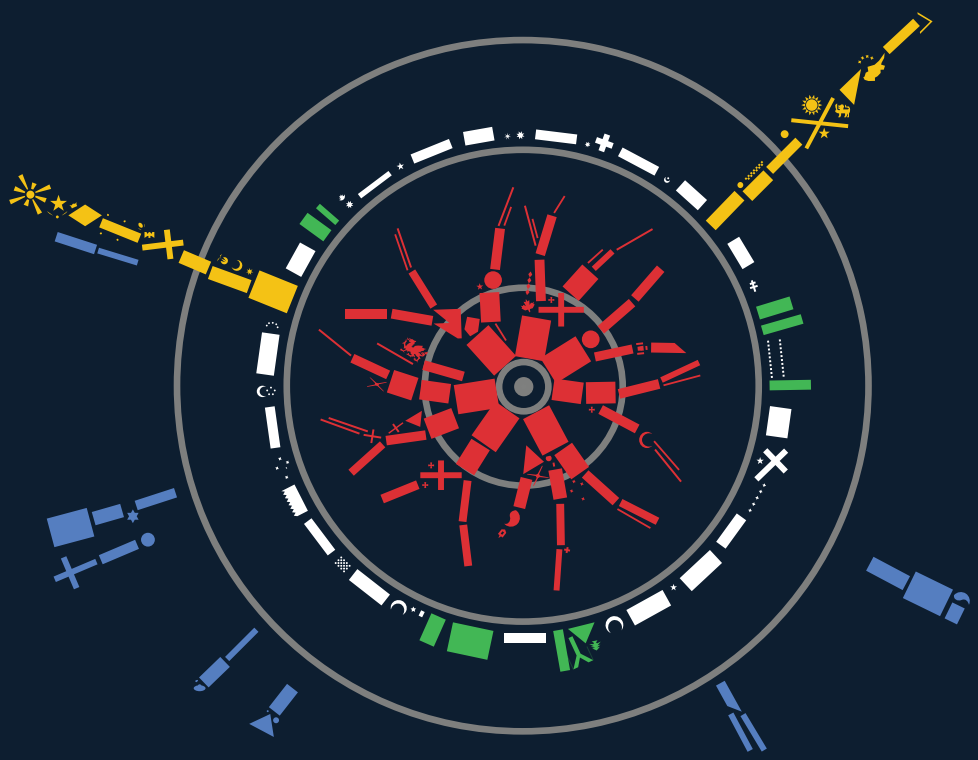
Resummation and Machine Learning Techniques

Towards Precision Phenomenology at the LHC



Tanjona R. Rabemananjara

UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata

Dipartimento di Fisica

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata

Ciclo XXXIV

Resummation and Machine Learning Techniques towards Precision Phenomenology at the LHC

Settore Scientifico Disciplinare FIS/02

Supervisore: Prof. Stefano Forte

Coordinatore: Prof. Matteo Paris

Tesi di Dottorato di:

Tanjona R. Rabemananjara

Anno Accademico 2020-2021

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the grant agreement number **740006**.

Commission of the final examination:

External Referees:

Dr. Marco Bonvini	INFN Sezione di Roma I
Dr. Riccardo Torre	INFN Sezione di Genova

External Members:

Prof. Tilman Plehn	Heidelberg University
Assoc. Prof. Juan Rojo	Vrije Universiteit Amsterdam

Internal Members:

Assoc. Prof. Stefano Carrazza	Università degli Studi di Milano

Final examination:

Date: December, 20th 2021

Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

To my mother and Angela whom I owe everything

Internal illustrations:

Tanjona R. Rabemananjara

Design:

Tanjona R. Rabemananjara

Cover illustration:

Higgs diphoton events in terms of flags from the CERN 2016 Summer School

MIUR subjects:

FIS/02

KEYWORDS:

QCD, fixed-order calculations, resummation, PDFs, machine learning, GANs

ABSTRACT

The present thesis focuses on three distinct yet complementary areas in QCD. The first area concerns the resummation of large logarithmic contributions appearing in transverse momentum distributions. In particular, we focus on the hadroproduction of colour singlet final states such as Higgs boson produced via gluon fusion and the production of a lepton-pair via Drell–Yan mechanism. We present phenomenological studies of a combined resummation formalism in which standard resummation of logarithms of Q/p_T is supplemented with the resummation of logarithmic contributions at large $x = Q^2/\hat{s}$. In such a formalism, small-p_T and threshold logarithms are resummed up to NNLL and NNLL* respectively. The second area concerns the construction of an approximate NNLO expression for the transverse momentum spectra of the Higgs boson production by exploiting the analytical structure of various resummed formulae in Mellin space. The approximation we construct relies on the combination of two types of resummations, namely threshold and high-energy (or small-x). Detailed phenomenological studies, both at the partonic and hadronic level, are presented. And finally, the third area concerns the a posteriori treatment of Parton Distribution Functions (PDFs), specifically the compression of Monte Carlo PDF replicas using techniques from deep generative models such as generative adversarial models (or GANs in short). We show that such a GAN-based methodology results in a compression methodology that is able to provide a compressed set with smaller number of replicas and a more adequate representation of the original probability distribution. The possibility of using this methodology to address the problem of finite size effects in PDF determination is also investigated.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

QCD at high energies

It is now an accepted fact that most of the visible matter in the universe is composed of *hadrons*. Hadrons, with the examples of protons and neutrons, are bound states of *quarks* and *gluons*. The interaction between these elementary constituents–quarks and gluons (collectively known as *partons*)–are dictated by the strong nuclear force whose mathematical description is formulated in a quantum field theory known as *Quantum Chromodynamics* (or QCD in short). In this picture, partons carry *colour charges* that transform under the (anti-) fundamental and adjoint representation of the group SU(3), but which at the level of hadrons lead to a neutral colour combination. QCD, in turns, constitutes a part of a larger theory called *Standard Model* (SM) which provides a theoretical framework to the description of all known interactions (electromagnetic, weak and strong interactions) except gravity.

One of the defining features of QCD is that as the energy grows, the elementary interactions taking place in a collision of hadrons occur at distances much smaller than the confinement scale, allowing for their descriptions to be formulated in terms of quasi-free quarks and gluons. In this respect, theoretical predictions of hard scatterings are calculated using perturbative QCD (or pQCD). In pQCD, the observable of interest is expanded as a power series in the QCD strong coupling α_s. In the case where the strong coupling is small compared to other scales involved in the process, truncating the perturbative series at some finite power of α_s is an accurate approximation to the full calculation. This is crucial since the exact computation of physical observables is not feasible within the realm of QCD. To the present day, several observables that are of interests to hadron colliders such as LHC have been computed to very high accuracy in perturbation theory both at the integrated [9–21] and differential [22–33] levels. Whilst pushing the frontier of currently known perturbative orders seems to be a gigantic task, it is natural to extract as much information as possible concerning the unknown orders from known orders. Hereinafter, we refer to this as the estimation of the *missing higher-order uncertainties*.

However, in order to compute measurable observables that can be compared to experimental measurements, the results from perturbative calculations have to be supplemented with the knowledge of the momentum distributions of the partons inside the colliding hadrons. Such information are encapsulated in what is known as *Parton Distribution Functions* (PDFs). This suggests that in order to perform precision calculations for hadron-initiated processes (henceforth referred to as *hadroproduction*), an accurate knowledge of the PDFs is crucial. As a matter of fact, not only PDFs are vital to high-precision physics at LHC, but as will be briefly discussed later, they are essential tools to interpret experimental measurements for a variety of hard processes in light of the quest for new physics (beyond the Standard Model or BSM). Since they contain information on long-distance interactions, parton distributions are non-perturbative objects, and hence they cannot be computed from first principle. Instead, they have to be determined through a fitting procedure in which theoretical predictions are compared to experimental measurements.

The problem of accurately determining PDFs has seen considerable advancements in the past decades. There exists thus far various PDF fitting groups [34–74] that implement different methodology and provide different estimation of the PDF errors. Remarkably, despite the fact the resulting PDFs may differ, they agree within a reasonable uncertainty. Roughly speaking, the underpinning idea behind a PDF determination consists in learning from experimental data a set of functions. That is, a parton density function is modelled in terms of a function whose parameters are adjusted during the fit to yield matching theoretical predictions. Such a function, for instance, can be defined in terms of a suitable polynomials whose functional form is chosen based on physical arguments such as *Regge theory* [75], *Brodsky-Farrar quark counting rules* [76], et cetera. This approach is used by all the PDF fitting groups mentioned previously except the NNPDF. The NNPDF collaboration takes a different approach by modelling the parton density functions in terms of *neural networks*. This approach turns out to be very robust as it removes the bias introduced in the choice of a particular functional form to fit the PDFs. Very recently [4,77], thanks to the adoption of state-of-the-art machine learning methodologies and the inclusion of new datasets, the NNPDF collaboration claims to reach–in a wide range of kinematic regions–*one-percent* relative uncertainties. This is a milestone toward reaching the objective of having smaller PDF errors, which so far represented one of the dominant sources of uncertainties in LHC processes such as the Higgs boson production. However, in order to consistently produce PDFs accurate at a percent-level, theoretical uncertainties [52, 53] (especially those arising from missing higher-orders) have to be taken into account. Hopefully, with the current NNPDF4.0 methodology [77] and evolution codes such as EKO [78] and YADISM [79], this will become possible in the very near future.

Returning back to the short-distance interactions, perturbative computations suffer from two main pathologies: first the appearance of large logarithmic enhanced terms in a multiscale process even if the observable in question is *Infrared and Collinear* (IRC) safe, second the non-existence of a robust and reliable method to estimate missing higher-order uncertainties. In the former, the smallness of the QCD coupling α_s is compensated by the largeness of $\ln \mu^2$ (where μ denotes an arbitrary ratio of scales) such that $\alpha_s \ln \mu^2 \sim \mathcal{O}(1)$. In such a scenario, the convergence of the perturbative series is spoiled and any truncation at a given power of α_s is meaningless. These types of logarithmic divergences are cured by performing an *all-order* computation through a procedure known as *resummation*. Generally, different resummation techniques are required for different classes of large logarithms. For instance, logarithmic enhancements arising when the invariant mass of the final-state system approaches the kinematic threshold are resummed in *threshold resummation*.

The issue related to finding an efficient method to estimate theoretical uncertainties associated with missing higher-order corrections is still arguably an open question. Currently, the most standard way to evaluate such uncertainties is by varying the unphysical scales involved in the process according to a *scale variation prescription*. This approach, however, has a number of caveats. Various approaches have recently emerged [80–82] in order to address the shortcomings related to the scale variation method. However, most of these approaches are only applicable to particular types of processes and observables.

All of this said, in order to push forward precision and discovery physics at the LHC, the three pillars of QCD–namely the fixed-order calculations, resummations, and PDFs–need to be determined at the highest accuracy possible. While significant efforts have been underway at the PDF level to achieve this objective, much works remain both in

terms of fixed-order computations (especially with regards to the estimation of missing higher-order uncertainties), and resummations. The present thesis attempts to fill the tiny gaps in those areas by combining theoretical computations and machine learning.

Structure of the thesis

In Chapter 1, we first establishes background notations and definitions on which the theoretical part of the present thesis is based upon. In particular, we review how theoretical predictions in QCD are computed. Specifically, we briefly review how matrix elements factorize in the collinear and soft limits and re-derive the expression of the splitting functions using non-classical approach such as spinor helicity formalism and MHV techniques. This is followed by a description of the underlying ideas behind a PDF determination. We then conclude the introductory part by commenting on the estimation of missing higher-order uncertainties in fixed-order computations and the road toward *one-percent* accuracy.

In Chapter 2, we detail how the large logarithmic contributions appearing in perturbative computations due to kinematic enhancements are cured using resummation techniques in the context of transverse momentum distributions. Specifically, we focus on the resummations of contributions that are enhanced at the partonic threshold and at small transverse momentum. We then provide detailed derivations on how subclasses of the threshold logarithms can be consistently incorporated into the transverse momentum resummation. In doing so, we extensively study the relation between the soft and collinear logarithms which drive the transverse momentum distribution.

In Chapter 3, we perform extensive phenomenological studies in which the soft-improved standard transverse momentum resummation is compared with the standard transverse momentum resummation in the context of Higgs boson production at the LHC and Z-boson production via Drell–Yan (DY) mechanism. This eventually requires us to construct prescriptions to perform the inverse Fourier and Mellin transforms in which the resummations are derived. Then, by studying the singularity structure of the resummed expressions (specifically threshold and high energy resummations), we try to construct an approximation to the NNLO transverse momentum distribution of the Higgs boson produced via gluon fusion in the HEFT approximation.

In Chapter 4, we introduce machine learning concepts that are relevant for the treatment of Monte Carlo PDF sets. For the sake of pedagogical clarity, we purposefully give a brief review of deep learning and analyse, through explicit illustrations the mechanism behind the training of a deep neural network. We then describe how deep learning can be used to *model* posterior probability distributions. In particular, we concentrate our attention to a specific deep generative model known as *Generative Adversarial Network* (or GANs in short). To conclude the chapter, we give a brief survey of how GANs have been applied so far to tackle problems in high energy physics.

In Chapter 5, we apply the concept of GANs to the problem of compressing Monte Carlo PDF sets. The need for a PDF compression can be stated as follows: the main published PDFs are typically based on a $N_R = 1000$ replicas fit, however, having to deal with such a large set of replicas is not ideal when performing Monte Carlo simulations, for example. For this reason, it is therefore advantageous to construct a set of replicas with smaller sizes

while maintaining the same underlying probability distribution as the prior. The role of the GAN is then to further enhance the statistics of a given PDF set by generating *synthetic* replicas before a compression is performed. This, as will be discussed further, results in a compression methodology that is able to provide a compressed set with smaller number of replicas and a more adequate representation of the original probability distribution. Before closing the chapter, we investigate the possibility of using the GANs to tackle *finite size effects* in PDFs.

Finally, we conclude by first summarizing the main guiding ideas and results presented in the present thesis. In particular, we highlight the relevance of resummation in perturbative computations and the possibility of combining different categories of resummation to construct approximate predictions for yet unknown higher-order contributions.

Part I

PROLOGUE

1

PARTON DISTRIBUTION FUNCTIONS AT COLLIDER PHYSICS

In the following chapter, we set the stage to the research topics studied in the present thesis. In particular, we highlight the role the QCD factorization theorem plays in perturbative calculations in order to separate short-distance interactions from long-distance ones. The concept of Mellin transforms is then introduced in the context of transverse momentum observables, which turns out to be crucial for all-order computations. As very well known in pQCD, scattering amplitudes factorize in presence of radiative (soft and/or collinear) emissions. These properties are reviewed in Sec. (1.1.1) using the *spinor helicity formalism* and *Maximally Helicity Violating* (MHV) techniques. Such computations eventually lead to the introduction of *splitting functions* that govern the DGLAP equations. We then explicitly show how the solutions to the DGLAP equations are computed. Such a derivation is relevant in the context of *soft-improved transverse momentum resummation* (discussed in Sec. (2.4)) where the PDF evolutions need to be treated in a particular way. While the μ^2-dependence of the PDFs is predicted at all values by the DGLAP equations, the x-dependence must be derived through a fit to the experimental data. Details on such determinations are briefly described in Sec. (1.1.3). To conclude the chapter, we discuss the main missing pieces in terms of PDF determination and perturbative computations toward achieving a *one-percent* accuracy.

1.1 Structure of QCD predictions

1.1.1 Perturbative computations

At hadron colliders such as LHC, the colliding particles are bound configurations of quarks and gluons. As mentioned in the introduction, these bound states are held together by the strong nuclear force whose mathematical descriptions are described by the Quantum Chromodynamics (or QCD in short). A proton, for example, can be thought as a bound state of one *down* quark (d) and two *up* quarks (u)[1]. At high-energies (greater than $\Lambda_{\text{QCD}}\sim$200 MeV) where hadrons behave as bunches of quasi-free partons (phenomenon known as *Asymptotic Freedom*), QCD effects modify the effective composition of the protons where they also contain gluons, and quarks & antiquarks of all flavours. In the parton model, interactions of hadrons taking place at such energy can be described through the interactions of partons via the *QCD factorization theorem*.

1 Recent studies of the contents of the proton by the NNPDF collaboration, however, seem to suggests that the bound state of the proton also contains a *charm* (c) quark. This is thought to be the case due to the presence a bump in the charm-quark distribution at $x\sim$0.3

Hadron collisions & QCD factorization theorem

Physical processes (specifically hadroproduction) typically studied at hadron colliders take place at very high energies[2] allowing for the separation of the long-distance physics and the short-distance ones. The assumption of the parton model states that physical observables can be computed in terms of perturbative objects such as partonic cross sections. Thus, for a collision of two hadrons h_1 and h_2 leading to the creation of a final-state system F, an observable $\mathcal{O}$ (typically cross section) can be factorized as follows

$$\mathcal{O} = \sum_{a,b} f_{a/1}(\mu) \otimes f_{b/2}(\mu) \otimes \hat{\mathcal{O}}_{ab}(\mu), \tag{1.1.1}$$

where the sum takes into account all the possible sub-partonic channels and $\otimes$ denotes the convolution. The function $f_{a/h}$ represents the Parton Distribution Function (PDF) and describes the fraction of momentum carried out by a parton a from a hadron h. The partonic version of the physical observable is represented by $\hat{\mathcal{O}}$. The scale μ–conventionally known as *factorization scale*–separates soft (long-distance) and hard (short-distance) physics. As physical quantities should not depend on the scale μ, Eq. (1.1.1) suggests that the scale dependence must cancel when the perturbative part is convoluted with the PDFs. However, in perturbative computations, only a few orders in the strong coupling are included in $\hat{\mathcal{O}}$, leaving the scale dependence present. Based on renormalization group invariance, such a scale dependence is, however, expected to decrease as the perturbative order increases. As a result, the residual scale dependence μ is typically used to estimate the size of the next unknown contribution in the perturbative expansion.

The PDFs, as opposed to the hard scattering part, are process-agnostic. They cannot be computed using perturbative QCD and therefore needs to be extracted from experimental data using some fitting procedure. In order for the PDFs to have a probabilistic interpretation, they need to satisfy a set of constraints known as *sum rules*. In other words, PDFs are constrained such that integrating a parton PDF over all the momentum fraction x must yield the number of valence partons present in the hadron. For instance, a proton must be composed of *one valence down* quark and *two valence up* quarks. In addition, for the remaining quark flavours ($\ell = s, c, b, t$), the valence must be equal to zero as these quark-antiquark pairs fluctuate in and out of existence within the proton. These constraints can be mathematically written as

$$\int_0^1 dx\,(f_\ell(x) - f_{\bar{\ell}}(x)) = 0, \int_0^1 dx\,(f_d(x) - f_{\bar{d}}(x)) = 1, \int_0^1 dx\,(f_u(x) - f_{\bar{u}}(x)) = 2. \tag{1.1.2}$$

Since there is no number conservation law for gluons, gluons are only constrained by momentum conservation, i.e. satisfy $\sum_i \int x\, f_i(x) = 1$ where i sums over the gluon and all quark & antiquark flavours.

The factorization formula given in Eq. (1.1.1) thus far has only been rigorously proven for sufficiently inclusive observable such as *Deep Inelastic Scatterings* (DIS) and its validity for more exclusive observables remain to be proved (especially at higher perturbative order). That is, for most of the physical observables that are of interests to LHC phenomenology,

2 The Run II of the LHC, for examples approached a center-of-mass energy of $\sqrt{s} \sim 14$ TeV.

such as differential distributions, the factorization theorem is simply assumed to hold true. In addition, Eq. (1.1.1) is only valid up to power corrections $\mathcal{O}\left(\Lambda_{\rm QCD}/Q\right)$ therefore neglecting hadronization corrections and *higher-twist contributions* which at the energy considered in this thesis are negligible.

Let us make a final comment on the partonic component of Eq. (1.1.1). As mentioned before, the hard parton scattering $\hat{\mathcal{O}}$ can be computed using perturbative approach. In the presence of collinear emission in either the initial or final sate, it exhibits collinear divergences. Such divergences can be factorized out from the hard scattering as

$$\hat{\mathcal{O}}_{ab}(\mu) = \sum_{i,j} \mathcal{C}_{i/a}(\mu) \otimes \mathcal{C}_{j/b}(\mu) \otimes \hat{\mathcal{O}}_{ij} \tag{1.1.3}$$

where $\mathcal{C}_{i/a}$ are referred to as *collinear counter-terms* describing the singular part of the parton-in-parton distribution due to higher order corrections to the parton a. In the case where the observable we are considering is both *Infrared and Collinear Safe* (IRC safe), the functions $\mathcal{C}_{i/a}$ and $\hat{\mathcal{O}}_{ij}$ are well defined (finite) perturbative functions. Details on this subject are explored in the subsequent sections.

Notations & Definitions for transverse momentum distributions

Consider the collision of two protons $p_1+p_2 \longrightarrow \mathrm{F}+X$ in which a colour singlet object F is produced with an invariant mass M and transverse momentum p_T. We can define a scaling variable τ as follows

$$\tau = \frac{\left(E_T + p_T\right)^2}{s} \tag{1.1.4}$$

where $s=(p_1+p_2)^2$ represent the hadronic center-of-mass energy squared. The analogue of τ and s at the partonic level is given by

$$x = \frac{\left(E_T + p_T\right)^2}{\hat{s}} \quad \text{where} \quad E_T = \sqrt{M^2 + p_T^2} \quad \text{and} \quad \hat{s} = x_1 x_2 s \tag{1.1.5}$$

with x_1 and x_2 representing the momentum fractions. By introducing a new variable $Q=E_T+p_T$, the hadronic and partonic scaling variables become $\tau=Q^2/s$ and $x=Q^2/\hat{s}$ respectively. Here, Q is the threshold energy, i.e. the minimum energy needed to produce the final state system F with an invariant mass M. This ensures that $0\leq\tau\leq1$. Using the factorization formula given in Eq. (1.1.1), the hadronic cross section for the transverse momentum distribution can be written as

$$\frac{1}{\tau}\frac{d\sigma}{d\xi_p}\left(\tau, \frac{M^2}{Q^2}, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) = \sum_{a,b} \int \frac{dx}{x} \mathcal{L}_{ab}\left(\frac{\tau}{x}, \mu_F^2\right) \frac{1}{x}\frac{d\hat{\sigma}_{ab}}{d\xi_p}\left(x, \frac{M^2}{Q^2}, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) \tag{1.1.6}$$

where we have defined the dimensionless variable $\xi_p=p_T^2/M^2$. For brevity, we have omitted the explicit dependence on the strong coupling α_s. It should always be kept in mind that $d\hat{\sigma}_{ab}/d\xi_p$ is a perturbative function admitting a series expansion in α_s. The luminosity function $\mathcal{L}_{ab}$ is expressed in terms of the two parton densities as

$$\mathcal{L}_{ab}\left(z, \mu_F^2\right) = \int_z^1 \frac{dy}{y} f_a(y) f_b\left(\frac{z}{y}\right). \tag{1.1.7}$$

In the context of all-order computations, it is clear that Eq. (1.1.6) is not suitable due to the presence of convolution. However, it turns out that in Mellin space, the convolution becomes a normal product. The Mellin transform of Eq. (1.1.6) can be constructed by defining the following transformations

$$\frac{d\sigma}{d\xi_p}\left(N,\frac{M^2}{Q^2},\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_F^2}\right)=\int_0^1 d\tau\,\tau^{N-1}\frac{d\sigma}{d\xi_p}\left(\tau,\frac{M^2}{M^2},\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_F^2}\right) \tag{1.1.8}$$

$$\frac{d\hat{\sigma}}{d\xi_p}\left(N,\frac{M^2}{Q^2},\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_R^2}\right)=\int_0^1 dx\,x^{N-1}\frac{d\hat{\sigma}}{d\xi_p}\left(x,\frac{M^2}{Q^2},\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_F^2}\right). \tag{1.1.9}$$

Using Eq. (1.1.8) and Eq. (1.1.9), it can be shown that Eq. (1.1.6) can be written as

$$\frac{d\sigma}{d\xi_p}\left(N,\frac{M^2}{Q^2},\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_F^2}\right)=\sum_{a,b}\mathcal{L}_{ab}\left(N+1,\mu_F^2\right)\frac{d\hat{\sigma}_{ab}}{d\xi_p}\left(N,\frac{M^2}{Q^2},\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_F^2}\right). \tag{1.1.10}$$

Notice that the cross sections and their Mellin transforms are denoted with the same symbol and only distinguished through their arguments. It is worth emphasizing that the Mellin transform of the hadronic cross section in Eq. (1.1.8) is taken w.r.t. the scaling variable τ while the Mellin transform of the hard partonic cross section in Eq. (1.1.9) is taken w.r.t. the partonic variable x.

Matrix element factorization in the collinear limit

To see how matrix elements factorize in the collinear limit, let us consider a 2 to 3 process where two of the outgoing particles are collinear. In particular, let us consider a process in which a quark and an antiquark annihilates and produces three gluons ($q\bar{q}\to ggg$), as depicted in Fig. 1. Using the spinor helicity formalism, each momentum involved in the process can be expressed in terms of a left and right-handed spinors $k_{a\dot{a}}=\lambda_a\tilde{\lambda}_{\dot{a}}$ (where a and $\dot{a}$ are spinor indices running from 1 to 2). For real-valued momenta, the holomorphic and anti-holomorphic spinors (λ_a and $\tilde{\lambda}_{\dot{a}}$ respectively) are complex conjugate of each other, $\lambda_a^\star=\tilde{\lambda}_{\dot{a}}$. In such a representation, massless tree-level amplitudes are straightforwardly computed by considering all the momenta to be either incoming or outgoing in which the leading non-vanishing contribution is given by the Maximally Helicity Violating (MHV) amplitudes [83–86]. By virtue of the colour kinematic decomposition [87–91], the full amplitude for a given helicity configuration of a process involving three gluons and a quark-antiquark pair is given by

$$\mathcal{M}_5\left(p,1,2,3,\bar{p}\right)=g_s^3\sum_{\mathcal{P}_3}\left(T_{a_1}T_{a_2}T_{a_3}\right)\tilde{\mathcal{M}}_5\left(p,1,2,3,\bar{p}\right) \tag{1.1.11}$$

where $\mathcal{P}_3$ sums over the permutations of all three gluons, g_s is the fine structure constant ($g_s^2=4\pi\alpha_s$), T_a are the generator matrices of $SU(3)$, and $\tilde{\mathcal{M}}$ are called *colour-ordered partial amplitudes* which have all the colour factors removed. It is worth emphasizing that each partial amplitude corresponds to a particular colour flow, which naively can be thought as the ordering in which the gluons are emitted. Let us first consider an MHV configuration in which only the quark p and gluon k_1 possess negative helicities. Using the Parke-Taylor

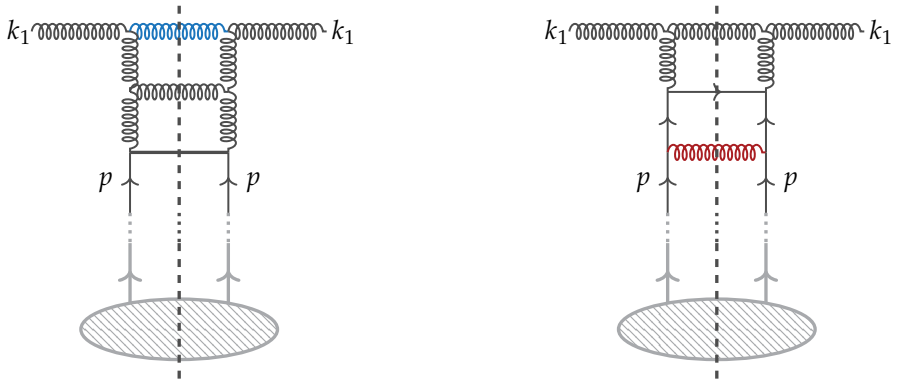


Figure 1 Sample of the squared amplitudes for the tree level diagrams that contribute to the five-point amplitude $qq \to ggg$. For both diagrams, the time runs from the bottom to the top. The dashed lines that divide the diagrams represent the *Cutkosky cut* [1–3]. The gluon that is emitted *collinearly* to the outgoing gluon k_1 is represented by the *blue-coloured* coiled lines. Similarly, the *soft* gluon that is emitted from the quark p is labelled by the *red-coloured* coiled line. The internal unlabelled lines represent the outgoing gluon k_2 and the incoming anti-quark q.

formula [92] (see discussions in App. 1.C), the MHV colour-ordered amplitude for our five-point process is given by the following expression

$$\tilde{\mathcal{M}}_5\left(p^-,1^-,2^+,3^+,\bar{p}^+\right) = \frac{\langle p1\rangle^3\langle\bar{p}1\rangle}{\langle p1\rangle\langle 12\rangle\langle 23\rangle\langle 3\bar{p}\rangle\langle\bar{p}p\rangle} \tag{1.1.12}$$

where the angle brackets are defined in terms of the scalar product of two momenta as $2(k_ik_j) = \left(\varepsilon^{ab}\lambda_a\mu_b\right)\left(\varepsilon^{\dot{a}\dot{b}}\tilde{\lambda}_{\dot{a}}\tilde{\mu}_{\dot{b}}\right) = \langle ij\rangle\,[ij]$. Therefore, the angle and square brackets representation of the spinors are defined as $\langle ij\rangle = \varepsilon^{ab}\lambda_a\mu_b$ and $[ij] = \varepsilon^{\dot{a}\dot{b}}\tilde{\lambda}_{\dot{a}}\tilde{\mu}_{\dot{b}}$. Hereinafter, we drop the dependence on p and $\bar{p}$ and only keep $\tilde{\mathcal{M}}_5\left(1^-,2^+,3^+\right)$. Indeed, due to the parity of the partial amplitudes, the helicity of the quark-antiquark line is immaterial and one only has to take into account in the full amplitude configurations where the helicities of the gluons change.

In the limit where the two gluons k_2 and k_3 are collinear, the spinor products $\langle 23\rangle$ and $[23]$ vanish. This can be used to simplify further the expression of the full amplitude since in this limit only terms that contain $\langle 23\rangle$ and/or $[23]$ in the denominator contribute. That is, the partial amplitudes $\tilde{\mathcal{M}}_5\left(1^-,2^+,3^+\right)$ and $\tilde{\mathcal{M}}_5\left(2^-,3^+,1^+\right)$ will contribute to the full amplitude while partial amplitudes with helicity configurations $\tilde{\mathcal{M}}_5\left(2^-,1^+,3^+\right)$ and $\tilde{\mathcal{M}}_5\left(3^-,1^+,2^+\right)$ will be suppressed. Introducing an intermediate momentum $q=(k_2+k_3)$, the collinearity condition implies that $k_2 = xq$ and $k_3 = (1-x)q$ where x is the fraction of momentum carried out by the gluon k_2. Defining the momenta to be $q = \mu_a\tilde{\mu}_{\dot{a}}$, $k_2 = \lambda_a\tilde{\lambda}_{\dot{a}}$, and $k_3 = \eta_a\tilde{\eta}_{\dot{a}}$, the aforementioned parametrization can be translated into spinor variables as follows

$$\begin{cases}\lambda_a = \sqrt{x}\mu_a\\ \tilde{\lambda}_{\dot{a}} = \sqrt{x}\tilde{\mu}_{\dot{a}}\end{cases}, \qquad \text{and} \qquad \begin{cases}\eta_a = \sqrt{1-x}\mu_a\\ \tilde{\eta}_{\dot{a}} = \sqrt{1-x}\tilde{\mu}_{\dot{a}}.\end{cases} \tag{1.1.13}$$

Plugging the above relations into Eq. (1.1.12) and exploiting the linearity of the angle brackets in order to factorize the fraction of momentum x, we can write the partial amplitude for the helicity configuration $\tilde{\mathcal{M}}_5\left(1^-,2^+,3^+\right)$ as follows

$$\tilde{\mathcal{M}}_5\left(1^-,2^+,3^+\right) = \frac{1}{\langle 23\rangle\sqrt{x(1-x)}}\frac{\langle p1\rangle^3\langle\bar{p}1\rangle}{\langle p1\rangle\langle 1q\rangle\langle q\bar{p}\rangle\langle\bar{p}p\rangle}, \tag{1.1.14}$$

where we have made explicit the term coming from the four-point partial amplitude $\tilde{\mathcal{M}}_4\left(1^-,q^+\right)$. Indeed, it is clear from Eq. (1.1.14) that in the collinear limit, the MHV amplitude for the helicity configuration considered factorizes into a four-point part and a collinear term, $\tilde{\mathcal{M}}_5 \sim f(2,3,x)\tilde{\mathcal{M}}_4$. The same exact calculations can be performed for all possible helicity configurations that lead to non-vanishing partial amplitudes. For completeness, we list the two remaining MHV configurations below

$$\tilde{\mathcal{M}}_5\left(1^+,2^-,3^+\right) = \frac{x^2}{\langle 23\rangle\sqrt{x(1-x)}}\tilde{\mathcal{M}}_5\left(1^+,q^-\right) \tag{1.1.15}$$

$$\tilde{\mathcal{M}}_5\left(1^+,2^+,3^-\right) = \frac{(1-x)^2}{\langle 23\rangle\sqrt{x(1-x)}}\tilde{\mathcal{M}}_5\left(1^+,q^-\right). \tag{1.1.16}$$

Notice that the anti-MHV partial amplitudes can be straightforwardly computed from the MHV configurations by realizing that $\tilde{\mathcal{M}}\left(\{h_i\}\right)$ and $\tilde{\mathcal{M}}\left(\{-h_i\}\right)$ (with h_i denoting the helicity) are related by a change of angle brackets into square brackets. That is,

$$\tilde{\mathcal{M}}_5\left(1^+,2^-,3^-\right) = \frac{1}{[23]\sqrt{x(1-x)}}\tilde{\mathcal{M}}_5\left(1^+,q^-\right) \tag{1.1.17}$$

$$\tilde{\mathcal{M}}_5\left(1^-,2^+,3^-\right) = \frac{x^2}{[23]\sqrt{x(1-x)}}\tilde{\mathcal{M}}_5\left(1^-,q^+\right) \tag{1.1.18}$$

$$\tilde{\mathcal{M}}_5\left(1^-,2^-,3^+\right) = \frac{(1-x)^2}{[23]\sqrt{x(1-x)}}\tilde{\mathcal{M}}_5\left(1^-,q^+\right) \tag{1.1.19}$$

In order to complete the computation of the full amplitude, for say the configuration $(1^-,2^+,3^+)$, we need to consider all the possible permutations of gluons as expressed in Eq. (1.1.11). This can be effortlessly done by realizing that: (a) due to the cyclic ordering of the gluons the partial amplitude $\tilde{\mathcal{M}}_5\left(2,3,1\right)$ is related to $\tilde{\mathcal{M}}_5\left(1,2,3\right)$ by a swap of k_1 and q in the four-point partial amplitude, (b) due to the antisymmetric property of the angle and square brackets $\tilde{\mathcal{M}}_5\left(1,2,3\right)$ and $\tilde{\mathcal{M}}_5\left(1,3,2\right)$ are related by an overall minus sign. Combining these observations and exploiting the normalization of the generator matrices $[T_a,T_b]=f_{abc}T_c$, we can finally write down the expression of the full amplitude for the helicity configuration $(1^-,2^+,3^+)$

$$\mathcal{M}_5\left(1^-,2^+,3^+\right) = -\frac{g_s f_{a_2a_3c}}{\langle 23\rangle\sqrt{x(1-x)}}\sum_{\mathcal{P}_2}\left(T_{a_1}T_c\right)\tilde{\mathcal{M}}_4\left(1^-,q^+\right), \tag{1.1.20}$$

where $\mathcal{P}_2$ sums over the permutation of the two gluons k_1 and q. We can now take the modulus square of the above amplitude and sum over the colours by promoting the generator matrices to structure constants and finally using $\sum f_{a_2a_3c}f_{a_2a_3d}=C_A\delta_{cd}$. Doing

so leads to an expression where the collinear term and the four-point amplitude are independent of each other

$$\left|\mathcal{M}_5\left(1^-,2^+,3^+\right)\right|^2 = \frac{g_s^2}{2(k_2k_3)}\frac{C_A}{x(1-x)}\left|\mathcal{M}_4\left(1^-,q^+\right)\right|^2 \tag{1.1.21}$$

The final step consists in summing the square amplitudes (as in Eq. (1.1.21)) over all the possible helicity configurations (including $\overline{\text{MHV}}$). Provided with Eqs. (1.1.15-1.1.19), this just amounts to a few algebraic simplifications. The final expression now writes as

$$\sum_{\text{hel.}}\left|\mathcal{M}_5\left(1,2,3\right)\right|^2 = \alpha_s\left(\frac{4\pi}{q}\right)^2\tilde{P}_{gg}^{(0)}(x)\sum_{\text{hel.}}\left|\mathcal{M}_4\left(1,q\right)\right|^2, \tag{1.1.22}$$

where we have defined the leading-order *Altarelli-Parisi* splitting function associated with the probability of the gluon to split into two gluons ($g \to gg$) to be

$$\tilde{P}_{gg}^{(0)}(x) = \frac{C_A}{\pi}\left[\frac{x}{1-x}+\frac{1-x}{x}+x(1-x)\right]. \tag{1.1.23}$$

Notice that the above expression is symmetric under the interchange of x and $(1-x)$ reflecting the fact that it does not matter which gluon carries the fraction x of the parent's momentum. It is then clear that there are different splitting functions associated with the different possible branching and each of them can be expanded as a series in powers of the strong coupling α_s. For instance, by performing the same procedure as described before, one can show that the splitting functions $\tilde{P}_{gq}^{(0)}(x)$ and $\tilde{P}_{qg}^{(0)}(x)$ are given by

$$\tilde{P}_{gq}^{(0)}(x) = \frac{C_F}{2\pi}\left[\frac{1}{x}+\frac{(1-x)^2}{x}\right], \qquad \tilde{P}_{qg}^{(0)}(x) = \frac{n_f}{2\pi}\left[x^2+(1-x)^2\right]. \tag{1.1.24}$$

Notice that we have used the shorthand notation $\tilde{P}_{ab}^{(0)}(x)$ to label splittings associated to the branching $a \to bX$. By inspection of $\tilde{P}_{gg}^{(0)}(x)$ in Eq. (1.1.23) one notice that the expression is singular when $x \to 1$. This reflects the double divergence that arise when the emitted gluon is both soft and collinear at the same time. In order to cure such a divergence, one needs to include virtual contributions depicted in Fig. 2. Since virtual contributions can only affect terms that are proportional to $\delta(1-x)$, the virtual corrections can be encoded in a function V generically written as

$$P_{gg}^{(0)}(x) = \tilde{P}_{gg}^{(0)}(x) + V\delta(1-x). \tag{1.1.25}$$

Notice that this does not apply to the $\tilde{P}_{gq}^{(0)}(x)$ and $\tilde{P}_{qg}^{(0)}(x)$ splitting functions since Eq. (1.1.24) is free of divergence when $x \to 1$ and therefore $P_{gq}^{(0)}(x) \equiv \tilde{P}_{gq}^{(0)}(x)$ and $P_{qg}^{(0)}(x) \equiv \tilde{P}_{qg}^{(0)}(x)$. Unlike for fermions, the computation of the $P_{gg}^{(0)}(x)$ splitting is slightly more difficult since

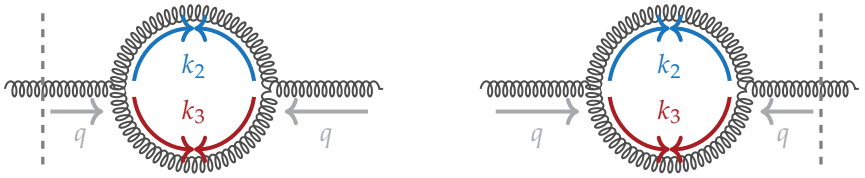


Figure 2 Amplitude squares arising from the virtual three-point contribution where two of the outgoing gluons are *collinear*. For both diagrams, the time runs from left to right. The dashed lines that slices both diagrams represent the *Cutkosky cuts* [1–3].

the number of gluons is not conserved and therefore one has to use as a constraint the conservation of momentum of the incoming gluon which can be written as follows[3]

$$\int_0^1 dx\, x \left(P_{gg}^{(0)}(x) + P_{qg}^{(0)}(x) \right) = 0. \tag{1.1.26}$$

The above momentum conservation can therefore be used to determine the value of the function V in Eq. (1.1.25). By isolating the singular part and performing the integration, one can express the gg-splitting function in terms of a *plus distribution*,

$$P_{gg}^{(0)}(x) = \left[\tilde{P}_{gq}^{(0)}(x) - x\delta(1-x) \int_0^1 d\tilde{x}\, \frac{C_A}{\pi(1-\tilde{x})} \right] + \beta_0 \delta(1-x) \tag{1.1.27}$$

$$= \frac{C_A}{\pi} \left[\frac{x}{(1-x)_+} + \frac{1-x}{x} + x(1-x) \right] + \beta_0 \delta(1-x) \tag{1.1.28}$$

where β_0 is the first coefficient of the QCD β-function. Explicit expressions of the QCD β-functions including higher-orders are given in App. 1.A. The $P_{qq}^{(0)}(x)$ splitting function can be derived in a similar way using instead the conservation of the number of fermions in which the total fraction of momentum carried away by an emitted quark $xP_{qq}^{(0)}(x)$ integrated over x vanishes. Repeating the same calculations as performed above, the splitting associated with the branching $q \to qg$ is given by

$$P_{qq}^{(0)}(x) = \frac{C_F}{2\pi} \left[\frac{1+x^2}{(1-x)_+} + \frac{3}{2}\delta(1-x) \right]. \tag{1.1.29}$$

The results presented above are only the leading-order coefficients of the perturbative splitting functions. To the present day, splitting functions are fully known up to three-loop in both the singlet and non-singlet case [93, 94], and up to four-loop in the planar [95–97] and small-x [98] limit. Since resummation of transverse momentum distributions are mainly performed in the Mellin space (as will be discussed later), the following thesis mainly focuses on the Mellin space version of the splitting functions known as *Anomalous Dimensions*. Their expressions up to three-loop are given in Refs. [93, 94].

3 The total fraction of momentum carried away by an emitted gluon is $\int dx\, x\tilde{P}_{gg}^{(0)}(x)$. On the other hand, a gluon can also split into a quark-antiquark pair where the total fraction of momentum carried away by the quark (or respectively the antiquark) is $\int dx\, xP_{qg}^{(0)}(x)$. Combining these probabilities with the case that the gluon does not radiate yields unity.

Matrix element factorization in the soft limit

In this section, we see how the matrix element factorize in the case where the gluon, say with momentum k_3, radiated from the quark or antiquark line goes soft. Let us start again with the colour kinematic decomposition of the amplitude as expressed in Eq. (1.1.11) and consider the helicity configuration as given in Eq. (1.1.12). The sum over the permutation $\mathcal{P}_3$ can be decomposed into a sum over the permutation $\mathcal{P}_2$,

$$\mathcal{M}_5\left(1^-,2^+,3^+\right) = g_s^3 \sum_{\mathcal{P}_2(1,2)} \frac{\langle p1\rangle^3\langle\bar{p}1\rangle}{\langle p1\rangle\langle 12\rangle\langle 2\bar{p}\rangle\langle\bar{p}p\rangle}\times \tag{1.1.30}$$
$$\left[\left(T_{a_3}T_{a_1}T_{a_2}\right)\frac{\langle p1\rangle}{\langle p3\rangle\langle 31\rangle} + \left(T_{a_1}T_{a_3}T_{a_2}\right)\frac{\langle 12\rangle}{\langle 13\rangle\langle 32\rangle} + \left(T_{a_1}T_{a_2}T_{a_3}\right)\frac{\langle 2\bar{p}\rangle}{\langle 23\rangle\langle\bar{p}1\rangle}\right]$$

where we recognize the partial amplitude for the four-point scattering $\tilde{\mathcal{M}}_4\left(1^-,2^+\right)$. Notice, however, that its colour structure is entangled with the factorized piece. In order to recover the full amplitude of the four-point process, one has to decompose the colour factors. The idea is to regroup the colour terms of the four-point amplitude in such a way that the colour term for the radiative process can be factorized out. This can be done by realizing that product of generator matrices can always be written in terms of (anti) commutators. One possible decomposition is given by the following

$$\mathcal{M}_5\left(1^-,2^+,3^+\right) = g_s^3 \sum_{\mathcal{P}_2(1,2)} \left(T_{a_1}T_{a_2}T_{a_3}\right)\left[\frac{\langle p1\rangle}{\langle p3\rangle\langle 31\rangle} + \frac{\langle 12\rangle}{\langle 13\rangle\langle 32\rangle} + \frac{\langle 2\bar{p}\rangle}{\langle 23\rangle\langle\bar{p}1\rangle}\right] \tag{1.1.31}$$
$$\times\tilde{\mathcal{M}}_4\left(1^-,2^+\right) + g_s^3 \sum_{\mathcal{P}_2(1,2)} \tilde{\mathcal{M}}_4\left(1^-,2^+\right)\left[\left[T_{a_3},T_{a_1}T_{a_2}\right]\frac{\langle p1\rangle}{\langle p3\rangle\langle 31\rangle}T_{a_1}\left[T_{a_3},T_{a_2}\right]\frac{\langle 12\rangle}{\langle 13\rangle\langle 32\rangle}\right]$$

Some comments are in order. First, we are here interested in the case where the radiative gluon is emitted from the quark or antiquark line (as depicted on the right-hand side of Fig. 1), therefore neglecting the possibility that the gluon is emitted from the outgoing gluon legs. In the MHV formalism, this corresponds to requiring the gauge fields of QCD to commute by setting the Casimir factor C_A to zero. Based on this observation, the second term in Eq. (1.1.31) can be neglected. Second, the colour structure of the radiative process has now been factorized out from the four-point amplitude. Indeed, the sum over the permutation $\mathcal{P}_2$ is independent of the soft gluon k_3 and therefore T_{a_3} can be extracted from the summation. Finally, the square brackets appearing in the first term can be simplified using the Schouten identity [85] $\langle ij\rangle\langle lk\rangle+\langle jk\rangle\langle il\rangle=\langle ik\rangle\langle lj\rangle$. As a result, in the soft limit, the expression of the five-point amplitude is given by the following

$$\mathcal{M}_5\left(1^-,2^+,3^+\right) = \left(g_s^2\sum_{\mathcal{P}_2(1,2)}\left(T_{a_1}T_{a_2}\right)\tilde{\mathcal{M}}_4\left(1^-,2^+\right)\right)\times g_s\frac{T_{a_3}\langle p\bar{p}\rangle}{\langle p3\rangle\langle 3\bar{p}\rangle}, \tag{1.1.32}$$

where one recognizes immediately the full amplitude for the four-point process with the helicity configuration $(1^-,2^+,3^+)$. The factorized terms is usually referred to as the (partial) *eikonal current* $\mathcal{J}_g$ for a single soft gluon emission. Therefore, writing Eq. (1.1.32) in a compact form yields

$$\mathcal{M}_5\left(1^-,2^+,3^+\right) = \mathcal{M}_4\left(1^-,2^+\right)\times\mathcal{J}_g(3^+) \quad \text{with} \quad \mathcal{J}_g(3^+) = g_s\frac{T_{a_3}\langle p\bar{p}\rangle}{\langle p3\rangle\langle 3\bar{p}\rangle}. \tag{1.1.33}$$

The expression of the eikonal current is interesting in two ways: first, it is invariant under the rescaling of $\langle p| \to a\langle p|$ and $|\bar{p}\rangle \to b|\bar{p}\rangle$; second it does not depend on the momentum of the two hard gluons. For the helicity configuration we are currently considering, we can now compute the modulus square of the five-point amplitude and sum over the colours. Multiplying Eq. (1.1.33) by its complex conjugate, summing over the colours, and using the cyclic property of the trace, we have

$$\left|\mathcal{M}_5\left(1^-,2^+,3^+\right)\right|^2 = g_s^2 C_F \frac{(p\bar{p})}{2(pk_3)(k_3\bar{p})} \times \left|\mathcal{M}_4\left(1^-,2^+\right)\right|^2 \tag{1.1.34}$$

We recall that taking the complex conjugate of the eikonal current just amounts to replacing the angle brackets to square brackets and use the definition $\langle ij\rangle\,[ij] = 2(k_i k_j)$ to recover the usual scalar products.

In order to get the full expression of the amplitude in the soft limit, we have to sum the amplitude in Eq. (1.1.34) over all possible helicity configurations. Let us recall that from the properties of the partial amplitudes, we have $\mathcal{M}_5\left(\{h_i\}\right) = \mathcal{M}_5^\star\left(\{-h_i\}\right)$ which obviously implies that the modulus squares are also equal. With this property, it is clear that one only has to compute in total three MHV amplitudes associated with the configurations $(1^-,2^+,3^+)$, $(1^+,2^-,3^+)$, and $(1^+,2^+,3^-)$. But even the first two configurations are related since flipping the helicity of the two gluons involved in the hard scattering is equivalent to simply swapping the gluons $(1 \leftrightarrow 2)$. In addition, one can notice immediately that in the limit $k_3 \to 0$, the amplitude arising from the third helicity configuration is subleading due to the presence of $\langle p3\rangle$ and $\langle 3\bar{p}\rangle$ in the numerator. Since the first helicity configuration has already been computed, the expression for $(1^+,2^-,3^+)$ can be read-off directly from Eq. (1.1.34) by swapping $1 \leftrightarrow 2$ which leaves the whole expression unchanged. Reading-off the $\overline{\text{MHV}}$ configurations $(1^+,2^-,3^-)$ and $(1^-,2^+,3^-)$ from the two MHV amplitudes and combining everything, we finally find

$$\sum_{\text{hel.}} \left|\mathcal{M}_5\left(1,2,3\right)\right|^2 = g_s^2 C_F \frac{2(p\bar{p})}{(pk_3)(k_3\bar{p})} \times \sum_{\text{hel.}} \left|\mathcal{M}_4\left(1,2\right)\right|^2 . \tag{1.1.35}$$

One should notice that in the soft limit, the Lorentz invariant phase space also factorizes. We can therefore deduce the expression of the gluon multiplicity distribution $d\omega_g^{(1)}$ that a single gluon is radiated from a quark-antiquark line. In terms of the transverse momentum k_T of the soft gluon, the distribution writes as

$$d\mathcal{W}_g^{(1)}(x,k_T^2) = \left(\frac{\alpha_s}{\pi}\right) C_F \frac{dk_T^2}{k_T^2} \frac{dx}{(1-x)}. \tag{1.1.36}$$

This distribution exhibits *soft singularity* when $x \to 1$ and $k_T \to 0$. As guarantied by the Kinoshita-Lee-Nauenberg (KNL) [99–101] theorem, such infrared divergences are cured by adding the soft limit of the virtual correction that cancels the singularities at the level of physical cross section. Indeed, at the level of the inclusive cross sections, the divergences appearing in the real contribution are cancelled exactly by those appearing in the virtual correction due to the fact that the integral over soft momenta are scaleless and such that the phase space integral and loop corrections vanish in pure dimensional regularization. However, this is not entirely the case for exclusive observables such as *transverse momentum distributions*. Whist the inclusion of the virtual corrections still keeps the cross section

Infrared (IR) finite, the sensitivity to some measurements yields leftover logarithms that can become large. Omitting the details of the calculations for the sake of brevity, the exact result after subtracting the infrared divergences is given by

$$\mathcal{W}_g^{(1)}(k_T^{\max},\mu) = -2C_F\left(\frac{\alpha_s}{\pi}\right)\ln^2\left(\frac{k_T^{\max}}{\mu}\right), \tag{1.1.37}$$

where μ is the scale induced by the dimensional regularization of the phase space integral and $k_T^{\max}$ is the maximum cut-off imposed on the k_T integral.

What about multiple radiative gluon emission? Within the framework of the MHV computations, it is straightforward to show that the eikonal current for n radiative gluon emission for a given helicity configuration can be decoupled from the four-point hard process. Analogous to the result of Eq. (1.1.33), the eikonal contribution in the case of multiple gluon emission is given by

$$\mathcal{J}_g^{(n)}\left(1^+,\cdots,n^+\right) = g_s^n\langle p\bar{p}\rangle\sum_{\mathcal{P}_n}\frac{(T_{a_1}\cdots T_{a_n})}{\langle p1\rangle\langle 12\rangle\cdots\langle n\bar{p}\rangle}. \tag{1.1.38}$$

Due to the non-commutativity of the QCD colour charges T_a, the eikonal current for multiple gluon emission cannot be factorized nicely as in the case for photons in Quantum Electrodynamics (QED). This reflects the fact that as opposed to photons, radiative gluons can be emitted from a hard gluon line. However, it can be shown that the most singular contribution to Eq. (1.1.38) in the soft limit is obtained when the emitted gluons are *strongly-ordered*. That is, the transverse momenta of the gluons satisfy $k_{1,T}^2 \ll k_{2,T}^2 \ll \cdots \ll k_{n,T}^2$ where the labels $1,\cdots,n$ on the transverse momenta represent the order of the emission. In this case, for instance, the first emitted gluon cannot resolve the remaining radiative emission from the parent (anti) quark line. Such an approximation implies that the following relations hold at the spinor levels when $k_{1,T}^2 \ll k_{2,T}^2$,

$$\frac{\langle pi\rangle}{\langle p1\rangle\langle 1i\rangle} \longrightarrow \frac{\langle p\bar{p}\rangle}{\langle p1\rangle\langle\bar{p}1\rangle} \qquad \text{and} \qquad \frac{\langle ij\rangle}{\langle i1\rangle\langle 1j\rangle} \longrightarrow 0. \tag{1.1.39}$$

This procedure can be performed iteratively by starting with the first emitted gluon. Once the first radiative gluon is factorized out, we can do the same for the second emitted gluon which in turn by virtue of the strong angular ordering cannot resolve the $(n-2)$ emission from the parent (anti) quark. Doing so up to the m-th gluon emission yields

$$\langle p\bar{p}\rangle\sum_{\mathcal{P}_n}\frac{(T_{a_1}\cdots T_{a_n})}{\langle p1\rangle\langle 12\rangle\cdots\langle n\bar{p}\rangle} = \frac{T_{a_1}\langle p\bar{p}\rangle}{\langle p1\rangle\langle 1\bar{p}\rangle}\cdots\frac{T_{a_m}\langle p\bar{p}\rangle}{\langle pm\rangle\langle m\bar{p}\rangle}\sum_{\mathcal{P}_{n-m}}\frac{\left(T_{a_{m+1}}\cdots T_{a_n}\right)\langle p\bar{p}\rangle}{\langle p(m+1)\rangle\cdots\langle n\bar{p}\rangle}. \tag{1.1.40}$$

When such a simplification is carried on until the n-th radiative gluon emission, one finds that the eikonal current for n gluon emission factorizes into a product of the individual eikonal current. Using Eq. (1.1.40), we can now compute the modulus square of the radiation current, average over the colours, and calculate the multiplicity distribution for the case of n radiative gluon emission

$$d\mathcal{W}_g^{(n)}\left(\{x_i,k_{i,T}^2\}\right) = \prod_{i=1}^{n} d\mathcal{W}_g^{(1)}\left(x_i,k_{i,T}^2\right) \tag{1.1.41}$$

with the single emission probability defined as in Eq. (1.1.36). Similar to the case for the single gluon emission as shown in Eq. (1.1.37), the exact result can be computed by subtracting the infrared divergences. Since in general, we do not know from which external leg the radiative gluon is emitted off, we have to sum over all external partons. Therefore, the part that contains the radiative contributions becomes

$$\mathcal{W}_g^{(n)}(k_T^{\max},\mu) = \sum_{n=0}^{\infty}\frac{1}{n!}\left[-2C_F\left(\frac{\alpha_s}{\pi}\right)\ln^2\left(\frac{k_T^{\max}}{\mu}\right)\right]^n \tag{1.1.42}$$

$$= \exp\left\{-2C_F\left(\frac{\alpha_s}{\pi}\right)\ln^2\left(\frac{k_T^{\max}}{\mu}\right)\right\}, \tag{1.1.43}$$

where the $1/n!$ comes from the fact that the gluons are indistinguishable. Eq. (1.1.43) shows that the soft contributions are resummed to all-order in the strong coupling in a space where the multiple emission kinematics factorizes w.r.t the LO hard process. Eq. (1.1.43) is evidently a first approximation as it only resums the leading-logarithmic (LL) power. Resummation of the soft contributions to higher logarithmic order is one of the main topics of the present thesis and will be amply detailed in the subsequent sections.

Infrared and Collinear safe observables

It was mentioned previously that whilst the KNL theorem ensures that higher-order corrections are well defined and finite for inclusive cross sections, this is not necessarily the case in practice. Therefore, in order for an observable to be computable (i.e. well defined and finite), one must define precisely the requirements that such an observable has to fulfil. Two requirements to properly define a physical observable has been proposed by Sterman and Weinberg [102]:

- *Infrared safety*: For an observable $\mathcal{O}_n$ evaluated on a set of n particles, if one of the particles is soft, then the observable $\mathcal{O}_n$ must equal the observable $\mathcal{O}_{n-1}$ obtained by neglecting the soft emission. Mathematically, this can be written as follows

$$\mathcal{O}_n(k_1,k_2,\cdots,k_n) \xrightarrow{k_1\to 0} \mathcal{O}_{m-1}(k_2,\cdots,k_n). \tag{1.1.44}$$

- *Collinear safety*: For an observable $\mathcal{O}_n$ evaluated on a set of n particles, if two of the particles are collinear to each other, then the observable $\mathcal{O}_n$ must equal the observable $\mathcal{O}_{n-1}$ obtained by replacing the two collinear momenta with a particle whose momentum equals the sum of the momenta of the two collinear particles. Mathematically, this can be written as follows

$$\mathcal{O}_n(k_1,k_2\cdots,k_n) \xrightarrow{k_1||k_2} \mathcal{O}_{m-1}(k_1+k_2,\cdots,k_n). \tag{1.1.45}$$

These two properties define what is known as *Infrared and Collinear Safety* (IRC) and guarantees that perturbative observables with radiative corrections are well defined and finite. Typically, two more ingredients are required in order to fully link the perturbative description of a given process and what is being observed experimentally. The first ingredient consists on replacing the final state plain partons in favour of *jets*. Conceptually, jets are defined as collimated bunches of particles composed of quarks and/or gluons

clustered according to a *jet algorithm*. Different jet algorithms might implement different definitions of how final state particles are grouped into jets. It should also be noted that the optimal jet definitions and methods may vary depending on the specific physics analysis tasks. For a review of the jet definitions and algorithms, refer to Refs. [103–105]. The second ingredient concerns the recombination of the partons or jets into hadrons, or *hadronization*. As opposed to the situation in QED where soft radiations can be resolved up to the detector resolution, in QCD, one never observes free partons due to the *confinement*. Indeed, once the shower of partons is terminated, the process enters a low-momentum transfer and long-distance regime in which non-perturbative effects can no longer be neglected. Various models such as the independent fragmentation model [106], string model [107–110], and cluster model [111, 112] have been proposed to simulate the transition of the partons into hadrons. These models have been implemented in various *Parton Shower Monte Carlo Event Generators* such as HERWIG [113], PYTHIA [114, 115], and SHERPA [116].

1.1.2 Solving the DGLAP equations

In the previous section, it was shown that long-distance effects are separated from short-distance scattering by the factorization scale μ (commonly denoted μ_F). Such a factorization allows for the μ-dependent hard function (partonic cross section) to be computable perturbatively. In the same way that physical observables should not depend on the renormalization scale μ_R (introduced through the renormalization procedure), they should not also depend on the factorization scale, i.e. $\partial\mathcal{O}/\partial\mu^2=0$. This implies that the parton distribution function has to satisfy some *Renormalization Group Equations* (RGEs) known as DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations.

In order to define the DGLAP equations in a more compact form while being explicit, let us define the vectors $\mathbf{q}(x,\mu^2)$ and $\bar{\mathbf{q}}(x,\mu^2)$ whose components are the different flavour of quark and antiquark distributions. That is, the k-th elements are respectively given by the following expressions

$$\Big[\mathbf{q}(x,\mu^2)\Big]_a = f_a(x,\mu^2), \qquad \Big[\bar{\mathbf{q}}(x,\mu^2)\Big]_a = f_{\bar{a}}(x,\mu^2). \tag{1.1.46}$$

Adopting the same notation for the gluon distribution, the all-order DGLAP evolution equations can be written as

$$\mu^2\frac{\partial}{\partial\mu^2}\begin{bmatrix}\mathbf{q}(x,\mu^2)\\ \bar{\mathbf{q}}(x,\mu^2)\\ g(x,\mu^2)\end{bmatrix} = \int_x^1\frac{dz}{z}\begin{bmatrix}\mathbf{P_{qq}}(x/z) & \mathbf{P_{q\bar{q}}}(x/z) & \mathbf{P_{qg}}(x/z)\\ \mathbf{P_{\bar{q}q}}(x/z) & \mathbf{P_{\bar{q}\bar{q}}}(x/z) & \mathbf{P_{\bar{q}g}}(x/z)\\ \mathbf{P_{gq}}(x/z) & \mathbf{P_{g\bar{q}}}(x/z) & \mathbf{P_{gg}}(x/z)\end{bmatrix}\begin{bmatrix}\mathbf{q}(z,\mu^2)\\ \bar{\mathbf{q}}(z,\mu^2)\\ g(z,\mu^2)\end{bmatrix} \tag{1.1.47}$$

where we for brevity we have omitted the dependence on the strong coupling $\alpha_s(\mu^2)$. The functions $\mathbf{P}$ are defined in terms of the splitting functions introduced in Sec. (1.1.1) and can be computed perturbatively $\mathbf{P}(\alpha_s(\mu^2),x/z)=\sum_{n=0}^{\infty}\alpha_p^{n+1}(\mu^2)\mathbf{P}^{(n)}(x/z)$. The functions given by $\mathbf{P_{qq}},\mathbf{P_{q\bar{q}}},\mathbf{P_{\bar{q}q}}$, and $\mathbf{P_{\bar{q}\bar{q}}}$ are n_f-dimensional square matrices; $\mathbf{P_{qg}}$ and $\mathbf{P_{\bar{q}g}}$ are column vectors with dimension n_f; $\mathbf{P_{gq}}$ and $\mathbf{P_{g\bar{q}}}$ are row vectors with dimension n_f; and finally $\mathbf{P_{gg}}$ is a one-dimensional matrix. Therefore, the big square matrix on the right-hand side of the equations has dimension $(2n_f+1)$. At leading-logarithmic accuracy,

each element of these matrices corresponds to the splitting function whose expression we derived in Sec. (1.1.1)

$$\left[\mathbf{P_{q\bar{q}}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{ab} = \left[\mathbf{P_{\bar{q}q}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{ab} = 0 \tag{1.1.48}$$

$$\left[\mathbf{P_{qg}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{a} = \left[\mathbf{P_{\bar{q}g}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{a} = \alpha_s P^{(0)}_{qg}\left(\frac{x}{z}\right) \tag{1.1.49}$$

$$\left[\mathbf{P_{gq}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{a} = \left[\mathbf{P_{g\bar{q}}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{a} = \alpha_s P^{(0)}_{gq}\left(\frac{x}{z}\right) \tag{1.1.50}$$

$$\left[\mathbf{P_{qq}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{ab} = \left[\mathbf{P_{\bar{q}\bar{q}}}\left(\alpha_s(\mu^2),\frac{x}{z}\right)\right]_{ab} = \alpha_s P^{(0)}_{qq}\left(\frac{x}{z}\right)\delta_{ab} \tag{1.1.51}$$

and the gluon distribution $\mathbf{P_{gg}} = \alpha_s P^{(0)}_{gg}$. The DGLAP equations in Eq. (1.1.47) is a system of coupled $(2nf+1)$ integro-differential equations, and therefore in order to solve such a differential equations, it is advantageous to decouple them. It turns out that one could find a combination of the parton distribution functions that can decouple the system of equations. This combination is given by the $(2n_f-1)$ independent combination of PDF matrices with dimension one known as *non-singlet*, and a single rank 2 matrix that couples a particular combination of quarks known as *singlet*.

Let us start with construction of the non-singlet sector. Using charge conjugation and flavour asymmetry constraints, the DGLAP equations related to the non-singlet decomposition are given by

$$\mu^2\frac{\partial}{\partial\mu^2}q^V_{\rm ns}(x,\mu^2) = \int_x^1\frac{dz}{z}P^V_{\rm ns}\left(\frac{x}{z}\right)q^V_{\rm ns}(z,\mu^2) \tag{1.1.52}$$

$$\mu^2\frac{\partial}{\partial\mu^2}q^{\pm}_{{\rm ns},a}(x,\mu^2) = \int_x^1\frac{dz}{z}P^{\pm}_{{\rm ns},a}\left(\frac{x}{z}\right)q^{\pm}_{\rm ns}(z,\mu^2) \tag{1.1.53}$$

where the non-singlet combination of the various (anti) quarks are given by the following linearly independent functions:

$$q^V_{\rm ns}(z,\mu^2) = \sum\nolimits_q\left(q(z,\mu^2)-\bar{q}(z,\mu^2)\right) \tag{1.1.54}$$

$$q^{\pm}_{{\rm ns},a}(z,\mu^2) = -a\left(\eta(a)\pm\bar{\eta}(a)\right)+\sum\nolimits_{k=1}^{a}\left(\eta(k)\pm\bar{\eta}(k)\right) \tag{1.1.55}$$

with $a=\{2,\cdots,n_f\}$ and η mapping a flavour integer representation $\mathcal{F}_I=\{1,\cdots,n_f\}$ into the set of quark flavours $\mathcal{F}_Q=\{u,\cdots,t\}$. The splitting function governing the evolution equation as expressed in Eq. (1.1.47) are derived using the general structure of the splitting functions and they are given by

$$P^{\pm}_{{\rm ns},a}(x) = P_{qq}(x)\pm P_{q\bar{q}}(x) \tag{1.1.56}$$

$$P^V_{\rm ns}(x) = P^-_{\rm ns}(x)+P^{\rm sea}_{\rm ns}(x) \tag{1.1.57}$$

where the $P^{\rm sea}_{\rm ns}=n_f(P_{qQ}-P_{q\bar{Q}})$ represents the *sea-combination* with Q denoting a quark with different flavour. It is worth emphasizing that only the flavour diagonal quantity P_{qq} contributing in the valence starts at $\mathcal{O}(\alpha_s)$. The function $P_{q\bar{q}}$ along with the flavour independent sea contributions start to contribute at $\mathcal{O}(\alpha_s^2)$.

The non-singlet sector described above entirely decoupled from the flavour singlet gluon density. One therefore needs to define a flavour singlet quark density that maximally

couples to the gluon. Such a construction leads to the DGLAP equations for the flavour singlet densities in which the splitting functions that govern the evolution equation is expressed in terms of a rank-2 matrix

$$\mu^2\frac{\partial}{\partial\mu^2}\begin{bmatrix}\Sigma(x,\mu^2)\\ g(x,\mu^2)\end{bmatrix}=\int_x^1\frac{dz}{z}\begin{bmatrix}P_{qq}(x/z) & P_{qg}(x/z)\\ P_{gq}(x/z) & P_{gg}(x/z)\end{bmatrix}\begin{bmatrix}\Sigma(z,\mu^2)\\ g(z,\mu^2)\end{bmatrix} \tag{1.1.58}$$

with the flavour singlet quark density defined as follows

$$\Sigma(x,\mu^2)\equiv\sum_{i=1}^{n_f}\Big(q_i(x,\mu^2)+\bar{q}_i(x,\mu^2)\Big)\,. \tag{1.1.59}$$

It is worth mentioning that processes such as DIS and Higgs boson production are only sensitive to the singlet distributions and therefore the various non-singlet distributions are not involved.

Due to the presence of convolutions in the DGLAP equations, it is convenient to solve the non-singlet and singlet evolution equations in Mellin space where the convolutions become normal products. We see that whilst the solution to the non-singlet equation possesses a closed exponential form at all-order, the singlet case beyond leading-order does not. As a matter of fact, the diagonalization of the singlet matrix in order to have two independent equations for the two linear combinations of gluon and singlet-quark Σ only works at leading-order. Therefore, for the singlet section, one has to compute higher-order solutions as a series expansion around the lowest order. In Mellin space, the evolution equations for the non-singlet and singlet sectors can be generically written in the following way

$$\mu^2\frac{\partial}{\partial\mu^2}\mathcal{Q}(N,\mu^2)=\Gamma(N)\mathcal{Q}(N,\mu^2) \tag{1.1.60}$$

where $\mathcal{Q}(N)$ denotes the Mellin version of one of the non-singlet $q^V_{\rm ns}$, $q^\pm_{{\rm ns},a}$, singlet Σ, and gluon g distributions while the function $\Gamma(N)$ represents the associated Mellin transform of one of the various splitting functions. Again, for the sake of brevity, we have omitted the dependence in the strong coupling. Recall that both Γ and $\mathcal{Q}$ are scalar quantities for the non-singlet case and a rank-2 matrix for the singlet. The differential equation in μ of Eq. (1.1.60) can be promoted into an evolution equation in α_s. This is advantageous since one can write Eq. (1.1.60) unambiguously as a series in α_s using the running of the coupling given in App. 1.A. Using the definition of the QCD β-function to take into account of the change of variables, Eq. (1.1.60) becomes

$$\frac{\partial}{\partial\alpha_s}\mathcal{Q}(N,\alpha_s)=\frac{\Gamma(N,\alpha_s)}{\beta(\alpha_s)}\mathcal{Q}(N,\alpha_s), \tag{1.1.61}$$

where for simplicity we have omitted the explicit dependence of strong coupling on the scale μ^2. Unless explicitly mentioned, throughout this section, we always consider $\alpha_s{\equiv}\alpha_s(\mu^2)$. Expanding both the β-function and the anomalous dimensions, we get

$$\frac{\partial}{\partial\alpha_s}\mathcal{Q}(N,\alpha_s)=-\frac{1}{\alpha_s\beta_0}\left[1+\sum_{n=1}^{\infty}\alpha_s^n\frac{\beta_n}{\beta_0}\right]^{-1}\left[\sum_{n=0}^{\infty}\alpha_s^n\Gamma^{(n)}(N)\right]Q(N,\alpha_s) \tag{1.1.62}$$

$$=-\frac{1}{\alpha_s}\left[\mathcal{R}_0(N)+\sum_{n=1}^{\infty}\mathcal{R}_n(N)\right]Q(N,\alpha_s) \tag{1.1.63}$$

where we have defined the recursive function $\mathcal{R}_n(N)$ as [94,117]

$$\mathcal{R}_0(N) = \frac{1}{\beta_0}\Gamma^{(0)}(N), \tag{1.1.64}$$

$$\text{and} \qquad \mathcal{R}_n(N) = \frac{1}{\beta_0}\Gamma^{(n)}(N) - \sum_{k=1}^{n}\frac{\beta_k}{\beta_0}\mathcal{R}_{n-k}(N) \quad (\text{if } n > 1) \tag{1.1.65}$$

The anomalous dimension quantities $\mathcal{R}_n$ are diagonalizable if and only if they commute, which is not the case for the singlet sector. On the contrary, since the functions $\Gamma(N)$ are just scalar quantities in the case of non-singlet, the anomalous dimension quantities $\mathcal{R}_n$ do commute and closed exponential form of the solution can be constructed. The leading-order solution to Eq. (1.1.63) for both non-singlet and singlet is given by

$$U_{\text{NS,S}}^{(\text{LO})}\left(N, \mu^2 \leftarrow \mu_0^2\right) \equiv \frac{Q\left(N, \alpha_s(\mu^2)\right)}{Q\left(N, \alpha_s(\mu_0^2)\right)} = \sum_{n=0}^{\infty}\frac{1}{n!}\left[-\mathcal{R}_0(N)\ln\left(\frac{\alpha_s(\mu^2)}{\alpha_s(\mu_0^2)}\right)\right]^n \tag{1.1.66}$$

where $U_{\text{NS,S}}^{(\text{LO})}$ is called an evolution function that evolves Q from scale μ_0^2 to μ^2. In the case where the function Q represents one of the non-singlet $q_{\text{ns}}^V, q_{\text{ns},a}^{\pm}$ distributions, higher-order solutions can be exactly computed from Eq. (1.1.63). At next-to-leading order, the non-singlet solution has a closed exponential form

$$U_{\text{NS}}^{(\text{NLO})}\left(N, \mu^2 \leftarrow \mu_0^2\right) = \exp\left\{\frac{\mathcal{R}_1(N)}{\left(\alpha_s(\mu_0^2) - \alpha_s(\mu^2)\right)^{-1}}\right\} U_{\text{NS}}^{(\text{LO})}\left(N, \mu^2 \leftarrow \mu_0^2\right). \tag{1.1.67}$$

Using the corresponding order of the β-function and anomalous dimensions, higher-order solutions to the non-singlet evolution equation can be exactly determined and expressed in a closed exponential form. The term in the denominator of the exponent of Eq. (1.1.67) can be written in terms of $\ln\left(\mu^2/\mu_0^2\right)$. In this sense, the evolution function can be reorganized in terms of the logarithmic contributions. That is, the all-order evolution function can be written as $U_{\text{NS}}(N) = \sum_{n=1}^{\infty}\alpha_s^{n-2}g_n(N)$ where g_1/α_s contains logarithms of the form $\alpha_s^n\ln^{n+1}\left(\mu^2/\mu_0^2\right)$, g_2 contains logarithms of the form $\alpha_s^n\ln^n\left(\mu^2/\mu_0^2\right)$, and so forth. Using jargon from resummation (see subsequent sections for details), the first term resums the leading logarithmic contributions (LL), the second resums the next-to-leading logarithmic contributions (NLL), et cetera.

Before moving to the higher-order solutions of the singlet evolution equation, let us first rewrite its leading order solution (Eq. (1.1.66)) in a more intuitive form by diagonalizing the leading-order singlet anomalous dimension $\Gamma^{(0)}(N)$. In light of doing so, we decompose the zeroth order anomalous dimension quantity $\mathcal{R}_0$ as follows

$$\mathcal{R}_0(N) = \lambda_-^{(0)}(N)\mathcal{E}_0^-(N) + \lambda_+^{(0)}(N)\mathcal{E}_0^+(N) \tag{1.1.68}$$

where $\lambda_\pm$ represent the (lower and higher) eigenvalues of $\mathcal{R}_0$ ($\Gamma^{(0)}$ to be precise) and $\mathcal{E}^\pm$ the corresponding projectors. Recall that the projectors $\mathcal{E}^\pm$, similar to the anomalous

dimension quantities $\mathcal{R}_n$, are rank-2 matrices while the eigenvalues $\lambda_\pm$ are scalar quantities. Their expressions are respectively given by

$$\lambda_\pm^{(0)} = \frac{1}{2}\left[\gamma_{qq}^{(0)} + \gamma_{gg}^{(0)} \pm \sqrt{\left(\gamma_{qq}^{(0)} - \gamma_{gg}^{(0)}\right)^2 + 4\gamma_{gq}^{(0)}\gamma_{qg}^{(0)}}\right] \tag{1.1.69}$$

$$\mathcal{E}_0^\pm = \frac{1}{\lambda_\pm^{(0)} - \lambda_\mp^{(0)}}\left[\beta_0\mathcal{R}_0 - \lambda_\mp^{(0)} \times \mathbb{I}\right], \tag{1.1.70}$$

where we have omitted the N dependence for brevity. By making use of the eigenvalues and projectors, the leading order evolution function for the singlet sector given in Eq. (1.1.66) can now be expressed as

$$U_{\rm S}^{(\rm LO)}\left(N, \mu^2 \leftarrow \mu_0^2\right) = \sum_{r=\pm} \mathcal{E}_0^r(N) \exp\left\{-\frac{\lambda_r^{(0)}(N)}{\beta_0} \ln\left(\frac{\alpha_s(\mu^2)}{\alpha_s(\mu_0^2)}\right)\right\}. \tag{1.1.71}$$

As repeatedly mentioned, the solutions to the singlet evolution equation cannot be written in a closed exponential form and have to be computed as a series expansion around the leading-order solution in Eq. (1.1.71). We therefore use the following ansatz [94, 117]

$$\begin{aligned} U_{\rm S}\left(N, \mu^2 \leftarrow \mu_0^2\right) &= V_{\rm S}\left(N, \alpha_s(\mu^2)\right) U_{\rm S}^{(\rm LO)}\left(N, \mu^2 \leftarrow \mu_0^2\right) V_{\rm S}^{-1}\left(N, \alpha_s(\mu_0^2)\right) \\ &= \left[1 + \sum_{n=1}^{\infty} \alpha_s^n(\mu^2) V_{\rm S}^{(n)}(N)\right] U_{\rm S}^{(\rm LO)}(N) \left[1 + \sum_{n=1}^{\infty} \alpha_s^n(\mu_0^2) V_{\rm S}^{(n)}(N)\right]^{-1} \end{aligned} \tag{1.1.72}$$

where the evolution matrices $V_{\rm S}^{(n)}$ are constructed from combinations of the anomalous dimension quantities $\mathcal{R}_n$. The inverse evolution matrix factor ensures that the evolution $U_{\rm S}$ reduces to unity when $\mu = \mu_0$. By plugging Eq. (1.1.72) into Eq. (1.1.63), one can recursively construct commutation relations between $V_{\rm S}^{(n)}$ and $\mathcal{R}_0$. Sorting in powers of α_s, we have

$$\sum_{n=1}^{\infty} \alpha_s^n \left[V_{\rm S}^{(n)}, \mathcal{R}_0\right] = \sum_{n=1}^{\infty} \alpha_s^n \left(n V_{\rm S}^{(n)} + \mathcal{R}_n\right) + \sum_{n,m=1}^{\infty} \alpha_s^m \alpha_s^n V_{\rm S}^{(m)} \mathcal{R}_n, \tag{1.1.73}$$

which order-by-order gives the following commutation relations

$$\left[V_{\rm S}^{(n)}, \mathcal{R}_0\right] = n V_{\rm S}^{(n)} + \mathcal{R}_n \sum_{m=1}^{n-1} V_{\rm S}^{(n-m)} \mathcal{R}_n \equiv n V_{\rm S}^{(n)} + \tilde{\mathcal{R}}_n. \tag{1.1.74}$$

These equations can be solved recursively using the eigenvalue decomposition of the leading-order singlet anomalous dimension matrix. Details of such computations are given in Refs. [118, 119], here we directly give the solutions

$$V_{\rm S}^{(n)} = \frac{\beta_0 \mathcal{E}_1^+ \tilde{\mathcal{R}}_n \mathcal{E}_0^-}{\lambda_-^{(0)} - \lambda_+^{(0)} - n\beta_0} + \frac{\beta_0 \mathcal{E}_0^- \tilde{\mathcal{R}}_n \mathcal{E}_0^+}{\lambda_+^{(0)} - \lambda_-^{(0)} - n\beta_0} - \frac{1}{n}\left(\mathcal{E}_0^+ \tilde{\mathcal{R}}_n \mathcal{E}_0^- + \mathcal{E}_0^- \tilde{\mathcal{R}}_n \mathcal{E}_0^+\right). \tag{1.1.75}$$

Notice that the expression of $V_{\rm S}^{(n)}(N)$ contains poles when the denominators $\lambda_-^{(0)} - \lambda_+^{(0)} \pm n\beta_0$ vanish. These poles, however, are cancelled by the inverse evolution matrix $V_{\rm S}^{-1}(N)$

in Eq. (1.1.72). This statement also holds for the truncated solutions where the inverse evolution matrix V_S^{-1} is not determined as an inverse of the evolution matrix V_S but rather from a series expansion. Plugging everything back into Eq. (1.1.63) using the form of the leading-order solution given in Eq. (1.1.76), the next-to-leading (NLO) order solution to the singlet evolution equation is given by

$$U_S^{(\mathrm{NLO})}\left(N,\mu^2\leftarrow\mu_0^2\right)=\sum_{r=\pm}\exp\left\{-\frac{\lambda_r^{(0)}(N)}{\beta_0}\ln\left(\frac{\alpha_s(\mu^2)}{\alpha_s(\mu_0^2)}\right)\right\}\mathcal{M}(r,-r) \tag{1.1.76}$$

where we have defined the matrix $\mathcal{M}$ as follows

$$\begin{aligned}\mathcal{M}(r,p)=\mathcal{E}_0^{(r)}&+\frac{\mathcal{E}_0^{(r)}\mathcal{R}_1\mathcal{E}_0^{(p)}}{\left(\alpha_s(\mu_0^2)-\alpha_s(\mu^2)\right)^{-1}}-\alpha_s(\mu_0^2)\frac{\beta_0\mathcal{E}_0^{(r)}\mathcal{R}_1\mathcal{E}_0^{(p)}}{\lambda_p^{(0)}-\lambda_r^{(0)}-\beta_0}\\&+\alpha_s(\mu^2)\frac{\beta_0\mathcal{E}_0^{(r)}\mathcal{R}_1\mathcal{E}_0^{(p)}}{\lambda_p^{(0)}-\lambda_r^{(0)}-\beta_0}\exp\left\{\frac{\lambda_r^{(0)}-\lambda_p^{(0)}}{\beta_0}\ln\left(\frac{\alpha_s(\mu^2)}{\alpha_s(\mu_0^2)}\right)\right\}.\end{aligned} \tag{1.1.77}$$

In the above equation, the poles $\lambda_\pm^{(0)}-\lambda_\mp^{(0)}=\beta_0$ are cancelled when the last two terms are combined. The analytic expressions of the truncated solutions to the singlet evolution equation up to three-loop can be found in Refs. [117, 118]. These expressions will be important when constructing the improved transverse momentum resummation for colour singlet observables, that will be discussed in the subsequent sections.

Before closing this section, let us emphasize the role that the DGLAP evolution equations play in QCD, but mainly in the determination of the PDFs. Parton distribution Functions have an intrinsic non-perturbative feature that manifest through the scale μ^2 and the fraction of momentum x. While the x-dependence must be derived through a fit to the experimental data, the μ^2-dependence is predicted at all values by the DGLAP evolution equations. This is very practical as it allows the PDFs to be fitted at an initial scale (usually chosen to be slightly below the *charm* quark mass if heavier quark PDFs are assumed to be generated by QCD radiation, i.e. $\mu_0\equiv Q_0\sim 1$ GeV, or slightly above if the *charm* PDFs are fitted on the same footing as light quarks, i.e. $\mu_0\equiv Q_0\sim 1.65$ GeV) and then afterward evolved to higher scales.

1.1.3 Determination of the proton PDFs

The non-perturbative nature of the parton distribution functions prevent their computation from first principle, instead, they have to be determined by comparing theoretical predictions of hadronic cross sections (as expressed by the key equation, Eq. (1.1.1)) with experimental measurements. That is, PDFs are modelled in terms of a collection of sensible parameters whose values are optimized w.r.t. a suitable goodness-of-fit measure through some minimization procedure in order to match experimental measurements.

To the present day, there exists various groups that implement different methodology and provide different estimation of the PDF errors. To mention just a few collaborations: ZEUS/H1 [34–39], NNPDF [40–53], ABM [54–58], CTEQ [59–68], and MSHT [69–74]. As the machine learning part in Sec. (5.5) is purely based on PDF sets produced using NNPDF methodology, we base the subsequent descriptions of how PDFs are parametrized

in terms of the NNPDF framework. While we refer the interested reader to the aforementioned references for a detailed review of the methodology, we briefly highlight in the following passage the main key ingredients. The determination of (unpolarized) proton PDFs relies primarily on the following elements:

- **Perturbative Calculations**

  The factorization formula given in Eq. (1.1.1) provide a framework in which perturbative and non-perturbative regimes are separated allowing for the short-distance (hard) interactions to be computed perturbatively. In order for the perturbative series (partonic cross section) to converge to its asymptotic result, computations of higher-order contributions are vital. Beyond leading-order, the expression of the partonic cross sections include loop momenta which when integrated over leads to *Ultraviolet* (UV) divergences. Such divergences are *renormalized* through the running of strong coupling with the introduction of the *renormalization scale* μ_R. Based on the renormalization group argument, the dependence of the partonic cross section on the renormalization scale decreases as the perturbative calculations are carried out to higher orders.

  Apart from the UV divergences, two classes of singularities also arise at higher orders: infrared divergences associated with virtual contributions which (as described in the previous section) cancel the corresponding soft divergences from the emission of real gluons; and collinear divergences which are subtracted by terms arising in the renormalization of the PDF operators. Depending on the observables, additional singularities might arise due to the appearance of large logarithmic terms that compensate the smallness of the strong coupling and spoil the convergence of the perturbative calculations. Such divergences have to resummed to all-order in the strong coupling.

  Higher order partonic cross sections, usually next-to-leading order (NLO) and next-to-next-to-leading order (NNLO), are generally available via Monte Carlo programs such as MADGRAPH [120], NNLOJET [121], and MCFM [122] that typically takes countless hours to accurately compute cross sections. In addition, since during the fitting procedure PDFs are constantly evolved and convoluted with the partonic cross sections, the parts which perform the DGLAP evolution and convolutions must be carried out quickly and accurately. Several proposals have been put forward in order to achieve this. The underlying idea behind these approaches consists on pre-computing the partonic cross sections convoluted with the evolution kernels and store the results into a suitable interpolation grid. The computation of the full hadronic cross sections thus reduces to the scalar product between the PDFs parameterized at the initial scale Q_0^2 and the interpolation tables. Each of the previously mentioned Monte Carlo Program has its own fast interpolating software: APPLFAST [123] as the MADGRAPH interface, FASTNLO [124, 125] for NNLOJET, and APPLGRID [126] interfaces both NNLOJET and MCFM. In addition, the NNPDF collaboration uses the FASTKERNEL methodology [42, 43] that permits a fast precomputation of evolved parton densities.

- **PDF Parametrization**

As highlighted in the previous section, thanks to the DGLAP evolution equations, proton PDFs can be parameterized at an initial scale Q_0^2 and then evolved to the energy of the experimental predictions. In the NNPDF methodology, each parton f_a is parameterized in terms of a neural network at an initial scale Q_0 as

$$xf_a(x, Q_0^2, \theta) = \mathcal{N}_a x^\alpha (1-x)^\beta \mathrm{NN}_a(x, \theta) \tag{1.1.78}$$

where $\mathcal{N}_a$ is a normalization constant that accounts for theoretical constraints enforced by the QCD sum rules. In the NNPDF jargon, the factors x^α and $(1-x)^\beta$ are called *preprocessing* whose purpose is to speed up the convergence. They respectively describe the small-x and large-x behaviours of the PDFs, in particular, the second factor ensures that the PDFs vanish at large-x. The exponents α and β are different for each PDF flavour and are determined from the data in global QCD analyses. The main core of the parametrization lies in the neural network function $\mathrm{NN}_a(x, \theta)$–with θ representing the set of neural network parameters–that is defined to interpolate between the low-x and high-x regions. For non-linear regression problems such as the extraction of PDFs from experimental data, the use of neural networks is suitable as it offers more flexibility and more importantly reduces the bias associated with the choice of a functional form. Due to the sensitivity of the data set to different species of PDFs, the extractions of the PDFs are typically performed in the *evolution basis* (constructed from a linear combination of independent PDFs, or the singlet and non-singlet distributions) instead of the *flavour basis*. However, the choice of parametrization basis is arbitrary and the results of the fit should be basis independent. Such an independence of the parametrization basis has been demonstrated in the NNPDF4.0 global analysis.

- **Optimization**

Once the PDFs have been computed at the initial scale Q_0, they can be evolved to the scale of the experimental data set using the DGLAP evolution equations presented in the previous section. The resulting PDFs can then be convoluted with the partonic cross section (computed at a given perturbative order) in order to obtain theoretical predictions for the hadronic cross section. Theoretical predictions are then compared to experimental predictions through a log-likelihood χ^2 measure. The optimal PDF parameters are thus obtained by minimizing the χ^2-figure of merit. For a given N_{dat} of experimental measurements D_i and a corresponding theoretical predictions T_i, the loss function to be minimized is defined as

$$\chi^2(\theta) \equiv \sum_{i,j}^{N_{\mathrm{dat}}} \left(D_i - T_i(\theta)\right) \mathrm{cov}_{ij}^{-1} \left(D_j - T_j(\theta)\right), \tag{1.1.79}$$

where i, j run over the experimental data points and cov_{ij} represents the covariance matrix. A typical fit typically only includes experimental information in the covariance matrix. However, recently, it has become possible to also include into the covariance matrix theoretical information due to the uncertainties related to missing

higher order in perturbation theory [52,53]. The experimental covariance matrix is defined as

$$\text{cov}_{ij} = \delta_{ij} \left(\delta D_i^{\text{UNC}}\right)^2 + \sum_{k=1}^{N_{\text{COR}}} \delta D_{k,i}^{\text{COR}} \delta D_{k,j}^{\text{COR}} \tag{1.1.80}$$

with $\delta D_{ij}^{\text{UNC}}$ and $\delta D_{k,i}^{\text{COR}}$ respectively denoting the *uncorrelated* and *correlated* uncertainties. It is worth pointing that in the absence of correlated uncertainties, the covariance matrix has non-zero entries only in the diagonal.

The outcome of a PDF fit largely depends on the optimization strategy used to minimize the figure of merit expressed in Eq. (1.1.79). In previous NNPDF determinations (NNPDF3.1 and priors), *stochastic algorithms* have been used for the training of neural networks. While these algorithms are less prone to end up in a local minima, they become unsuitable when the parameter space is large, which is the case for PDF extractions due to the large covariance matrices. The NNPDF4.0 methodology [4,77] instead considers (deterministic) *Stochastic Gradient Descent* algorithms which not only is more computationally efficient but most importantly provides faster and stable convergence. Indeed, due to frequent updates, the steps taken by stochastic gradient descent algorithms towards the minima of the loss function contain oscillations that can help to get out of local minima.

With fast and efficient minimization strategies also comes the risk of over-training. This is also often referred to as *overfitting*. In order to avoid fitting noises in the data, it is important to conceive a suitable stopping criterion. The NNPDF methodology implements a widely used method in literature known as *cross-validation* in which the data points are randomly divided into a *training* and *validation* sets. While the χ^2 is computed for both sets, the training optimization is only performed on the training set. The stopping algorithms keeps track of the values of both χ^2 and terminates when the value of the validation χ^2_{val} does not improve for a certain number of steps (while the training χ^2_{tr} continues to decrease).

Let us finally comment on the propagation of PDF uncertainty. State-of-the-art PDF determinations (including NNPDF) only includes experimental uncertainties on the data used to extract the PDFs. Theoretical uncertainties associated to missing higher order corrections are usually neglected as they are assumed to be negligible compared to experimental and possibly methodological uncertainties. Although proof of concepts on how theoretical uncertainties can be systematically included in the PDF determination already exist, PDF sets are delivered accounting only experimental (and possibly methodological) uncertainties.

There exists two commonly used approaches to propagate experimental uncertainties into the PDFs: the Monte Carlo and Hessian methods. For the determination of unpolarized PDFs, the latter has been used in CT18 [67] and MSHT20 [74] while the former has been used by the NNPDF and JAM [127] collaborations. While both of these methods have their own advantages, the Monte Carlo method is more robust. As a matter of fact, the Monte Carlo approach outweighs the traditional Hessian approach in two main ways. First, the Monte Carlo method does not require the selection of a functional form removing the bias associated to the PDF parametrization. Second, the Monte Carlo approach does

not assume that the underlying PDF uncertainties follow a Gaussian distribution and thus it does not require the linearity of the observables in order to propagate the uncertainties to the PDFs. The Monte Carlo technique, instead, consists on generating *Monte Carlo replicas* according to a Multi-Gaussian distribution centered on the central values D_i and with variances based on the experimental covariance matrices. The distribution of the data replicas is thus mapped into the confidence interval in the space of PDFs by performing a fit to each of the replicas. A fit to a replicas consists in minimizing the χ^2_R where R denotes a given replica index. In order to reproduce the statistical properties of the original data, the number of replicas needs to be sufficiently large. Once fit results for the whole set of Monte Carlo replicas are available, post-fit checks are required in order to ensure that each replica satisfies a set of criteria. For instance, replicas with too large χ^2 values or that do not pass positivity constraints [128] must be removed.

1.2 Evaluating MHOUs with scale variations

A reliable estimation of the theoretical uncertainties is of foremost importance for high precision physics of the kind aimed for at the LHC. The main source of theoretical uncertainties stem from the missing higher-order corrections in perturbative computations. The variation of the unphysical scales appearing in the calculation is thus far the most used and widespread method to estimate such uncertainties. In the following, we first describe how missing higher-order uncertainties coming from the hard scattering and PDFs are evaluated independently using scale variations. Then, we discuss how these uncertainties can be simultaneously assessed at the level of physical observables, still using the method of scale variations. Our derivations follow closely Refs. [52, 53]

For the sole reason that the present thesis is mainly concerned about hadroproduction processes, let us consider as illustration the factorized expression given in Eq. (1.1.1)

$$\mathcal{O}\left(\alpha_s(\mu^2), \frac{Q^2}{\mu^2}\right) = \mathcal{L} \otimes \hat{\mathcal{O}}\left(\alpha_s(\mu^2), \frac{Q^2}{\mu^2}\right) \tag{1.2.81}$$

where for the time being, we assume that the PDFs (expressed through the luminosity $\mathcal{L} = f \otimes f$) are evaluated at ta fixed scale. Thus, the scale μ^2 in Eq. (1.2.81) represents an arbitrary normalization scale. For the sake of brevity, the sum over the partonic (sub) channels have been omitted and assumed to be implicit.

1.2.1 Scale variation for partonic cross sections

As introduced in the previous section, the *renormalization group* (RG) invariance states that all-order predictions are independent of the renormalization scale. That is,

$$\mu^2 \frac{\partial}{\partial \mu^2} \mathcal{O}\left(\alpha_s(\mu^2), \frac{Q^2}{\mu^2}\right) = \mathcal{L} \otimes \mu^2 \frac{\partial}{\partial \mu^2} \hat{\mathcal{O}}\left(\alpha_s(\mu^2), \frac{Q^2}{\mu^2}\right) = 0 \tag{1.2.82}$$

Eq. (1.2.82) shows that the renormalization group invariance of the physical observable (henceforth cross section) requires the renormalization group invariance of the perturbative observable. Thus, hereinafter, let us mainly focus on the hard scattering partonic part $\mu^2 \partial \hat{\mathcal{O}} / \partial \mu^2 = 0$. For the purpose of what will follow, let us define the following observables:

$Q^2 = k\mu^2, \kappa = -\ln k, t = \ln(Q^2/\Lambda^2_{\rm QCD})$ such that $\alpha_s(\mu^2)$ is a function of $(t+\kappa)$. Assuming that the function $\hat{\mathcal{O}}$ is analytic in α_s and κ, it can be shown that the renormalization group equation in Eq. (1.2.82) writes as [52, 53]

$$\frac{\partial}{\partial t}\hat{\mathcal{O}}\left(\alpha_s(t+\kappa),\kappa\right) = -\frac{\partial}{\partial \kappa}\hat{\mathcal{O}}\left(\alpha_s(t+\kappa),\kappa\right) \tag{1.2.83}$$

Notice that the derivative w.r.t. t (in the first term) is taken keeping κ fixed, and in a similar way, the derivative w.r.t. κ (in the second terms) is performed keeping α_s fixed. By Taylor expanding the partonic cross section $\hat{\mathcal{O}}$ around $\kappa = 0$ (i.e. $\mu^2 = Q^2$) while keeping $\alpha_s(t+\kappa)$ fixed, it can be shown that scale-dependent κ cross section can be expressed in terms of the cross section evaluated at the central scale while keeping the κ-dependence in the argument of the coupling α_s. This can be written as follows

$$\hat{\mathcal{O}}\left(\alpha_s(t+\kappa),\kappa\right) = \sum_{n=0}^{\infty}\frac{\kappa^n}{n!}\frac{\partial}{\partial \kappa^n}\hat{\mathcal{O}}\left(\alpha_s(t+\kappa),0\right) \tag{1.2.84}$$

$$= \sum_{n=0}^{\infty}(-1)^n\frac{\kappa^n}{n!}\frac{\partial}{\partial t^n}\hat{\mathcal{O}}\left(\alpha_s(t+\kappa),0\right) \tag{1.2.85}$$

where to go from the first to the second line, we used the renormalization group invariance condition given in Eq. (1.2.83) $\partial/\partial\kappa = -\partial/\partial t$. Notice that all the derivatives in Eq. (1.2.84) and Eq. (1.2.85) are taken keeping α_s and κ fixed respectively. Let us emphasize that when $\kappa = 0$, the left and right-handed side of Eqs. (1.2.84-1.2.85) are exactly the same at every order in perturbation theory. It is clear from Eq. (1.2.85) that derivatives w.r.t. t always add one power of the running of the coupling. That is, the $\mathcal{O}(\kappa^n)$-term in Eq. (1.2.84) is $\mathcal{O}(\alpha_s^n)$ w.r.t. the leading term $\hat{\mathcal{O}}(\alpha_s, 0)$. By defining the cross section evaluated at the central scale $\bar{\mathcal{O}}(t) \equiv \hat{\mathcal{O}}\left(\alpha_s(t), 0\right)$, the missing higher order uncertainty can be defined in the following way

$$\Delta_{\rm N^mLO}(t,\kappa) = \hat{\mathcal{O}}_{\rm N^mLO}\left(\alpha_s(t+\kappa),\kappa\right) - \bar{\mathcal{O}}_{\rm N^mLO}(t) \tag{1.2.86}$$

where $m \in \mathbb{N}$ denotes the order at which the perturbative computation is truncated. If we assume that the hard partonic cross section admits the following perturbative expansion

$$\bar{\mathcal{O}}(t) = \sum_{m=0}^{\infty}\alpha_s^{m+n}(t)\bar{\mathcal{O}}_n \quad \text{with} \quad n \in \mathbb{N}, \tag{1.2.87}$$

then the derivatives that appear in Eqs. (1.2.84-1.2.85) can be computed exactly using the definition of the β-function. For instance, by regrouping terms w.r.t. the coupling constant, it follows that the first derivative of Eq. (1.2.85) is given by

$$\frac{\partial}{\partial t}\bar{\mathcal{O}}(t) = n\alpha_s^{n-1}(t)\beta(\alpha_s)\bar{\mathcal{O}}_0 + (n+1)\alpha_s^{n+1}(t)\beta(\alpha_s)\bar{\mathcal{O}}_1 + \cdots \tag{1.2.88}$$

$$= n\alpha_s^{n+1}\beta_0\bar{\mathcal{O}}_0 + \alpha_s^{n+2}\left(n\beta_1\bar{\mathcal{O}}_0 + (n+1)\beta_0\bar{\mathcal{O}}_1\right) + \cdots \tag{1.2.89}$$

Hence, by computing up to the second derivatives and regrouping the terms w.r.t. the strong coupling α_s, it can be shown that up to NLO order the renormalization scale variation of the hard partonic cross section is given by the following expression

$$\hat{\mathcal{O}}\left(\alpha_s(t+\kappa),\kappa\right) = \alpha_s^n(t+\kappa)\bar{\mathcal{O}}_0 + \alpha_s^{n+1}(t+\kappa)\left(\bar{\mathcal{O}}_1 - n\kappa\beta_0\bar{\mathcal{O}}_0\right) + \cdots \tag{1.2.90}$$

Upon convoluted with the luminosity (as shown in Eq. (1.2.82)), we can derive the renormalization scale-dependent hadronic function

$$\mathcal{O}\left(\alpha_s(t+\kappa),\kappa\right) = \alpha_s^n(t+\kappa)\mathcal{L}\otimes\bar{\mathcal{O}}_0 + \alpha_s^{n+1}(t+\kappa)\mathcal{L}\otimes\left(\bar{\mathcal{O}}_1 - n\kappa\beta_0\bar{\mathcal{O}}_0\right) + \cdots \tag{1.2.91}$$

In order to estimate the missing higher-order uncertainty at the hadronic level, it suffices to simply replace $\hat{\mathcal{O}}$ by $\mathcal{O}$ with the renormalization scale-dependent hadronic function given in Eq. (1.2.91). The estimation of the missing higher-order uncertainty just amounts to varying κ, typically $\kappa \in [-\ln 4, \ln 4]$. In practice, the evaluation of the derivatives appearing in Eq. (1.2.85) cannot always be performed analytically, especially when the dependence on α_s is non-trivial as is often the case in resummed (all-order) computations. Therefore, it is often convenient to perform the derivatives numerically. While the formalism presented here were described in terms of the hadroproduction, it can be used to estimate missing higher-order uncertainty for any (partonic) cross sections [52, 53].

1.2.2 Scale variation for parton evolutions

In the previous section, we assumed that the factorization scale appearing in Eq. (1.1.1)–that separates short-distances from long-distance physics–is fixed. However, since such a factorization scale persists at the level of the physical observables (similar to the renormalization scale), it can be considered as a source of missing higher-order uncertainty and the ambiguity of its definition must be assessed. As argued in Refs. [52, 53], missing higher-order uncertainties associated with the parton evolutions can be assessed using scale variations (similar to the case for the partonic cross section) by varying the scale either at the level of the anomalous dimensions, or at the level of the PDFs, or at the level of the hard scattering partonic coefficient functions. For its simplicity, here, we only present the second approach. We refer the reader to Refs. [52, 53] for details related to the third approach. As we shall see, the derivation of the second approach can be inferred from the results obtained using the first approach.

As our starting point, let us consider the PDF evolution equation defined in Mellin space as given by Eq. (1.1.60)

$$\mu^2\frac{\partial}{\partial\mu^2}\mathcal{Q}(\mu^2) = \Gamma\left(\alpha_s(\mu^2)\right)\mathcal{Q}(\mu^2) \tag{1.2.92}$$

where as before $\mathcal{Q}(N)$ denotes the Mellin version of one of the non-singlet $q_{\rm ns}^V, q_{{\rm ns},a}^{\pm}$, singlet Σ, and gluon g distributions while the function $\Gamma(N)$ represents the associated Mellin transform of one of the various splitting functions. For brevity, in this section, the argument of $\mathcal{Q}$ and Γ are presented assuming Mellin space formalism. Here, μ represents the factorization scale and is related to the scale of the process via $Q^2 = k\mu^2$. As it was the case for the partonic cross section, the factorization scale dependence μ^2 of the anomalous dimension can be made explicit order-by-order in perturbation theory using the the renormalization group invariance. Repeating the procedure in Eqs. (1.2.83-1.2.90) assuming that Γ admits a series expansion $\bar{\Gamma}(t) = \sum_{n=0}^{\infty}\alpha_s^{n+1}(t)\bar{\Gamma}_0$, where we defined the anomalous dimension computed at the central scale $\bar{\Gamma}(t)\equiv\Gamma(\alpha_s(t),0)$, it follows that

$$\Gamma\left(\alpha_s(t+\kappa),\kappa\right) = \alpha_s(t+\kappa)\bar{\Gamma}_0 + \alpha_s^2(t+\kappa)\left(\bar{\Gamma}_1 - \kappa\beta_0\bar{\Gamma}_0\right) + \cdots. \tag{1.2.93}$$

Eq. (1.2.93) represents the key element for the estimation of the missing higher-order uncertainties of the parton evolution through a variation of the factorization scale on which the anomalous dimensions depend. In order to incorporate the above results into the PDFs, we need to integrate Eq. (1.2.92). Based on the derivation performed in Sec. (1.1.2), the scale varied parton density function can be expressed as

$$\mathcal{Q}\left(\alpha_s(t+\kappa),\kappa\right) = \exp\left\{\int_{t_0}^{t} dq\,\Gamma\left(\alpha_s(t+\kappa),\kappa\right)\right\}\bar{\mathcal{Q}}(t_0) \tag{1.2.94}$$

where as before we defined the PDF computed at the central scale to be $\bar{\mathcal{Q}}(t_0)\equiv\mathcal{Q}(\alpha_s(t),0)$. The exponential term in Eq. (1.2.94) can be simplified using Eqs. (1.2.84-1.2.85) and replacing $\hat{\mathcal{O}}$ with Γ. After performing some algebraic simplifications, we have

$$\exp\left\{\int_{t_0}^{t} dq\,\Gamma\left(\alpha_s(t+\kappa),\kappa\right)\right\} = \left[1-\kappa\bar{\Gamma}(t+\kappa)+\cdots\right]\exp\left\{\int_{t_0}^{t+\kappa} dq\,\bar{\Gamma}(q)\right\}. \tag{1.2.95}$$

Plugging the above equation back into Eq. (1.2.94), we can derive the PDF obtained by varying the renormalization scale in the anomalous dimensions with a fixed $\bar{\mathcal{Q}}(t_0)$ evaluated at the initial scale μ_0. Eq. (1.2.94) thus becomes

$$\mathcal{Q}\left(\alpha_s(t+\kappa),\kappa\right) = \left[1-\kappa\bar{\Gamma}(t+\kappa)+\cdots\right]\bar{\mathcal{Q}}(t+\kappa). \tag{1.2.96}$$

The above equation gives the estimation of the missing higher-order uncertainty of the parton evolutions through a variation of the factorization scale μ^2 in the PDFs. Similar to the case for the (partonic) cross section, the missing higher order uncertainty can be conveniently computed using numerical methods with κ varying in the range $[-\ln 4, \ln 4]$.

1.2.3 Combined scale variations

After having described how missing higher-order uncertainties coming from the hard cross section and parton evolutions are estimated independently, let us now discuss how such estimations can be combined in a consistent way. Consider the factorization formula in Mellin space for the hadroproduction cross section:

$$\mathcal{O}\left(\alpha_s(\mu^2),\frac{Q^2}{\mu_F^2},\frac{Q^2}{\mu_R^2}\right) = \mathcal{L}\left(\alpha_s(\mu_F^2),\frac{Q^2}{\mu_F^2}\right)\hat{\mathcal{O}}\left(\alpha_s(\mu_R^2),\frac{Q^2}{\mu_R^2}\right) \tag{1.2.97}$$

where the dependence of the perturbative functions on N are assumed to be implicit. In addition, for the sake of brevity, we have omitted the summation over all possible (sub) channels. Here, μ_R denotes the *renormalization scale* whose dependence is entirely embodied in the hard scattering partonic cross section, and whose variation estimates the missing higher-order uncertainties from due to truncation of the partonic cross section. In a similar way, μ_F represents the *factorization scale* whose dependence is entirely embodied in the anomalous dimensions, and whose variation estimates the missing higher-order uncertainties in the parton evolutions.

Given that the physical cross section in Eq. (1.2.97) factorizes the scale dependence, the renormalization scale μ_R and the factorization scale μ_F can be chosen independently. For what will follow, let us defined the following variables:

$$k_R\mu_R^2 = Q^2, \quad k_F\mu_F^2 = Q^2, \quad \text{and} \quad \kappa_R = -\ln k_R, \quad \kappa_F = -\ln k_F \tag{1.2.98}$$

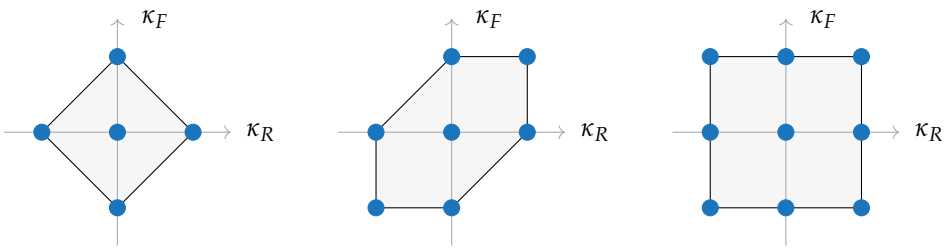


Figure 3 Various prescriptions indicating the sampled values of the factorization scale κ_R and renormalization scale κ_F. From left to right, we show the *five-point* prescription, the *seven-point* prescription, and the *nine-point* prescription. The are spanned by the variation of scales are represented by the gray shapes. The origin of coordinates corresponds to the central scale.

such that the coupling α_s appearing in the hard cross section and in the PDFs are functions of $t_R = t + \kappa_R$ and $t_F = t + \kappa_F$ respectively with $t = \ln\left(Q^2/\Lambda^2_{\rm QCD}\right)$. In terms of these variables, the scale-varied factorized hadronic cross section is expressed as

$$\mathcal{O}\left(t, \kappa_F, \kappa_R\right) = \mathcal{L}\left(\alpha_s(t_F), \kappa_F\right) \hat{\mathcal{O}}\left(\alpha_s(t_R), \kappa_R\right) \tag{1.2.99}$$

Based on the derivations performed in Secs. 1.2.1-1.2.2, the all-order scale varied PDFs (here expressed in terms of the luminosity) and the partonic cross section are respectively given by the following expressions

$$\mathcal{L}\left(\alpha_s(t_F), \kappa_F\right) = \sum_{n=0}^{\infty} (-1)^n \frac{\kappa_F^n}{n!} \frac{\partial}{\partial t_F^n} \bar{\mathcal{L}}(t_F) \tag{1.2.100}$$

$$\hat{\mathcal{O}}\left(\alpha_s(t_R), \kappa_R\right) = \sum_{n=0}^{\infty} (-1)^n \frac{\kappa_R^n}{n!} \frac{\partial}{\partial t_R^n} \bar{\mathcal{O}}(t_R). \tag{1.2.101}$$

The luminosity and partonic cross sections evaluated at the central scales, namely $\mu^2 = Q^2$, are respectively defined as $\bar{\mathcal{L}}(t_F) = \mathcal{L}(\alpha_s(t_F), 0)$ and $\bar{\mathcal{O}}(t_R) = \hat{\mathcal{O}}(\alpha_s(t_R), 0)$. Taking into account that $\partial/\partial t$ is $\mathcal{O}(\alpha_s^2)$, the NLO scale-varied hadronic cross section is given by

$$\mathcal{O}_{\rm NLO}\left(t, \kappa_F, \kappa_R\right) = \bar{\mathcal{L}}\left(t_F\right) \bar{\mathcal{O}}\left(t_R\right) - \left[\kappa_R \frac{\partial}{\partial t} \bar{\mathcal{L}}\left(t_F\right) \bar{\mathcal{O}}\left(t_R\right) + \kappa_F \bar{\mathcal{L}}\left(t_F\right) \frac{\partial}{\partial t} \bar{\mathcal{O}}\left(t_R\right)\right] \tag{1.2.102}$$

The above expression suggests that scale variations w.r.t κ_R can be determined by taking the derivatives of the hadronic cross section w.r.t t_R while keeping t_F fixed, and vice-versa. Up to NLO, the scale-varied hadronic cross section can finally be written as

$$\mathcal{O}_{\rm NLO}\left(t, \kappa_F, \kappa_R\right) = \tilde{\mathcal{O}}(t_F, t_R) - \left(\kappa_R \frac{\partial}{\partial t_R} \tilde{\mathcal{O}}(t_F, t_R) + \kappa_F \frac{\partial}{\partial t_R} \tilde{\mathcal{O}}(t_F, t_R)\right), \tag{1.2.103}$$

where $\tilde{\mathcal{O}}(t_F, t_R) = \bar{\mathcal{L}}\left(t_F\right) \bar{\mathcal{O}}\left(t_R\right)$ is the hadronic cross section evaluated at the central scale. The missing higher-order uncertainties of the hadronic cross section are then estimated by varying independently in a symmetric way, typically $|\kappa_F|, |\kappa_R| \leq \ln 4$. One of the ways to achieve such a double scale variations is a prescription known as *fived-point* scale variation

where the scale choices are given by $(\kappa_F, \kappa_R) = (0,0), (\pm \ln 4, 0), (0, \pm \ln 4)$. Alternative prescriptions are shown in Fig. 3. Throughout this thesis, we adopt the *seven-point* scale variation, details on such a prescription are described in Sec. (3.3).

1.3 The road toward 1% accuracy

Chief among the goals of precision phenomenology of the kind studied at hadron colliders such as LHC is to produce theoretical predictions accurate at percent-level to be compared to an ever increasing range of high precision experimental measurements. The road to such an accuracy requires the following aspects: an understanding of the meaning of the residual theoretical uncertainty given by unknown higher-orders in perturbation theory, and a better determination of the parton density functions in order to push their uncertainties to lower values. In the following section, we summarize the advancements as well as the issues that currently persist toward achieving that goal. In particular, we report in the first part the current state-of-the-art on PDF determination and their accuracies. The second part is dedicated to the discussion of the estimation of missing higher-order uncertainties including a discussion on how resummed calculations can be used to estimate these unknown contributions.

1.3.1 Methodological aspects

Considerable advancements have happened in the area of PDF determination with the release of the NNPDF4.0 methodology [4, 77]. The improvements in comparison to NNPDF3.1 manifest in two main aspects: the systematic inclusions, for the first time, of LHC Run II data at a center-of-mass energy $\sqrt{s} = 13$ TeV, enlarging significantly the experimental data used to constrain PDFs; and a new fitting procedure based on state-of-the-art machine learning techniques, whose impacts improve significantly the fitting time and provide more accurate descriptions of the PDF uncertainties.

In terms of experimental data set, the NNPDF4.0 PDF set includes new LHC measurements at $\sqrt{s} = 13$ TeV for processes already present in NNPDF3.1, but also new processes such as single top production [129], dijets [130], and W+jet [131]. In terms of methodology, not only the minimization algorithm has been changed from *Genetic Algorithm* to *Stochastic Gradient Descent*, but also the whole methodology itself is fitted using the *hyperparameter optimization*. The hyperparameter optimization if performed through a procedure called κ-folding where the effectiveness of a given methodology (here methodology means a particular set of hyperparameters) is checked on sets of data excluded in turn from the fit. Despite the fact that fitting procedure does not account for the methodological uncertainties, such ambiguities can be kept under control using: *closure tests* [40] which assesses the faithfulness of the uncertainties, and *future tests* which tests the backward and forward compatibility of the data.

Thanks to these improvements, both in terms of experimental data and methodology, the PDFs are reliable and accurate at *one*-percent level across a wide range of kinematic regions. This is illustrated in Fig. 4 where we plot the relative parton luminosities as a function of the invariant mass m_X and the rapidity y for various channels at a center-of-mass energy $\sqrt{s} = 14$ TeV. The results from NNPDF3.1 are shown on the left while

results from NNPDF4.0 are shown on the right. For all the channels considered, it is apparent that the NNPDF4.0 methodology yields smaller PDF uncertainties reaching up to *one*-percent accuracy for most of the kinematic ranges.

In terms of PDF determination, the main missing piece towards achieving a *one*-percent accuracy is the inclusion of *theoretical uncertainties*, which is dominated by missing higher-order uncertainties in perturbative QCD computations. The inclusion of the theoretical uncertainties into the PDF fit will become increasingly crucial with the insatiable quests for new physics and the advent of future high energy colliders [132]. Despite the fact that a proof of concept of how this could be achieved already exists [52,53], PDF sets are still delivered without accounting for theoretical uncertainties.

1.3.2 Theoretical aspects

On the theory side, one of the main obstacles toward approaching a percent-level accuracy is the unknown higher-order corrections in perturbative calculations. Whilst considerable progress has been made in performing higher-order perturbative computations for a large number of processes of interest at the LHC, the need for a reliable estimation of theory uncertainties associated with missing higher contributions is becoming more pressing. As mentioned in previous sections, thus far, the most standard and widespread approach of estimating theoretical uncertainties associated with missing higher-orders in perturbative calculations is to perform a variation of the unphysical scales involved in the process, namely the factorization and renormalization scales. In Sec. (1.2.3), we presented various prescriptions on how the two scales could be varied. The uncertainty envelope is then computed by taking the minimum and the maximum from the scale variations.

From the practical point of view, the use of scale variation to estimate missing higher-order uncertainties presents a number of advantages. First, due to the fact that the scale dependence of the strong coupling and PDFs are universal, scale variations can be used to estimate theoretical uncertainties for any perturbative process. Second, the constraint imposed by the renormalization group invariance ensures that as the order order of perturbative calculations increases, the scale dependence decreases. Third, estimation of missing higher-order uncertainties resulting from scale variations produces smooth functions of the kinematics, incorporating strong correlations in the nearby regions of the phase space. However, the procedure of estimating missing higher-order uncertainty using a scale variation also presents some shortcomings, chief among which is the fact that it does allow for a probabilistic interpretation, therefore removing any attempt to estimate the *degree of belief*. In addition, there is the ambiguity in defining the central scale around which the variation should be performed and the ranges at which the scales are allowed to vary. Most importantly, scale variation misses uncertainties associated to new singularities appearing at higher-orders but not present at lowers-orders.

Recently, several methods have been proposed in order to address some of the caveats of the scale variations. In Ref. [80], Cacciari and Houdeau proposed a new approach of estimating missing higher-order uncertainties using a Bayesian model. In short, the method consists on adopting some assumptions on the progression of the perturbative expansion, then based on the knowledge of the first few orders, one can infer on the *hidden parameters* that are assumed to bound the structure of the perturbative coefficients, allowing for an inference on the unknown subsequent contributions. While this approach

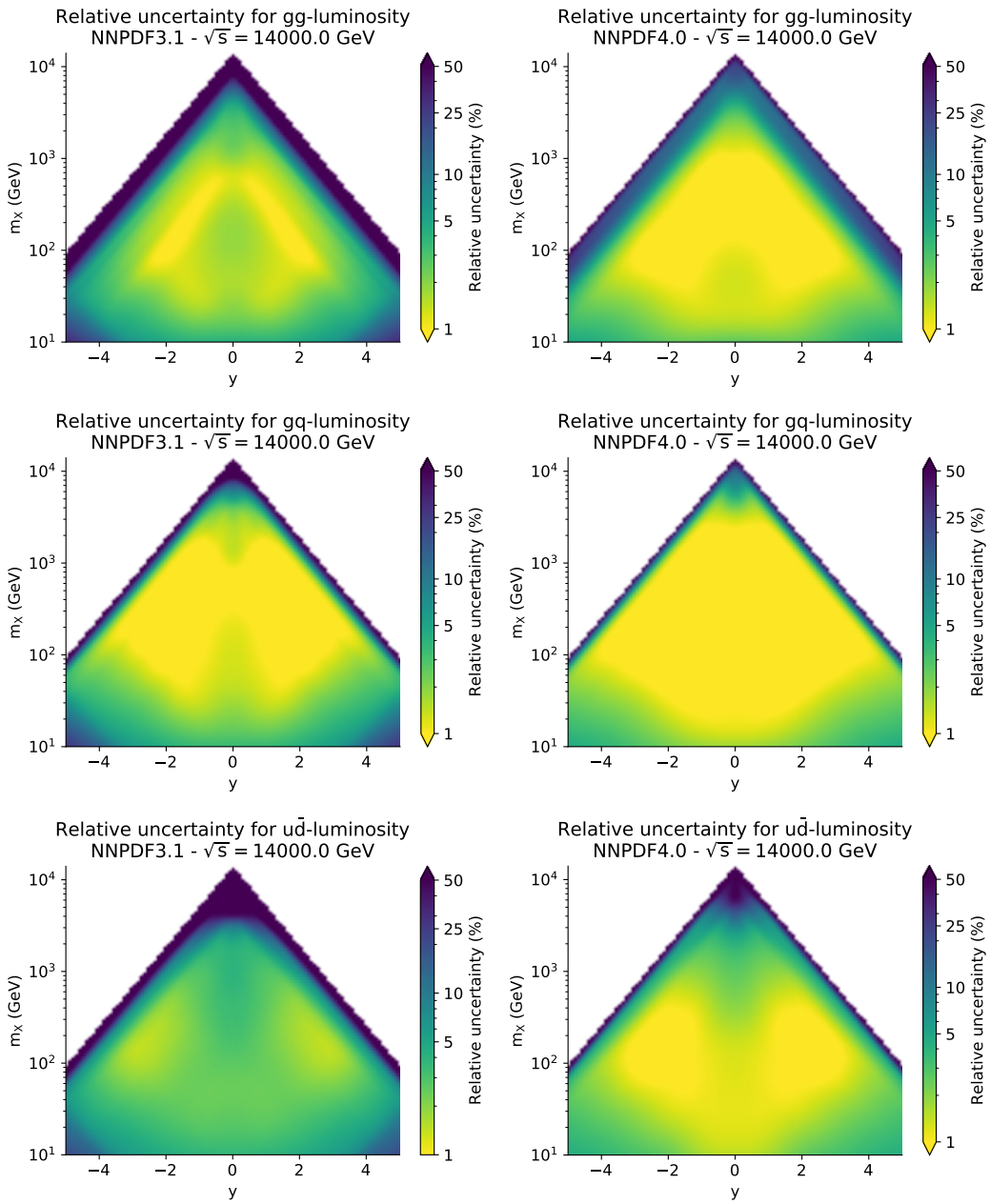


Figure 4 The relative parton luminosities as a function of the invariant mass m_X and the rapidity y for various channels, namely gg (top), gq (middle), and $u\bar{d}$ (bottom). The results are shown the the two families of NNPDF sets: NNPDF3.1 (left) and NNPDF4.0 (right). The colour-bars on the right represent the relative uncertainty in terms of percentage. Plots are taken from Ref. [4].

has proved to perform quite well for QCD observables at e^+e^- colliders [80], its reliability when it comes to proton-proton collider observables is subject to questions [22, 81]. In Ref. [82], Bonvini built upon Cacciari-Houdeau's work to construct more general, flexible and robust models. The various models have been validated on some observables at LHC that are known to very high accuracy.

A possible way to estimate the impacts of unknown higher-orders in perturbation theory is the information provided by resummed calculations. Indeed, information on the various kinematic limits that appear in fixed order calculations are contained at all-order in resummed expressions. Thus, it should be possible to *consistently* combine these various limits to predict the subsequent unknown contributions in the perturbative computations, in the same way as it was done in Ref. [133]. In Sec. (3.4) we give a proof of concept by combining high energy and threshold resummations to approximate the NNLO Higgs transverse momentum distribution.

1.4 Summary

We have devoted this chapter to the introduction of the concepts for making theoretical predictions in QCD. In particular, we highlighted the importance of the factorization property that allows one to perform perturbative computations by separating long-distance physics from short-distance interactions. As a result, it is possible to represent measurable cross sections by folding the partonic cross sections with the PDFs. At the partonic level, we emphasized the efficiency of the spinor helicity formalism and the MHV techniques in computing massless scattering amplitudes. This allowed us to introduce the splitting functions that govern the DGLAP evolution equations. At the PDF level, we investigated the relation between the scale dependence of the PDFs and the DGLAP evolution equations whose solutions we have explicitly computed up to NLO in the singlet and non-singlet sector. In particular, we provided a brief overview of the determination of the PDFs in terms of the momentum fractions using as a reference the NNPDF methodology. We then raised the issue that plagues perturbative computations, namely the estimation of missing higher-order corrections and the appearance of large logarithms. These subjects will be the topics of the next chapters.

1.A The running of α_s and asymptotic freedom

The applicability of perturbative computations in QCD relies on the smallness of the strong coupling constant α_s. Indeed, perturbative treatment of a given observable can only be reasonable if the coupling of quarks and gluons are small. As introduced in the previous sections, the strength of the interaction between these partons are dictated by $\alpha_s(\mu^2) = g_s(\mu^2)/4\pi$, where the the renormalization scale μ^2 is a remnant of the renormalization of UV divergences. Since physical observables should not depend on an unphysical scale, we can derive a differential equation that describes the dependence of α_s on μ^2. This is known as the *renormalization group* equation and writes as

$$\mu^2 \frac{\partial}{\partial \mu^2} \alpha_s(\mu^2) = \beta\left(\alpha_s(\mu^2)\right), \tag{A.1}$$

where the QCD beta function β is a perturbative function that admits an expansion in α_s. At all-order, its expression is given by $\beta(\alpha_s) = \sum_{n=0}^{\infty} -\beta_n \alpha_s^{n+2}$. Each coefficient appearing in that series can be computed perturbatively. In $\overline{\text{MS}}$, the first three coefficients are given by the following expressions

$$\beta_0 = \frac{1}{12\pi}\left(11C_A - 2n_f\right), \qquad \beta_1 = \frac{1}{24\pi^2}\left(17C_A^2 - 5C_A n_f - 3C_F n_f\right), \tag{A.2}$$

$$\beta_2 = \frac{1}{64\pi^3}\left(\frac{2857}{54}C_A^3 - \frac{1415}{54}C_A^2 n_f - \frac{205}{18}C_A C_F n_f + \frac{78}{54}C_A n_f^2 + \frac{11}{9}C_F n_f^2\right). \tag{A.3}$$

The QCD factors C_A and C_F are respectively defined as $C_A = N_c$ and $C_F = (N_C^2 - 1)/2N_c$ where $N_c = 3$ for QCD. By introducing a reference scale μ_0^2, we can derive the solution to the renormalization group equation in Eq. (A.1). Such a solution generally takes the following form

$$\alpha_s\left(\mu^2\right) = \sum_{n=1}^{\infty} \alpha_0^n f_n\left(\frac{\mu^2}{\mu_0^2}\right), \tag{A.4}$$

where for simplicity we defined $\alpha_0 = \alpha_s(\mu_0^2)$. Including contributions up to β_2, the functions f_n are given by

$$f_1(X) = \frac{1}{X}, \qquad f_2(X) = -\frac{\beta_1}{\beta_0}\frac{\ln X}{X^2} \tag{A.5}$$

$$f_3(X) = \frac{1}{X^3}\left[\frac{\beta_1^2}{\beta_0^2}\left(\ln^2 X - \ln X - (1-X)\right) + \frac{\beta_2}{\beta_0}(1-X)\right] \tag{A.6}$$

where the scale dependent variable X is defined to be $X = 1 + \alpha_0\beta_0 \ln(\mu^2/\mu_0^2)$. Eq. (A.4) then allows one to compute the coupling constant at scale μ^2 in case its value is known at a reference scale μ_0^2. Notice that solving the renormalization group equation sums up logarithms to all order in perturbation theory. The running of $\alpha_s(\mu^2)$ where functions up to f_3 are included, for instance, resums logarithms up to NNLL.

As a final side remark, it is worth emphasizing the presence of Landau pole. This pole is given by $\mu_L^2 = \Lambda^2 = \mu^2 \exp(-1/(\alpha_0\beta_0))$ and is present at all logarithmic order. Such a singularity appears also in resummed calculations in Fourier and Mellin space making the conversion to the direct space a complicated task.

1.B Integral transforms

In the following section, we describe various integral transforms that appear often in the reset of the manuscript, especially in the context of all-order computations. In particular, we describe some of the properties of the Fourier and Mellin transforms, mainly in the context of a differentiable function.

1.B.1 Fourier transform

Let us first briefly review the concept of *two-dimensional* Fourier transform which will be used on the context of small-p_T resummation. Such a transform is required in order to

factorize the delta constraints that enforces the conservation of transverse momentum. Using a function that is differential in $d\vec{p}_T = dp_T^x dp_T^y$ as an example, the two-dimensional Fourier transform is defined as:

$$\mathcal{F}\left[f(\vec{p}_T);\vec{p}\right] \equiv \frac{df}{d\vec{p}_T}(\vec{b}) = \int d^2\vec{p}_T \exp\left(i\vec{p}_T\vec{b}\right) \frac{df}{d\vec{p}_T}\left(\vec{p}_T\right). \tag{B.1}$$

In order to be consistent with previous notations, we denote the differential function and its Fourier transform with the same variable. Recall that they are only distinguished by the arguments. As a result, the inverse Fourier transform is derived by integrating back $\vec{b}$:

$$\frac{df}{d\vec{p}_T}\left(\vec{p}_T\right) = \int \frac{d^2\vec{b}}{(2\pi)^2} \exp\left(i\vec{p}_T\vec{b}\right) \frac{df}{d\vec{p}_T}(\vec{b}). \tag{B.2}$$

In the case where the differentiable function f is *azimuthally symmetric*, i.e. for $p_T = |\vec{p}_T|$ and $b = |\vec{b}|$, if

$$\frac{df}{d\vec{p}_T}(\vec{p}_T) = \frac{1}{2\pi}\frac{df}{p_T dp_T}(p_T), \quad \text{and} \quad \frac{df}{d\vec{p}_T}(\vec{b}) = \frac{df}{dp_T}(b), \tag{B.3}$$

then the azimuthal angle can be performed, leading to the *Hankel transforms*:

$$\frac{df}{dp_T}(b) = \int_0^\infty dp_T\, J_0(bp_T)\frac{df}{dp_T}(p_T), \tag{B.4}$$

$$\frac{df}{dp_T}(p_T) = p_T \int_0^\infty db\, b\, J_0(bp_T)\frac{df}{dp_T}(b), \tag{B.5}$$

where $J_0(x)$ is the zeroth-order Bessel function of the first kind.

1.B.2 Mellin transform

Definition & Example

Let $f(t)$ be a function defined on a positive real axis $0 < t < \infty$ (in QCD, the value of t is often restricted in $(0,1)$). The Mellin transform of f is defined on the complex plane as:

$$\mathcal{M}\left[f;N\right] \equiv f(N) = \int_0^\infty dt\, t^{-N} f(t). \tag{B.6}$$

In general, the above integral only exists for complex values $N = a + ib$ such that $c < a < d$ for some values of c and d depending on $f(t)$. Such a domain is often referred to as *strip* $\mathcal{S}(c,d)$. In order for the integral to converge, the function $f(t)$ can grow as much as e^{ct} when $t \to \infty$.

As an illustration, let us consider a straightforward function defined by $f(t) = e^{-pt}$ $\forall p > 0$. Using the definition given in Eq. (B.6), it is immediate to see that the Mellin transform is given in terms of a Gamma function, $f(N) = p^{-N}\Gamma(N)$. Recalling that the Gamma function is analytic in the region $\mathfrak{Re}(N) \equiv a > 0$, it follows that the strip of holomorphy is the positive half-plane in complex space.

Relation with two-sided Laplace transform

The Mellin transform defined in Eq. (B.6) is closely related to the Laplace transform upon performing the change of variable $t=e^{-x}$. the integral in Eq. (B.6) now becomes:

$$f(N) = \int_{-\infty}^{\infty} dx\ \exp\left(-Nx\right) f\left(\exp(-x)\right). \tag{B.7}$$

By defining a new function $g(x) \equiv f(e^{-x})$, one can recover the standard expression of the extended *two-sided Laplace transform*:

$$\tilde{\mathcal{L}}\left[g;N\right] \equiv \int_{-\infty}^{\infty} dx\ \exp\left(-Nx\right) g(x). \tag{B.8}$$

It is worth mentioning that the extended two-sided Laplace transform in Eq. (B.8) and the Mellin transform in Eq. (B.6) are related through the following relation

$$\mathcal{M}\left[f(t);N\right] = \tilde{\mathcal{L}}\left[f\left(\exp(-x)\right);N\right]. \tag{B.9}$$

Inverse Mellin transform

The inverse Mellin transform can be generally computed by defining the Mellin moment as $N=a+2\pi i\beta$ and realizing that (as it was the case for the two-sided Laplace transform) the Mellin and Fourier transforms are related by

$$\mathcal{M}\left[f(t);a+2\pi i\beta\right] = \mathcal{F}\left[\exp(-ax)f\left(\exp(-x)\right);\beta\right]. \tag{B.10}$$

The above relation implies that the function $f(x)$ is related to the Mellin function $f(N)$ according to the following relation:

$$\exp(-ax)f\left(\exp(-x)\right) = \int_{-\infty}^{\infty} d\beta\ f(N)\exp\left(2\pi\beta\right). \tag{B.11}$$

Going back to the definition in terms of t with a change of variable and performing some algebraic simplifications, we arrive at the expression of the Mellin transform

$$f(t) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} dN\ t^{-N} f(N), \tag{B.12}$$

where the integration is performed along the vertical line $\mathfrak{Re}(N) \equiv a$.

The uniqueness of the inverse Mellin transform depends on the strips of holomorphy. That is, a function $f(N)$ defined in Mellin space can map to different inverse Mellin transforms depending on the regions of holomorphy considered. This is related to how the value of a should be chosen.

To illustrate this, let us consider our previous example, $f(N)=\Gamma(N)$ (with $p=1$). For $q \equiv \mathfrak{Re}(N) > 0$, $f(t)=e^{-t}$ is shown to be the Mellin transform. Using the inverse formula in Eq. (B.12), one has the following integral representation

$$\exp(-t) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} dN\ t^{-N}\Gamma(N). \tag{B.13}$$

As mentioned earlier, the Γ-function is known to be analytically continuable in the left half-place of the complex space, expect at zero and infinite negative integers where poles exist. Nevertheless, the contour integration can be shifted to the left such that the integral only picks up the values of the residues at each pole. thus, introducing $-m < \tilde{a} < -(m-1)$ for an integer m, the integral in Eq. (B.13) can be written as

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} dN\, t^{-N}\Gamma(N) = \sum_{n=0}^{m-1}\frac{(-1)^n}{n!}t^n + \frac{1}{2\pi i}\int_{\tilde{a}-i\infty}^{\tilde{a}+i\infty} dN\, t^{-N}\Gamma(N). \tag{B.14}$$

Hence, the inverse Mellin transform of $\Gamma(N)$ in the region of strip $\mathcal{S}(-m,-m+1)$ is given by the following expression:

$$\frac{1}{2\pi i}\int_{\tilde{a}-i\infty}^{\tilde{a}+i\infty} dN\, t^{-N}\Gamma(N) = e^{-t} - \sum_{n=0}^{m-1}\frac{(-1)^n}{n!}t^n. \tag{B.15}$$

The above result is different from the original function $f(t)=e^{-t}$. However, it can be proved using the *Stirling formula* that the right-hand side of the above integral transform vanishes in the limit $m \to \infty$.

1.C Spinor Helicity Formalism

This section is devoted to a brief review of the spinor helicity formalism. This formalism was first introduced when it was realized that the Lorentz group $SO(1,3)$ ib four dimensions is isomorphic to the group $SL(2)\otimes SL(2)$ and hence the finite-dimensional representations are classified as (m,n) where m and n are integers or half-integers.

Spinor variables

Massless particles have remarkable properties when expressed in the helicity basis. As a matter of fact, the spinor helicity formalism renders the analytic expression of scattering amplitudes in a more compact and simple form as compared to the standard formulation using momentum four-vectors. The underpinning idea behind the spinor helicity formalism is the mapping from a four-vector into a two-by-two matrix given by the following expression

$$k_{a\dot{a}} = k_\mu \left(\sigma^\mu\right)_{a\dot{a}} = \begin{pmatrix} k_0 - k_3 & -k_1 + ik_2 \\ -k_1 - ik_2 & k_0 + k_3 \end{pmatrix}, \tag{C.1}$$

where $\sigma^\mu = \left(\sigma^0, \vec{\sigma}\right)$ are the usual Pauli matrices. In the high energy limit where particles are quasi-massless, the determinant of the of $k_{a\dot{a}}$ vanishes. Hence, for a lightlike momentum k, the momentum can be written as a product of two vectors $k_{a\dot{a}} = \lambda_a \tilde{\lambda}_{\dot{a}}$ where λ_a and $\tilde{\lambda}_{\dot{a}}$ are respectively the left and right-handed spinors. Scattering amplitudes possess interesting properties when the momentum $p_{a\dot{a}}$ is complex. In such a case, the spinors λ_a and $\tilde{\lambda}_{\dot{a}}$ are independent of each other. For the purpose of this review, let us stick to the treatment of real-valued momenta. As mentioned before, the scalar product between two momenta can be expressed as $2(k_i k_j) = \langle ij\rangle\,[ij]$ where the angle and square brackets are defined in

terms of the spinors as $\langle ij\rangle = \varepsilon^{ab}\lambda_a\mu_b$ and $[ij] = \varepsilon^{\dot{a}\dot{b}}\tilde{\lambda}_{\dot{a}}\tilde{\mu}_{\dot{b}}$ respectively. Since spinors are simple mathematical objects, they exhibit algebraic properties. For instance, one can use the convention for left and right-handed spinors to show that the above definitions imply $\langle ii\rangle = [ii] = 0$. In addition, all other product of angle and square brackets vanish, for instance, $\langle ij] = [ij\rangle = = 0$. Notice also that these spinor products are antisymmetric, i.e. $\langle ij\rangle = -\langle ji\rangle$ and $[ij] = -[ji]$. Another useful relation is that square and angle brackets are related to each other according to the relation $\langle ij\rangle^\star = [ij]$.

For spin-1 bosons, the polarization can also be expressed in terms of a pair of massless spinors. In order to construct an expression of the polarization in terms of the square and angle brackets, the expression must satisfy $\epsilon^\star_\mu(p_i,k)\epsilon^\mu(p_i,k) = -1$ and the transverse condition $p_\mu\epsilon^\mu = 0$. Expressions of the polarization that satisfy those constraints are given by the following [85]

$$\epsilon^-_\mu(p_i,k) = -\frac{1}{\sqrt{2}}\frac{[k|\gamma_\mu|i\rangle}{[ki]} \quad \text{and} \quad \epsilon^+_\mu(p_i,k) = \frac{1}{\sqrt{2}}\frac{\langle k|\gamma_\mu|i]}{\langle ki\rangle}, \tag{C.2}$$

where k is a reference lightlike momentum that can be chosen arbitrarily. This freedom in choosing a reference momentum reflects the *gauge invariance*. The only constraint is that the reference momentum k cannot be aligned to the momentum p_i since we do not want the product $\langle ik\rangle$ or $[ik]$ to vanish.

The little group scaling

In the following section, we show how any three-point amplitude can be trivially computed using the *little group scaling*. With the knowledge of the three-point amplitude, all higher-order MHV ($\overline{\text{MHV}}$) amplitudes can be computed using the BCFW recursion relations [134]. This is possible because the spinor helicity variables make manifest the symmetry obeyed by the scattering amplitudes.

Roughly speaking, the little group is the group of transformation that leaves invariant the momentum of an *on-shell* particle [85]. From the previous definitions, it is straightforward to deduce that the momentum of a given particle is given by $k_i = |i\rangle[i|$. The little group transformation is therefore the scaling

$$|i\rangle \longrightarrow t|i\rangle \quad \text{and} \quad |i] \longrightarrow t^{-1}|i] \tag{C.3}$$

By virtue of this rescaling, one can setup some ground rules for the external fermions and spin-1 bosons. These rules are defined as follows:

- Spinors for fermion scales as t^{-2h} where h denotes the helicity of the fermion.

- Polarization vectors for spin-1 bosons also scale as t^{-2h} but for $h = \pm 1$. By rescaling the spinors $|i\rangle$ and $|i]$, one can check in the expression of the polarization (Eq. (C.2)) that this rule hold true.

$$\epsilon^-_\mu \to t^2\epsilon^-_\mu = -\frac{t^2}{\sqrt{2}}\frac{[k\,|\gamma_\mu|\,i\rangle}{[ki]} \quad \text{and} \quad \epsilon^+_\mu \to t^{-2}\epsilon^+_\mu = \frac{t^{-2}}{\sqrt{2}}\frac{\langle k\,|\gamma_\mu|\,i]}{\langle ki\rangle} \tag{C.4}$$

The above descriptions imply that the little group scaling of each polarization encodes its helicity. Thus, a scattering amplitude with n external state particles will then be multi-linear in the corresponding polarization

$$\mathcal{M}\left(1^{h_1},\cdots,n^{h_n}\right)=\epsilon^{h_1}_{\mu_1}\cdots\epsilon^{h_n}_{\mu_n}\mathcal{M}^{\mu_1\cdots\mu_n}(1,\cdots,n). \tag{C.5}$$

The tensorial object on the right-hand side of the above equation is the usual amplitude computed from Feynman diagrams. The object on the left-hand side instead represent a set o functions corresponding to different helicity configurations. These objects are known as the *true* scattering amplitudes in the sense that they are gauge invariant. Therefore, for on-shell amplitudes, the scaling of a particle labelled by i gives rise to a weight $(-2h_i)$ where again h_i labels the helicity of the particle i. That is,

$$\mathcal{M}\left(\cdots,\left\{t_i|i\rangle,t_i^{-1}|i\right]\right\},\cdots\Big)=t^{-2h_i}\mathcal{M}(\cdots,\{|i\rangle,|\,i]\},\cdots\Big)\,. \tag{C.6}$$

The above expression allows one to compute any three-point amplitude in QCD and subsequently scattering amplitudes with larger number of external legs. In the next section, we demonstrate how trivial it is to compute three-point MHV amplitudes using the concept of little group scaling.

Computation of a three-point amplitude using the little group scaling

Consider a three-point amplitude where we have one gluon and a pair of quark-antiquark. This is a simpler version of the process studied in Sec. (1.1.1). Let us start with the observation that in order for three-point amplitudes to not vanish, they can only depend on either the angle or square brackets. To demonstrate this, consider a three point amplitude where the momentum of the particles are labelled by k_i $(i=1,2,3)$. Assuming all the momenta to be incoming, the momentum conservation writes as $\sum_{i=1}^{3}k_i=0$, which implies that $(k_1+k_2)^2=p_3^2=\langle 12\rangle\,[12]=0$. This tells us that either $\langle 12\rangle$ or $[12]$ must be equal to zero. Let us first suppose that $\langle 12\rangle$ does not vanish, then we have $\langle 12\rangle\,[23]=\langle 1|p_2|3]=-\langle 1|(p_1+p_3)|3]=0$, and similarly $\langle 12\rangle\,[13]=\langle 2|(p_2+p_3)|3]=0$. This shows that $[12]$, $[13]$, and $[23]$ vanish if we require $\langle 12\rangle$ to be non-zero. In a similar way, one can easily show that $\langle 13\rangle$ and $\langle 23\rangle$ vanish if we consider the case where $\langle 12\rangle$ vanish while $[12]$ does not.

We are now ready to compute the partial three-point amplitude $\tilde{\mathcal{M}}_3(p,k,\bar{p})$ where as before p and $\bar{p}$ denote the quark and antiquark with helicity $(-1/2)$ and $(1/2)$ respectively. In order to have an MHV amplitude, let us choose the gluon k to have a *negative* helicity. Based on the above observation, the partial amplitude $\tilde{\mathcal{M}}_3$ can therefore be expressed in terms of angle brackets as follows

$$\tilde{\mathcal{M}}_3\left(h_p,h_k,h_{\bar{p}}\right)=\langle pk\rangle^x\langle p\bar{p}\rangle^y\langle k\bar{p}\rangle^z. \tag{C.7}$$

The values of x,y and z are then fixed by the little group scaling via the equations

$$\begin{aligned}x+y&=-2h_p\\x+z&=-2h_k\\y+z&=-2h_{\bar{p}}.\end{aligned} \tag{C.8}$$

Recalling that the helicities are given by $h_p = -1/2$, $h_{\bar{p}} = 1/2$ and $h_k = 1$, the above system of equations can be straightforwardly solved in order to find that $(x, y, z) = (2, -1, 0)$. Hence, the three-point amplitude given in Eq. (C.7) can be expressed as:

$$\tilde{\mathcal{M}}_3\left(p^-, k^-, \bar{p}^+\right) = \frac{\langle pk\rangle^3 \langle \bar{p}k\rangle}{\langle pk\rangle\langle k\bar{p}\rangle\langle \bar{p}p\rangle}. \tag{C.9}$$

The above computations clearly show the simplicity of the spinor helicity formalism and the the MHV technique. Using *on-shell recursion relations* such as the BCFW recursion relation in which a complex component is added to the momentum, all higher-order amplitudes can be recursively derived from the most basic three-point ones. For instance, the general expression for an MHV amplitude involving a pair of quark-antiquark and an arbitrary number of gluons is given by

$$\tilde{\mathcal{M}}^{\text{MHV}}_{n+2}\left(p^-, 1^+, \cdots, i^-, \cdots, n^+, \bar{p}^+\right) = \frac{\langle pi\rangle^3 \langle \bar{p}i\rangle}{\langle p1\rangle\langle 12\rangle\cdots\langle n\bar{p}\rangle\langle \bar{p}p\rangle} \tag{C.10}$$

The $\overline{\text{MHV}}$ counterpart of the above partial amplitude can be read off directly from that expression by simply replacing the angle brackets into square brackets.

1.D Splitting functions & Anomalous dimensions

In the following section, we give the explicit expressions of the splitting functions and anomalous dimensions at leading order. As we shall discuss in Sec. (2), expression of the splitting functions (anomalous dimensions) up to NLO is enough in the context of computing NNLL standard transverse momentum resummation, while the large-N limit of three loop anomalous dimensions are required in order to correctly account for the soft logarithms in the context of the improved transverse momentum resummation.

Let us start by collecting the LO expression of the splitting functions that we computed in Sec. (1.1.1). They were first computed in Ref. [135] and are given by the following

$$P^{(0)}_{gg}(x) = \frac{C_A}{\pi}\left[\frac{x}{(1-x)_+} + \frac{1-x}{x} + x(1-x)\right] + \beta_0\delta(1-x), \tag{D.1}$$

$$P^{(0)}_{qq}(x) = \frac{C_F}{2\pi}\frac{1+x^2}{(1-x)_+} + \frac{3C_F}{4\pi}\delta(1-x), \tag{D.2}$$

$$P^{(0)}_{qg}(x) = \frac{n_f}{2\pi}\left[x^2 + (1-x)^2\right], \qquad P^{(0)}_{gq}(x) = \frac{C_F}{2\pi}\left[\frac{1+(1-x)^2}{x}\right]. \tag{D.3}$$

Notice that the above expressions can be used to derive the non-singlet splitting functions, $P^{\pm}_{\text{ns},a}$ and P^{V}_{ns}, which at leading-order vanish. For the NLO splitting functions, they have been computed in Refs. [136–144]. As for the LO, the expression of the splitting functions for the singlet and non-singlet combinations can be derived from the above results. Let us turn to the Mellin version (anomalous dimensions) of the above splitting functions. The Mellin transform of these functions were mostly computed in Refs. [141, 142] and their

continuation for complex Mellin moments in Refs. [119, 145–149]. For the sake of brevity, we only collect below the LO expressions:

$$\gamma_{qq}^{(0)}(N) = \frac{C_F}{2\pi}\left[\frac{3}{2} - \frac{1}{N+1} - \frac{1}{N+2} - 2\psi(N+1) - 2\gamma_E\right] \tag{D.4}$$

$$\gamma_{gq}^{(0)}(N) = \frac{C_F}{2\pi}\left[\frac{2}{N} - \frac{2}{N+1} + \frac{1}{N+2}\right] \tag{D.5}$$

$$\gamma_{qg}^{(0)}(N) = \frac{n_f}{2\pi}\left[\frac{1}{N+1} - \frac{2}{N+2} + \frac{2}{N+3}\right] \tag{D.6}$$

$$\gamma_{gg}^{(0)}(N) = \frac{C_A}{\pi}\left[\frac{1}{N} - \frac{2}{N+1} + \frac{1}{N+2} - \frac{1}{N+3} - \psi(N+1) - \gamma_E\right] + \beta_0 \tag{D.7}$$

Part II

A TALE OF SCALES

2

ALL-ORDER RESUMMATION OF TRANSVERSE MOMENTUM DISTRIBUTIONS

The use of perturbative computations, described in the previous chapter, is mainly justified if the strong coupling α_s is small compared to other scales involved in the process. However, as mentioned before, kinematic situations occur where the smallness of the coupling is compensated by the largeness of logarithmic enhanced terms. In this scenario, truncating the perturbative expansion at any fixed order will not yield correct results. In order to produce reliable results, these logarithmically enhanced terms must be resummed to all-order. Depending on the types of logarithms to be resummed, there exists different categories of resummation. In the case of transverse momentum distribution–which is the observable we are mainly interested here and in the subsequent section–we have: (i) the *transverse momentum* (or *small-p_T*) *resummation* that resums logarithmic contributions of the form $\ln(p_T^2/Q^2)$ which are enhanced when $p_T^2/Q^2 \to 0$, (ii) the *soft/threshold* (or *large-x*) *resummation* that resums logarithmic terms that are enhanced when the invariant mass of the final system approaches the kinematic threshold, and (iii) the *high energy* (or *small-x*) *resummation* that resums logarithmic contributions arising when the center-of-mass energy of the system is large compared to the scale Q^2.

The following chapter starts by describing in a succinct way the kinematic regions that are relevant to the different classes of resummation mentioned previously while highlighting the need to combine these resummations in a consistent way. We then separately review the standard formulation of the threshold and transverse momentum resummations up to NNLL. This will allow us to introduce the concept of *soft-improved transverse momentum resummation* in which a subclass of the soft contributions are incorporated into the small-p_T resummation. For the purpose of the phenomenological studies that will be presented in the next chapter, explicit resummed expressions using this improved transverse momentum resummation are shown in App. 2.B in the context of the Higgs and Z (via DY mechanism) boson production. As it is not discussed here, the high energy resummation of the Higgs transverse momentum spectra will be discussed in the next chapter.

2.1 Which regions to resum?

Using the definitions given in Sec. (1.1.1), one can map the region of the phase space where resummations are required. In Fig. 5, we show the regions of the phase space where the three types of resummation mentioned earlier–namely transverse momentum, threshold, and high energy resummations–are relevant. Usually, the construction of the resummed expressions for each of these types of resummation are performed separately. In the past

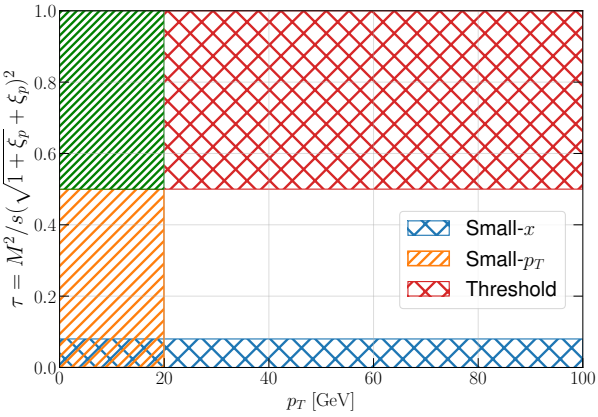


Figure 5 Phase space available for the production of a final-state system F with an invariant mass M in the three kinematic limits: threshold, collinear, and high energy. The hadronic variable τ is defined in terms of the center-of-mass energy $\sqrt{s}$ and the scale Q defined according to the conventions in Sec. (1.1.1).

few years, several attempts have been made to construct a resummed expression in which two kinematic limits are take into account. The formulation of a double large-x and small-x resummation , for instance, has been developed in Ref. [150] where the large-x logarithms are resummed up to N3LL while the small-x ones up to LL accuracy. Another example is the combination of the small-p_T and small-x resummations, which was formulated in Ref. [151] where the small-p_T logarithms are resummed up to NNLL accuracy. The combination of the small-p_T and the large-x resummations is one of the main subjects of the following chapter.

To the present day, the current state-of-the-art accuracy in the resummed calculations of transverse momentum distributions is N3LL matched with fixed order predictions (usually NNLO). This has been achieved for a variety of processes including the Higgs boson production [28, 152, 153] and DY [32, 154–158]. The motivation to include threshold resummed contributions to the transverse momentum resummation boils down to the fact that large logarithmic corrections present in transverse momentum and threshold resummations both originate from the emission of soft gluons. There have also been attempts to construct a joint formalism that simultaneously resum logarithmic contributions enhanced at small-p_T and at partonic threshold [159, 160]. Such joint resummation have been successful in producing phenomenological results at NLL accuracy for various processes including Higgs [161] and vector boson production via DY mechanism [162] which has been recently extended to NNLL [163]. While the aforementioned resummations were done in Fourier-Mellin space, joint resummation in direct space has also been achieved up to NNLL accuracy using SCET [164]. It was shown recently that, in the case of processes with colourless final-state, standard resummation of logarithms of p_T/Q can be consistently supplemented with the resummation og logarithmic contributions at large $x = Q/\hat{s}$ [165]. Such an improved resummation has the following features: first, it repro-

duces the correct behaviour to any logarithmic order in the limit $p_T \to 0$ for fixed x, and second, it reproduces the total cross section at any given logarithmic order in the threshold limit upon integration over the transverse momentum. Details on the construction of such a resummation is described in Sec. (2.4).

2.2 Transverse momentum resummation

As mentioned in the previous section, in a process involving multiple scales, the convergence of the perturbative series might be spoiled due to the presence of large logarithms multiplying the running of the coupling α_s. For transverse momentum distributions, such large logarithmic enhancements are of the form $\ln(\xi_p)/\xi_p$. This quantity indeed diverges as the transverse momentum $\xi_p = p_T^2/Q^2$ goes to zero. In order to obtain reliable prediction in this situation, the logarithms of $\ln(\xi_p)/\xi_p$ must be resummed through the transverse momentum (or small-p_T) resummation.

In the following section, we briefly review two formulations of the transverse momentum resummation for transverse momentum distributions, namely the Collin-Soper-Sterman (CSS) formalism and the universal transverse momentum resummation formalism by Catani-deFlorian-Grazzini (CFG).

2.2.1 The Collin-Soper-Sterman formulation

Let us first review the formulation of the transverse momentum resummation introduced by Collin, Soper, and Sterman [166]. Consider the Mellin space expression of the hadronic cross-section

$$\frac{d\sigma}{d\xi_p}\left(N, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) = \sum_{a,b} f_a\left(N+1, \mu_F^2\right) f_b\left(N+1, \mu_F^2\right) \frac{d\hat{\sigma}_{ab}}{d\xi_p}\left(N, \xi_p, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) \quad (2.2.1)$$

where for brevity we have omitted the scale Q^2 and the dependence of the coupling α_s on the renormalization scale μ_R^2. One should always keep in mind, unless stated otherwise, that the hadronic and partonic cross sections are perturbative series in the strong coupling. The all-order expansion of the hard part is given by

$$\frac{d\hat{\sigma}_{ab}}{d\xi_p}\left(N, \xi_p, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) = \sigma_{\mathrm{F}}^{\mathrm{Born}}\left(\alpha_s^p(\mu_R^2)\right) \sum_{n=1}^{\infty} \alpha_s^n(\mu_R^2)\, \Sigma_{ab}^{(n)}\left(N, \xi_p, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right), \quad (2.2.2)$$

where p denotes the lowest power of α_s needed in order for the process to occur and $\sigma_{\mathrm{F}}^{\mathrm{Born}}$ is the leading order Born cross section. Since the final state F is colourless, the lowest order partonic cross section is initiated either by gluons or $q\bar{q}$ annihilation. Later on, we will make explicit the dependence of $\sigma_{\mathrm{F}}^{\mathrm{Born}}$ on the incident partons.

At this stage, we can decompose the partonic cross section into a regular and singular part as the transverse momentum $p_T \to 0$ (or equivalently $\xi_p \to 0$):

$$\Sigma_{ab}^{(n)}\left(N, \xi_p, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) = \Sigma_{ab}^{\mathrm{R},(n)}\left(N, \xi_p, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) + \Sigma_{ab}^{\mathrm{S},(n)}\left(N, \xi_p, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) \quad (2.2.3)$$

where $\Sigma_{ab}^{\mathrm{R},(n)}$ contains all the terms that are less singular than $\delta(p_T^2)$ and p_T^2 (or equivalently ξ_p) in the limit $p_T \to 0$. Its hadronic expression truncated at a given order α_s^m in the perturbative expansion is given by

$$\frac{d\sigma}{d\xi_p}\left(N, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right) = \sigma_{\mathrm{F}}^{\mathrm{Born}}\left(\alpha_s^p(\mu_R^2)\right) \sum_{a,b} f_a\left(N+1, \mu_F^2\right) \times \tag{2.2.4}$$
$$f_b\left(N+1, \mu_F^2\right) \sum_{n=1}^{m} \alpha_s^n(\mu_R^2)\, \Sigma_{ab}^{\mathrm{R},(n)}\left(N, \xi_p, \frac{M^2}{\mu_R^2}, \frac{M^2}{\mu_F^2}\right).$$

This term corresponds exactly to the difference between the full fixed-order result and its asymptotic in the small-p_T limit. The asymptotic piece contains terms that are at least as singular as $1/\xi_p$ up to NNLO in the limit $\xi_p \to 0$.

The singular term $\Sigma_{ab}^{\mathrm{S},(n)}\left(N, \xi_p\right)$ contains the $\delta(p_T^2)$ term and all the logarithmic enhanced terms when $\xi_p \to 0$. These logarithmic terms are resummed to all-order in the strong coupling α_s:

$$\Sigma_{ab}^{\mathrm{S},(n)}\left(N, \xi_p\right) = \Sigma_{ab}^{\delta,(n)}\left(N\right)\delta(p_T^2) + \sum_{m=1}^{2n-1} \Sigma_{ab}^{(n,m)}\left(N\right) \frac{\ln^m(\xi_p)}{\xi_p}. \tag{2.2.5}$$

where the functions $\Sigma(N)$ are ξ_p-independent and therefore behave as constants when $\xi_p \to 0$. In order to fully take into account the kinematic constraints on the conservation of transverse momentum, it is advantageous to work in Fourier space where the factorization of the multiple gluon emission is performed more naturally. Denoting by b the impact parameter–defined as the conjugate variable of p_T in Fourier space–the resummed cross section in Fourier space can be written as

$$\frac{d\sigma^{\mathrm{css}}}{d\xi_p}\left(N, b\right) = \sigma_{\mathrm{F}}^{\mathrm{Born}}\left(\alpha_s^p(\mu_R^2)\right) \sum_{a,b} \sum_{n=1}^{m} \alpha_s^n(\mu_R^2) f_a\left(N+1, \mu_F^2\right) f_b\left(N+1, \mu_F^2\right) \tag{2.2.6}$$
$$\int_0^\infty \frac{d^2p_T}{M^2} \exp(-i\,\vec{b}\vec{p}_T) \left[\Sigma_{ab}^{\delta,(n)}\left(N\right)\delta(p_T^2) + \sum_{m=1}^{2n-1} \Sigma_{ab}^{(n,m)}\left(N\right) \frac{\ln^m(\xi_p)}{\xi_p}\right].$$

As mentioned previously, the above equations obeys the renormalization group equations. Eq. (2.2.6) can therefore be determined by solving an evolution equation [167]. Using the results presented in Refs. [166, 168–171], the full resummed cross section in Eq. (2.2.6) can be written in a more compact form:

$$\frac{d\sigma^{\mathrm{css}}}{d\xi_p}\left(N\right) = \sum_{a,b} \int_0^\infty d\hat{b}\frac{\hat{b}}{2} J_0\left(\hat{b}\sqrt{\xi_p}\right) f_a\left(N+1, \frac{C_1}{C_2\hat{b}}\right) f_b\left(N+1, \frac{C_1}{C_2\hat{b}}\right) \frac{d\hat{\sigma}_{ab}^{\mathrm{css}}}{d\xi_p}(N, \hat{b}). \tag{2.2.7}$$

The partonic cross section that embodies the short range interaction is given by

$$\frac{d\hat{\sigma}_{ab}^{\mathrm{css}}}{d\xi_p}(N, \hat{b}) = \sum_c \sigma_{\mathrm{F},c\bar{c}}^{\mathrm{Born}}\left(\alpha_s^p(\mu_R^2)\right) C_{ca}\left(N, \alpha_s\left(\frac{C_1}{C_2\hat{b}}\right)\right) C_{\bar{c}b}\left(N, \alpha_s\left(\frac{C_1}{C_2\hat{b}}\right)\right) S_c(\hat{b}) \tag{2.2.8}$$

where $J_0(x)$ is the zeroth order Bessel function of the first kind while $\hat{b} = bM$ represents the modified conjugate variable of ξ_p in Fourier space. The partonic function $d\hat{\sigma}_{ab}^{\mathrm{css}}/d\xi_p$

embodies the all-order resummation of logarithms of the form $\ln(\hat{b})$. It contains the coefficient functions C_{ca} and the Sudakov form factor S_c. Both the coefficient functions and the Sudakov exponent are process dependent.

The coefficient functions C_{ca} can be computed perturbatively. In Mellin space, they are given in terms of the following series expansion:

$$C_{ca}\left(N, \alpha_s(\mu^2)\right) = \delta_{ca} + \sum_{n=1}^{\infty} \alpha_s(\mu^2)\, C_{ca}^{(n)}(N). \tag{2.2.9}$$

These coefficient functions are sometimes referred to as collinear factor for the hadron h_a. They contain both the full splitting functions and the hard process-independent contribution coming from the finite part of the virtual loop corrections. The one and two-loop coefficient functions have been computed for various processes and various resummation schemes, see Refs. [167, 172–177] for reference. For instance, for the production of a Higgs boson via gluon fusion, the leading-order coefficient functions can be defined as

$$C_{gg}^{(1)}(N) = 0, \qquad C_{gq}^{(1)}(N) = C_{qg}^{(1)}(N) = \frac{C_F}{2\pi}\frac{1}{N+1}. \tag{2.2.10}$$

The Sudakov form factor which resums all logarithmic contributions can be shown to have the following integral representation [168, 169, 178, 179]

$$S_c(\hat{b}) = \exp\left\{ -\int_{C_3^2/\hat{b}^2}^{C_2^2} \frac{dq^2}{q^2}\left[A_c\left(\alpha_s(q^2)\right)\ln\left(\frac{C_2^2 M^2}{q^2}\right) + B_c\left(\alpha_s(q^2)\right)\right]\right\}. \tag{2.2.11}$$

The functions A_c and B_c are perturbative series in $\alpha_s(q^2)$. In particular, the function A_c which is related to the soft-radiation is process-independent while the function B_c which is related to the flavour conserving collinear radiation is process-dependent from the two-loop contribution. They are perturbatively given by

$$A_c(\alpha_s) = \sum_{n=1}^{\infty} \alpha_s^n A_c^{(n)}, \qquad B_c(\alpha_s) = \sum_{n=1}^{\infty} \alpha_s^n B_c^{(n)}. \tag{2.2.12}$$

The coefficients $A_c^{(n)}$ and $B_c^{(n)}$ can be computed either by solving the corresponding renormalization group equation or by comparing the expanded result to fixed-order computations. The three loop contributions which are required to compute NNLL accuracy have been computed both for the gluon [180] and quark [178, 179] initiated process. In this section, we only presents results up to two loops. At leading-order, the process-independent parts are given by:

$$A_c^{(1)} = \frac{C_c}{\pi}, \quad A_c^{(2)} = \frac{C_c}{2\pi}\mathrm{K}, \quad B_g^{(1)} = -2\beta_0, \quad B_q^{(1)} = -\frac{3}{2}\frac{C_F}{\pi^2}, \tag{2.2.13}$$

where $\mathrm{K} = C_A(67/18 - \zeta_2) - 5n_f/9$, $\beta_0 = (11C_A - 2n_f)/12\pi$, and C_c which is equal to C_A if $c = g$ and C_F if $c = q$, is the QCD colour factor. The process-dependent part $B_c^{(2)}$ is generally expressed as follows [181]

$$B_c^{(2)} = -2\delta P_{cc}^{(2)} + \beta_0\left[\frac{2}{3}\pi + \mathcal{A}_c^{\mathrm{F}}(\phi)\right], \tag{2.2.14}$$

where $\delta P_{cc}^{(2)}$ is the coefficient of the $\delta(1-x)$ term in the two-loop splitting functions. The finite part $\mathcal{A}_c^F$ comes from the one-loop virtual contributions, it depends on the kinematics of the final-state colour singlet particle. Its expression has been computed for both Higgs [172] and DY [182, 183] and they are given by

$$\mathcal{A}_g^{\rm H} = \frac{C_A}{\pi}\left(\frac{2}{3}\pi^2 - \frac{C_F}{C_A} + 5\right), \quad \mathcal{A}_q^{\rm DY} = \frac{C_F}{\pi}\left(\frac{2}{3}\pi^2 - 8\right). \tag{2.2.15}$$

As a result, the $B_c^{(2)}$-coefficient has been computed for Higgs [181] and DY [167, 181]:

$$\begin{aligned} B_g^{(2),{\rm H}} &= \left(\frac{C_A}{\pi}\right)^2\left(\frac{23}{6} + \frac{22}{9}\pi^2 - 6\zeta_3\right) \\ &\quad + \frac{4}{\pi^2}n_f C_F T_R - \frac{11}{2\pi^2}C_A C_F - \frac{C_A}{\pi^2}n_f T_R\left(\frac{2}{3} + \frac{8}{9}\pi^2\right), \end{aligned} \tag{2.2.16}$$

$$\begin{aligned} B_q^{(2),{\rm DY}} &= \left(\frac{C_F}{\pi}\right)^2\left(\pi^2 - \frac{3}{4} - 12\zeta_3\right) \\ &\quad + \frac{C_F}{\pi^2}T_R\left(\frac{17}{3} - \frac{4}{9}\pi^2\right) + \frac{C_A C_F}{\pi^2}\left(\frac{11}{9}\pi^2 - \frac{193}{12} + 6\zeta_3\right), \end{aligned} \tag{2.2.17}$$

where $T_R = 1/2$ is the $SU(3)$ QCD colour factor. The results for the coefficient $B_c^{(3)}$ can be found in Ref. [184].

We finally comment on the constants C_i that enter in Eq. (2.2.6) and (2.2.11) and that so far were left arbitrary. These constants arise from solving the renormalization group equations in which the two scales M and b are separated. The variation of these constants can also be used to estimate higher-order corrections. In the CSS formalism, they are chosen to be $C_1 = C_3 = b_0 M$ and $C_2 = 1$ where $b_0 = 2\exp(-\gamma_E)$ with γ_E denoting the Euler-Mascheroni constant.

The p_T distribution given by Eq. (2.2.6) presents some subtleties as it probes non perturbative effect for some particular values of b. Indeed, when b is large, specifically $b \gtrsim 1/\Lambda_{\rm QCD}$, the running of the coupling α_s and the PDFs which both explicitly depend on the scale b enter the non-perturbative regime. In order to prevent the scale b to probe such a regime, the $b_\star$ prescription is introduced where

$$b_\star = \frac{b}{\sqrt{1 + (b/b_{max})^2}}, \tag{2.2.18}$$

with b_{max} chosen such that $1/b_{max} \sim \Lambda_{\rm QCD}$. This modifies the lower bound of the Sudakov integral in Eq. (2.2.11) to $C_1^2/\hat{b}_\star$. This ensures that $d\hat{\sigma}_{ab}^{\rm css}/d\xi_p(b) \sim d\hat{\sigma}_{ab}^{\rm css}/d\xi_p(b_\star)$ when $b \lesssim b_{max}$. The non-perturbative dependence is then moved into an auxiliary function $d\hat{\sigma}_{ab}^{\rm NP}/d\xi_p$ defined as

$$\frac{d\hat{\sigma}_{ab}^{\rm css}}{d\xi_p}(N,\hat{b}) = \frac{d\hat{\sigma}_{ab}^{\rm css}}{d\xi_p}(N,\hat{b}_\star) \times \frac{d\hat{\sigma}_{ab}^{\rm NP}}{d\xi_p}(N,\hat{b}). \tag{2.2.19}$$

This ensures that the non-perturbative behaviour is not part of the resummed expression. The function $d\hat{\sigma}_{ab}^{\rm NP}/d\xi_p$ is chosen such that $d\hat{\sigma}_{ab}^{\rm NP}/d\xi_p \to 1$ when $\hat{b} \to 0$. An example of

such a non-perturbative function is given in Refs. [166,168,185,186]. The correction factor $d\hat{\sigma}_{ab}^{\rm NP}/d\xi_p$ generally takes the form $\exp(-g(\hat{b}))$ where $g(x)$ is just some polynomial. Since the large $\hat{b}$ region is out of control of perturbation theory, the parameters that enter $g(x)$ needs to be adjusted to the experimental data. For DY, different forms of the correction factor are available from global fits [167,185–188].

Some comments are in order concerning the disadvantages of the CSS resummation formalism. Indeed, despite the fact that the formalism has been used to produce theoretical predictions at various level of perturbative accuracies, it presents some drawbacks:

- First, the behaviour of the soft gluons that is captured by the Sudakov form factor should be independent of the hard process. Indeed, the Sudakov exponent is supposed to be universal and should only depend on the initial state partons (either quark or gluon). However, as wee see in Eqs. (2.2.14, 2.2.16, 2.2.17), the Sudakov form factor in Eq. (2.2.11) is process-dependent through B_c. The same can also be said about the coefficient functions C_{ca}.
- Second, the fact that the parton distributions in Eq. (2.2.7) is parameterized at the scale b_0/b rather than μ_F^2 raises some complications. With such a value of scale, the PDFs will have to be interpolated in the non-perturbative region.

The above issues can be tackled by performing resummation in momentum (or direct) space. However, despite the fact that the problem of the resummation of the transverse momentum distribution in direct space has received considerable attention in the past few years [189–192], it has not been possible to find beyond leading logarithmic accuracy a closed analytic expression which is simultaneously free of subleading logarithmic contributions and spurious singularities at finite values of p_T until very recently [152,193]. An alternative approach consists in re-arranging the terms in the impact parameter space $\hat{b}$ using the renormalization group identity, this is known as the universal transverse momentum resummation formalism.

2.2.2 The Catani-deFlorian-Grazzini formulation

In the Catani-deFlorian-Grazzini (CFG) formalism, the issue related to the universality of the Sudakov form factor and the coefficient functions is addressed factorizing out the process-independent term into a hard function $\mathrm{H}_c^{\rm F}$ where the superscript F is used to indicate that the hard function depends on the kinematics of the final-state system F. This is achieved by expressing the C_{ca}-functions and the B_c-functions as [194]

$$C_{ca}^{\rm CSS}(N,\alpha_s) = \sqrt{\mathrm{H}_c^{\rm F}(\alpha_s)} \times C_{ca}^{\rm CFG}(N,\alpha_s) \tag{2.2.20}$$

$$B_c^{\rm CSS}(N,\alpha_s) = B_c^{\rm CFG}(N,\alpha_s) - \beta(\alpha_s)\frac{d\ln \mathrm{H}_c^{\rm F}(\alpha_s)}{d\ln\alpha_s}, \tag{2.2.21}$$

where $\beta(\alpha_s)$ is the ACD β-function that describes the running of the coupling α_s. Notice that one recovers the CSS expressions by setting $\mathrm{H}_c^{\rm F}=1$. The hard function satisfies the renormalization group identity:

$$\mathrm{H}_c^{\rm F}(\alpha_s(M^2)) = \mathrm{R}_c\left(\alpha_s\left(M^2 \leftarrow \frac{b_0^2}{b^2}\right)\right)\mathrm{H}_c^{\rm F}\left(\alpha_s\left(\frac{b_0^2}{b^2}\right)\right) \tag{2.2.22}$$

where the evolution factor R_c is defined as

$$\mathrm{R}_c\left(\alpha_s\left(M^2 \leftarrow \frac{b_0^2}{b^2}\right)\right) = \exp\left\{-\int_{b_0^2/b^2}^{M^2} \frac{dq^2}{q^2}\beta\left(\alpha_s\left(q^2\right)\right)\frac{d\ln \mathrm{H}_c^F\left(\alpha_s\left(q^2\right)\right)}{d\ln \alpha_s\left(q^2\right)}\right\}. \quad (2.2.23)$$

The hard function is a perturbative function and therefore can be expressed as a series expansion in α_s as $\mathrm{H}_c^{\mathrm{F}}(\alpha_s) = 1 + \sum_{n=1}^{\infty} \alpha_s^n \, \mathrm{H}_c^{\mathrm{F},(n)}$. The expressions of $\mathrm{H}_c^{\mathrm{F},(n)}$ up to $\mathcal{O}(\alpha_s^3)$ are explicitly given in Refs. [195–199] for Higgs and in Refs. [200, 201] for DY.

Thanks to the introduction of the hard function $\mathrm{H}_c^{\mathrm{F}}$, the coefficient functions C and the Sudakov form factor S_c are now universal and depend only on the the flavour and colour charges of the radiating partons. The resummed expression can now be organized in terms of process-independent contribution in which the logarithmic contributions are controlled by a set of perturbative functions. The order at which these perturbative functions are known determine the accuracy at which the resummed formulae is performed.

The issue related to the parametrization of the PDFs at a scale $\hat{b}^2/b_0^2$ can be addressed by substituting the parton densities $f_a\left(N+1, b_0^2/b^2\right)$ with the same parton densities but evaluated at μ_F^2. This is done using the scale dependence relation [202–205]

$$f_a\left(N+1, \frac{b_0^2}{b^2}\right) = \sum_b U_{ab}\left(N, \frac{b_0^2}{b^2} \leftarrow \mu_F^2\right) f_b\left(N+1, \mu_F^2\right). \quad (2.2.24)$$

where U_{ab} is a component of the evolution operator matrix obtained by solving the DGLAP equation to the required accuracy. It satisfies the following evolution equation:

$$\frac{\partial U_{ab}}{\partial \ln \mu}(N, \mu^2 \leftarrow \mu_0^2) = \sum_c \gamma_{ac}(N, \alpha_s(\mu^2))\, U_{cb}(N, \mu^2 \leftarrow \mu_0^2) \quad (2.2.25)$$

with γ_{ac} denoting the usual Mellin moments of the Altarelli-Parisi splitting functions.

Henceforth, we will omit the superscript CFG that denotes the process-dependent quantity as given by the replacements in Eqs.(2.2.20, 2.2.21) for simplicity. Using those same equations, Eqs. (2.2.20, 2.2.21), the resummed expression in Eq. (2.2.7) now becomes

$$\frac{d\sigma^{\mathrm{cfg}}}{d\xi_p}(N) = \sum_{a,b}\int_0^{\infty} d\hat{b}\frac{\hat{b}}{2} J_0\left(\hat{b}\sqrt{\xi_p}\right) f_a\left(N+1, \mu_F^2\right) f_b\left(N+1, \mu_F^2\right) \frac{d\hat{\sigma}_{ab}^{\mathrm{cfg}}}{d\xi_p}(N, \hat{b}) \quad (2.2.26)$$

where now the partonic part is given by:

$$\frac{d\hat{\sigma}_{ab}^{\mathrm{cfg}}}{d\xi_p}(N, b) = \sum_c \sigma_{\mathrm{F},c\bar{c}}^{\mathrm{Born}}\, \mathrm{H}_c^{\mathrm{F}}(\alpha_s(M^2)) S_c(\hat{b}) \sum_{k,l} C_{ck}\left(N, \alpha_s\left(\frac{b_0^2}{b^2}\right)\right) \times \quad (2.2.27)$$
$$C_{\bar{c}l}\left(N, \alpha_s\left(\frac{b_0^2}{b^2}\right)\right) U_{ka}\left(N, \frac{b_0^2}{b^2} \leftarrow \mu_F^2\right) U_{lb}\left(N, \frac{b_0^2}{b^2} \leftarrow \mu_F^2\right).$$

The above expression can be written in a more compact form by introducing a perturbative function $\mathcal{H}^{\mathrm{F}}$ that collects all the coefficient functions and the evolution operators,

$$\mathcal{H}_{ab\to c\bar{c}}^{\mathrm{F}}(N, \alpha_s) = \mathrm{H}_c^{\mathrm{F}}(\alpha_s(M^2)) \sum_{k,l} C_{ck}\left(N, \alpha_s\left(\frac{b_0^2}{b^2}\right)\right) \times \quad (2.2.28)$$

$$C_{\bar{c}l}\left(N,\alpha_s\left(\frac{b_0^2}{b^2}\right)\right)U_{ka}\left(N,\frac{b_0^2}{b^2}\leftarrow\mu_F^2\right)U_{lb}\left(N,\frac{b_0^2}{b^2}\leftarrow\mu_F^2\right).$$

The notation $ab\to c\bar{c}$ is a shorthand for the subprocess $ab\to c\bar{c}\to \mathrm{F}+X$. By inspection of the right-hand side of Eq. (2.2.28), one notices that the scale at which the running of the coupling α_s is computed is not set to a unique value. While the coefficient functions C_{ab} are evaluated at $\alpha_s(b_0^2/b^2)$, the α_s in $\mathrm{H}_c^{\mathrm{F}}$ is evaluated at M^2 which reflects the process-dependent nature of the hard function. The presence of two different arguments of α_s can be attributed to the physical origin of the various perturbative functions [166, 196, 206]. As mentioned previously, the hard function $\mathrm{H}_c^{\mathrm{F}}$ contains contributions from the hard momentum region of the virtual corrections to the lowest subprocess $c\bar{c}\to\mathrm{F}$ while the coefficient function C_{ca} refer to the subprocess $ca\to\mathrm{F}+X$ where the momenta of the partons in the final-state system X are almost collinear to the the momentum of the initial-state parton a. The function $\mathcal{H}^{\mathrm{F}}$ can therefore be regarded as the hard-collinear partonic function. The b-dependence of the perturbative function $\mathcal{H}^{\mathrm{F}}$ can be factorized out by computing the evolution operators and the coefficient functions at the renormalization scale. The difference has to be compensated by an evolution factor that evolves the C and U-functions from b_0^2/b^2 to μ_R^2. Details of such a procedure is given in details in Sec. (2.4.3). Now that the perturbative hard-collinear function is independent of the impact parameter b, it only embodies terms that behave as constants when $b\to\infty$ (or equivalently $p_T\to 0$). Therefore, it can be computed perturbatively,

$$\mathcal{H}^{\mathrm{F}}_{ab\to c\bar{c}}\left(N,\alpha_s(\mu_R^2),\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_F^2}\right)=\delta_{ac}\delta_{b\bar{c}}+\sum_{n=1}^{\infty}\alpha_s^n(\mu_R^2)\mathcal{H}^{\mathrm{F},(n)}_{ab\to c\bar{c}}\left(N,\frac{M^2}{\mu_R^2},\frac{M^2}{\mu_F^2}\right). \quad (2.2.29)$$

As opposed to the Sudakov form factor, the perturbative function $\mathcal{H}^{\mathrm{F}}_{ab\to c\bar{c}}$ depends on the factorization scheme through the coefficient functions, the anomalous dimensions, and on the final state system F through the hard function $\mathrm{H}_c^{\mathrm{F}}$. In addition, the hard function $\mathrm{H}_c^{\mathrm{F}}$ and the coefficient functions C_{ca} are resummation scheme dependent. The resummation scheme just amounts to unambiguously define the two functions $\mathrm{H}_c^{\mathrm{F}}$ and C_{ca}. However, it can be proved using the renormalization group identity that the such dependence is cancelled at the level of $\mathrm{H}_c^{\mathrm{F}}\times C_{ca}$ therefore making the hard-collinear function $\mathrm{H}_c^{\mathrm{F}}$ resummation scheme independent.

Putting everything together, we can now write down in a compact form the expression of the partonic resummed cross section

$$\frac{d\hat{\sigma}^{\mathrm{cfg}}_{ab}}{d\xi_p}(N,\hat{b})=\sum_c\sigma^{\mathrm{Born}}_{\mathrm{F},c\bar{c}}\mathcal{H}^{\mathrm{F},\{\mathrm{S}\}}_{ab\to c\bar{c}}(N,\alpha_s(\mu_R^2),\mu_F^2)\exp\left(\mathcal{S}_c(N,\lambda_{\hat{b}})+\mathcal{R}_{\{\mathrm{S}\}}(N,\lambda_{\hat{b}})\right), \quad (2.2.30)$$

where $\lambda_{\hat{b}}=\bar{\alpha}_s\ln(\hat{b}^2/b_0^2)$ with $\bar{\alpha}_s=\alpha_s\beta_0$, $\{\mathrm{S}\}$ labels a set of flavour indices, $\mathcal{S}_c=\ln(S_c)$ is the universal Sudakov form factor, and $\mathcal{R}$ contains all the evolution factors coming from the coefficient functions C and the evolution operators U. The terms in the exponent contain all the dependence in $\hat{b}$, and in particular, embodies to all-order in α_s the logarithmic divergent terms when $\hat{b}\to\infty$. In Ref. [195], the resummation variable is expressed in terms of $\lambda_{\tilde{b}}=\bar{\alpha}_s\ln(\tilde{b}^2/b_0^2)$ instead where the $\tilde{b}=bQ$. Indeed, the argument of the logarithms can always be rescaled as $\ln(\hat{b}^2)=\ln(\tilde{b}^2)+\ln(\hat{b}^2/\tilde{b}^2)$ provided that Q is independent

of $\hat{b}$ and $\ln(\hat{b}^2/\tilde{b}^2) \sim \mathcal{O}(1)$. This takes into account the degree of arbitrariness in the factorization [207]. Since physical cross sections should not depend on the resummation scale Q, its value should be set to M. However, varying Q around this central value should provide an estimate of yet unknown logarithmic corrections at higher-orders. Throughout this thesis, we always choose $Q=M$ and hence $\hat{b}=\tilde{b}$.

The universal Sudakov exponent $\mathcal{S}_c$ has the following logarithmic expansion

$$\mathcal{S}_c(N,\lambda_{\hat{b}}) = \sum_{n=1}^{\infty} \bar{\alpha}_s^{n-2} g_n(N,\lambda_{\hat{b}}), \qquad \text{where} \qquad \bar{\alpha}_s \equiv \bar{\alpha}_s(\mu_R^2). \tag{2.2.31}$$

The g_i functions categorize the logarithms into classes and they should be defined in such a way that $g_i = 0$ when $\lambda_{\hat{b}} = 0$. The first term $\bar{\alpha}_s^{-1} g_1$ contains LL contributions that are enhanced as $\alpha_s^n \ln^{n+1}(\hat{b})$, the NLL contributions $\alpha_s^n \ln^n(\hat{b})$ are contained in g_2, the NNLL contributions $\alpha_s^n \ln^{n-1}(\hat{b})$ are contained in $\alpha_s g_3$, and so forth. By inspecting the right-hand side of Eq. (2.2.31), we notice that unphysical behaviour appears when $\hat{b} \to 0$ which corresponds to the large-p_T region. This is expected since the resummation formulae only assures correct predictions in the region when p_T is small. However, as suggested in Ref. [195], the impact of the unjustified contributions can be reduced by shifting the argument of the logarithms by 1,

$$\lambda_{\hat{b}} = \ln\left(\frac{\hat{b}^2}{b_0^2}\right) \longrightarrow \tilde{\lambda}_{\hat{b}} = \ln\left(1+\frac{\hat{b}^2}{b_0^2}\right). \tag{2.2.32}$$

In the resummation region $\hat{b} \gg 1$, the two definitions of resummation variable coincide since we have $\tilde{\lambda}_{\hat{b}} = \lambda_{\hat{b}} + \mathcal{O}(1/\hat{b}^2)$ and thus the replacement is fully justified. However, in the region $\hat{b} \ll 1$, the two definitions exhibit different behaviours, indeed, as $\tilde{\lambda}_{\hat{b}} \to 0$ the exponent $\exp(\mathcal{S}_c) \to 1$ as opposed to being divergent in the case if $\lambda_{\hat{b}}$ was chosen. In particular, in the case $b = 0$ which corresponds to the total inclusive cross section, we see that the integral over ξ_p of the resummed p_T-distribution is given by

$$\frac{d\hat{\sigma}_{ab}^{\text{cfg}}}{d\xi_p}(N,\hat{b}=0) = \int_0^\infty d\xi_p \frac{d\hat{\sigma}_{ab}^{\text{cfg}}}{d\xi_p}(N,\xi_p) = \sigma_{\text{F}}^{\text{Born}} \mathcal{H}_{ab}^{\text{F}}(N). \tag{2.2.33}$$

When matched to fixed-order calculations, this puts constraints on the total inclusive cross section [195], i.e. upon integration over ξ_p the resummed distribution at NLL+LO (or respectively NNLL+NLO) exactly reproduces the total cross section at NLO (or respectively NNLO) [1].

Upon performing the replacement $\lambda_{\hat{b}} \to \tilde{\lambda}_{\hat{b}}$ (henceforth, fo brevity, $\lambda_{\hat{b}}$ will always refer to $\tilde{\lambda}_{\hat{b}}$), the resummed expression in Eq. (2.2.30) can be truncated at a given logarithmic order. At LL accuracy, the function g_1 in the Sudakov exponent $\mathcal{S}_c$ is included and the hard-collinear function $\mathcal{H}^{\text{F}}$ is approximated by the Born cross section σ^{LO}. At NLLL, we include both g_1 and g_2 in the exponent $\mathcal{S}_c$ and the coefficient $\mathcal{H}^{\text{F},(1)}$ from the hard contribution. At NNLL, the function g_3 is also included along with $\mathcal{H}^{\text{F},(2)}$. The evolution factor $\mathcal{R}$ only start to contribute from NLL. This is summarized in Table. 1 where the logarithmic accuracy is also explicitly shown.

1 Notice that the perturbative accuracy is counted in different ways at the level of the p_T-distribution and the total inclusive cross section

Approx.	g_i up to	$\mathcal{H}^{\rm F}$ up to	$\mathcal{R}$ up to	$\bar{\alpha}_s \ln^k(1+\hat{b}^2/b_0^2)$
LL	g_1	1	-	$k=2n$
NLL	g_2	$\mathcal{H}^{\rm F,(1)}$	h_1	$2n-2 \leq k \leq 2n$
NNLL	g_3	$\mathcal{H}^{\rm F,(2)}$	h_2	$2n-4 \leq k \leq 2n$

Table 1 Orders of logarithmic approximations and accuracy of the considered logarithms up to NNLL. The last column represents the predicted power of the logarithms upon expansion of the resummed formulae as a series in α_s. In a similar way as for the Sudakov exponent $\mathcal{S}_c$, the logarithmic contributions coming from the evolution factor can also be organized in terms of classes.

2.3 Threshold resummation

We have seen in the previous section that under certain circumstances, the smallness of the strong coupling α_s does not justify the usage of perturbation theory as the presence of large logarithms ($\alpha_s \ln(\mu) \sim \mathcal{O}(1)$) may spoil the convergence. Another example of such a situation is when the center of mass energy $\sqrt{s}$ approaches its minimum or *threshold* (minimum here means the smallest amount of energy needed to produce the final state F) leading to the phase space for gluon bremsstrahlung to vanish. In our definitions of kinematics variables, this corresponds to $\sqrt{s}=M$ and $\tau=1$ (or in terms of its partonic version, $x \to 1$). To put this into perspective, consider the inclusive production of Higgs boson via gluon fusion in HEFT [182]

$$\hat{\sigma}^{\rm NLO}_{gg\to \rm H}(x) = \sigma^{\rm Born}_{\rm F} \times \left(\frac{\alpha_s}{\pi}\right) \left\{ \frac{C_A}{3\pi}\delta(-x) \right. \tag{2.3.1}$$
$$\left. +4C_A \left(x^2-\tau+1\right)^2 \left(\frac{\ln(1-x)}{1-x}\right) - 2C_A \frac{\left(x^2-x+1\right)^2}{(1-x)_+}\ln(x) - \frac{11}{6}C_A(1-x)^3 \right\}$$

where for brevety we chose $\mu_r = \mu_f = m_H = M$. We see that Eq. (2.3.1) exhibits several distributions that are singular at threshold $x \to 1$ which spoils the convergence of the perturbative series even if we are well in the perturbative regime where $\alpha_s \ll 1$. It can be shown that such logarithms appear at all-order in the perturbative series and generally can be written as

$$\sum_{n=0}^{\infty}\sum_{k=1}^{2n} \alpha_s^n \, C^{(n,k)} \times \left(\frac{\ln^{2n-k}(1-x)}{1-x}\right)_+, \tag{2.3.2}$$

where $C^{(n,k)}$ just represent the perturbative coefficients. One can notice from Eq. (2.3.2) that for each n-th order in the perturbative expansion there is a *double-logarithmic* structure, i.e. for every new order in the running of the coupling there is two power of logarithms arising. Similar to the case for transverse momentum resummation, these logarithms must be resummed in order to produce reliable predictions.

Threshold resummation is usually performed by factorizing the hadronic cross-section in Mellin space in which the conjugate transform of the appropriate kinematic variable (x throughout this thesis) is taken [180,208–210]. Upon performing such a transformation, the

threshold logarithms is mapped into logarithms of the conjugate variable $\bar{N} = N\exp(\gamma_E)$ where again the γ_E denotes the Euler constant.

As common to all resummation formalisms, threshold resummation relies on two main ingredients, namely the factorization of the phase space and the matrix element, and the exponentiation of the logarithmically enhanced terms to the latter either through a factorization [180] or eikonal [208] approaches. The NLL threshold resummation has been computed since long time ago both for colour-singlet and non-colour-singlet in the final state. For color-singlet observables, NNLL results are available for Higgs [211–213] and more recently for DY [214]. With the recent computations of up to three-loop of all threshold distributions, resummed results for up to $\mathrm{N^3LL}$ have become available [215,216]. At present, no full NNLL threshold resummation is available for processes characterized by non-colour-singlet final state due to the complexity of the colour structure. Nevertheless, progress has been made toward constructing threshold resummations beyond NLL for coloured observables. So far, resummation studies beyond NLL have been developed for single-particle inclusive hadroproduction at high transverse momentum [217], top-pair production [218], and squark and gluino production [219].

In the following parts, we sketch a simplified picture of the derivation of a threshold resummed expression using the renormalization group argument [213]. From the derivation, a resummed formulae for a generic colour-singlet final state will be presented up to NNLL. Finally, explicit expressions for Higgs and DY–which are the two processes of interest in this thesis–will be given.

2.3.1 Resummation from renormalization-group evolution

For the sake of sketching our argument toward a derivation of a resummed formula using the renormalization group argument, let us consider a generic DIS-like observable $\mathcal{O}$ which in Mellin space is defined as just the product of a partonic function and a parton distribution function,

$$\mathcal{O}(N, Q^2) = \hat{\mathcal{O}}\left(N, \alpha_s(\mu^2), \frac{Q^2}{\mu^2}\right) f(N, \mu^2), \tag{2.3.3}$$

where for brevety we chose $\mu^2 = \mu_R^2 = \mu_F^2$. The standard expression of the renormalization-group improved of the factorized observable $\mathcal{O}$ is derived by setting $\mu^2 = Q^2$ on the right-hand side of Eq. (2.3.3) yielding

$$\mathcal{O}(N, Q^2) = \hat{\mathcal{O}}\left(N, \alpha_s(\mu^2), 1\right) f(N, Q^2). \tag{2.3.4}$$

By introducing what is known as the *physical anomalous dimension*, the observable $\mathcal{O}$ can be written in a derivative form. Such a form is convenient since resummed expressions take the form of an exponent,

$$Q^2 \frac{\partial \mathcal{O}(N, Q^2)}{\partial \mathcal{O}} = \gamma_{\mathrm{AD}}\left(N, \alpha_s(Q^2)\right) \mathcal{O}(N, Q^2), \tag{2.3.5}$$

where the physical anomalous dimension γ_{AD} is related to the standard Altareli-Parisi anomalous dimensions γ through the relation

$$\gamma_{\mathrm{AD}}\left(N, \alpha_s(Q^2)\right) = \gamma\left(N, \alpha_s(Q^2)\right) + \frac{\partial \ln \mathcal{O}\left(N, \alpha_s(Q^2), 1\right)}{\partial \ln Q^2} \tag{2.3.6}$$

with again the Altareli-Parisi anomalous dimension γ satisfying the following renormalization group identity:

$$\mu^2 \frac{\partial f(N,\mu^2)}{\partial \mu^2} = \gamma\left(N, \alpha_s(\mu^2)\right) f(N,\mu^2). \tag{2.3.7}$$

By solving the renormalization group equation in Eq. (2.3.4), one resums all large logarithmic contributions to $\mathcal{O}$. Whilst solving renormalization group equations can be straightforward for regular functions such as hard functions, this is no longer the case when $\mathcal{O}$ (or specifically its partonic version $\hat{\mathcal{O}}$) contains distributions. This is mostly the case in our scenario where the observable in question is given by the transverse momentum spectrum. Let us shortly comment on the relationship between the physical and standard anomalous dimensions. According to Eq. (2.3.6), the two definitions coincide at LO since the second term starts at order α_s^2, from α_s^2 the physical anomalous dimension receives both contribution from γ and $\mathcal{O}$.

By solving the renormalization group equation in Eq. (2.3.5) we have

$$\mathcal{O}(N,Q^2) = \mathcal{E}(N,Q^2 \leftarrow Q_0^2)\mathcal{O}(N,Q_0^2), \tag{2.3.8}$$

where $\mathcal{E}$ is known as the evolution kernel with the constraint $\mathcal{E}(N,Q_0^2 \leftarrow Q_0^2) = 1$ and the function $\mathcal{O}(N,Q_0^2)$ can be thought as a boundary condition. The evolution kernel takes into account the evolution from scale Q_0^2 to Q^2. It is expressed in an exponential form as follows

$$\mathcal{E}(N,Q^2 \leftarrow Q_0^2) = \exp\left\{\int_{Q_0^2}^{Q^2} \frac{dq^2}{q^2} \gamma_{\rm AD}(N,\alpha_s(q^2))\right\}. \tag{2.3.9}$$

Eq. (2.3.9) exponentiates all logarithmic contributions and hence predicts their all-order behaviour in $\mathcal{O}(N,Q^2)$. For DIS-like kinematics, the evolution factor up to NLL can be written as follows [142, 200, 220–223]

$$\mathcal{E}(N,Q^2 \leftarrow Q_0^2) = \int_0^1 dx \frac{x^{N-1}-1}{1-x} \int_{Q_0^2(1-x)}^{Q^2(1-x)} \frac{dq^2}{q^2} \hat{g}(\alpha_s(q^2)) + \mathcal{K}, \tag{2.3.10}$$

where $\mathcal{K}$ just denotes the non-logarithmic terms and the function $\hat{g}$ is expressed as a series expansion in the running of coupling which up to NLL is expressed as

$$\hat{g}(\alpha_s(q^2)) = \sum_{n=1}^{2} \alpha_s^n(q^2)\hat{g}_n. \tag{2.3.11}$$

By relating Eq. (2.3.10) and Eq. (2.3.10) using the definition of the $\hat{g}$ function given by Eq. (2.3.11), we can express the physical anomalous dimension as follows:

$$\gamma_{\rm AD}(N,\alpha_s(q^2)) = \int_0^1 dx \frac{x^{N-1}-1}{1-x} \sum_{n=1}^{2} \alpha_s^n\left(q^2(1-x)\right)\hat{g}_n, \tag{2.3.12}$$

If we let the coupling α_s run in Eq. (2.3.12),

$$\alpha_s\left(q^2(1-x)\right) = \sum_{n=1}^{\infty} \alpha_s(q^2)\hat{f}_n\left(\alpha_s(q^2)\ln(1-x)\right), \tag{2.3.13}$$

where $\hat{f}_n$ are just analytical functions in their arguments (see App. 1.A for details), we realize that the terms that are going to be integrated are of the form $\ln(1-x)$. These are exactly the terms that we are interested in as they are the ones that are enhanced at threshold $x \to 1$. In Mellin space, these logarithms translate into logarithms of the Mellin moment N, specifically, the region $x \sim 1$ is mapped onto the large-N region. This can be seen by writing the integral in Eq. (2.3.12) as

$$\int_0^1 d\tau \frac{x^{N-1}-1}{1-x} \ln^n(1-x) = \frac{1}{n+1}\left(\ln\frac{1}{N} - \gamma_E\right)^{n+1}. \tag{2.3.14}$$

From Eq. (2.3.14), it follows that Eq. (2.3.12) now becomes

$$\gamma_{\rm AD}(N, \alpha_s(q^2)) = \sum_{n=1}^{2} \alpha_s^n(q^2)\, \gamma_{\rm AD}^{(n+1)} \left(\alpha_s(q^2)\ln N\right) + \mathcal{O}\left(\alpha_s^{k+2}(q^2)\ln^k N\right) + \mathcal{K}, \tag{2.3.15}$$

where k resums to all-order α_s, i.e. k runs from 1 to infinity. Notice also that the series has been truncated to order α_s as these are the only terms relevant for NLL resummed expression. As shown in Eq. (2.3.15), the physical anomalous dimension admits a series expansion in α_s for fixed $\alpha_s^m \ln^m N$. That is, the LL refers to the contribution where the coefficient of the physical anomalous dimension $\gamma_{\rm AD}^{(1)}$ is the sum of terms (all-order in α_s) where the powers of α_s and $\ln N$ are the same. Therefore, the NLL embodies sum of terms to all-order in α_s where α_s has one more power w.r.t. $\ln N$.

As mentioned before, beyond leading order, the physical anomalous dimension differs from the standard one as the former receives an extra-contribution from the coefficient function as given by Eq. (2.3.6). In order to separate the contribution that originates from the Altareli-Parisi anomalous dimension, we can take a form analogue to Eq. (2.3.6) to write the evolution kernel as

$$\ln \mathcal{E}(N, Q^2) = \int_0^1 dx \frac{x^{N-1}-1}{1-x} \left\{ \int_{Q_0^2(1-x)}^{Q^2(1-x)} \frac{dq^2}{q^2} A(\alpha_s(q^2)) + \left[B\left(\eta^2(1-x)\right)\right]_{\eta=Q_0}^{\eta=Q} \right\} \tag{2.3.16}$$

where both up to NLL the perturbative functions are defined as $A(\alpha_s) = \alpha_s A^{(1)} + \alpha_s^2 A^{(2)}$ and $B(\alpha_s) = \alpha_s B^{(1)}$. The coefficients of α_s and α_s^2 represent the one and two-loop coefficients of $\ln N$ in the standard anomalous dimension respectively. In Eq. (2.3.16), the coefficient B accounts for the difference between the physical and the standard anomalous dimensions. It is therefore defined to be,

$$\hat{g}(\alpha_s(q^2)) = A(\alpha_s(q^2)) + \frac{\partial B(\alpha_s(q^2))}{\partial \ln(q^2)}. \tag{2.3.17}$$

By putting back the Q^2 dependence in the partonic component $\mathcal{O}$, it is now possible to rewrite the resummed formulae for the factorized observable in Eq. (2.3.3) as

$$\mathcal{O}^{\rm res}(N, Q^2) = \hat{\mathcal{O}}^{\rm res}\left(N, \alpha_s(\mu^2), \frac{Q^2}{\mu^2}\right) f(N, \mu^2), \tag{2.3.18}$$

where the partonic resummed component is given by

$$\ln \hat{\mathcal{O}}^{\text{res}}\left(N, \alpha_s(\mu^2), \frac{Q^2}{\mu^2}\right) = \int_0^1 dx \frac{x^{N-1}-1}{1-x}\left\{\int_{\mu^2}^{Q^2(1-x)} \frac{dq^2}{q^2} A(\alpha_s(q^2)) + B\left(Q^2(1-x)\right)\right\}. \tag{2.3.19}$$

Depending on the resummation scheme, the resummed component can be part of either the partonic function $\mathcal{O}$ or the evolution of the parton density f. The function $A(\alpha_s)$ embodies contributions from the emission that are both soft and collinear, in our case, it only resums logarithms of the form $\alpha_s^k \ln^{k+1} N$ and $\alpha_s^k \ln^k N$. The function $B(\alpha_s)$ contains corrections from the emissions that are only soft. The logarithmic counting is as follows. At LL, only $A^{(1)}$ contribute while at NLL one has to also include $A^{(2)}$ and $B^{(1)}$. The coefficients of these perturbative functions can be obtained by comparing fixed-order computations of physical anomalous dimensions with the fixed-order order expansion of Eq. (2.3.19) as described in Ref. [224].

2.3.2 Beyond NLL resummation for colour singlet observables

Since the factorization of the hadronic cross-section in Mellin space is understood from previous sections, here and in the following, we will be mainly working at the level of partonic partonic. From Refs. [173, 213], it was shown that a natural choice to perform the resummation is $\tilde{Q} = Qp_T{}^2$. Since the general form of the threshold resummed formulae is very well known, thanks to the exponentiation of the logarithmic behaviours [225, 226], here we simply give the expression [227, 228] without detailing the derivations

$$\begin{aligned}\frac{d\hat{\sigma}_{ab}^{\text{th}}}{d\xi_p}\left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) &= C_{abc}^{(1)}(N) g_{ab}\left(\alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) \Delta_a\left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) \\ &\Delta_b\left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) J_c\left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) \Delta_{abc}\left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right)\end{aligned} \tag{2.3.20}$$

where the subscript abc is just a shorthand for $ab \to c\text{F}$. For brevity, we have also omitted all the ξ_p-dependence. Notice that each radiative factor Δ and J is expressed in terms of an exponential. The radiative factor Δ_a represents the soft and collinear gluon radiation associated to the initial parton a, the factor J_c takes into account the collinear emission from the unobserved parton c that recoils against the final state F, and finally the function Δ_{abc} accounts for the large-angle soft gluon emission. While the functions Δ_a and J_c are universal, the function Δ_{abc} is process-dependent. The N-independent g_{ab} coefficient is known as the *matching coefficient*, it ensures that at every order, the resummed cross section reproduces the exact fixed-order results, up to corrections suppressed by $\mathcal{O}(1-\tau)$. Although it is not apparent from the indices, the definition of the function g_{ab} depends on the nature of the process. Finally, the function $C_{abc}^{(1)}$ embodies the LO ξ_p distribution, and therefore is process-dependent. It can be shown that in the large-N limit, this term behaves as $1/\sqrt{N}$. In addition, $C_{abc}^{(1)}$ only exhibits soft-behaviour (i.e. non-soft for fixed-ξ_p)

2 In order to relate Ref. [213] with Ref. [213], one indeed need to modify Q into Qp_T.

in the small-ξ_p limit where it behaves as $(\ln \xi_p/\xi_p)_+$ which is equivalent to $\ln^2 \hat{b}$ in Fourier space.

The radiative factors are given in terms of integrals over the running of the coupling,

$$\ln \Delta_a \left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) = \int_0^1 dx \frac{x^{N-1}-1}{1-x} \int_{\mu_F^2}^{\tilde{Q}(1-x)^2} \frac{dq^2}{q^2} A_a \left(\alpha_s \left(q^2 \right) \right) \tag{2.3.21}$$

$$\ln J_c \left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) = \int_0^1 dx \frac{x^{N-1}-1}{1-x} \int_{\tilde{Q}(1-x)^2}^{\tilde{Q}(1-x)} \frac{dq^2}{q^2} \mathcal{J}_c \left(\alpha_s \left(q^2 \right), \tilde{Q} \right) \tag{2.3.22}$$

$$\ln \Delta_{abc} \left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) = \int_0^1 dx \frac{x^{N-1}-1}{1-x} \mathcal{D}_{abc}^{\mathrm{F}} \left(\alpha_s \left(\tilde{Q}(1-x)^2 \right) \right), \tag{2.3.23}$$

where the hard-collinear radiation is contained in

$$\mathcal{J}_c \left(\alpha_s \left(q^2 \right), \tilde{Q} \right) = A_c \left(\alpha_s \left(q^2 \right) \right) + B_c \left(\alpha_s \left(\tilde{Q}(1-x) \right) \right). \tag{2.3.24}$$

The coefficients A_c and B_c are computed perturbatively and their expressions are closely related to the ones appearing in transverse momentum resummation (see App. 2.A and App. 2.B). The lower and upper bounds of the q^2-integration in Eq. (2.3.22) represent the soft and collinear limits of the q^2 component of the emitted gluon respectively. They ensures that the logarithmic contributions that appear at fixed-order are fully reproduced. The process dependent $\mathcal{D}_{abc}^{\mathrm{F}}$ is also a power series in the running of the coupling, $\mathcal{D}_{abc}^{\mathrm{F}} = \sum_{n=1}^{\infty} \alpha_s^n \mathcal{D}_{abc}^{\mathrm{F},(n)}$. Its first coefficient in the expansion is just proportional to a combination of the colour factors for each parton involved in the process [213, 227]

$$\mathcal{D}_{abc}^{(1)} = (A_a + A_b - A_c) \ln \left(\frac{Q}{p_T} \right). \tag{2.3.25}$$

Notice that while the functions Δ_a and J_c start to contribute at LL, the radiative factors that embodies the contributions from the large-angle only start to contribute at NLL. With regards to the matching function, only the first two order terms in the expansion, namely $g_{ab}^{(1)}$ and $g_{ab}^{(2)}$, are required to compute up to NNLL predictions. These coefficients are computed by comparing the expansion of the resummed formulae with the exact fixed-order results.

Due to the fact that the radiative factors Δ_a and J_c are process-independent, we are going to compute their expressions first and present the results for the process-dependent parts later. In a similar way as for the transverse momentum resummation, let us define the logarithmic variable $\lambda_{\bar{N}} = \bar{\alpha}(\mu_R^2) \ln \bar{N}^2$ and organize the classes of logarithms as follows

$$\ln \Delta_a \left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) = \frac{1}{\alpha_s} h_a^{(1)}(\lambda_{\bar{N}}) + h_a^{(2)} \left(\lambda_{\bar{N}}, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) + \alpha_s h_a^{(3)} \left(\lambda_{\bar{N}}, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right)$$

$$\ln J_c \left(N, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) = \frac{1}{\alpha_s} f_c^{(1)}(\lambda_{\bar{N}}) + f_c^{(2)} \left(\lambda_{\bar{N}}, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right) + \alpha_s f_c^{(3)} \left(\lambda_{\bar{N}}, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2} \right)$$

In order to determine the expressions of the two functions $h_a^{(n)}$ and $f_c^{(n)}$, it is required to compute the integrals in Eqs. (2.3.21, 2.3.22). There are various ways to compute such

double integrals, here we follow the approach presented in Ref. [211]. The procedure consists on fixing the order at which the strong coupling α_s is run and then interchanging the two integrals by redefining the parameter for the Mellin integration. The resulting integrals of such a procedure can now be computed in a closed form. For illustration purposes, let us compute the the LL contribution $(h_a(1))$ from Eq. (2.3.21). At such a logarithmic order, the perturbative coefficient $A_a(\alpha_s)$ is truncated at $\mathcal{O}(\alpha_s)$. Expressing the argument of α_s in terms of the renormalization scale using the definition of the running of the coupling, we observe that the integral over q^2 becomes

$$A_a^{(1)} \int_{\mu_F^2}^{\tilde{Q}(1-x)^2} \frac{dq^2}{q^2} \alpha_s\left(q^2\right) = A_a^{(1)} \int_{\mu_F^2}^{\tilde{Q}(1-x)^2} \frac{dq^2}{q^2} \frac{\alpha_s(\mu_R^2)}{1+\bar{\alpha}_s(\mu_R^2)\ln(q^2/\mu_R^2)}. \tag{2.3.26}$$

Leaving out the coefficient $A_a^{(1)}$, we can introduce the following relation

$$x^{N-1} - 1 = -\exp\left\{\sum_{n=2}^{\infty}(-)^n \frac{\zeta(n)}{n}\left(-\frac{\partial}{\partial \ln \bar{N}}\right)^n\right\} + \mathcal{O}\left(\frac{1}{\bar{N}}\right), \tag{2.3.27}$$

which at the logarithmic order we are interested in is just equal to (-1) after expanding the exponential. Inserting Eq. (2.3.27) back into Eq. (2.3.21) changes the boundary of the Mellin integral. We therefore end up with the following expression

$$-\int_0^{1-1/\bar{N}} dx \frac{1}{1-x} \int_{\mu_F^2}^{\tilde{Q}(1-x)^2} \frac{dq^2}{q^2} \frac{\alpha_s(\mu_R^2)}{1+\bar{\alpha}_s(\mu_R^2)\ln(q^2/\mu_R^2)} \tag{2.3.28}$$

The order of the x and q-integral can now be interchanged after which the integral over x can be performed exactly in a closed form. Finally, the integral that we need to compute is expressed as

$$-\frac{1}{2}\int_{\tilde{Q}}^{\tilde{Q}/\bar{N}^2} \frac{dq^2}{q^2} \frac{\alpha_s(\mu_R^2)}{1+\bar{\alpha}_s(\mu_R^2)\ln(q^2/\mu_R^2)} \ln\left(\frac{\bar{N}^2 q^2}{\tilde{Q}}\right) - \int_{\mu_F^2}^{\tilde{Q}} \frac{dq^2}{q^2} \frac{\alpha_s(\mu_R^2)}{1+\bar{\alpha}_s(\mu_R^2)\ln(q^2/\mu_R^2)} \ln \bar{N} \tag{2.3.29}$$

The integral over q of Eq. (2.3.29) can be easily performed after which the ensuing expression can be expanded as a series in $\alpha_s(\mu_R^2)$ by fixing $\lambda_{\bar{N}} = \bar{\alpha}_s(\mu_R^2)\ln \bar{N}^2$. Notice that the last term in Eq. (2.3.29) is constructed such that it exactly vanishes when $\tilde{Q} = \mu_F^2$. Retaining only the leading logarithm terms and after performing some algebraic simplification, the LL contribution to $\Delta_a(N)$ is given by

$$\ln \Delta_a^{\mathrm{LL}}\left(N, \alpha_s(\mu_R^2)\right) = \frac{A_a^{(1)}}{\beta_0 \lambda_{\bar{N}}}\left((1-\lambda_{\bar{N}})\ln(1-\lambda_{\bar{N}}) + \lambda_{\bar{N}}\right). \tag{2.3.30}$$

The same procedure applies for the computation of the functions Δ_a, J_c and Δ_{abc} at higher logarithmic orders. In order to obtain $\mathrm{N}^k\mathrm{LL}$ the perturbative functions A_a, B_c and $\mathcal{D}_{abc}$ have to be included and they should be included up to $\mathcal{O}(\alpha_s^{k+1})$ while the running of the strong coupling have to be evaluated at $(k+1)$-loops.

The procedure described above permits us to compute any logarithmic accuracy provided that the perturbative coefficients are known. Up to NNLL, the functions $h_a^{(i)}$ are given by the following expressions

$$h_a^{(1)}(\lambda_{\bar{N}}) = \frac{A_a^{(1)}}{\beta_0^2 \ln \bar{N}^2}\left((1-\lambda_{\bar{N}})\ln(1-\lambda_{\bar{N}}) + \lambda_{\bar{N}}\right) \tag{2.3.31}$$

$$h_a^{(2)}\left(\lambda_{\bar N},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=-\frac{A_a^{(2)}}{2\beta_0^2}\left(\lambda_{\bar N}+\ln\left(\lambda_{\bar N}\right)\right)+\frac{A_a^{(1)}}{2\beta_0}\left(\lambda_{\bar N}+\ln\left(1-\lambda_{\bar N}\right)\right)\ln\left(\frac{Q^2}{\mu_R^2}\right)$$
$$-\frac{A_a^{(1)}}{2\beta_0}\lambda_{\bar N}\ln\left(\frac{Q^2}{\mu_F^2}\right)\frac{A_a^{(1)}}{2\beta_0^3}\beta_1\left(\lambda_{\bar N}+\ln\left(1-\lambda_{\bar N}\right)\frac{1}{2}\ln^2\left(1-\lambda_{\bar N}\right)\right) \quad (2.3.32)$$

$$h_a^{(3)}\left(\lambda_{\bar N},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=-\frac{A_a^{(2)}}{2\beta_0^3}\frac{\beta_1}{1-\lambda_{\bar N}}\left(\lambda_{\bar N}+\ln\left(1-\lambda_{\bar N}\right)\frac{\lambda_{\bar N}^2}{2}\right)+\frac{A_a^{(3)}}{\beta_0^2}\frac{\lambda_{\bar N}^2}{4\left(1-\lambda_{\bar N}\right)}$$
$$+\frac{A_a^{(1)}}{2\beta_0^4}\frac{\beta_1^2}{1-\lambda_{\bar N}}\left(\frac{\lambda_{\bar N}^2}{2}+\lambda_{\bar N}\ln\left(1-\lambda_{\bar N}\right)+\frac{1}{2}\ln^2\left(1-\lambda_{\bar N}\right)\right)-\frac{A_a^{(1)}\lambda_{\bar N}^2}{2\left(1-\lambda_{\bar N}\right)}$$
$$\times\ln\left(\frac{Q^2}{\mu_R^2}\right)+\frac{A_a^{(1)}\beta_2}{2\beta_0^3}\left(\lambda_{\bar N}+\ln(1-\lambda_{\bar N})+\frac{\lambda_{\bar N}^2}{2\left(1-\lambda_{\bar N}\right)}\right)+\frac{A_a^{(1)}\zeta_2\lambda_{\bar N}}{1-\lambda_{\bar N}}$$
$$-\frac{A_a^{(2)}}{2\beta_0}\lambda_{\bar N}\ln\left(\frac{Q^2}{\mu_F^2}\right)+\frac{A_a^{(1)}}{2}\lambda_{\bar N}\ln^2\left(\frac{Q^2}{\mu_F^2}\right)+\frac{A_a^{(1)}}{2}\lambda_{\bar N}\ln\left(\frac{Q^2}{\mu_F^2}\right)\ln\left(\frac{Q^2}{\mu_R^2}\right)$$
$$+\frac{1}{1-\lambda_{\bar N}}\left(\frac{A_a^{(1)}\beta_1}{2\beta_0^2}\left(\lambda_{\bar N}+\ln\left(1-\lambda_{\bar N}\right)\right)-\frac{A_a^{(2)}}{2\beta_0}\lambda_{\bar N}^2\right)\ln\left(\frac{Q^2}{\mu_R^2}\right). \quad (2.3.33)$$

The functions $f_c^{(i)}$ that enter into the definition of J_c can be expressed in terms of the $h_a^{(i)}$ functions above. Indeed, Eq. (2.3.21) and Eq. (2.3.22) are very similar with the difference that the latter embodies the hard-collinear contribution and the lower boundary of the q-integration is $\tilde Q(1-x)^2$ instead of μ_F^2. They are therefore expressed as follows

$$f_c^{(1)}\left(\lambda_{\bar N}\right)=2h_c^{(1)}\left(\frac{\lambda_{\bar N}}{2}\right)-h_c^{(1)}(\lambda_{\bar N})$$
$$f_c^{(2)}\left(\lambda_{\bar N},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=2h_c^{(2)}\left(\frac{\lambda_{\bar N}}{2},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)-h_c^{(2)}\left(\lambda_{\bar N},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)+\frac{B_c^{(1)}}{2\beta_0}\ln\left(1-\frac{\lambda_{\bar N}}{2}\right)$$
$$f_c^{(3)}\left(\lambda_{\bar N},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=2h_c^{(3)}\left(\frac{\lambda_{\bar N}}{2},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)-h_c^{(3)}\left(\lambda_{\bar N},\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)+\frac{3}{2}\zeta_2A_c^{(1)}\frac{\lambda_{\bar N}}{2-\lambda_{\bar N}}$$
$$+\frac{\beta_1B_c^{(1)}}{2\beta_0^2}\frac{\lambda_{\bar N}+2\ln(1-\lambda_{\bar N}/2)}{2-\lambda_{\bar N}}+\frac{B_c^{(1)}}{2}\frac{\lambda_{\bar N}}{2-\lambda_{\bar N}}\ln\left(\frac{\tilde Q^2}{\mu_R^2}\right)-\frac{B_c^{(2)}}{2\beta_0}\frac{\lambda_{\bar N}}{2-\lambda_{\bar N}}. \quad (2.3.34)$$

Provided with the procedure to compute integrals of form given by Eqs. (2.3.21, 2.3.22) as described above, the computation of the radiative factor that embodies the contribution arising from large-angle soft gluon emission is now straightforward. Since, as mentioned above, the perturbative function $\mathcal{D}_{abc}$ is process dependent, the expression of its coefficients are given in App. 2.B. For the same reason, thee expression of the leading-order p_T-distribution $C_{abc}^{(1)}$ and the matching function g_{ab} are also given in App. 2.B for the two colour singlet processes we are interested in, namely the production of a Higgs boson production via gluon fusion and the production of a weak boson that decays into two leptons via DY mechanism.

2.3.3 Threshold resummation and the Landau pole

Some comments are in order w.r.t the resummed formulae derived in the previous section. First, notice that the expression in Eq. (2.3.20) are written in Mellin N-space and have to be inverted back into the momentum space. However, as we shall see, the resulting momentum space resummed expression turns out to be ill-defined. This is because the contributions from the resummed expression at any finite logarithmic accuracy corresponds to a divergent series of contributions to the partonic cross section. Such a divergence, which is of perturbative origin, can be traced to the *Landau pole* in the running of the strong coupling. By inspection of Eq. (2.3.20) we notice that a pole is present at $\bar{N}_L = \exp(1/2\alpha_s\beta_0)$, as a matter of fact, this actually corresponds to a branch cut as it extends from N_L to positive infinity. On the other hand, since the hadronic cross section is obtained by convoluting the partonic part with the luminosity, the integration over the momenta always intercepts the region $x \to 1$ therefore integrating the scale of the running of coupling down to zero. This results as a divergence of the expansion in the coupling α_s. To make this more tangible, let us only focus on the LL part from the exponent of Eq. (2.3.20) from which the Mellin inversion can be performed analytically [226]

$$\frac{d\hat{\sigma}^{\text{th}}_{ab}}{d\xi_p}(x) \propto \frac{d}{dx}\left[\exp\left\{\sum_{n=1}^{\infty}\alpha_s^n \hat{f}^{(n)}\left(\ln^{n+1}\left(\frac{1}{1-x}\right)\right)\right\} + \mathcal{O}\left(\alpha_s^m \ln^m(1-x)\right)\right], \tag{2.3.35}$$

where $\mathcal{O}\left(\alpha_s^m \ln^m(1-x)\right)$ denotes the subleading NLL contributions. This expression clearly blows up when $x \to 1$. For the sake of illustrating the above point, let us consider a toy PDF whose form is chosen such that when integrated over the momentum it suppresses the integrand in Eq. (2.3.35) when $x \to 1$. For simplicity, here we choose $f(x) = (1-x)^a$. Combining this with Eq. (2.3.35), we have

$$\int_0^1 dx\,(1-x)^a \exp\left(\ln^b\left(\frac{1}{1-x}\right)\right) = \sum_{k=1}^{\infty}\frac{(bk)!}{k!(1+a)^{bk+1}}. \tag{2.3.36}$$

The right-hand side of Eq. (2.3.36) diverges for $b \geq 2$, which is always the case, regardless of whether or not the Landau pole is present. This factorial divergence is attributed to the fact that subleading terms were neglected in the inversion of the resummation exponent. Thus, the divergence can be removed by addition of subleading terms which are more suppressed than any power of $1/Q^2$ [226, 229]. Several prescriptions have been proposed to treat these divergent effects, the most used one is the *Minimal Prescription* which was proposed long ago in Ref. [226]. The benefit of the Minimal Prescription is that it does not only handle the Landau pole but also properly treats the subleading terms. Alternatively, a prescription based on *Borel summation* was proposed in Ref. [230] and refined in Refs. [231–234]. Details on these prescriptions will be presented in Sec. (3.1).

2.4 Soft-improved small-p_T (SIPT) resummation

2.4.1 New factorization & Generating Functions

In the following section, we review the soft-improved transverse momentum (SIPT) resummation formalism first proposed in Ref. [165]. Here, we only highlight the main points

of the formalism, thus for further details on the derivation refer to the aforementioned paper. This new resummation formalism relies on a new factorization of the phase space in which soft and collinear contributions are treated at the same footing. Let us recall that the factorization of the phase space exhibits terms of the following form [165]

$$\frac{1}{\sqrt{(1-\tau)^2-4\tau\xi_p}}. \tag{2.4.1}$$

The interplay between the threshold and small-p_T scaling variables is clearly apparent from Eq. (2.4.1). In the context of CSS and CFG, it sufficed to just take the small-p_T limit of Eq. (2.4.1) leading to

$$\lim_{\xi_p\to 0}\frac{1}{\sqrt{(1-\tau)^2-4\tau\xi_p}}=\left(\frac{1}{1-\tau}\right)_+-\frac{1}{2}\delta(1-\tau)\ln\xi_p. \tag{2.4.2}$$

In Fourier-Mellin space, this is equivalent to taking the large-$\hat{b}$ limit while keeping N fixed. This is justifiable as long as the scaling variable ξ_p is the smallest scale involved in the process. However, in the region where threshold contributions are also relevant, i.e. $N/b\sim 1$, oversimplification such as Eq. (2.4.2) neglect soft radiations that are emitted at small angles. In the context of SIPT, this can be dealt by using a distribution identity in order to highlight the divergence at threshold

$$\frac{1}{\sqrt{(1-\tau)^2-4\tau\xi_p}}=\left(\frac{1}{\sqrt{(1-\tau)^2-4\tau\xi_p}}\right)_+-\frac{1}{2}\delta(1-\tau)\left[\ln\xi_p-\ln(1+\xi_p)\right]. \tag{2.4.3}$$

Notice that when using the standard Eq. (2.4.2) or the modified Eq. (2.4.3) factorization of the phase space, the matrix element is always the same. This is justified by the fact that unlike the phase space, the squared amplitude does not display further infrared singularities as $\tau\to 1$ since power counting arguments ensure that all infrared singularities only arise in the collinear limit. This means that terms of the form $N\sqrt{\xi_p}$ (or $N/\hat{b}$) only arise in phase space and it is sufficient to expand the amplitude in powers of ξ_p at fixed N. With Eq. (2.4.3), the two limits are kept for $N/b\sim 1$. in conjugate Fourier space, this corresponds to taking the large-b limit for fixed N/b. As we shall see later, the two variables N and b are now embodied into one single scaling variable. In such a case, logarithms of the form $\ln(\hat{b})\ln^k(\bar{N})$ and $\ln^{k+1}(\hat{b})$ are considered of the same logarithmic order. The SIPT then differs from the standard resummation formalism by terms which are power-suppressed in the large-$\hat{b}$ limit at fixed N.

It is now possible to construct a transverse momentum resummation that takes into account the soft limit by combining the new factorized phase space in Eq. (2.4.3) with the factorized amplitude. Notice that here, we are only interested in the leading-power resummed expression, i.e. up to corrections that are suppressed by powers of $1/\hat{b}$. At the partonic level, the soft-improved transverse momentum resummation is defined as

$$\frac{d\hat{\sigma}^{\rm th}_{ab}}{d\xi_p}\left(N,\alpha_s(\mu_R^2),\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=\sigma_{\rm F}^{\rm Born}\mathcal{H}_{ab\to c\bar{c}}\left(N,\alpha_s(Q^2)\right)a^N(\xi_p)\int_0^\infty d\hat{b}\frac{\hat{b}}{2}J_0\left(\hat{b}\sqrt{\xi_p}\right)$$

$$\exp\left\{\int_0^\infty d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right)\int_0^1 d\tau\,\tau^{N-1}\mathcal{A}_c(\alpha_s(Q^2))\mathcal{D}_0(\xi,\tau)+\mathcal{O}\left(\frac{1}{\hat{b}}\right)\right\}$$
$$\exp\left\{\int_0^\infty d\xi\, a^{-N}(\xi)J_0\left(\hat{b}\sqrt{\xi}\right)\left(\frac{1}{\xi}\right)_+\mathcal{B}_c\left(N,\alpha_s(Q^2\xi)\right)+\mathcal{O}\left(\frac{1}{\hat{b}}\right)\right\}, \tag{2.4.4}$$

where $a(\xi)=\left(\sqrt{1+\xi}+\sqrt{\xi}\right)^2$ and the distribution $\mathcal{D}_0$ is defined as a more general class of the term that appears in the phase space of Eq. (2.4.3)

$$\mathcal{D}_k(\xi,\tau)=\left(\frac{\ln^k\xi}{\xi}\right)_+\left[\left(\frac{1}{\sqrt{(1-\tau)^2-4\tau\xi_p}}\right)_+-\frac{1}{2}\delta(1-\tau)\left[\ln\xi_p-\ln(1+\xi_p)\right]\right]. \tag{2.4.5}$$

The resummed formula Eq. (2.4.4) resums small-p_T and threshold logarithms up to NNLL and NNLL* respectively. By construction, it reduces to the CFG resummation formalism in the limit $b\to\infty$. In addition, it reproduces up to NNLL* the threshold resummation for the total inclusive cross section upon integration over ξ_p, or equivalently by setting $b=0$ in Fourier space. By expanding the perturbative functions $\mathcal{A}_c$ and $\mathcal{B}_c$ in powers of α_s, one notice that all the integrals appearing in the expansion can be categorized as,

$$\mathfrak{G}_{1,k}(N,\hat{b})=\int_0^\infty d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right)\int_0^1 d\tau\,\tau^{N-1}\mathcal{D}_k(\xi,\tau) \tag{2.4.6}$$

$$\mathfrak{G}_{2,k}(N,\hat{b})=\int_0^\infty d\xi\, a^{-2N}(\xi)J_0\left(\hat{b}\sqrt{\xi}\right)\left(\frac{\ln\xi}{\xi}\right)_+, \tag{2.4.7}$$

where one can immediately notice that the functions $\mathfrak{G}_{1,k}$ and $\mathfrak{G}_{2,k}$ are the terms that are proportional to the coefficients $\mathcal{A}_c^{(n)}$ and $\mathcal{B}_c^{(n)}$ respectively. The construction of the final soft-improved transverse momentum resummation amounts to determining the perturbative coefficients and computing the integrals given by $\mathfrak{G}_{i,k}$. The functions $\mathfrak{G}_{i,k}$ defined in Eqs. (2.4.6, 2.4.6) can be obtained by taking the k-th derivatives w.r.t. ϵ at $\epsilon=0$ of the following generating functions

$$\mathfrak{G}_{1,k}(N,\hat{b})=\left[\frac{\partial^k}{\partial\epsilon^k}\mathcal{G}_1(N,\hat{b},\epsilon)\right]_{\epsilon=0}, \tag{2.4.8}$$

$$\mathfrak{G}_{2,k}(N,\hat{b})=\left[\frac{\partial^k}{\partial\epsilon^k}\mathcal{G}_2(N,\hat{b},\epsilon)\right]_{\epsilon=0}. \tag{2.4.9}$$

Since the information concerning the soft behaviour is embodied in Eq. (2.4.7) we will mostly focus on the computation of $\mathfrak{G}_{1,k}$. Before doing so, let us first study its behaviour in various kinematic limits. This will provide us insights on what kind of contributions are contained in the soft-improved resummation when compared to the standard transverse momentum resummation formalism.

2.4.2 Large logarithms at the level of generating functions

In the following parts, we study the behaviours of the generating function given by Eq. (2.4.8) in the following limits: (a) large-$\hat{b}$ limit at fixed N which corresponds to the standard

transverse momentum resummation, (b) large-N limit at fixed $\hat{b}$ which corresponds to the soft limit, (c) $\hat{b} \to 0$ limit which corresponds to the total inclusive resummed cross section, and (d) large-$\hat{b}$ at fixed N/b limit which is the limit we are mainly interested in.

Let us start with the expression of the generating function $\mathcal{G}_1$ which is defined as

$$\mathcal{G}_1(N,\hat{b},\epsilon) = \int_0^\infty d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right) \int_0^1 d\tau\, \tau^{N-1} \mathcal{D}_0(\xi,\tau,\epsilon) \tag{2.4.10}$$

$$= \int_0^1 d\tau\, \tau^{N-1} \int_0^{\xi_{\max}} d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right) \mathcal{D}_0(\xi,\tau,\epsilon) \tag{2.4.11}$$

where $\xi_{\max} = (1-\tau)^2/4\tau$ and $\mathcal{D}_0(\xi,\tau,\epsilon)$ is given by the following expression

$$\mathcal{D}_0(\xi,\tau,\epsilon) = \left(\frac{1}{\xi^{1-\epsilon}}\right)_+ \left[\left(\frac{1}{\sqrt{(1-\tau)^2 - 4\tau\xi_p}}\right)_+ - \frac{1}{2}\delta(1-\tau)\left[\ln\xi_p - \ln(1+\xi_p)\right]\right]. \tag{2.4.12}$$

Using the following distribution identities

$$\left(\frac{1}{\xi^{1-\epsilon}}\right)_+ = \frac{1}{\xi^{1-\epsilon}} - \frac{1}{\epsilon}\delta(\xi) \qquad \text{and} \qquad \left(\frac{1}{\xi^{1-\epsilon}}\right)_+ \ln\xi = \frac{1}{\xi^{1-\epsilon}}\ln\xi + \frac{1}{\epsilon^2}\delta(\xi), \tag{2.4.13}$$

the generating function in Eqs. (2.4.10, 2.4.11) can be therefore simplified as

$$\mathcal{G}_1(N,\hat{b},\epsilon) = -\frac{1}{2\epsilon^2} + \int_0^1 d\tau\, \tau^{N-1}\left[\frac{1}{\epsilon}\left(\frac{1}{\tau-1}\right)_+ + \int_0^{\xi_{\max}} d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right) \frac{\xi^{\epsilon-1}}{\sqrt{(1-\tau)^2 - 4\tau\xi_p}}\right]. \tag{2.4.14}$$

The Mellin transform that appears in the first term of the τ integration Eq. (2.4.14) can be explicitly expressed in terms of a logarithm in the large-N limit

$$\int_0^1 d\tau\, \tau^{N-1}\left(\frac{1}{\tau-1}\right)_+ = -\left(\psi(N) + \gamma_E\right) \xrightarrow{N\to\infty} -\ln(\bar{N}). \tag{2.4.15}$$

This allows us to simplify Eq. (2.4.14) further

$$\mathcal{G}_1(N,\hat{b},\epsilon) = \frac{1}{\epsilon}\ln\bar{N} - \frac{1}{2\epsilon^2} + \int_0^1 d\tau\, \tau^{N-1} \int_0^{\xi_{\max}} d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right) \frac{\xi^{\epsilon-1}}{\sqrt{(1-\tau)^2 - 4\tau\xi_p}}. \tag{2.4.16}$$

Large-N limit at fixed $\hat{b}$:

In this section, let us study the logarithmic structure of the generating function, Eq. (2.4.16), in the limit where N is large and ξ fixed. This will give us an idea on the structure of $\ln\bar{N}$ that appears in the resummed exponent at threshold when $\xi \gg 0$. For simplicity, we can isolate the integrals in Eq. (2.4.16) from the rest and only consider $\tilde{\mathcal{G}}_1(N,\hat{b})$ where

$\tilde{\mathcal{G}}_1(N,\hat{b},\epsilon) = \mathcal{G}_1(N,\hat{b},\epsilon) - \ln\bar{N}/\epsilon + 1/2\epsilon^2$. By expanding the zeroth-th order Bessel function in powers of its argument

$$J_0(\hat{b}\sqrt{\xi}) = \sum_{p=0}^{\infty} \frac{(-1)^p}{\Gamma^2(p+1)} \left(\frac{\hat{b}^2}{4}\right)^p \xi^p, \tag{2.4.17}$$

we can interchange the sum and the integrals in order to perform the integration for each power of p. Doing so yields

$$\tilde{\mathcal{G}}_1(N,\hat{b},\epsilon) = \frac{1}{2}\sum_{p=0}^{\infty}\left[\left(-\frac{\hat{b}^2}{4}\right)^p \frac{\Gamma^2(p+\epsilon)\Gamma(N-p-\epsilon)}{\Gamma^2(p+1)\Gamma(N+p+\epsilon)}\right], \tag{2.4.18}$$

where the duplication formula for Gamma functions has been used. The first term in the series ($p=0$) is independent of $\hat{b}$ and therefore it is more convenient to separate it off from the sum. Hence, Eq. (2.4.18) now writes as

$$\tilde{\mathcal{G}}_1(N,\hat{b},\epsilon) = \frac{1}{2}\Gamma^2(\epsilon)\frac{\Gamma(N-\epsilon)}{\Gamma(N+\epsilon)} + \frac{1}{2}\sum_{p=1}^{\infty}\left(-\frac{\hat{b}^2}{4}\right)^p \frac{\Gamma^2(p+\epsilon)}{\Gamma^2(p+1)}\frac{\Gamma(N-p-\epsilon)}{\Gamma(N+p+\epsilon)}. \tag{2.4.19}$$

In the large-N limit, the ratio of Gamma functions can be simplified using

$$\frac{\Gamma(N-p-\epsilon)}{\Gamma(N+p+\epsilon)} \xrightarrow{N\to\infty} \left(\frac{1}{N^2}\right)^{p+\epsilon}\left(1+\mathcal{O}\left(\frac{1}{N}\right)\right) \tag{2.4.20}$$

from which it follows that

$$\tilde{\mathcal{G}}_1(N\to\infty,\hat{b},\epsilon) = \frac{\Gamma^2(1+\epsilon)}{2\epsilon^2 N^{2\epsilon}} + \frac{1}{N^{2\epsilon}}\sum_{p=1}^{\infty}\left(-\frac{\hat{b}^2}{4N^2}\right)^p \frac{\Gamma^2(p+\epsilon)}{\Gamma^2(p+1)}. \tag{2.4.21}$$

We can notice that the second term of the above equation vanishes in the large-N limit. Thus, the generating function in Eq. (2.4.16) becomes $\hat{b}$-independent and is expressed as

$$\mathcal{G}_1(N\to\infty,\epsilon) = \frac{1}{\epsilon}\ln\bar{N} - \frac{1}{2\epsilon^2} + \frac{\Gamma^2(1+\epsilon)}{2\epsilon^2 N^{2\epsilon}}. \tag{2.4.22}$$

By taking the k-th derivative of the above expression w.r.t. ϵ as defined by Eq. (2.4.8), one notices that the singular terms appearing in Eq. (2.4.22) when $\epsilon\to 0$ cancels out and one can safely set $\epsilon = 0$. As a result, the function $\mathfrak{G}_{1,k}$ is just a polynomial in $\ln N$,

$$\mathfrak{G}_{1,k}(N\to\infty) = \frac{k!}{2}\sum_{m=0}^{k+2}\frac{2^m(-1)^{k-m}}{m!(2+k-m)!}(\ln N)^{2+k-m}. \tag{2.4.23}$$

Large-$\hat{b}$ limit at fixed N:

Let us know study the logarithmic structure of the generating functionEq. (2.4.16) in the large-$\hat{b}$ limit while keeping N fixed. As shown in Refs. [165, 235] at the resummed level in both Fourier-Mellin and direct space, such a limit reproduces the standard CFG resummed formula.

Let us start from the integrand of Eq. (2.4.10)

$$\mathcal{D}_0(\xi,\tau,\epsilon) = \left(\frac{1}{\xi^{1-\epsilon}}\right)_+ \left[\left(\frac{1}{\sqrt{(1-\tau)^2-4\tau\xi_p}}\right)_+ - \frac{1}{2}\delta(1-\tau)\left[\ln\xi_p - \ln(1+\xi_p)\right]\right]. \tag{2.4.24}$$

Due to the square root term in the denominator, Eq. (2.4.24) vanishes for all ξ greater than ξ_{max}. This means that the function $\mathcal{D}_0$ only exhibits singular behaviour when $\xi = 0$ due to the factor $1/\xi$. In the small-ξ limit, the last term $\ln(1-\xi)$ behaves as ξ cancelling the singularity and can therefore be neglected. For a fixed value of τ, Eq. (2.4.24) now becomes

$$\mathcal{D}_0(\xi\to 0,\tau,\epsilon) = \left(\frac{1}{\xi^{1-\epsilon}}\right)_+ \left\{\left(\frac{1}{1-\tau}\right)_+ - \frac{1}{2}\delta(1-\tau)\ln\xi_p\right\}. \tag{2.4.25}$$

The form of Eq. (2.4.25) allows us to decouple the Mellin integral from the Fourier. Performing the integral over the scaling variable τ, we are left with

$$\mathcal{G}_1(N,\hat{b}\to\infty,\epsilon) = -\int_0^{\xi_{\max}} d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right)\left(\frac{1}{\xi^{1-\epsilon}}\right)_+ \left(\ln\bar{N} + \frac{1}{2}\ln\xi\right). \tag{2.4.26}$$

The computation of the first term in Eq. (2.4.26) to all logarithmic order can be done in a closed form using results from Ref. [232]. The approach consists in using the polar coordinates form of $\hat{b}$ and expressing the Bessel function in terms of its integral representation. Applying the results to our case, the first term of Eq. (2.4.26) reads

$$\int_0^{\infty} d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right)\left(\frac{1}{\xi^{1-\epsilon}}\right)_+ = \frac{1}{\epsilon}\left[\frac{\Gamma(1+\epsilon)}{\Gamma(1-\epsilon)}\left(\frac{\hat{b}^2}{4}\right)^{-\epsilon} - 1\right], \tag{2.4.27}$$

where we have extend the upper boundary of the integration to $(+\infty)$ in order to relate Eq. (2.4.26) with Ref. [232]. As mentioned before, this does not change the integral since contributions for $\xi > \xi_{max}$ are zero. The second term in Eq. (2.4.26) can be computed from Eq. (2.4.27) by realizing that

$$\int_0^{\xi_{\max}} d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right)\left(\frac{1}{\xi^{1-\epsilon}}\right)_+ \ln\xi = \frac{\partial}{\partial\epsilon}\int_0^{\infty} d\xi\, J_0\left(\hat{b}\sqrt{\xi}\right)\left(\frac{1}{\xi^{1-\epsilon}}\right)_+. \tag{2.4.28}$$

Putting Eqs. (2.4.27, 2.4.28) back into Eq. (2.4.26), performing some algebraic simplifications, and reorganizing the expression, we find

$$\mathcal{G}_1(N,\hat{b}\to\infty,\epsilon) = \frac{1}{2\epsilon^2}\left(\mathcal{M}(\epsilon) - \epsilon\mathcal{M}^{(1)}(\epsilon) - \mathcal{M}(0) - 2\epsilon(\psi(N)+\gamma_E)(\mathcal{M}(\epsilon)-1)\right) \tag{2.4.29}$$

where

$$\mathcal{M}(\epsilon) = \left(\frac{\hat{b}^2}{4}\right)^2 \frac{\Gamma(1+\epsilon)}{\Gamma(1-\epsilon)} = \sum_{k=0}^{\infty}\frac{\epsilon^k}{k!}\mathcal{M}^{(k)}(0) \qquad \text{with} \qquad \mathcal{M}^{(k)}(\epsilon) = \frac{\partial^k}{\partial\epsilon^k}\mathcal{M}(\epsilon). \tag{2.4.30}$$

Replacing $\mathcal{M}$ in Eq. (2.4.29) using its Taylor expansion in Eq. (2.4.30), taking the k-th derivative of the resulting expression, and putting everything back into Eq. (2.4.9), we arrive at the following expression

$$\mathfrak{G}_{1,k}(N,\hat{b}\to\infty) = -\frac{1}{2}\left[\frac{M^{(k+2)}(0)}{k+2} + 2\ln\bar{N}\frac{M^{(k+1)}(0)}{k+1}\right]. \tag{2.4.31}$$

Using the definition of $\mathcal{M}^{(k)}$ from Eq. (2.4.30) one observes that such a function is expressed in as a polynomials in $\ln\hat{b}$. This is expected as those logarithms are the ones that are resummed in the Sudakov exponent. Keeping only the highest power, we have

$$\mathfrak{G}_{1,k}(N,\hat{b}\to\infty) = (-1)^{k+1}\left[\frac{1}{2k+4}\ln^{k+2}\left(\frac{\hat{b}^2}{b_0^2}\right) - \frac{\ln\bar{N}}{k+1}\ln^{k+1}\left(\frac{\hat{b}^2}{b_0^2}\right)\right]. \tag{2.4.32}$$

Once multiplied with the coefficient functions $\mathcal{A}_c$ as in Eq. (2.4.4), the above logarithmic structure corresponds exactly to the ones that appear in the standard formulation of the transverse momentum.

Large-$\hat{b}$ limit at fixed N/b:

The main point of having constructed the two functions Eqs. (2.4.6, 2.4.7) was to capture the behaviours of the soft radiations that are emitted at very small angle. This is the main idea of the soft-improved transverse momentum resummation. In order to make such a behaviour manifest at the exponent level, the generating functions in Eqs. (2.4.7, 2.4.9) have to be taken in the limit where $\hat{b}$ is large but N/b is fixed. In this section, we study the behaviours of these generating functions in the aforementioned limit and compare the results with Eq. (2.4.32).

Let us start with Eq. (2.4.18) where no limit has been taken yet

$$\tilde{\mathcal{G}}_1(N,\hat{b},\epsilon) = \frac{1}{2}\sum_{p=0}^{\infty}\left[\left(-\frac{\hat{b}^2}{4}\right)^p \frac{\Gamma^2(p+\epsilon)\Gamma(N-p-\epsilon)}{\Gamma^2(p+1)\Gamma(N+p+\epsilon)}\right]. \tag{2.4.33}$$

Since the Gamma functions do not depend on $\hat{b}$, their expressions can be replaced by their asymptotic expansions as in Eq. (2.4.20). This leads us to the exact same expression as in Eq. (2.4.21) taking into account that the label $N\to\infty$ only corresponds to the N-independent of the expression. In contrast to Eq. (2.4.22), the sum in Eq. (2.4.21) has to be performed explicitly. Doing so leads to

$$\tilde{\mathcal{G}}_1(N,\hat{b},\epsilon) = \frac{1}{2}\frac{\Gamma^2(\epsilon)}{N^{2\epsilon}}{}_2F_1\left(\epsilon,\epsilon;1;-\frac{\hat{b}^2}{4N^2}\right). \tag{2.4.34}$$

The simplification of the Hypergeometric function in Eq. (2.4.34) has to be done taking into account that the generating function has to reproduce the standard transverse momentum resummation in the large-$\hat{b}$ limit and reproduce the inclusive threshold momentum resummation upon integration over the transverse momentum. The former can be achieved by: expanding the Hypergeometric function as a series in ϵ, evaluating the large-$\hat{b}$ limit, and

retaining the highest power in $\ln \hat{b}$. The latter is achieved by letting $\hat{b} = 0$ in the Hypergeometric function, expanding the resulting expression as a series in ϵ, and retaining only the first terms in the expansion. The soft-improved expression can then be constructed by matching the two expressions.

The expansion of the Hypergeometric function is straightforward using [236, 237]

$$ {}_2F_1\left(\epsilon, \epsilon; 1; -z\right) = 1 - \epsilon^2 \mathrm{Li}_2(-z) + \mathcal{O}(\epsilon^3). \tag{2.4.35}$$

On the other hand, for a large parameter z, the Hypergeometric is given by [236, 237]

$$ {}_2F_1\left(\epsilon, \epsilon; 1; -z\right) = \frac{1}{z^\epsilon} \frac{\ln z - \psi(1-\epsilon) - \psi(\epsilon) - 2\gamma_E}{\Gamma(\epsilon)\Gamma(1-\epsilon)} + \mathcal{O}\left(\frac{1}{z}\right). \tag{2.4.36}$$

By shifting $\hat{b}$-dependent variable $z \to z+1$ in Eq. (2.4.36), we arrive at an expression that is regular when $z \to 0$ but still coincides with Eq. (2.4.36) when $z \to \infty$. This allows us to take the small-ϵ expansion of Eq. (2.4.36) as follows

$$ {}_2F_1\left(\epsilon, \epsilon; 1; -z \to -(z+1)\right) = 1 - \epsilon^2 \left(\frac{1}{2}\ln^2(z+1) + \zeta_2\right) + \mathcal{O}(\epsilon^3). \tag{2.4.37}$$

In order to interpolate the Hypergeometric function in Eq. (2.4.34), we first relate Eq. (2.4.37) with Eq. (2.4.35), focusing only on the terms that multiply ϵ^2, and using the definition of the dilogarithm for complex argument in Ref. [238], we have

$$ \mathrm{Li}_2\left(\frac{1}{1-z}\right) = \mathrm{Li}_2(z) + \zeta_2 + \ln(-z)\ln(1-z) - \frac{1}{2}\ln^2(1-z), \tag{2.4.38}$$

where in the limit $z \to 1$, the last two terms of Eq. (2.4.38) can be neglected. This is fully justified since the large-z behaviour is fully described with Eq. (2.4.36). Combining everything together, we can write a more malleable form of Eq. (2.4.34) that is different from it by subleading terms. Up to NNLL, we therefore have

$$ {}_2F_1\left(\epsilon, \epsilon; 1; -\frac{\hat{b}^2}{4N^2}\right) = \frac{\ln(\hat{b}^2/4N^2) - \psi(1-\epsilon) - \psi(\epsilon) - 2\gamma_E}{(4N)^{-2}(N^2+\hat{b}^2)\Gamma(\epsilon)\Gamma(1-\epsilon)} + \epsilon^2 \mathrm{Li}_2\left(\frac{4N^2}{N^2+\hat{b}^2}\right). \tag{2.4.39}$$

Putting Eq. (2.4.39) back into Eq. (2.4.34) and adding back the ϵ-dependent term that we left out prior to the integration, the derivative w.r.t ϵ can now be easily taken. After setting $\epsilon = 0$ and retaining only terms that contribute up to NNLL, we finally have

$$ \mathfrak{G}_{1,k}\left(\hat{b} \to \infty, \text{ fixed } \frac{N}{b}\right) = (-1)^{k+1}\left[\frac{\ln^{k+2}\chi}{2k+4} - \frac{\ln \bar{N}\ln^{k+1}\chi}{k+1} - \ln^k \bar{N}\mathrm{Li}_2\left(\frac{\bar{N}}{\chi}\right)\right], \tag{2.4.40}$$

where we have introduced a new variable $\chi = \bar{N}^2 + \hat{b}^2/b_0^2$. By proceeding exactly in the same way, one can also compute the expression of $\mathfrak{G}_{2,k}$. The difference only resides in the fact that the expansion of the Hypergeometric function is done up to order ϵ. This is because the perturbative function $\mathcal{B}_c$ that multiplies $\mathfrak{G}_{2,k}$ starts at NLL. The final result is given by

$$ \mathfrak{G}_{2,k}\left(\hat{b} \to \infty, \text{ fixed } \frac{N}{b}\right) = (-1)^{k+1}\ln^{k+1}\chi. \tag{2.4.41}$$

The two functions $\mathfrak{G}_{1,k}$ and $\mathfrak{G}_{2,k}$ preserve the same logarithmic accuracy as the standard transverse momentum resummation. In addition, $\mathfrak{G}_{1,k}$ contains NNLL soft-contributions that at the level of the total inclusive cross section reproduces the threshold resummation up to some subleading corrections. Comparing Eq. (2.4.40) to Eq. (2.4.32), we notice two main differences. First, the argument of the logarithms has changed from $\hat{b}^2/b_0^2$ to $\bar{N}^2 + \hat{b}^2/b_0^2$. This takes into account soft contributions that are not taken into in the standard transverse momentum formalism. This also suggests that $\ln N$ are resummed at the same footing as $\ln \hat{b}$. Second, Eq. (2.4.40) contains a dilogarithmic contribution that multiplies the highest power of $\ln N$. As expected, Eq. (2.4.40) exactly reproduces Eq. (2.4.32) in the large-$\hat{b}$ limit since $\chi \to \hat{b}^2/b_0^2$ and $\mathrm{Li}_2(\bar{N}^2/\chi) \to 0$ in that limit. In particular, it is interesting to see that the combined resummation variable χ leads to the modified $(1 + \hat{b}^2/b_0^2)$ of the CFG when $\bar{N} = 1$. Notice, however, that the above expression cannot be exactly compared to Eq. (2.4.23) since the large-$\hat{b}$ and large-N limits do not commute.

2.4.3 From the generating functions to the resummed formulae

The previous section provided a description on how to consistently combine soft and recoil effects in transverse momentum resummation up to NNLL for the process $h_1 + h_2 \to \mathrm{F} + \mathrm{X}$. It was shown that the structure of the large logarithms are governed by two generating functions, Eqs. (2.4.8, 2.4.9). Having studied and performed the integral of the generating functions in various kinematic limits, we can now write down the expression of the partonic resummed formula. Using Eqs. (2.4.40, 2.4.41), it can be shown that the partonic cross section can be written in a similar form as Eq. (2.2.27)

$$\frac{d\hat{\sigma}^{\mathrm{tr}}_{ab}}{d\xi_p}\left(N,\chi,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right) = \sum_c \sigma^{\mathrm{Born}}_{\mathrm{F},c\bar{c}}\,\bar{\mathrm{H}}_c^F(N,\chi)\mathcal{S}_c(N,\chi)\sum_{k,l} C_{ck}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right),\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)$$
$$C_{\bar{c}l}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right),\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right) U_{ka}\left(N,\frac{Q^2}{\chi}\leftarrow\mu_F^2,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right) U_{lb}\left(N,\frac{Q^2}{\chi}\leftarrow\mu_F^2,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right). \tag{2.4.42}$$

The soft-improved resummed expression in Eq. (2.4.42) is different from Eq. (2.2.27) in the following ways. First, the Sudakov form factor $\mathcal{S}_c$ depends on the resummation variable χ which, as mentioned previously, depends both on the Mellin moment N and the impact parameter b. As a consequence, the exponent of $\mathcal{S}_c$ now vanishes when $\chi = 1$. Second, the modified hard function $\bar{\mathrm{H}}_c$ now depends on χ, nonetheless, it still behaves as a constant in the large-$\hat{b}$ limit. It contains constant terms that do not vanish in the exponent when $\chi = 1$. They originate from the dilogarithms appearing in Eq. (2.4.40). Therefore, up to NLO, $\bar{\mathrm{H}}_c$ and H_c are related through $\bar{\mathrm{H}}_c(N,\chi) = \mathrm{H}_c(\alpha_s) + \alpha_s \mathrm{Li}_2(\bar{N}^2/\chi)$. Finally, the coefficient functions and the evolution kernels are evaluated at a scale Q^2/χ instead of $\hat{b}^2/b_0^2$, hence the resulting evolution factors also depend on χ. The fact that the evolution kernels explicitly depend on the scale Q^2/χ entails a new way of counting the logarithms since the anomalous dimensions can generate $\ln \bar{N}$ in the large-N limit. To make this statement

tangible, let us consider for the sake of argument the singlet case where the solution of the evolution kernel is given by

$$U_{ka}\left(N,\frac{Q^2}{\chi}\leftarrow Q^2\right)=\sum_{i,j}V^N_{ki}\left(\alpha_s\left(\frac{Q^2}{\chi}\right)\right)U^{(\mathrm{LO})}_{ij}\left(N,\frac{Q^2}{\chi}\leftarrow Q^2\right)\tilde{V}^N_{ja}(\alpha_s(Q^2)),\quad(2.4.43)$$

where for brevity we omitted the remaining scale dependence and expressed the N-dependence of V as a superscript. As detailed in Sec. (1.1.2), $U^{(\mathrm{LO})}$ and V represent the lowest and higher perturbative order solution to the evolution equation respectively. The function $\tilde{V}_{ja}$ denotes the (j,a)-element of the inverse matrix $\mathbf{V}^{-1}$ in the flavour space. Here and in the following, we shall use the boldface notation to denote the representation in flavour matrix space. From Eq. (2.4.42) to Eq. (2.4.43) we have evolved the parton densities from scale Q^2/χ to Q^2 and then from Q^2 to μ_F^2 which in case $Q=\mu_F$ leads to $U(Q^2\leftarrow\mu_F^2)=1$. The lowest perturbative order solution $U^{(\mathrm{LO})}$ can be derived by diagonalizing the LO anomalous dimension matrix $\gamma^{(0)}$. Following Refs. [117, 195],

$$U^{(\mathrm{LO})}_{ij}\left(N,\frac{Q^2}{\chi}\leftarrow Q^2\right)=\sum_{r=\pm}\mathcal{E}^{(r)}_{ij}(N)\exp\left(\mathcal{Q}_r(N,\chi)\right),\quad(2.4.44)$$

where the exponent $\mathcal{Q}$ that resums both soft and recoil contributions is given by

$$\mathcal{Q}_r(N,\chi)=-\frac{\lambda^{(0)}_r(N)}{\beta_0}\ln\left(\frac{\alpha_s(Q^2/\chi)}{\alpha_s(Q^2)}\right).\quad(2.4.45)$$

As described in Sec. (1.1.2), $\lambda_\pm$ and $\mathcal{E}_\pm$ represent the eigenvalues of the singlet matrix $\gamma^{(0)}$ and its projectors respectively. In the context of standard transverse momentum resummation, the leading-order solution to the evolution equation contributes with a single logarithm in $\ln\hat{b}$ and therefore starts to contribute at NLL. However, in the context of soft-improved transverse momentum resummation, at leading-logarithm the flavour-diagonal anomalous dimensions carry an additional contribution of the form $A^{(1)}_c\ln\bar{N}^2$. Indeed, the lowest order contribution to $\mathcal{Q}$ is given by

$$\mathcal{Q}_\pm(N,\chi)=\frac{1}{\beta_0}\gamma^{(0)}_{gg/qq}\ln(1-\lambda_\chi)\qquad\text{with}\qquad\lambda_\chi=\bar{\alpha}_s\ln\chi.\quad(2.4.46)$$

Taking the large-N limit of the anomalous dimensions in the above expression leads to the following result

$$\frac{1}{\beta_0}\gamma^{(0)}_{gg/qq}\ln(1-\lambda_\chi)\xrightarrow{N\to\infty}-\frac{A^{(1)}_{g/q}}{\beta_0}\ln(\bar{N})\ln(1-\lambda_\chi)=-\frac{A^{(1)}_{g/q}}{2\bar{\alpha}_s\beta_0}\lambda_{\bar{N}}\ln(1-\lambda_\chi),\quad(2.4.47)$$

which exhibits soft logarithms through $\lambda_{\bar{N}}=2\bar{\alpha}_s\ln\bar{N}$ with $\bar{\alpha}_s\equiv\bar{\alpha}_s(\mu_R^2)$. This suggests that in order to compute N^{k}LL soft-improved transverse momentum resummation, the large-N behaviour of the evolution has to be included up to N^{k}LL, which is not the case in the standard transverse momentum resummation as N^{k-1}LL is enough. That is, at NNLL, the solution to the evolution equation is performed up to NNLO accuracy but with the NNLO anomalous dimensions replaced by their large-N behaviours. To illustrate this

further, still in the context of singlet which is the only important for the Higgs production, let us put Eq. (2.4.42) into the same form as Eq. (2.2.30)

$$\frac{d\hat{\sigma}^{\rm tr}_{ab}}{d\xi_p}\left(N,\chi,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=\sum_c \sigma^{\rm Born}_{{\rm F},gg}(\alpha_s(\mu_R^2))\bar{\mathcal{H}}^{{\rm F},\{\rm S\}}_{ab\to c\bar{c}}\left(N,\chi,\alpha_s(\mu_R^2),\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right) \tag{2.4.48}$$
$$\exp\left\{\mathcal{S}_c\left(N,\chi,\alpha_s(\mu_R^2),\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)+\mathcal{R}_{\{\rm S\}}\left(N,\chi,\alpha_s(\mu_R^2),\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)\right\},$$

where here $\{\rm S\}$ labels a set of flavour indices. As pointed out above, the function $\bar{\mathcal{H}}$ receives contribution from the constant terms coming from the generating functions. Despite the fact that it depends explicitly on $\hat{b}$ through χ, it behaves as a constant in the limit where $\hat{b}$ is large since ${\rm Li}_2(\bar{N}^2/\chi)\to 0$.

The Sudakov exponent is fully determined by the generating functions. Putting back the expression of the generating functions in Eqs. (2.4.40, 2.4.41) into Eq. (2.4.4), expressing the perturbative functions $\mathcal{A}_c$ and $\mathcal{B}_c$ in terms of the standard A_c and B_c, it follows that

$$\mathcal{S}_c(N,\chi,Q^2)=\int_{\frac{Q^2}{\bar{N}^2}}^{Q^2}\frac{dq^2}{q^2}D_c(\alpha_s(q^2))-\int_{\frac{Q^2}{\chi}}^{Q^2}\frac{dq^2}{q^2}\left[A_c(\alpha_s(q^2))\ln\frac{Q^2}{q}+B_c(\alpha_s(q^2))\right] \tag{2.4.49}$$

where up to NNLL the additional perturbative function D_c contributes up to order α_s^2. With the first coefficient being zero, $D_c^{(2)}=A_c^{(1)}\beta_0{\rm Li}_2(\bar{N}^2/\chi)$. The function D_c has its origin from the dilogarithm part appearing in $\mathfrak{G}_{1,k}$. When comparing the second term of Eq. (2.4.49) with Eq. (2.2.11), putting aside the process-dependent part, the difference only resides in the definition of resummation scale. With Q^2/χ, the improved transverse momentum resummation captures soft contributions that were not present in the standard resummation formalism which are crucial in order to reproduce the total inclusive threshold resummed expression. As for the construction of the soft-improved resummed transverse momentum expression, the universal form factor can be organized in terms of classes of logarithms as $\mathcal{S}_c(N,\chi)=\sum_{n=-1}^{\infty}\bar{\alpha}_s^n g_{n+2}(\lambda_{\bar{N}},\lambda_\chi)$. The explicit expressions of the functions g_i are given by

$$g_1(\lambda_\chi)=\frac{A_g^{(1)}}{\beta_0^2}\left(\lambda_\chi+\ln(1-\lambda_\chi)\right), \tag{2.4.50}$$

$$g_2\left(\lambda_\chi,\frac{Q^2}{\mu_R^2}\right)=\frac{A_g^{(1)}\beta_1}{\beta_0^3}\left(\frac{\lambda_\chi+\ln(1-\lambda_\chi)}{1-\lambda_\chi}+\frac{1}{2}\ln(1-\lambda_\chi)^2\right)+\frac{B_g^{(1)}}{\beta_0}\ln(1-\lambda_\chi)$$
$$-\frac{A_g^{(2)}}{\beta_0^2}\frac{\lambda_\chi+(1-\lambda_\chi)\ln(1-\lambda_\chi)}{1-\lambda_\chi}+\frac{A_g^{(1)}}{\beta_0}\frac{\lambda_\chi+(1-\lambda_\chi)\ln(1-\lambda_\chi)}{1-\lambda_\chi}\ln\left(\frac{Q^2}{\mu_R^2}\right), \tag{2.4.51}$$

$$g_3\left(\lambda_{\bar{N}},\lambda_\chi,\frac{Q^2}{\mu_R^2}\right)=\frac{A_g^{(1)}\beta_1^2}{2\beta_0^4}\left(\frac{\lambda_\chi+\ln(1-\lambda_\chi)}{(1-\lambda_\chi)^2}\left(\lambda_\chi+(1-2\lambda_\chi)\ln(1-\lambda_\chi)\right)\right)$$
$$+\frac{A_g^{(1)}\beta_2}{\beta_0^3}\left(\frac{(2-3\lambda_\chi)\lambda_\chi}{2(1-\lambda_\chi)^2}+\ln(1-\lambda_\chi)\right)+\frac{B_g^{(1)}\beta_1}{\beta_0}\frac{\lambda_\chi+\ln(1-\lambda_\chi)}{1-\lambda_\chi}+A_g^{(1)}$$

$$\times \mathrm{Li}_2\left(\frac{\bar{N}^2}{\chi}\right)\frac{\lambda_{\bar{N}}}{1-\lambda_{\bar{N}}} - \frac{A_g^{(2)}\beta_1}{\beta_0^3}\left(\frac{(2-3\lambda_\chi)\,\lambda_\chi}{2\,(1-\lambda_\chi)^2} + \frac{(1-2\lambda_\chi)\ln(1-\lambda_\chi)}{(1-\lambda_\chi)^2}\right)$$
$$- \frac{A_g^{p_{\mathrm{T}},(3)}}{2\beta_0^2}\frac{\lambda_\chi^2}{(1-\lambda_\chi)^2} - \frac{B_g^{(2)}}{\beta_0}\frac{\lambda_\chi}{1-\lambda_\chi} + B_g^{(1)}\frac{\lambda_\chi}{1-\lambda_\chi}\ln\frac{m_{\mathrm{H}}^2}{\mu_{\mathrm{R}}^2} + \frac{A_g^{(2)}}{\beta_0}\frac{\lambda_\chi^2}{(1-\lambda_\chi)^2}$$
$$\times \ln\left(\frac{Q^2}{\mu_R^2}\right) + \frac{A_g^{(1)}\beta_1}{\beta_0^2}\left(\frac{\lambda_\chi\,(1-\lambda_\chi) + (1-2\lambda_\chi)\ln(1-\lambda_\chi)}{(1-\lambda_\chi)^2}\right)\ln\left(\frac{Q^2}{\mu_R^2}\right)$$
$$- \frac{A_g^{(1)}}{2}\frac{\lambda_\chi^2}{(1-\lambda_\chi)^2}\ln^2\left(\frac{Q^2}{\mu_R^2}\right). \tag{2.4.52}$$

The function $\mathcal{R}$ embodies three types of contributions up to NNLL: the exponent of the lowest order solution to the evolution given by Eq. (2.4.45), the evolution of the coefficient functions, and the evolution of the χ-dependent V-function in Eq. (2.4.43). Expressing the argument of the strong coupling in terms of Q^2, the evolution factors for the coefficient and V functions are respectively given by

$$C_{ck}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right)\right) = \exp\left\{R_{ck}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\leftarrow Q^2\right)\right)\right\}C_{ck}(N,Q^2) \tag{2.4.53}$$
$$V_{ki}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\right)\right) = \exp\left\{E_{ki}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\leftarrow Q^2\right)\right)\right\}V_{ki}(N,Q^2) \tag{2.4.54}$$

where terms in the exponents are defined as

$$R_{ck}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\leftarrow Q^2\right)\right) = -\int_{Q^2/\chi}^{Q^2}\frac{\mathrm{d}q^2}{q^2}\frac{\beta\left(\alpha_s\left(q^2\right)\right)}{\alpha_s\left(q^2\right)}\left[\frac{\mathrm{d}\ln C_{ck}\left(N,\alpha_s\left(q^2\right)\right)}{\mathrm{d}\ln\alpha_s\left(q^2\right)}\right] \tag{2.4.55}$$
$$E_{ki}\left(N,\alpha_s\left(\frac{Q^2}{\chi}\leftarrow Q^2\right)\right) = -\int_{Q^2/\chi}^{Q^2}\frac{\mathrm{d}q^2}{q^2}\frac{\beta\left(\alpha_s\left(q^2\right)\right)}{\alpha_s\left(q^2\right)}\left[\frac{\mathrm{d}\ln V_{ki}\left(N,\alpha_s\left(q^2\right)\right)}{\mathrm{d}\ln\alpha_s\left(q^2\right)}\right]. \tag{2.4.56}$$

Therefore, the function $\mathcal{R}$ that enter in Eq. (2.4.48) can now be written as

$$\mathcal{R}(N,\chi) = \mathcal{Q}_r(N,\chi) + \mathcal{Q}_p(N,\chi) + R_{ck}(N,\chi) + R_{\bar{c}l}(N,\chi) + E_{ki}(N,\chi) + E_{lj}(N,\chi) \tag{2.4.57}$$

where $\{r,p\} = \pm$ and for brevity we have neglected the explicit dependence on the strong coupling. In the standard transverse momentum resummation formalism, the logarithmic expansion of $\mathcal{S}_c$ starts at LL accuracy while the expansion of $\mathcal{Q}$ starts at NLL accuracy. However, in the context of soft-improved transverse momentum, the large-N behaviour of $\mathcal{Q}$ already contribute at LL and has to be included in order to account for the soft behaviour. Similarly to the standard resummation, the flavour off-diagonal terms of $\tilde{R}_{ci} = R_{ck} + E_{ki}$ (i.e. $c \neq i$) starts to contribute at NLL. The terms in the flavour-off diagonal, instead, start to contribute at NLL as opposed to the standard resummation procedure where they start to contribute at NNLL.

In analogy to Eq. (2.2.28), we can collect all the terms that behave as constants in the limit $\hat{b} \to \infty$. Expressing the results in terms of running $\alpha_s(Q^2)$

$$\bar{\mathcal{H}}_{ab\to c\bar{c}}^{\mathrm{F},\{\mathrm{S}\}}(N,\chi) = \bar{\mathrm{H}}_c(N,\chi)\,\mathcal{W}_{ca}^{(r)}(N,\alpha_s(Q^2))\,\mathcal{W}_{\bar{c}b}^{(p)}(N,\alpha_s(Q^2)), \tag{2.4.58}$$

where we have defined

$$\mathcal{W}_{ca}^{(r)}(N,\alpha_s(Q^2)) = C_{ck}(N,\alpha_s(Q^2))\, V_{ki}(N,\alpha_s(Q^2))\, \mathcal{E}_{ij}^{(r)}(N)\, \tilde{V}_{ja}(N,\alpha_s(Q^2)), \tag{2.4.59}$$

$$\mathcal{W}_{\bar{c}b}^{(p)}(N,\alpha_s(Q^2)) = C_{\bar{c}l}(N,\alpha_s(Q^2))\, V_{lm}(N,\alpha_s(Q^2))\, \mathcal{E}_{mn}^{(p)}(N)\, \tilde{V}_{nb}(N,\alpha_s(Q^2)). \tag{2.4.60}$$

We recall that the definition of $\bar{\mathrm{H}}$ is related to the H_c from the standard transverse momentum resummation with an additional dilogarithm function. In addition, it is worth mentioning that the definition of C-functions are exactly the same as the ones appearing in the CFG formalism.

The hard-collinear function given in Eq. (2.4.58) can be expressed in a different form. Indeed, the hard and collinear functions can be separated from the terms coming from the evolutions. That is, Eq. (2.4.58) can be written as follows

$$\bar{\mathcal{H}}^{\mathrm{F},\{\mathrm{S}\}}_{ab\to c\bar{c}} = \tilde{\mathcal{H}}^{\mathrm{F}}_{ab\to c\bar{c}} \left(V_{ki}\mathcal{E}_{ij}^{(r)}\tilde{V}_{ja}\right)\left(V_{lm}\mathcal{E}_{mn}^{(p)}\tilde{V}_{nb}\right) \quad \text{with} \quad \tilde{\mathcal{H}}^{\mathrm{F}}_{ab\to c\bar{c}} = \bar{\mathrm{H}}_c \mathrm{C}_{ck}\mathrm{C}_{\bar{c}l}, \tag{2.4.61}$$

where for brevity we have removed the explicit dependence on the Mellin moment N and the resummation variable χ. The explicit expressions of the process dependent hard functions in App. 2.B are given in terms of $\tilde{\mathcal{H}}^{\mathrm{F}}_{ab\to c\bar{c}}$, specifically in term of $\mathcal{H}^{\mathrm{F}}_{ab\to c\bar{c}} = \mathrm{H}_c\mathrm{C}_{ck}\mathrm{C}_{\bar{c}l}$ where as mentioned before $\bar{\mathrm{H}}_c$ and H_c are related through a di-logarithm function.

2.5 Summary

The aim of this chapter was to address the large logarithms appearing in fixed order computations that spoil the convergence of the perturbative series. In particular, we focused on logarithmic terms that are enhanced at small transverse momentum and at partonic threshold. After reviewing the standard formulations of the threshold and transverse momentum resummations separately, we reviewed a formalism in which these two resummations are combined in a consistent way. This led us to the formulation of the soft-improved transverse momentum resummation in which a subclass of the soft logarithms (not contained in the standard formulation) are included in the transverse momentum resummation. Such a formulation of the transverse momentum resummation entails a new way to include anomalous dimensions in the evolutions. To briefly summarize, the truncation of the soft-improved transverse momentum resummation at a given logarithmic accuracy is defined in the following way:

- At **LL**, we approximate the process-dependent hard function $\bar{\mathcal{H}}^{\mathrm{F}}_{ab\to c\bar{c}}$ by its lowest perturbative order and include g_1 in the Sudakov exponent. The $\mathcal{Q}_\pm$ functions are included up to LO with $\gamma^{(1)}$ replaced by its large-N behaviour while the $\tilde{R}$ functions are approximated to 1.

- At **NLL**, we include $\bar{\mathcal{H}}^{\mathrm{F},(1)}_{ab\to c\bar{c}}$ in $\bar{\mathcal{H}}^{\mathrm{F}}_{ab\to c\bar{c}}$ with the function g_2 in the Sudakov exponent. In addition, we include the complete LO term in $\mathrm{Q}_\pm$ together with the NLO term where the anomalous dimensions are replaced the their large-N behaviour. The full expression of $\tilde{R}_{gq}$ is also included while we only include the large-N behaviour of $V^{(1)}(N)$ in $\tilde{R}_{gg}$.

- At **NNLL**, we include in $\bar{\mathcal{H}}^{\mathrm{F}}_{ab\to c\bar{c}}$ the coefficient $\bar{\mathcal{H}}^{\mathrm{F},(2)}_{ab\to c\bar{c}}$ and the function g_3 in the Sudakov exponent. Similarly, we also include the complete NLO term in $\mathrm{Q}_\pm$ together with the NNLO term where the anomalous dimensions are replaced by their large-N behaviour. Finally, we add to $\tilde{R}_{gg}$ the large-N behaviour of $V^{(2)}(N)$.

Phenomenological studies assessing the impacts of the improved transverse momentum resummation is the subject of the next chapter using the Higgs and Z-boson production as test cases. We recall that herein and in the following chapter, both the small-p_T and large-x logarithms are resummed up to NNLL. In the following Appendix, we give the explicit expressions of the process-dependent functions that appear in the resummed expressions for the two processes mentioned earlier. We present analytical expressions first for the threshold resummation, and then for the soft-improved transverse momentum resummation.

2.A Process-dependent functions in Threshold resummation

As described in Sec. (2.3.2), the threshold resummed expression contains four terms that are process-dependent: the Born level cross section $\sigma^{\mathrm{Born}}_{\mathrm{F},c\bar{c}}$, the leading-order p_T distribution $C^{(1)}_{abc}$, the matching function g_{ab}, and the $\mathcal{D}_{abc}$ that appear in the exponent of the large-angle radiative factor.

Before presenting the explicit expressions of the four process-dependent terms mentioned earlier, let us first present the explicit expressions of the coefficients of the cusp anomalous dimensions $A_c(\alpha_s)$ and $B_c(\alpha_s)$. So far, we used the same notation of cusp anomalous dimensions for the threshold and (soft-improved) small-p_T resummations. However, at three loop, the coefficient $A^{(3)}_c$ appearing in the threshold resummation differs from the one appearing in the (soft-improved) small-p_T resummations. Refer to App. 2.B for the relation between the cusp anomalous dimensions in the two resummations.

In order to reach NNLL accuracy, cusp anomalous dimensions need to be determined up to three loop. In the context of threshold resummation, the coefficients of the cups anomalous dimension A_c are given by:

$$A^{(1)}_c = \frac{C_c}{\pi}, \qquad A^{(2)}_c = \frac{C_c}{2\pi^2}\left[C_A\left(\frac{67}{18}-\frac{\pi^2}{6}\right)-\frac{5}{9}n_f\right] \tag{A.1}$$

$$\begin{aligned} A^{(3)}_c = \frac{C_c}{4\pi^3}\Bigg[&C_A^2\left(\frac{245}{24}-\frac{67}{9}\zeta_2+\frac{11}{6}\zeta_3+\frac{11}{5}\zeta_2^2\right)\\ &+n_f C_F\left(-\frac{55}{24}+2\zeta_3\right)+n_f C_A\left(-\frac{209}{108}+\frac{10}{9}\zeta_2-\frac{7}{3}\zeta_3\right)-\frac{1}{27}n_f^2\Bigg], \end{aligned} \tag{A.2}$$

while the coefficients of the cusp anomalous dimension B_c are given by:

$$B^{(1)}_g = -\beta_0, \qquad B^{(1)}_q = -\frac{3}{4}\frac{C_F}{\pi} \tag{A.3}$$

$$B^{(2)}_g = \frac{1}{16\pi^2}\left[C_{\mathrm{A}}^2\left(-\frac{611}{9}+\frac{88}{3}\zeta_2+16\zeta_3\right)+C_{\mathrm{A}}n_f\left(\frac{428}{27}-\frac{16}{3}\zeta_2\right)+2C_{\mathrm{F}}n_f-\frac{20}{27}N_f^2\right]$$

$$B^{(2)}_q = \frac{1}{16\pi^2}\left[C_{\mathrm{F}}^2\left(-\frac{3}{2}+12\zeta_2-24\zeta_3\right)+C_{\mathrm{F}}C_{\mathrm{A}}\left(\frac{44}{3}\zeta_2+40\zeta_3\right)+C_{\mathrm{F}}N_f\left(\frac{247}{27}-\frac{8}{3}\zeta_2\right)\right]$$

2.A.1 Higgs production via gluon fusion in HEFT

In this section, we present the analytical expressions of the process-dependent functions that enter in Eq. (2.3.20) in the case of the Higgs boson production. The Born level cross section for the Higgs production via gluon fusion has been computed long time ago in Refs. [239,240] in the infinite top-quark mass limit and is given by

$$\sigma^{\text{Born}}_{\text{H},gg} = \left(\frac{\alpha_s}{\pi}\right) \frac{\pi\sqrt{2}G_F}{576}, \tag{A.4}$$

where G_F is the Fermi's constant. The Mellin space version of the leading-order p_T-distribution $C^{(1)}_{abc}$ has been first computed in Ref. [227] and recomputed in Ref. [165]. Recall that in our notation, the subscript (abc) is shorthand notation for $ab \to c\,\text{H}$. They can be written in the following generic form:

$$C^{(1)}_{abc}(N,\xi_p) = \frac{2\alpha_s C_A}{\xi_p \pi} \sum_{m=0}^{4} (-1)^m f^{(m)}_{ab}\left(\xi_p\right) \frac{\Gamma(N+m)}{\Gamma(N+m+1/2)} \times \Gamma\left(\frac{1}{2}\right) {}_2F_1\left(\frac{1}{2}, N+m, N+m+\frac{1}{2}, a^2\right), \tag{A.5}$$

where $a \equiv a(\xi_p) = \left(\sqrt{1+\xi_p}+\sqrt{\xi_p}\right)^{-2}$ and the functions $f^{(m)}_{ab}$ fully depend on the dimensionless variable ξ_p. For the various partonic channels, they are defined as follow

- $(gg \to g\,\text{H})$-channel:

$$f^{(0)}_{gg}\left(\xi_p\right) = 1 \tag{A.6}$$
$$f^{(1)}_{gg}\left(\xi_p\right) = 2a(\xi_p)\left(1+\xi_p\right) \tag{A.7}$$
$$f^{(2)}_{gg}\left(\xi_p\right) = a^2(\xi_p)\left(1+\xi_p\right)\left(3+\xi_p\right) \tag{A.8}$$
$$f^{(3)}_{gg}\left(\xi_p\right) = 2a^3(\xi_p)\left(1+\xi_p\right) \tag{A.9}$$
$$f^{(4)}_{gg}\left(\xi_p\right) = a^4(\xi_p) \tag{A.10}$$

- $(gq \to g\,\text{H})$-channel:

$$f^{(0)}_{gq}\left(\xi_p\right) = 1 \tag{A.11}$$
$$f^{(1)}_{gq}\left(\xi_p\right) = a(\xi_p)\left(1+\xi_p\right) \tag{A.12}$$
$$f^{(2)}_{gq}\left(\xi_p\right) = 3a^2(\xi_p)\left(1+\xi_p\right)\left(3+\xi_p\right) \tag{A.13}$$
$$f^{(3)}_{gq}\left(\xi_p\right) = a^3(\xi_p)\left(1+\xi_p\right) \tag{A.14}$$
$$f^{(4)}_{gq}\left(\xi_p\right) = 0 \tag{A.15}$$

- $(qq \to g\,\text{H})$-channel:

$$f^{(0)}_{qq}\left(\xi_p\right) = 1 \tag{A.16}$$

$$f_{qq}^{(1)}\left(\xi_p\right) = 2a(\xi_p)\left(1+\xi_p\right) \tag{A.17}$$

$$f_{qq}^{(2)}\left(\xi_p\right) = a^2(\xi_p)\left(1+\xi_p\right)\left(3+\xi_p\right) \tag{A.18}$$

$$f_{qq}^{(3)}\left(\xi_p\right) = f_{gq}^{(4)}\left(\xi_p\right) = 0 \tag{A.19}$$

The matching function g_{ab} that collects the N-independent contributions has been first determined in Ref. [227] by comparing the expansion of the threshold resummed expression with the fixed-order calculations. That is, once the resummed expression and the fixed-order results are truncated at the same perturbative order, one removes all the N-dependent terms and collects the rest into the function g_{ab}. The expression of g_{ab} has been later recomputed using the definition of kinematics introduced in Sec. (1.1.1). In momentum space, the matching function is perturbatively defined as

$$g_{ab}^{\mathrm{H}} = 1 + \sum_{n=1}^{\infty}\left(\frac{\alpha_s}{\pi}\right)^n g_{ab}^{(n),\mathrm{H}}, \tag{A.20}$$

where we have made the dependence on the final system H explicit. At the logarithmic accuracy we are interested in, i.e. NNLL, only the first coefficient $g_{ab}^{(1),\mathrm{H}}$ is required. Their expression for the different channels are given by

$$\begin{aligned}
g_{gg}^{(1),\mathrm{H}}\left(\xi_p\right) = {} & \frac{67}{36}\mathrm{C_A} - \frac{5}{18}n_f + \mathrm{C_A}\zeta_2 - \beta_0 \ln\frac{\xi_p}{1+\xi_p} - \frac{1}{8}\mathrm{C_A}\ln^2\frac{\xi_p}{1+\xi_p} \\
& + 2\mathrm{C_A}\mathrm{Li}_2\left(1-\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) + \mathrm{C_A}\ln\left(1-\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)\ln\frac{\xi_p}{1+\xi_p} \\
& - \frac{1}{2}\mathrm{C_A}\ln\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)\ln\frac{\xi_p}{1+\xi_p} + \frac{1}{2}\mathrm{C_A}\ln^2\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) \\
& + \mathrm{C_A}\mathrm{Li}_2\left(\frac{2\sqrt{\xi_p}}{\sqrt{1+\xi_p}+\sqrt{\xi_p}}\right) - \left(\mathrm{C_A}-n_f\right)\frac{\sqrt{\xi_p}\sqrt{1+\xi_p}\left(1+\xi_p\right)}{6\left(1+8\xi_p+9\xi_p^2\right)} \\
& + 2\beta_0\ln^2\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) + \left(\mathrm{C_A}-n_f\right)\frac{2\xi_p+\xi_p^2}{6\left(1+8\xi_p+9\xi_p^2\right)}
\end{aligned} \tag{A.21}$$

$$\begin{aligned}
g_{gq}^{(1),\mathrm{H}}\left(\xi_p\right) = {} & -\frac{7}{4}\mathrm{C_F} + \frac{134}{36}\mathrm{C_A} - \frac{20}{36}n_f - 8\mathrm{C_F}\zeta_2 + 12\mathrm{C_A}\zeta_2 - 4\beta_0\ln\frac{\xi_p}{1+\xi_p} \\
& - \frac{1}{2}\mathrm{C_A}\ln^2\frac{\xi_p}{1+\xi_p} + 4\left(\mathrm{C_F}+\mathrm{C_A}\right)\mathrm{Li}_2\left(2,1-\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) + \frac{3}{2}\mathrm{C_F}\ln\frac{\xi_p}{1+\xi_p} \\
& + \frac{2\left(\mathrm{C_A}-\mathrm{C_F}\right)\left(1+3\xi_p+3\sqrt{\xi_p}\sqrt{1+\xi_p}\right)}{2\sqrt{\xi_p}\sqrt{1+\xi_p}+1+3\xi_p} + 8\beta_0\ln\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) \\
& - 3\mathrm{C_F}\ln\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) + 2\mathrm{C_F}\ln\left(1-\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)\ln\frac{\xi_p}{1+\xi_p} \\
& + 2\mathrm{C_A}\ln\left(1-\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)\ln\frac{\xi_p}{1+\xi_p} - 2\mathrm{C_F}\ln\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)\ln\frac{\xi_p}{1+\xi_p}
\end{aligned}$$

$$-2C_F \ln^2\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)+4C_F \mathrm{Li}_2\left(\frac{2\sqrt{\xi_p}}{\sqrt{1+\xi_p}+\sqrt{\xi_p}}\right) \tag{A.22}$$

$$\begin{aligned} g_{qq}^{(1),H}(\xi_p) = &-\frac{9}{2}C_F+\frac{79}{12}C_A-\frac{5}{6}n_f+12C_F\zeta_2-10C_A\zeta_2-\frac{(C_F-C_A)\sqrt{1+\xi_p}}{\sqrt{\xi_p}} \\ &+4C_F \mathrm{Li}_2\left(1-\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)-\frac{3}{4}C_F \ln\frac{\xi_p}{1+\xi_p}-\beta_0 \ln\frac{\xi_p}{1+\xi_p}+\frac{1}{4}C_A \ln^2\frac{\xi_p}{1+\xi_p} \\ &-\frac{1}{2}C_F \ln^2\frac{\xi_p}{1+\xi_p}+2C_F \ln\left(1-\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)\ln\frac{\xi_p}{1+\xi_p}+\frac{3}{2}C_F \ln\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) \\ &+2\beta_0 \ln\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)+C_A \ln^2\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right) \\ &-C_A \ln\left(1+\frac{\sqrt{\xi_p}}{\sqrt{1+\xi_p}}\right)\ln\frac{\xi_p}{1+\xi_p}+2C_A \mathrm{Li}_2\left(\frac{2\sqrt{\xi_p}}{\sqrt{1+\xi_p}+\sqrt{\xi_p}}\right) \end{aligned} \tag{A.23}$$

2.A.2 Vector boson production via DY mechanism

In this section, we describe the approach we took in order to implement the threshold resummed expressions for the production of a Z boson via DY mechanism. The Born level cross section for such a process is given by:

$$\sigma^{\mathrm{Born}}_{Z,q\bar{q}} = \frac{Q^4}{\pi m_Z}\frac{\Gamma_Z B\left(Z\to \ell\bar{\ell}\right)}{\left(Q^2-m_Z^2\right)^2+m_Z^2\Gamma_Z^2}, \tag{A.24}$$

where m_Z denotes the mass of the Z boson, $\Gamma_Z = 2.4952$ GeV represents the distribution width, finally $\mathrm{B}\left(Z\to\ell\bar{\ell}\right)$ represents the branching ratio and depends on the decay mode. For instance, $B\left(Z\to e^+e^-\right)=0.03363$, $B\left(Z\to\mu^+\mu^-\right)=0.03366$.

The expressions of the leading-order p_T distributions can be extracted from [241]. These can be used to compute the analytical expression of the $C^{(1)}_{abc}$-functions in Mellin space in the same way as for the Higgs boson case. However, these Mellin space expressions can also be computed numerically using the DYQT [242, 243] code using similar procedure as the one described in App. 3.B. In the subsequent studies, we opted for the second approach.

2.B Process-dependent functions in SIpT

In this section, we write down explicitly the expression of the process-dependent coefficients that enter into the definition of the soft-improved transverse momentum resummation as given in Eq. (2.4.42) or Eq. (2.4.48). Specifically, the process-dependent terms are collected in the perturbative function $\bar{\mathcal{H}}^{\mathrm{F}}_{ab\to c\bar{c}}$ which depends on the hard function $\bar{\mathrm{H}}_c$ and the coefficient functions C_{ab}. Recall that the partonic functions C_{ab} are process independent as a consequence of the universality of the QCD collinear radiation. Due to the charge invariance and the flavour symmetry of QCD, they satisfy the following relations

$$C_{q_iq_j}(z) = C_{\bar{q}_i\bar{q}_j}(z) = C_{qq}\delta_{ij}+C_{qq'}(1-\delta_{ij}) \tag{B.1}$$

$$C_{q_i\bar{q}_j}(z) = C_{\bar{q}_i q_j}(z) = C_{q\bar{q}}\delta_{ij} + C_{q\bar{q}'}(1-\delta_{ij}) \tag{B.2}$$

As pointed out in Sec. (2.2.2), the independent determination of the hard function H_c and the coefficient functions C_{ab} requires the specification of a resummation scheme as the two functions are not separately computable in an unambiguous way. However, the resulting $\mathrm{H}_c \times C_{ac} \times C_{b\bar{c}}$ must be scheme independent. In the followings, the expression of the hard and coefficients will be presented in direct space x, their Mellin transform can be computed either analytically or numerically using various routines. In addition, it was stressed that the soft-improved transverse momentum resummation is expressed in terms of $\bar{\mathrm{H}}_c$ instead of H_c. In the following derivations, however, we will only present results in terms of H_c taking into account that the difference between the two expressions is only a dilogarithm function that depend on the impact parameter but vanishes in the large-b limit.

As in the previous section, we are first going to present results for the Higgs boson production and then afterward present results for the DY case. But first, let us explicitly present the expressions of the coefficients of the cusp anomalous dimensions that appear in the (soft-improved) small-p_T resummation. We emphasized in the previous descriptions that those coefficients are closely related to the ones appearing in the threshold resummation. In the context of transverse momentum resummation, they are related to coefficients appearing in App. 2.A as follows:

$$\begin{aligned}
B_{c,\star}^{(1)} &= 2B_c^{(1)}, \quad A_{c,\star}^{(3)} = A_{c,\star}^{(3)} - \frac{C_c}{\pi^3}\left[C_A^2\left(\frac{77}{24}\zeta_3 - \frac{1111}{324}\right) + n_f C_A\left(\frac{89}{81} - \frac{7}{12}\zeta_3\right) + \frac{11}{162}n_f^2\right] \\
B_{g,\star}^{(2)} &= B_g^{(2)} - \frac{1}{\pi^2}\left[C_A^2\left(\frac{5}{2}\zeta_3 + \frac{11}{12}\zeta_2 - \frac{419}{144}\right) - n_f C_A\left(\frac{\zeta_2}{6} - \frac{71}{108}\right) - \frac{n_f}{216}\left(27C_F + 10n_f\right)\right] \\
B_{q,\star}^{(2)} &= B_q^{(2)} - \frac{1}{\pi^2}\left[C_F^2\left(\frac{3}{2}\zeta_3 - \frac{3}{4}\zeta_2 + \frac{3}{32}\right) + C_A C_F\left(\frac{35}{16} - \frac{11}{12}\zeta_2 - \frac{3}{2}\zeta_3\right)\right. \\
&\qquad \left. + n_f C_A\left(\frac{5}{2}\zeta_3 + \frac{5}{288}\zeta_2 - \frac{209}{6912}\right) + n_f C_F\left(\frac{229}{432} - \frac{\zeta_2}{3}\right)\right],
\end{aligned} \tag{B.3}$$

The above expressions represent the relation between the coefficients appearing in the (soft-improved) small-p_T resummation (indicated by the $\star$ subscript) and the the ones appearing in the threshold resummation as given in App. 2.A.

2.B.1 Higgs production via gluon fusion in HEFT

In the heavy top quark mass limit, the leading order hard and coefficient functions satisfy the following relation [197, 244]

$$\mathcal{H}_{ab\to gg}^{\mathrm{H},(1)}(x) = \delta_{ga}\delta_{gb}\delta(1-x)\mathrm{H}_g^{\mathrm{H}(1)} + \delta_{ga}C_{gb}^{(1)}(x) + \delta_{gb}C_{ga}^{(1)}(x), \tag{B.4}$$

where $\mathcal{H}_{ab\to gg}^{\mathrm{H},(1)}$ is the leading order perturbative coefficient of the $\mathcal{H}_{ab\to gg}^{\mathrm{H}}$ function (given in Eq. (2.4.61)) which in literature is often referred to as the hard-collinear coefficient function. The determination of the $\mathcal{H}_{ab\to gg}^{\mathrm{H},(n)}$ coefficient is unambiguous as it does not depend on the resummation scheme. For example, its leading order coefficient gg-contribution is given by

$$\mathcal{H}_{gg\to gg}^{\mathrm{H},(1)}(x) \equiv \left[\left(\frac{5+\pi^2}{2\pi}\right)C_A - \frac{3}{2\pi}C_F\right]\delta(1-x) = \delta(1-x)\mathrm{H}_g^{(1)} + 2C_{gg}^{(1)}(x). \tag{B.5}$$

We can for instance choose a resummation scheme where the hard function $\mathrm{H}_g^{(1)}$ vanishes in which case the coefficient function is equal to $\mathcal{H}^{\mathrm{H},(1)}_{gg\to gg}(x)$. Notice that the leading-order coefficient of the gq-channel is process-independent and is expressed as $C^{(1)}_{gq}(x) = xC_F/2$. This suggests that with the knowledge of the hard-collinear function at $\mathcal{O}(\alpha_s^n)$, we can extract the expression of the coefficients C_{ab} up to $\mathcal{O}(\alpha_s^n)$. Neglecting the spin correlations, the NLO hard-collinear coefficient function is expressed as

$$\begin{aligned}\mathcal{H}^{\mathrm{H},(2)}_{ab\to gg}(x) = \delta_{ga}\delta_{gb}\delta(1-z)\mathrm{H}_g^{\mathrm{H},(2)} + \delta_{ga}C^{(2)}_{gb}(z) \\ +\delta_{gb}C^{(2)}_{ga}(z) + \mathrm{H}_g^{\mathrm{H},(1)}\left(\delta_{ga}C^{(1)}_{gb}(z) + \delta_{gb}C^{(1)}_{ga}(z)\right).\end{aligned} \tag{B.6}$$

Knowing both Eq. (B.5) and Eq. (B.7) allows one to compute the NLO coefficient functions. The two coefficients for the gg and gq-channels satisfy the following relations

$$\begin{aligned}C^{(2)}_{gq}(x) + \frac{1}{4}x\mathrm{H}_g^{\mathrm{H},(1)}C_F = \mathcal{H}^{\mathrm{H},(2)}_{gq\to gg}(x) + \frac{3}{8}xC_F^2 + C_FC_A\frac{1}{x}\times \\ \left\{(1+x)\ln x + 2(1-x) - \frac{5+\pi^2}{8}x^2\right\},\end{aligned} \tag{B.7}$$

$$\begin{aligned}2C^{(2)}_{gg}(x) + \delta(1-x)\left\{\mathrm{H}_g^{\mathrm{H},(2)} - \frac{3}{4}\left(\mathrm{H}_g^{\mathrm{H},(1)}\right)^2 + \frac{\left(5+\pi^2\right)C_A - 3C_F}{4}\mathrm{H}_g^{\mathrm{H},(1)}\right\} = \\ \mathcal{H}^{\mathrm{H}(2)}_{gg\to gg}(x) - \delta(1-x)\left\{\frac{\left(5+\pi^2\right)C_A - 3C_F}{4}\right\}^2 + C_A^2\left\{\left(\frac{1+x}{x}\right)\ln x + 2\left(\frac{1-x}{2}\right)\right\},\end{aligned} \tag{B.8}$$

where the hard-collinear coefficient functions are expressed as [197]

$$\mathcal{H}^{\mathrm{H},(2)}_{qq\to gg}(x) = -C_F^2\left\{\frac{2(1-x)}{x} + \frac{(2+x)^2}{4x}\ln x\right\} \tag{B.9}$$

$$\begin{aligned}\mathcal{H}^{\mathrm{H},(2)}_{gq\to gg}(x) = \frac{C_F^2}{\pi^2}\left\{\frac{1}{48}(2-x)\ln^3 x - \frac{1}{32}(3x+4)\ln^2 x + \frac{5}{16}(x-3)\ln x + \frac{5}{8} - \frac{13}{16}x\right. \\ + \frac{1}{12}\left(\frac{1}{x} + \frac{x}{2} - 1\right)\ln^3(1-x) + \frac{1}{16}\left(x + \frac{6}{x} - 6\right)\ln^2(1-x) + \left(\frac{5x}{2} + \frac{8}{x} - 8\right) \\ \left.\times\frac{1}{4}\ln(1-x)\right\} + n_f\frac{C_F}{\pi^2}\left\{\frac{1}{24x}\left(1+(1-x)^2\right)\ln^2(1-x) + \frac{1}{18}\left(x + \frac{5}{x} - 5\right)\right. \\ \left.\times\ln(1-x) - \frac{14}{27} + \frac{14}{27x} + \frac{13}{108}x\right\} + \frac{C_FC_A}{\pi^2}\left\{-\frac{\left(1+(1+x)^2\right)}{2x}\mathrm{Li}_3\left(\frac{1}{1+x}\right)\right. \\ + \left(\frac{1}{2} - \frac{5}{2x} - \frac{5}{4}x\right)\mathrm{Li}_3(x) - \frac{3}{4x}\left(1+(1+x)^2\right)\mathrm{Li}_3(-x) + \left(\frac{1}{12x} + \frac{(1+x)^2}{12x}\right) \\ \times\ln^3(1+x) + \left(2 - \frac{11}{6x} - \frac{x}{2} + \frac{x^2}{3} + \left(-\frac{1}{2} + \frac{3}{2x} + \frac{3x}{4}\right)\ln x\right)\mathrm{Li}_2(x) + \mathrm{Li}_2(-x) \\ \times\left(\frac{x}{4} + \frac{\left(1+(1+x)^2\right)}{4x}\ln(x)\right) - \left(\frac{1}{24x} + \frac{(1+x)^2}{24x}\left(\frac{\ln^2}{3}x + \zeta_2\right) - x^2\ln x\right) \\ \left.\times\ln(1+x) - \frac{\left(1+(1-x)^2\right)}{24x}\ln^3(1-x) + \frac{6}{48x}\left(\left(1+(1-x)^2\right)\ln x - \frac{5}{6}x^2\right)\right.\end{aligned}$$

$$\times \ln^2(1-x) - \frac{22}{48x}(1-x)\ln^2(1-x) + \frac{1}{72x}\left(-152+152x-43x^2-132\right)$$
$$+ \frac{6}{72x}\left(\left(24x-9x^2+4x^3\right)\ln x\right)\ln(1-x) + \frac{9\ln^2 x}{72x}\left(1+(1-x)^2\right)\ln(1-x)$$
$$- \frac{1}{12}\left(1+\frac{x}{2}\right)\ln^3 x + \frac{1}{48}\left(36+9x+8x^2\right)\ln^2 x + \frac{1007}{108} - \frac{5}{2}\zeta_3 - \frac{\pi^2}{3}\ln x$$
$$\times\left(-\frac{107}{24} - \frac{1}{x} + \frac{x}{12} - \frac{11}{9}x^2\right) + \frac{1}{x}\left(4\zeta_3 - \frac{503}{54} + \frac{11}{36}\pi^2\right) + x^2\left(\frac{38}{27} - \frac{\pi^2}{18}\right)$$
$$+x\left(\frac{\pi^2}{3} + 2\zeta_3 - \frac{133}{108}\right)\Bigg\}, \tag{B.10}$$

$$\mathcal{H}^{\mathrm{H},(2)}_{gg\to gg}(x) = \left(\left(-\frac{101}{27}+\frac{7}{2}\zeta_3\right)C_A^2 + \frac{14}{27}C_A n_f\right)\left(\frac{1}{1-x}\right)_+ - \delta(1-x)\left(\frac{145}{24}+\frac{3}{4}\pi^2\right)$$
$$\times C_A C_F + \delta(1-x)\left(C_A^2\left(\frac{3187}{288}+\frac{157}{72}\pi^2+\frac{13}{144}\pi^4-\frac{55}{18}\zeta_3\right)+\frac{9}{4}C_F^2-\frac{5}{96}C_A\right.$$
$$\left.-\frac{1}{12}C_F - C_A n_f\left(\frac{287}{144}+\frac{5}{36}\pi^2+\frac{4}{9}\zeta_3\right)+C_F n_f\left(-\frac{41}{24}+\frac{1}{2}L_t+\zeta_3\right)\right)+2C_A^2$$
$$\times\left(\frac{\left(1+x+x^2\right)^2}{x(1+x)}\left(\mathrm{Li}_3\left(\frac{x}{1+x}\right)-\mathrm{Li}_3(-x)\right)+\frac{2-17x-22x^2-10x^3-12x^4}{2x(1+x)}\right.$$
$$\times\zeta_3 - \frac{5-x+5x^2+x^3-5x^4+x^5}{x(1-x)(1+x)}\left(\mathrm{Li}_3(x)-\zeta_3\right)+\mathrm{Li}_2(x)\frac{\ln(x)}{1-x}\frac{3-x+3x^2}{x(1+x)}$$
$$+\frac{\left(1+x+x^2\right)^2}{x(1+x)}\left(\ln(x)\mathrm{Li}_2(-x)-\frac{1}{3}\ln^3(1+x)+\zeta_2\ln(1+x)\right)+\mathrm{Li}_2(x)\frac{\ln(x)}{1-x}$$
$$\times\frac{x^3-3x^4+x^5}{x(1+x)}+\frac{1-x}{3x}\left(11-x+11x^2\right)\mathrm{Li}_2(1-x)+\frac{1}{12}x\ln(1-x)-\frac{1}{6}\frac{\ln^3(x)}{1-x}$$
$$+\ln^2(x)\left(\frac{\left(1-x+x^2\right)^2}{2x(1-x)}\ln(1-x)-\frac{\left(1+x+x^2\right)^2}{2x(1+x)}\ln(1+x)+\frac{25-11x+44x^2}{24}\right)$$
$$-\mathrm{Li}_2(x)\frac{\ln(x)}{1-x}\frac{\left(2+2x-x^2\right)^2}{1+x}+\ln(x)\left(\frac{\left(1+x+x^2\right)^2}{x(1+x)}\ln^2(1+x)+\frac{\left(1-x+x^2\right)^2}{2x(1-x)}\right.$$
$$\left.\times\ln^2(1-x)-\frac{72+773x+149x^2+536x^3}{72x}\right)+\frac{517}{27}-\frac{449}{27x}-\frac{380x}{27}+\frac{835x^2}{54}\Bigg)+$$
$$n_f C_F\left(\frac{1+x}{12}\ln^3(x)+\frac{1}{8}(3+x)\ln^2(x)+\frac{3}{2}(1+x)\ln(x)-\frac{1-x}{6x}\left(1-23x+x^2\right)\right) \tag{B.11}$$

2.B.2 Vector boson production via DY mechanism

Let us now move to the DY case where as before we consider the production of a vector boson V (in the subsequent application, $V=Z$). Similar to the case of the Higgs boson production, one has to define a resummation scheme in which the hard and coefficient

functions are determined simultaneously. At leading-order, the two perturbative functions H_c and C_{ab} fulfill the following relations

$$\mathcal{H}^{\mathrm{DY},(1)}_{ab\to q\bar{q}}(x) = \delta_{qa}\delta_{\bar{q}b}\delta(1-x)\mathrm{H}^{\mathrm{DY},(1)}_{q} + \delta_{qa}C^{(1)}_{\bar{q}b}(x) + \delta_{\bar{q}b}C^{(1)}_{qa}(x) \tag{B.12}$$

from which one can derive the subsequent definitions [201]

$$C^{(1)}_{gq}(x) = \frac{1}{2}xC_F, \qquad C^{(1)}_{qq}(x) = \frac{1}{x}C_F(1-x), \qquad C^{(1)}_{qg}(x) = \frac{1}{2}x(1-x), \tag{B.13}$$

where the remaining coefficient functions, i.e. $q\bar{q}, qq'$ and $q\bar{q}'$ vanish. It is worth emphasizing that the expression of the gq-coefficient is exactly the same as for the case of Higgs. Using the definition of $\mathcal{H}^{\mathrm{DY},(1)}_{ab\to q\bar{q}}$ found in Ref. [201], it follows directly that the hard function $\mathrm{H}^{\mathrm{DY},(1)}_{q}$ can be expressed in the following form

$$\mathrm{H}^{\mathrm{DY},(1)}_{q} = C_F\left(\frac{\pi}{4}-4\right). \tag{B.14}$$

Proceeding exactly as before, the coefficient functions can be computed from the hard-collinear functions. From the perturbative expansion of $\tilde{\mathcal{H}}^{\mathrm{DY}}_{ab\to q\bar{q}}$, it was shown that the second-order contribution satisfy the following relations

$$\begin{aligned}\mathcal{H}^{\mathrm{DY},(2)}_{ab\to q\bar{q}}(x) = {}& \delta_{qa}\delta_{\bar{q}b}\delta(1-x)\mathrm{H}^{\mathrm{DY},(2)}_{q} + \delta_{qa}C^{(2)}_{\bar{q}b}(x) + \delta_{\bar{q}b}C^{(2)}_{qa}(x) \\ &+ \mathrm{H}^{\mathrm{DY},(1)}_{q}\left(\delta_{qa}C^{(1)}_{\bar{q}b}(x) + \delta_{\bar{q}b}C^{(1)}_{qa}(x)\right) + \left(C^{(1)}_{qa}\otimes C^{(1)}_{\bar{q}b}\right)(x),\end{aligned} \tag{B.15}$$

where the symbol $\otimes$ is used to denote a convolution that defines the integral

$$\left(C^{(1)}_{qa}\otimes C^{(1)}_{\bar{q}b}\right)(x) \equiv \int_0^1 dx_1\int_0^1 dx_2\,\delta\left(x-x_1x_2\right)C^{(1)}_{qa}\left(x_1\right)C^{(1)}_{\bar{q}b}\left(x_2\right). \tag{B.16}$$

One can then recast the expression of the second-order contributions of the hard and coefficient functions. They can be collected from Ref. [201] and are given by

$$\begin{aligned}C^{(2)}_{qg}(x) + \frac{1}{4}\mathrm{H}^{\mathrm{DY},(1)}_{q}x(1-x) = {}& \mathcal{H}^{\mathrm{DY},(2)}_{qg\to q\bar{q}}(x) - \frac{C_F}{4}\times \\ & \left[x\ln(x) + \frac{1}{2}\left(1-x^2\right) + \left(\frac{\pi^2}{2}-4\right)x(1-x)\right],\end{aligned} \tag{B.17}$$

$$2C^{(2)}_{qq}(x) + \delta(1-x)\left[\mathrm{H}^{\mathrm{DY},(2)}_{q} - \frac{3}{4}\left(\mathrm{H}^{\mathrm{DY},(1)}_{q}\right)^2 + \frac{C_F}{4}\left(\pi^2-8\right)\mathrm{H}^{\mathrm{DY},(1)}_{q}\right] + \frac{1}{2}C_F\times \tag{B.18}$$

$$\mathrm{H}^{\mathrm{DY},(1)}_{q}(1-z) = \mathcal{H}^{\mathrm{DY},(2)}_{q\bar{q}\to q\bar{q}}(x) - \frac{C_F^2}{4}\left[\delta(1-x)\frac{\left(\pi^2-8\right)^2}{4} + \left(\pi^2-10\right)(1-x) - (1+x)\ln z\right]$$

Notice that the terms on the right-hand side of the above equations are resummation-scheme independent. The flavour-off diagonal quark-antiquark channels are given by the following expressions

$$C^{(2)}_{q\bar{q}}(x) = \mathcal{H}^{\mathrm{DY},(2)}_{qq\to q\bar{q}}(x), \qquad C^{(2)}_{qq'}(x) = \mathcal{H}^{\mathrm{DY},(2)}_{q\bar{q}'\to q\bar{q}}(x), \qquad C^{(2)}_{q\bar{q}'}(x) = \mathcal{H}^{\mathrm{DY},(2)}_{qq'\to q\bar{q}}(x) \tag{B.19}$$

where the next-to-leading order hard function is given by

$$\mathrm{H}_q^{(2)} = C_F C_A \left(\frac{59\zeta_3}{18} - \frac{1535}{192} + \frac{215\pi^2}{216} - \frac{\pi^4}{240}\right) + \frac{1}{864} C_F n_f \left(192\zeta_3 + 1143 - 152\pi^2\right) \frac{1}{4} C_F^2 \left(-15\zeta_3 + \frac{511}{16} - \frac{67\pi^2}{12} + \frac{17\pi^4}{45}\right). \tag{B.20}$$

For the case of a vector boson production via DY mechanism, the hard-collinear coefficients of the quark-antiquark initiated processes at NLO is given by five different initial states coming from the partonic channels $qq, qq', q\bar{q}, qg$ and $q\bar{q}'$. They are defined in Refs. [184,201] and reproduced here for convenience

$$\mathcal{H}^{\mathrm{DY},(2)}_{gg\to q\bar{q}}(x) = -\frac{x}{2}\left(1 - x + \frac{1}{2}(1+x)\ln(x)\right) \tag{B.21}$$

$$\begin{aligned}
\mathcal{H}^{\mathrm{DY},(2)}_{q\bar{q}\to q\bar{q}}(x) &= C_A C_F \left\{\left(\frac{7\zeta_3}{2} - \frac{101}{27}\right)\left(\frac{1}{1-x}\right)_+ + \left(\frac{59\zeta_3}{18} - \frac{1535}{192} + \frac{215\pi^2}{216} - \frac{\pi^4}{240}\right)\tilde{\delta}_x \right. \\
&+ \frac{1+x^2}{1-x}\left(\mathrm{Li}_3(x) - \frac{\mathrm{Li}_3(1-x)}{2} - \frac{\mathrm{Li}_2(x)\ln(x)}{2} - \frac{\mathrm{Li}_2(x)}{2}\ln(1-x) - \frac{\ln(x)}{2}\ln^2(1-x)\right. \\
&\left. - \frac{1}{24}\ln^3(x) + \frac{\pi^2}{12}\ln(1-x) - \frac{\pi^2}{8}\right) + \frac{1}{1-x}\left(\frac{(3x^2-11)}{4}\zeta_3 - \left(\frac{12z - x^2 + 11}{48}\right)\ln^2(x)\right. \\
&\left. - \left(\frac{83x^2 - 36x + 29}{36}\right)\ln(x) + \frac{\pi^2 x}{4}\right) + (1-x)\left(\frac{\mathrm{Li}_2(x)}{2} + \ln(1-x)\frac{\ln(x)}{2}\right) + \frac{100}{27} + \frac{x}{27} \\
&\left. + \frac{x}{4}\ln(1-x)\right\} + C_F n_f \left\{\frac{14}{27}\left(\frac{1}{1-x}\right)_+ + \left(\frac{192\zeta_3 + 1143 - 152\pi^2}{864}\right)\delta(1-x) - \frac{19x+37}{108}\right. \\
&\left. + \frac{(1+x^2)}{72(1-x)}\ln(x)(3\ln(x)+10)\right\} + C_F^2\left[\frac{1}{4}\left(-15\zeta_3 + \frac{511}{16} - \frac{67\pi^2}{12} + \frac{17\pi^4}{45}\right)\delta(1-x)\right. \\
&+ \frac{1+x^2}{1-x}\left(\frac{\mathrm{Li}_3(1-x)}{2} - \frac{5\mathrm{Li}_3(x)}{2} + \frac{\mathrm{Li}_2(x)}{2}\ln(1-x) + \frac{3\mathrm{Li}_2(x)\ln(x)}{2} + \frac{\ln(x)}{4}\ln^2(1-x)\right. \\
&\left. + \frac{\ln^2(x)}{4}\ln(1-x) - \frac{\pi^2}{12}\ln(1-x) + \frac{5\zeta_3}{2}\right) + \left(-\mathrm{Li}_2(x) - \frac{\ln(x)}{2}\ln(1-x) + \frac{2\pi^2}{3} - \frac{29}{4}\right) \\
&\times (1-x) + \frac{\ln^3(x)}{24}(1+x) + \frac{1}{1-x}\left(\left(-2x^2+2x+3\right)\frac{\ln^2(x)}{8} + \left(17x^2 - 13x + 4\right)\frac{\ln(x)}{4}\right) \\
&\left. - \frac{x}{4}\ln(1-x)\right] + C_F\left\{\frac{(1-x)}{x}\left(2x^2 - x + 2\right)\left(\frac{\mathrm{Li}_2(x)}{6} + \frac{\ln(x)}{6}\ln(1-x) - \frac{\pi^2}{36}\right) + \frac{(1-x)}{216x}\right. \\
&\left.\left(136x^2 - 143x + 172\right) - \frac{\left(8x^2+3x+3\right)}{48}\ln^2(x) + \frac{\left(32x^2 - 30z + 21\right)}{36}\ln(x) + \frac{(1+x)}{24}\ln^3(x)\right\}
\end{aligned} \tag{B.22}$$

$$\begin{aligned}
\mathcal{H}^{\mathrm{DY},(2)}_{qq\to q\bar{q}}(x) &= C_F\left(C_F - \frac{C_A}{2}\right)\left\{\frac{1+x^2}{1+x}\left(\frac{3\mathrm{Li}_3(-x)}{2} + \mathrm{Li}_3(x) + \mathrm{Li}_3\left(\frac{1}{1+x}\right) - \frac{\mathrm{Li}_2(x)\ln(x)}{2}\right.\right. \\
&\left.\left. - \frac{\mathrm{Li}_2(-x)\ln(x)}{2} - \frac{1}{24}\ln^3(x) - \frac{1}{6}\ln^3(1+x) + \frac{1}{4}\ln(1+x)\ln^2(x) + \frac{\pi^2}{12}\ln(1+x) - \frac{3\zeta_3}{4}\right)\right.
\end{aligned}$$

$$+(1-x)\left(\frac{\mathrm{Li}_2(x)}{2}+\frac{1}{2}\ln(1-x)\ln(x)+\frac{15}{8}\right)-\frac{1}{2}(1+x)\left(\mathrm{Li}_2(-x)+\ln(x)\ln(1+x)\right)+$$
$$+\frac{\pi^2}{24}(x-3)+\frac{1}{8}(11x+3)\ln(x)\Big\}+C_F\left\{\left(2x^2-x+2\right)\left(\mathrm{Li}_2(x)+\ln(1-x)\ln(x)-\frac{\pi^2}{6}\right)\right.$$
$$\times\frac{1}{12x}(1-x)+\frac{1}{432x}(1-x)\left(136x^2-143x+172\right)-\frac{1}{96}\left(8x^2+3x+3\right)\ln^2(x)+\frac{\ln^3(x)}{48}$$
$$\left.\times(1+x)+\frac{1}{72}\left(32x^2-30x+21\right)\ln(x)\right\} \tag{B.23}$$

$$\mathcal{H}^{\mathrm{DY},(2)}_{q\bar{q}'\to q\bar{q}}(x)=C_F\left\{\frac{(1-x)}{12x}\left(2x^2-x+2\right)\left(\mathrm{Li}_2(x)+\ln(1-x)\ln(x)-\frac{\pi^2}{6}\right)+\frac{(1-x)}{432x}\right.$$
$$\left.\times\left(136x^2-143x+172\right)+\frac{(1+x)}{48\ln^{-3}(x)}-\frac{\left(8x^2+3x+3\right)}{96\ln^{-2}(x)}+\frac{\left(32x^2-30x+21\right)}{72\ln^{-1}(x)}\right\} \tag{B.24}$$

$$\mathcal{H}^{\mathrm{DY},(2)}_{qg\to q\bar{q}}(x)=C_A\left\{-\frac{(1-x)}{12x}\left(11x^2-x+2\right)\mathrm{Li}_2(1-x)-\left(\frac{44x^2-12x+3}{96}\right)\ln^2(x)\right.$$
$$+\left(2x^2-2x+1\right)\left(\frac{\mathrm{Li}_3(1-x)}{8}-\frac{1}{8}\mathrm{Li}_2(1-x)\ln(1-x)+\frac{1}{48}\ln^3(1-x)\right)+\frac{43}{108x}$$
$$+\left(\frac{2x^2+2x+1}{4}\right)\left(\frac{3\mathrm{Li}_3(-x)}{4}+\mathrm{Li}_3\left(\frac{1}{1+x}\right)-\frac{\mathrm{Li}_2(-x)\ln(x)}{2}-\frac{1}{6}\ln^3(1+x)+\frac{\pi^2}{12}\right.$$
$$\left.\times\ln(1+x)+\frac{1}{4}\ln^2(x)\ln(1+x)\right)+\frac{1}{4}x(1+x)\mathrm{Li}_2(-x)+x\mathrm{Li}_3(x)-\frac{x\ln(x)}{2}\mathrm{Li}_2(1-x)$$
$$-x\mathrm{Li}_2(x)\ln(x)-\frac{3}{8}\left(2x^2+1\right)\zeta_3-\frac{149x^2}{216}+\frac{1}{72}\left(68x^2+6\pi^2x-30x+21\right)\ln(x)$$
$$+\frac{\pi^2x}{24}+\frac{43x}{48}+\frac{1}{48}(2x+1)\ln 3(x)-\frac{1}{2}x\ln(1-x)\ln^2(x)-\frac{1}{8}(1-x)x\ln^2(1-x)$$
$$+\frac{1}{4}x(1+x)\ln(1+x)\ln(x)+\frac{1}{16}(3-4x)x\ln(1-x)-\frac{35}{48}\Big\}+C_F\left\{(2x^2-2x+1)\right.$$
$$\times\left(\zeta_3-\frac{\mathrm{Li}_3(1-x)}{8}-\frac{\mathrm{Li}_3(x)}{8}+\frac{1}{8}\mathrm{Li}_2(1-x)\ln(1-x)+\frac{\mathrm{Li}_2(x)\ln(x)}{8}-\frac{1}{48}\ln^3(1-x)\right.$$
$$\left.+\frac{1}{16}\ln(x)\ln^2(1-x)+\frac{1}{16}\ln^2(x)\ln(1-x)\right)-\frac{3x^2}{8}-\frac{1}{96}\left(4x^2-2x+1\right)\ln^3(x)+\frac{1}{64}$$
$$\times\left(-8x^2+12x+1\right)\ln^2(x)+\frac{1}{32}\left(-8x^2+23x+8\right)\ln(x)+\frac{5}{24}\pi^2(1-x)x+\frac{11x}{32}+\frac{1}{8}$$
$$\left.\times(1-x)x\ln^2(1-x)-\frac{1}{4}(1-x)x\ln(1-x)\ln(x)-\frac{1}{16}(3-4x)x\ln(1-x)-\frac{9}{32}\right\} \tag{B.25}$$

Notice that for different quark flavours, $\tilde{\mathcal{H}}^{\mathrm{DY},(2)}_{qq'\to q\bar{q}}=\tilde{\mathcal{H}}^{\mathrm{DY},(2)}_{q\bar{q}'\to q\bar{q}}$. The above expressions therefore complete our definition of the soft-improved transverse momentum resummation for the case of a Vector boson production via DY mechanism. As an important remark, in addition to performing the Mellin transforms of the above expressions, the change of H_q into $\bar{\mathrm{H}}_q$ (or equivalently $\mathcal{H}^{\mathrm{DY}}_{ab\to q\bar{q}}$ into $\bar{\mathcal{H}}^{\mathrm{DY}}_{ab\to q\bar{q}}$) is required.

3

IMPROVED PREDICTIONS FOR TRANSVERSE MOMENTUM DISTRIBUTIONS

The previous chapter introduced a modified transverse momentum resummation–that we referred to as soft-improved transverse momentum resummation–in which a subclass of soft logarithms are taken into account. One of the main features of such an improved resummation is the fact that it reproduces the threshold resummation of the total cross section upon integration over the transverse momentum. In the following chapter, we perform phenomenological studies in order to estimate the effects of such a resummation on transverse momentum distributions. In order for this to be achieved, we have to deal with issues that arise from the particular construction of the improved resummation. In the standard approach [140, 166, 168, 195, 196, 200, 208, 227, 245–250] (refer also to Sec. (2.2) for a brief review) where resummed expressions are formulated in Fourier-Mellin space, the problem related the *Landau pole* prevents the existence of an inverse Mellin. The situation, however, is more complicated in the case of soft-improved transverse momentum resummation due to the presence of new singularities that arise in the interplay between the Mellin moment N and the impact parameter b in the argument of the logarithms. In order to address this issue, we adopt the Borel method that was recently studied in Refs. [232, 233] in the context of threshold resummation for DY cross sections. As shall be demonstrated in the sequel, this approach leads to a well-behaved resummed series that can be implemented numerically.

As alluded in the previous chapter, the improved transverse momentum resummation formalism does not include all the relevant soft logarithms as it does not account, for instance, soft gluons that are emitted at large angles. Such contributions can be *manually* added to the improved resummation by providing to it the pure threshold resummed expression using a profile matching function. This ensures that the combined expression reduces to the threshold resummation in the soft limit for fixed p_T up to power corrections in $(1-x)$ (or respectively $1/N$ in Mellin space).

The chapter is organized as follows. First, we describe how the inverse Mellin transform is performed in the context of threshold resummation where a Fourier back transform is not required. We then introduce the Borel method as a prescription to perform the inverse Fourier-Mellin in the case of soft-improved transverse momentum resummation. As a sanity check, we show, using a numerical approach, that the improved resummation indeed reduces to the standard transverse momentum resummation in the limit $\hat{b} \to \infty$. After investigating the soft-behaviour of the new resummation in the small-p_T limit, we finally move to the phenomenological studies in which two processes–namely Higgs and DY–are considered. The effects of the soft-improved and combined resummations are separately assessed both at the level of the pure resummed results and matched to

fixed-order predictions. We close the chapter by presenting results in which the NNLO transverse momentum distribution of the Higgs boson is approximated using threshold and high energy resummations.

3.1 From Fourier-Mellin to direct space

We have seen in the previous chapter that resummed formulae are computed in Fourier and/or Mellin space. Therefore, in order to get physical cross section, one needs to compute the inverse Fourier-Mellin transform of the $(N\text{-}\hat{b})$ space expressions. Because the resummation of transverse momentum observables corresponds to the asymptotic sum of a divergent series of p_T-space contributions, the Fourier-Mellin integral contains spurious singularities. While there are vast literatures that describe how to perform the inverse Fourier-Mellin inverse transform for the case of standard transverse momentum resummation, the situation is a bit a more peculiar for the soft-improved transverse momentum resummation. Indeed, the improved resummation contains additional singularities due to the fact that the Mellin moment N and the impact parameter b are jointly resummed through a single variable χ.

In this section, we examine the origin of these divergencies and present prescriptions to treat them. In particular, we introduce the *Minimal Prescription* for the computation of the inverse Mellin transform and discuss various methods to deal with the Mellin space computation of the parton densities. Finally, we present the *Borel summation* as a prescription to deal with the computation of the inverse Fourier transform.

3.1.1 Minimal prescription

For the sake of discussing the computation of the Mellin integral, let us first assume that the resummed expression is given in ξ_p-space and ignore about the Fourier space. The resummed hadronic cross section in physical space can be written as

$$\frac{d\sigma^{\text{res}}}{d\xi_p}(\tau,\xi_p,\alpha_s) = \sum_{a,b}\int_{N_0-i\infty}^{N_0+i\infty}\frac{dN}{2\pi i}\,\tau^{-N} f_a(N+1) f_b(N+1)\frac{d\hat{\sigma}^{\text{res}}_{ab}}{d\xi_p}(N,\alpha_s), \tag{3.1.1}$$

where for brevity we have omitted all the scale dependence. As briefly introduced before, the appearance of a branch a branch cut at $\mathfrak{Re}(N) > \exp(1/2\bar{\alpha}_s - \gamma_E)$ prevents the existence of a parameter N_0 that is larger than the real part of the rightmost singularity of the integrand therefore spoiling the convergence.

The Mellin prescription provides a way of tacking this issue of convergence analytically by simply modifying the path of integration,

$$\frac{d\sigma^{\text{res}}}{d\xi_p}(\tau,\xi_p,\alpha_s) = \sum_{a,b}\int_{\text{MP}}\frac{dN}{2\pi i}\,\tau^{-N} f_a(N+1) f_b(N+1)\frac{d\hat{\sigma}^{\text{res}}_{ab}}{d\xi_p}(N,\alpha_s), \tag{3.1.2}$$

where the subscript MP denotes the contour modification. It has to be chosen such that the parameter N_0 seats to the left of the branch cut but to right of all the singularities. As shown in Fig. 6, a path that guarantees convergence consists in choosing a line that is rotated counter-clockwise in the upper plane $\mathfrak{Im}(N) > 0$ and clockwise in the lower plane $\mathfrak{Im}(N) < 0$. This ensures that the integral in Eq. (3.1.2) is finite.

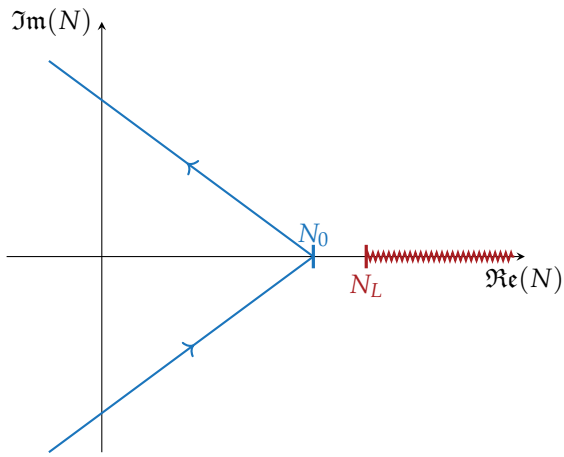


Figure 6 A possible path in complex Mellin space that can be chosen to evaluate the inverse Mellin transform. The branch cut related to the Landau pole is plotted in orange while the path of the Minimal prescription is plotted in blue. N_0 is the parameter of the integration. The path comes at some angle ϕ from negative complex infinity, crosses the real axis at N_0, and goes to positive complex infinity at a symmetric angle.

Given such a choice of a contour, the integral given in Eq. (3.1.2) can be written as

$$\int_{N_0-i\infty}^{N_0+i\infty} \frac{dN}{2\pi i}\,\tau^{-N}\frac{d\sigma^{\rm css}}{d\xi_p}(N,\xi_p) = \mathfrak{Im}\left[\int_{N_0}^{N_0+i\infty} dN\,\tau^{-N}\frac{d\sigma^{\rm css}}{d\xi_p}(N,\xi_p)\right] \tag{3.1.3}$$

$$= \mathfrak{Im}\left[(i+r)\int_0^1 \frac{du}{u}\,\tau^{-N(u)}\frac{d\sigma^{\rm css}}{d\xi_p}(N(u),\xi_p)\right]. \tag{3.1.4}$$

In the first line, we split the contour integral into two pieces, the first above the real axis and the second below. However, since $d\sigma/d\xi_p(\tau^\star) = \left(d\sigma/d\xi_p(\tau)\right)^\star$, the second piece turns out to be the negative of the complex conjugate of the first and hence can be combined to the former. In order to take into account the MP-contour, in the second line, we used the substitution $N(u) = N_0 + (r+i)\ln u$ with $r > 0$ being an arbitrary real and positive parameter which controls the slope of the path that enhances the numerical convergence. This parameter must be positive, so that for large imaginary values of N (i.e. small values of u), the prefactor $\exp(N(u))\ln(1/\tau)$ converges to zero requiring that the real part of $N(u)$ to be negative for $u \to 0$ (i.e. $N_0 - r\ln(1/u) < 0$). Notice that the following is also an equivalent transformation

$$\int_{N_0-i\infty}^{N_0+i\infty} \frac{dN}{2\pi i}\,\tau^{-N}\frac{d\sigma^{\rm css}}{d\xi_p}(N,\xi_p) = \mathfrak{Im}\left[(r-i)\int_0^1 \frac{du}{u}\,\tau^{-N(u)}\frac{d\sigma^{\rm css}}{d\xi_p}(N(u),\xi_p)\right], \tag{3.1.5}$$

where in this case8 $N(u) = N_0 + (r-i)\ln u$. We highlight that, recently, prescription based on Borel summation of the divergent series of the order-by-order inverse Mellin transform has been proposed in Refs. [232–234]. As a matter of fact, a variation of such a prescription is used in Sec. (3.1.3) for the treatment of divergencies appearing in the soft-improved transverse momentum resummed expression.

3.1.2 PDFs in Mellin space

By inspection of Eqs. (3.1.2, 3.1.4, 3.1.5), we notice that the Minimal Prescription is well-defined if the parton distributions are analytic within the region of integration. Parton distribution functions are usually available through the LHAPDF [5] as an interpolation over the momentum fraction. This raises some complications since the numerical evaluation of the their Mellin transform does not converge along the path of integration. Indeed, while the contour deformation of the Minimal Prescription probes values of N in the negative real axis, as illustrated in Fig. 6, the numerical Mellin transform of the parton densities converges only for $\mathfrak{Re}(N) > 0$. One may use parton distributions whose Mellin transform can be computed analytically at the initial scale [69]. Such sets assumes a functional N-space input form for the parton densities at some initial scale μ_0 from which it starts the evolution of scales. However, doing so significantly restricts the sets of PDFs that can be used.

One possible solution consists in fitting the parton densities at a fixed scale to a standard function. One usually chooses a specific generic parametrization

$$f(x,\mu_0^2) = Ax^\alpha(1-x)^\beta(1+\gamma\sqrt{x}+\delta x+\cdots) \tag{3.1.6}$$

with an appropriate number of free parameters. Once the parameters are determined, the computation of the Mellin transform is straightforward,

$$f(N,\mu_0) = A\left[\beta\left(N+\alpha,\beta+1\right)+\gamma\beta\left(N+\alpha+\frac{1}{2},\beta+1\right)+\delta\beta\left(N+\alpha+1,\beta+1\right)+\cdots\right],$$

where $\beta(a,b)$ is the Beta function. It can be clearly seen that the quality of the fit strongly depends on the choice of parametrization. While having too many degrees of freedom can be extremely computationally expensive, having too few can run into the risk of loosing accuracy therefore spoiling the result of the Mellin transform.

Another possible alternative consists in applying the Minimal prescription to the partonic cross-section and then convoluting the result with the parton densities in momentum space. This is achieved by realizing that the luminosity in Eq. (3.1.2) can be factorized out from the Mellin integral

$$\frac{d\sigma^{\text{css}}}{d\xi_p}(\tau,\xi_p,\alpha_s) = \sum_{a,b}\int_0^1\frac{d\xi}{\xi}\mathcal{L}_{ab}(\xi)\int_{\text{MP}}\frac{dN}{2\pi i}\left(\frac{x}{\xi}\right)^{-N}\frac{d\hat{\sigma}_{ab}^{\text{res}}}{d\xi_p}(N,\alpha_s). \tag{3.1.7}$$

This approach, however, is numerically very unstable due to the fact that the resummed component oscillates wildly in the region $x \sim 1$. In order to get rid of the oscillations, one needs to resort to some ad hoc prescriptions such as adding and subtracting the results of the Minimal Prescription evaluated with a fake luminosity.

In the next section, we describe an alternative approach that can be applied to any set provided by state-of-the-art PDF determination.

Expansion on the basis of Chebyshev polynomials:

An efficient way to approximate the Mellin transform of the parton densities is to expand the PDFs on a basis of polynomials whose Mellin transform may be computed directly. An

example of such polynomials, that has shown to yield numerical stability, is the *Chebyshev polynomials* [216].

Parton distributions provided through the LHAPDF interface are available in the form of $\tilde{f}(x) = xf(x)$ as a grid in (x, Q^2, a) space. In our derivation, we consider a fixed scale and since the procedure is also independent of the flavour of the parton, we drop both the scale dependence and the flavour index. Introducing a new variable u defined in terms of $x_{\rm max}$ $x = x_{\rm max}e^u$, where $x_{\rm max}$ is usually set to one, the parton distribution functions is now expressed as

$$\tilde{f}(u) = x_{\rm max}e^u f\left(x_{\rm max}e^u\right), \tag{3.1.8}$$

where for the momentum fraction varies between zero and $x_{\rm max}$, the new variable u is mapped into $(-\infty)$ to zero. The lower bound of u corresponds to $x_{\rm min} = x_{\rm max}e^{u_{\rm min}}$. From the practical point of view, $u_{\rm min}$ can just be chosen to be some finite negative number. Using Eqs. (A.1, A.5), we can write down the the expansion of $\tilde{f}$ on the basis of the Chebyshev polynomials

$$\tilde{f} = (u)\sum_{k=0}^{n} \tilde{c}_k \left(1 - 2\frac{u}{u_{\rm min}}\right)^k \qquad \text{with} \qquad \tilde{c}_k = -\frac{c_0}{2}\delta_{k0} + \sum_{i=k}^{n} T_{ik}, \tag{3.1.9}$$

where n and T_{ik} denote the degree and the coefficients of the polynomials respectively. The coefficients T_{ik} can be determined to all-order using the recursive relations given in Eqs. (A.1-A.5). The approximation of the proton distribution in x-space can now be extracted from Eq. (3.1.8) and Eq. (3.1.9) leading to

$$f(x) = \frac{1}{x}\sum_{p=0}^{n}(-2)^p u_{\rm min}^{-p} \ln\left(\frac{x}{x_{\rm max}}\right)\sum_{k=p}^{n}\binom{k}{p}\tilde{c}_k. \tag{3.1.10}$$

The order-by-order computation of the Mellin transform is now straightforward

$$f(N) = x_{\rm max}^{N-1}\sum_{p=0}^{n}\frac{2^p}{u_{\rm min}}\left(\frac{1}{N-1}\right)^{p+1}\sum_{k=p}^{n}\tilde{c}_k\frac{k!}{(k-p)!}. \tag{3.1.11}$$

For consistency check, in Fig. 7, we show a comparison between the approximation given in Eq. (3.1.10) and a NNLO PDF set with $\alpha_s = 0.118$ where we chose $n = 10$, $x_{\rm max} = 1$, and $u_{\rm min} = -15$ (which is equivalent to $x \sim 10^{-7}$). The samples of parton distributions are shown for $Q^2 = 1.65$ GeV. It is clear from Fig. 7 that the expansion on the basis of the Chebyshev polynomials provide a good approximation to the baseline PDFs. Slight deviations from the baseline PDFs can be seen for vary small values of x. However, since the PDF errors in that region are large, due to lack of experimental data, these deviations are not that significant. We have checked the approximation for various PDF sets and different values of scales and observed that the same conclusions hold. Therefore, in the subsequent phenomenological analyses, this is the approach that we are going to adopt.

3.1.3 Borel prescription

Fourier inverse transform of resummed transverse momentum distributions generally involves some complications since the computation of the integral (Eq. (2.4.4) for example)

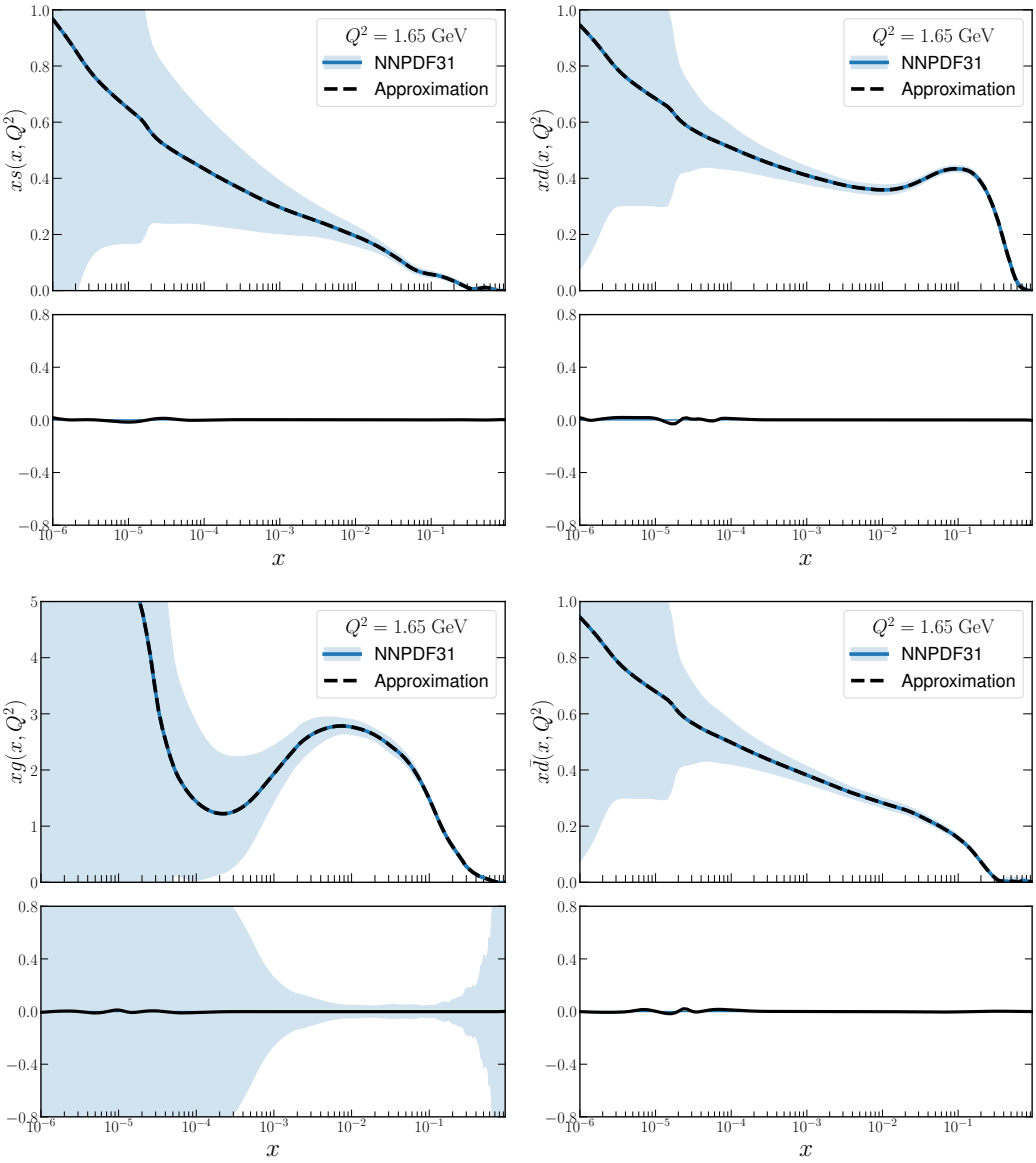


Figure 7 Comparison of the baseline set NNPDF3.1 with its expansion on the basis of the Chebyshev polynomials for the gluon and quark $(s, d, \bar{d})$ flavours at $Q^2 = 1.65$ GeV. The solid blue line represent the central value of the baseline PDF with the blue band representing its standard deviation 1σ-error. The dashed dark line represent the Chebyshev approximation. The lower panels shows the absolute difference, $\left(xf_{\text{cheb}} - xf\right)/xf$, between the two central values.

introduces an integration over the region of the impact parameters where the strong coupling α_s is ill-defined due to the existence of the Landau pole. This is because for a

physical process characterized by a hard scale Q^2, the standard resummation of logarithms of p_T, in Fourier space, re-scales the argument of the strong coupling to $\alpha_s(b_0^2/\hat{b}^2)$.

In order to obtain resummed predictions for physical observables, one therefore needs a prescription to cure this behaviour. To dates, several prescriptions have been proposed to perform the inverse Fourier transform. The first example of such a prescription is the $b_\star$-prescription introduced earlier (see Eq. (2.2.18)) in the context of CSS formalism. This has been widely used for phenomenological studies. However, the arbitrariness in choosing the value of $b_{\max}$ can be a drawback. Despite the fact that such a parameter can be used to estimate non-perturbative effects, it has been shown that truncating the value of the impact parameter at some $b_{\max}$ leads to numerical instabilities when matching with fixed-order calculations [189]. Another possibility is an analogue of the Minimal Prescription introduced in Sec. (3.1.1) in the context of threshold resummation [162, 163, 251]. The prescription amounts to finding a $\hat{b}$ contour that avoids the branch cut related to the Landau pole. Such a prescription leads to a finite results that is free of numerical and perturbative instabilities. However, in contrast to say the $b_\star$-prescription, this does not allow one to estimate the ambiguities related to the resummation procedure. A third alternative consists in expanding the Fourier-space resummed expression as a series in α_s and compute the Fourier-inverse integral order by order. This procedure, however, can only consistently resum logarithms at LL since results beyond that accuracy are very unstable [189, 190]. The final alternative prescription consists in expanding the resummed expression in $\hat{b}$-space as powers in α_s and resumming the ensuing series using the Borel summation. The divergence can then be removed in the Borel inversion by the inclusion of higher-twist terms. This permits the unambiguous computation of the asymptotic result the series is tending to that is free of numerical and perturbative instabilities.

The prescriptions cited above, however, were mainly studied and applied to the standard transverse momentum formalism. As a matter of fact, the situation is much more complicated in the context of soft-improved transverse momentum resummation due to the interplay between the soft and recoil variables. As shown previously, the argument of the strong coupling in SIPT is re-scaled to $\alpha_s(Q^2/\chi)$ where again $\chi = \bar{N}^2 + \hat{b}^2/b_0^2$. This then generates logarithms of the form $\ln\chi$. It is clear from this that in addition to the branch cut related to the Landau pole, soft-improved transverse momentum resummation contains additional singularities when

$$\mathfrak{Re}\left(\bar{N}^2 + \frac{\hat{b}^2}{b_0^2}\right) \leq 0, \qquad \text{and} \qquad \mathfrak{Im}\left(\bar{N}^2 + \frac{\hat{b}^2}{b_0^2}\right) = 0. \tag{3.1.12}$$

This means that the singularities on $\bar{N}$ depends on $\hat{b}$ and vice-versa. As a result, it is non-trivial to find a contour deformation that avoid all the singularities and branch cuts. An alternative approach consists in modifying the Borel method introduced in Ref. [230] in the context of standard transverse momentum resummation and in Refs. [216, 232]. Similar to the original derivations, the steps are as follows. First, we expand the resummed component as series in $\bar{\alpha}_s \ln\chi$ and resum the series using the Borel method. The divergence is then cured when performing the Borel back-transform. This leads to a N-space expression that can be inverted using the Minimal Prescription.

Let us start by writing Eq. (2.4.4) in a more intuitive form

$$\frac{d\hat{\sigma}^{\rm tr}_{ab}}{d\xi_p}(N,\xi_p,\alpha_s) = \sigma_0\left(\sqrt{1+\xi_p}+\sqrt{\xi_p}\right)^{2N}\int_0^\infty \mathrm{d}\hat{b}\,\frac{\hat{b}}{2}J_0\left(\hat{b}\sqrt{\xi_p}\right)\Sigma_{ab}\left(N,\lambda_\chi,\alpha_s\right) \tag{3.1.13}$$

where the large logarithms are collected in Σ_{ab} according to the series expansion

$$\Sigma_{ab}\left(N,\lambda_\chi,\alpha_s\right) = \sum_{k=0}^{\infty} h_k\left(N,\alpha_s\right)\bar{\alpha}_s^k \ln^k\left(\bar{N}^2+\frac{\hat{b}^2}{b_0^2}\right). \tag{3.1.14}$$

The zeroth order coefficient h_0 is just a constant and therefore contains terms that are not logarithmically enhanced when $\hat{b}\to\infty$. Putting back Eq. (3.1.14) into the resummed partonic expression in Eq. (3.1.13), exchanging the integral and the sum, one can perform the inverse Fourier transform order-by-order. This can be achieved by defining the following generating function

$$\mathcal{M}(N,\xi_p) = \left[\frac{\partial^k}{\partial\epsilon^k}\tilde{\mathcal{M}}(N,\xi_p,\epsilon)\right]_{\epsilon=0}, \tag{3.1.15}$$

$$\text{where}\qquad \tilde{\mathcal{M}}(N,\xi_p,\epsilon) = \frac{1}{b_0^{2\epsilon}}\int_0^\infty d\hat{b}\,\frac{\hat{b}}{2}J_0\left(\hat{b}\sqrt{\xi_p}\right)\left(4N^2+\hat{b}^2\right)^\epsilon. \tag{3.1.16}$$

For convenience, we have factorized out b_0 from the integrand such that the generating function is a function of N instead of $\bar{N}$. This allows us to easily perform the integral in Eq. (3.1.16) and subsequently derive the expression of the Fourier back-transform. Indeed, Eq. (3.1.16) now reads

$$\tilde{\mathcal{M}}(N,\xi_p,\epsilon) = \frac{2}{b_0^{2\epsilon}}\left(\frac{N}{\sqrt{\xi_p}}\right)^{1+\epsilon}\frac{\mathrm{K}_{1+\epsilon}\left(2N\sqrt{\xi_p}\right)}{\Gamma(-\epsilon)}, \tag{3.1.17}$$

where K denotes the modified Bessel function of the second kind. The form of Eq. (3.1.17) is interesting as it shows that in momentum space, the large-$\hat{b}$ limit at fixed $N/\hat{b}$ has been mapped to the small-ξ_p limit at fixed $N\sqrt{\xi_p}$. Finally, the derivative and the limit in Eq. (3.1.15) can be removed by performing a contour integral

$$\mathcal{M}(N,\xi_p) = \left[\frac{\partial^k}{\partial\epsilon^k}\tilde{\mathcal{M}}(N,\xi_p,\epsilon)\right]_{\epsilon=0} = \frac{k!}{2\pi i}\oint_H\frac{d\xi}{\xi^{k+1}}\tilde{\mathcal{M}}(N,\xi_p,\xi), \tag{3.1.18}$$

where H represents a contour enclosing the singularity at $\xi=0$.

Having expanded the function Σ_{ab} as a series in $\bar{\alpha}_s\ln\chi$ allowed us to perform the inverse Fourier transform of the resummed expression order-by-order, but it leaves us with a series representation of the result. From the numerical point of view, it would be convenient to trade the sum in favor of an integral. This calls for an analytic continuation which in our case can be achieved by summing the series using the Borel method and subsequently performing a Borel back-transformation. Writing down the partonic resummed cross section as

$$\frac{d\hat{\sigma}^{\rm tr*}_{ab}}{d\xi_p}(N,\xi_p,\alpha_s) \equiv \mathcal{A}(\bar{\alpha}_s) = \sum_{k=0}^{\infty} h_k\left(N,\alpha_s\right)\frac{k!}{2\pi i}\oint_H\frac{\mathrm{d}\xi}{\xi}\tilde{\mathcal{M}}\left(N,\xi_p,\xi\right)\left(\frac{\bar{\alpha}_s}{\xi}\right)^k, \tag{3.1.19}$$

where $d\hat{\sigma}^{\mathrm{tr}*}/d\xi_p$ is just $d\hat{\sigma}^{\mathrm{tr}}/d\xi_p$ without a kinematic factor $(\sqrt{1+\xi_p}+\sqrt{\xi_p})^{2N}$ and the Born level cross section $\sigma_{\mathrm{F}}^{\mathrm{Born}}$. The corresponding Borel sum is therefore given by

$$\mathcal{B}\left[\mathcal{A}(w)\right] = \frac{1}{2\pi i}\sum_{k=0}^{\infty} h_k\left(N,\alpha_{\mathrm{s}}\right)\oint_H \frac{d\xi}{\xi}\tilde{\mathcal{M}}\left(N,\xi_p,\xi\right)\left(\frac{w}{\xi}\right)^k \tag{3.1.20}$$

$$= \frac{1}{2\pi i}\oint_H \frac{d\xi}{\xi}\tilde{\mathcal{M}}\left(N,\xi_p,\xi\right)\Sigma\left(N,\frac{w}{\xi},\alpha_s\right), \tag{3.1.21}$$

where again we have interchanged the order of the summation and the integration in order to recover the form of the resummed function Σ_{ab} using its definition in Eq. (3.1.14). The inverse Borel transform can now be defined as

$$\frac{d\hat{\sigma}_{ab}^{\mathrm{tr}*}}{d\xi_p}(N,\xi_p,\alpha_s) = \int_0^{\mathrm{C}} dw\,\mathrm{e}^{-w}\mathcal{B}\left[\mathcal{A}\left(w\bar{\alpha}_s\right)\right] = \frac{1}{\bar{\alpha}_s}\int_0^{\infty} dw\,\mathrm{e}^{-w/\bar{\alpha}_s}\mathcal{B}\left[\mathcal{A}\left(w\right)\right] \tag{3.1.22}$$

$$= \frac{1}{\bar{\alpha}_s}\int_0^{\infty} dw\,\mathrm{e}^{-w/\bar{\alpha}_s}\frac{1}{2\pi i}\oint_H \frac{d\xi}{\xi}\tilde{\mathcal{M}}\left(N,\xi_p,\xi\right)\Sigma\left(N,\frac{w}{\xi},\alpha_s\right), \tag{3.1.23}$$

which represents our final resummed expression when adding the kinematic factor and the Born level cross section that we left out from the beginning. This expression is finite for arbitrary large values of C which cuts off the w-integration. At the level of the partonic cross section, the inverse Fourier transform of our soft-improved transverse momentum resummation can now be written as

$$\frac{d\hat{\sigma}_{ab}^{\mathrm{tr}}}{d\xi_p} = \frac{\sigma_{\mathrm{F}}^{\mathrm{Born}}}{\bar{\alpha}_s}\left(\sqrt{1+\xi_p}+\sqrt{\xi_p}\right)^{2N}\int_0^{\infty} dw\,\frac{\mathrm{e}^{-w/\bar{\alpha}_s}}{2\pi i}\oint_H \frac{d\xi}{\xi}\tilde{\mathcal{M}}\left(N,\xi_p,\xi\right)\Sigma\left(N,\frac{w}{\xi},\alpha_s\right). \tag{3.1.24}$$

Notice that despite the fact that there is some arbitrariness in choosing the value of the cutoff C, its value cannot be chosen freely. As a matter of fact, the choice of C generates correction terms of order $\mathcal{O}\left(\exp(-1/\bar{\alpha}_s)\right)$ that are higher-twist in nature. In other words, the inclusion or the exclusion of subleading terms in the Borel-resummed expression Eq. (3.1.24) is regulated by the cutoff C which can be used to asses the ambiguity of the resummation procedure. The larger the value of C gets, the more we include correction terms in the series. As the value of C, however, gets very large the integral in Eq. (3.1.24) becomes unstable and varies a lot eventually spoiling the accuracy of the result. This is due to the fact that the series in Eq. (3.1.14) is not Borel-summable. That said, the value of the cutoff C needs to be large enough such that the Borel Prescription is at least leading-twist but still free of numerical instability. In the context of standard transverse momentum resummation, the Borel and Minimal Prescription differ in the way higher-twist behaviours of the resummed series is handled. However, in practice and as will be demonstrated in the phenomenological sections, this difference is negligible at collider energies we are interested in. In particular, we show in Sec. (3.3) that unless one choose an unnaturally large (or respectively small) value of C, the resummed cross section is independent of the value of the cutoff.

From the perspective of the numerical implementation, it is convenient to map the contour integral onto an integral over the domain $[0,1]$. This is done by parametrizing the contour in terms of a new variable φ

$$\frac{1}{2\pi i}\oint_H d\xi\, f(\xi) = R\int_0^1 d\varphi\,\mathrm{e}^{2i\pi\varphi} f(\xi(\varphi)), \tag{3.1.25}$$

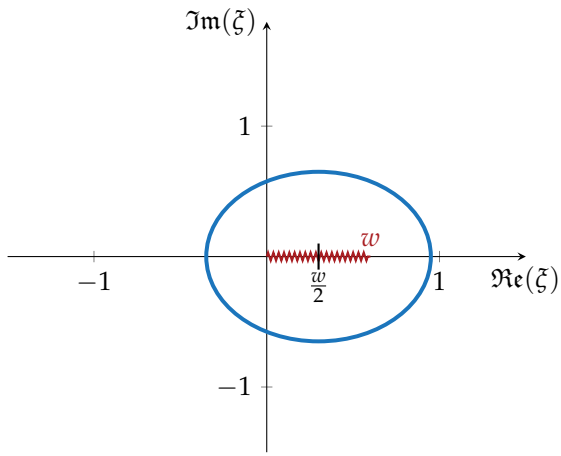


Figure 8 The integral contour H that encloses the branch cut (plotted in red) and the pole at the origin $\xi = 0$. The contour is usually taken to be a circle, here centered at $\xi_0 = w/2$ and with radius $R = \xi_0 + 1/2$. In practice, we noticed that the numerical integration converge faster if instead of a circle one considers an ellipse where the major axis is chosen to be along the real axis.

where f represents the integrand of the ξ-integration in Eq. (3.1.24). The contour is parameterized by $\xi(\varphi) = \xi_0 + R\exp(2i\pi\varphi)$ which is a circle around the real point $\xi = \xi_0 \in \mathbb{R}$ with a radius $R > 0$, in the positive direction around ξ_0. We can see that the branch cuts of f are now mapped in terms of the variable ξ into a branch cut on the real axis from zero to w (see Fig. 8). In practice, we choose $\xi_0 = w/2$ and $R = w/2 + \zeta$ ($\zeta \in \mathbb{R}^{*+}$). The value of ζ and the cutoff C must be chosen in such a way that the circle does not intersect the branch cut in ξ-space. Based on the *twist-4* argument [234], the value of the cutoff should be chosen according to $\mathrm{C} \geq 2$. Throughout our implementation, we chose the minimal value for C: i.e $\mathrm{C} = 2$. We, nevertheless, show in the phenomenological section that various values of C lead to the same results upon taking into consideration the numerical errors coming from the integration.

3.2 Large-N behaviour at small-p_T for SIPT

Before presenting phenomenological results, let us first check that the soft-improved transverse momentum resummation does indeed capture the correct soft behaviour at small-p_T, i.e. contributions from initial-state soft-gluons that are emitted at very small angles. As a test case, let us consider the case of the Higgs boson production via gluon fusion. For simplicity, let us first check the leading-order transverse momentum distribution with leading-logarithm resummed expression for the gg-channel. Such a check amounts to expanding the soft-improved resummed expression in Eq. (2.4.48) and compare the result with the leading-order p_T-distribution in Eq. (2.3.20).

As mentioned before, at such an accuracy, we approximate the hard function $\bar{\mathcal{H}}_c$ and the coefficient C_{gg} to one and we only include in the Sudakov exponent the first function

g_1. The parton evolutions that are described by the functions $\mathcal{Q}_r$ are included up to LO with the leading anomalous dimension $\gamma_{gg}^{(0)}$ approximated by its large-N behaviour while the evolution factors that are embodied in $\tilde{R}$ that evolve the coefficient functions are approximated to one. Thus, the resummed partonic cross section in Eq. (2.4.42) now becomes

$$\frac{d\hat{\sigma}_{gg}^{\mathrm{tr}*}}{d\xi_p}(N,\chi) = \left[U_{gg}^{(\mathrm{LO})}\left(N, \frac{Q^2}{\chi} \leftarrow Q^2\right)\right]^2 \exp\left(g_1(N,\chi)\right), \tag{3.2.1}$$

where we recall that $d\hat{\sigma}^{\mathrm{tr}*}/d\xi_p$ is just $d\hat{\sigma}^{\mathrm{tr}}/d\xi_p$ without the kinematic factor and the Born level cross section $\sigma_{\mathrm{H}_{gg}}^{\mathrm{Born}}$. According to Eqs. (2.4.46, 2.4.47), the large-N limit of the LO order solution to the DGLAP evolution equation can be written as

$$U_{gg}^{(\mathrm{LO})}\left(N\to\infty, \frac{Q^2}{\chi} \leftarrow Q^2\right) = -\frac{A_g^{(1)}}{2\bar{\alpha}_s\beta_0}\lambda_{\bar{N}}\ln(1-\lambda_\chi), \tag{3.2.2}$$

which therefore leads to a simple form of the resummed expression Eq. (3.2.1)

$$\frac{d\hat{\sigma}_{gg}^{\mathrm{tr}*}}{d\xi_p}(N,\chi) = \exp\left\{\frac{A_g^{(1)}}{\bar{\alpha}_s\beta_0}\left(\lambda_\chi + (1-\lambda_{\bar{N}})\ln(1-\lambda_\chi)\right)\right\}. \tag{3.2.3}$$

Expanding the above equation as a series in α_s and retaining only the α_s-term in the expansion, we arrive at the following expression

$$\frac{d\hat{\sigma}_{gg}^{\mathrm{tr}*}}{d\xi_p}(N,\chi) = -\alpha_s A_g^{(1)}\left(\frac{\ln^2\chi}{2} - 2\ln\bar{N}\ln\chi\right). \tag{3.2.4}$$

Eq. (3.2.4) can now be inserted back into Eq. (2.4.42) by adding the pre-factors that we left out. The inverse Fourier transform of the ensuing expression can be straightforwardly computed using the expression of the generating functions in Eq. (3.1.15) for fixed k. The final leading-logarithms (N,ξ_p)-space resummed expression is therefore given by

$$\frac{d\hat{\sigma}_{gg}^{\mathrm{tr}}}{d\xi_p}(N,\xi_p) = \sigma_{\mathrm{H}_{gg}}^{\mathrm{Born}}\frac{2\alpha_s A_g^{(1)}}{a^N(\xi_p)}\frac{N}{\sqrt{\xi_p}}\left[\mathrm{K}_1^{(1)}\left(2N\sqrt{\xi_p}\right) - \left(\ln\left(N\sqrt{\xi_p}\right)+\gamma_E\right)\mathrm{K}_1\left(2N\sqrt{\xi_p}\right)\right] \tag{3.2.5}$$

where $\mathrm{K}_1^{(1)}$ denotes the first derivative of the modified Bessel function w.r.t its first argument. Recall that we are here interested to check the large-N behaviour of the soft-improved transverse momentum resummation while keeping $N\sqrt{\xi_p}$ fixed in which case the expression above cannot be simplified further. In this case, in order to compare Eq. (3.2.5) with Eq. (2.3.20), we have to simplify the leading-order p_T distribution $C_{gg}^{(1)}$ in the aforementioned limit. In such a limit, it is enough to consider the most dominant term, i.e. $m=4$. Using the properties of the Gauss Hypergeometric functions, we can adopt the following form

$${}_2F_1\left(\frac{1}{2}, N+4, N+\frac{9}{2}, a^2\right) = \frac{1}{\sqrt{1-a^2}}\,{}_2\tilde{F}_1\left(N+\frac{9}{2}; \frac{a^2}{a^2-1}\right) \tag{3.2.6}$$

where for brevity we have defined $a \equiv a(\xi_p)$ and

$$ {}_2\tilde{F}_1\left(N+\frac{9}{2};\frac{a^2}{a^2-1}\right) \equiv {}_2F_1\left(\frac{1}{2},\frac{1}{2},N+\frac{9}{2},\frac{a^2}{a^2-1}\right). \tag{3.2.7} $$

In the limit were are interested in, the Hypergeometric function Eq. (3.2.6) can be simplified further using results from Refs. [252,253] in which the ${}_2\tilde{F}_1$ is expressed as a sum of Kummer functions U. The expression is given by

$$ \begin{aligned} {}_2\tilde{F}_1\left(N+\frac{9}{2};\frac{a^2}{a^2-1}\right) =& \frac{\Gamma(N+9/2)}{\Gamma(N+4)}\sum_{s=0}^{\infty} g_s\left(\frac{a^2}{a^2-1}\right) \\ & \frac{\Gamma(s+1/2)}{\Gamma(1/2)}\ln^s\frac{1}{a^2}\mathrm{U}\left(s+\frac{1}{2},s+1,-2N\ln a\right). \end{aligned} \tag{3.2.8} $$

We stress that this series expansion only holds if $N \to \infty$ uniformly w.r.t large values of $a^2/(a^2-1)$, which is indeed the case since $a^2/(a^2-1) \to \infty$ as $\xi_p \to 0$. In the limit we are interested in, it is sufficient to only consider the first term in the sum $(s=0)$ where the function g_0 and the second Kummer function U are defined as

$$ g_0\left(\frac{a^2}{a^2-1}\right) = \sqrt{\ln\left(\frac{1}{a^2}\right)} \tag{3.2.9} $$

$$ \mathrm{U}\left(\frac{1}{2},1,2\tilde{a}\right) = \frac{1}{\Gamma(1/2)}\tilde{a}\exp(z)\mathrm{K}_1^{(1)}(\tilde{a}), \tag{3.2.10} $$

where we have defined $\tilde{a} \equiv -N\ln a$ for brevity. We can now use Eqs. (3.2.9, 3.2.10) to rewrite the expression of ${}_2F_1$ in Eq. (3.2.6). Plugging the resulting equation into the LO p_T distribution, we arrive to the following expression

$$ \left[C_{gg}^{(1)}(N,\xi_p)\right]_{m=4} = 2\alpha_s\frac{\sigma_0}{\xi_p}\frac{C_A}{\pi}\frac{\tilde{a}(\xi_p)}{a^{N+2}(\xi_p)}\frac{\sqrt{2\ln a(\xi_p)}}{\sqrt{1-a^2(\xi_p)}}\mathrm{K}_1^{(1)}(\tilde{a}(\xi_p)) \tag{3.2.11} $$

The $a^N(\xi_p)$ in the denominator comes from rewriting the exponential in Eq. (3.2.10) in terms of the explicit expression of $a(\xi_p)$. The limit $\xi_p \to 0$ can now be taken safely in the terms that do not depend on N. This yields to the following simplifications $\sqrt{2\ln a} \sim 2\sqrt[4]{\xi_p}$ and $\sqrt{1-a^2} \sim 2\sqrt[4]{\xi_p}$. Finally, the large-$N$ limit just amounts to replacing $(N+2)$ with N. Putting everything together, rearranging some terms, and performing some algebraic simplifications, we get

$$ \left[C_{gg}^{(1)}(N,\xi_p)\right]_{m=4} = 4\alpha_s\frac{C_A}{\pi}\frac{\sigma_0}{a^N(\xi_p)}\frac{N}{\sqrt{\xi_p}}\mathrm{K}_1^{(1)}(2N\sqrt{\xi_p}). \tag{3.2.12} $$

In the limit $N\sqrt{\xi_p} \to 0$, the above expression is *asymptotic* to expansion of the modified resummation in Eq. (3.2.5). In order to generate K_1-terms from the leading-order transverse momentum spectra $C_{gg}^{(1)}$, the sum in Eq. (3.2.8) has to be performed beyond $s=0$. This is computationally challenging as the form of the Kummer U does not lead directly to compact Bessel functions. However, as it can be seen from the numerical results in Fig. 9, the soft-improved transverse momentum resummation captures best the soft behaviours from the fixed-order calculation while the standard transverse momentum resummation deviates largely from it. This feature can be noticed both at LO and NLO.

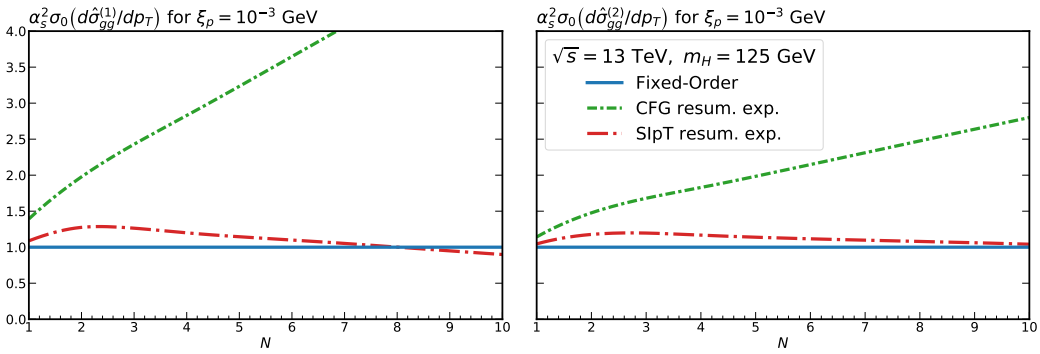


Figure 9 Ratio between the terms appearing in the expansion of the resummed expressions (standard and improved transverse momentum resummations) and the fixed-order results. The plot on the left shows the ratio between the first order term in the expansion and the LO predictions, while the plot on the right shows the ratio between the second order term in the expansion and the second order term appearing in the fixed order computations. The plots were produced with a center-of-mass energy $\sqrt{s}=13$ TeV with a Higgs mass $m_H=125$ GeV.

3.3 Towards a combined resummation for Higgs and Drell–Yan

In the previous sections, we have presented a resummation formalism for transverse momentum distributions that takes into account soft behaviours in the small-p_T limit. That is, the soft-improved transverse momentum formalism we presented does not take into account soft gluons that are emitted at large angles. The need for such a resummed expression is crucial in order to (a) elucidate the relation between soft and collinear logarithms that provide the main bulk of the contributions to the transverse momentum spectrum and the total inclusive cross section in the soft limit and (b) use the knowledge from resummation to estimate corrections from missing higher-orders. While a *consistent* combination of the transverse momentum and threshold resummations seems unreachable to present day due to the non-commutativity of the large-N and small-p_T limits, it is possible to combine our soft-improved formalism with the pure threshold resummation in an unambiguous way using some matching function.

In the following sections, we discuss the concept of consistently combining the two resummations, namely transverse momentum and threshold, using a profile matching function. We then present detailed phenomenological studies that assess the impact of the soft-improved transverse momentum and combined resummation in the case of Higgs and Z-boson production. Specifically, our aim is to estimate the impacts of: (a) the improved transverse momentum resummation through the inclusion of soft contributions which we expect to be more apparent in the small-p_T region, and (b) the additional soft resummation at fixed-p_T which we expect to improve the matching to fixed-order results in the intermediate and large-p_T regions. In doing so, we first check that, from the numerical perspective, the improved resummation reproduces exactly the standard formalism. This will be done using the Borel prescription method which will allow us to check the validity

of our resummation procedure and to assess the dependence of the resummed cross section on the cutoff C.

All the results presented here are produced using the NNPDF31_nnlo_as_0118 set of parton distributions from the NNPDF3.1 [8] through the LHAPDF [5] interface by expanding the parton densities in the basis of the Chebyshev polynomials as described in Sec. (3.1.2) [233, 234]. In order to estimate missing higher-order corrections, we will perform the commonly used factorization and renormalization scale variations of all the unresummed, resummed, and matched cross sections. Indeed, by performing scale variations of a cross section computed at $\mathcal{O}(\alpha_s^n)$, one generate subleading terms of $\mathcal{O}(\alpha_s^{n+1})$. Specifically, we will use the seven-point envelope prescription where the renormalization and factorization scales are independently varied by a factor of two around the central choice which in our case could either be the mass of the Higgs or Z boson. That is, we compute the combinations of the following choice of scales

$$\mu_R = \frac{1}{2}M, \qquad \mu_R = M, \qquad \mu_R = 2M \tag{3.3.1}$$

$$\mu_F = \frac{1}{2}M, \qquad \mu_F = M, \qquad \mu_F = 2M \tag{3.3.2}$$

with the requirement that $0.5 \leq \mu_R/\mu_F \leq 2$ which ensures that the logarithmic terms that vary with the scales are at most $\ln 2$. The uncertainty bands are then computed by taking the minimum and the maximum from the possible configuration of scales. As shown in our previous calculations, varying the scales in this way will modify the strong coupling, the evolution of parton densities and the partonic cross section which altogether ensures that no artificially large scale ration are generated. Notice, however, that throughout our results, scale variations related to the resummation scale are not taken into account.

3.3.1 SIPT in the large–$\hat{b}$ limit

In this section, let us check that the soft-improved transverse momentum resummation indeed reproduces the standard resummation when the soft contributions are switched off. Although this has been analytically verified to be the case, here we perform the check at the numerical level. Besides being a consistency check, this allows us to verify that our resummation procedure derived using the Borel prescription agrees with the Minimal Prescription and does not introduce spurious subleading terms. Taking the large-$\hat{b}$ limit means that we should perform the following replacements throughout the expressions that enter into the resummed partonic cross section in Eq. (2.4.48)

$$\chi = \left(\bar{N}^2 + \frac{\hat{b}^2}{b_0^2}\right) \longrightarrow \left(1 + \frac{\hat{b}^2}{b_0^2}\right), \qquad \text{and} \qquad \frac{Q}{\chi} \longrightarrow \frac{b_0^2}{b^2}. \tag{3.3.3}$$

Notice that in order to reduce the impact of unjustified logarithms in the large-p_T regions (similar to the CFG formalism), we replaced $\bar{N}$ by one. The computation of the inverse Fourier transform is achieved following Sec. (3.1.2) where the generating function now has the following form

$$\tilde{\mathcal{M}}^\star(N, \xi_p, \epsilon) = 2\exp\left(2\gamma_E\epsilon\right)\left(\frac{1}{\sqrt{\xi_p}}\right)^{1+\epsilon}\frac{K_{1+\epsilon}\left(2\sqrt{\xi_p}\right)}{\Gamma(-\epsilon)}. \tag{3.3.4}$$

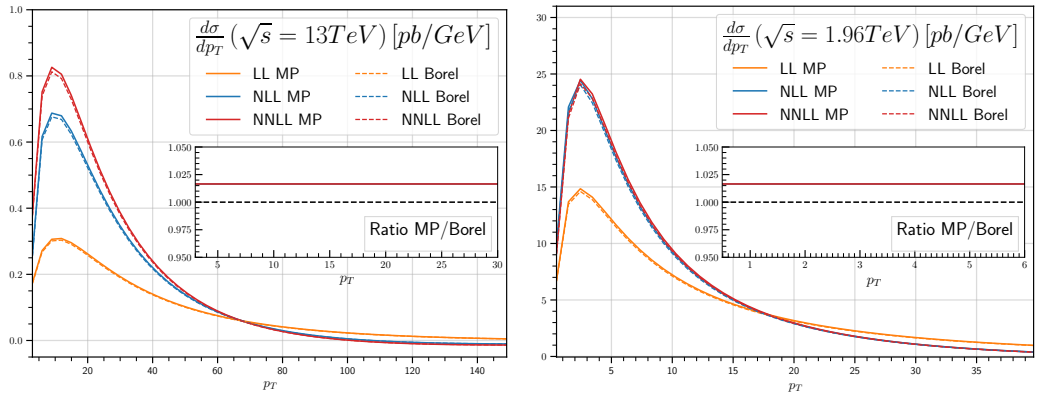


Figure 10 Comparison of the Borel and Minimal Prescription for the transverse momentum distribution of the Higgs (left) and Z boson (right). The dashed lines represent the results of taking the large-$\hat{b}$ limit of the soft-improved transverse momentum resummation and using the Borel method to perform the inverse Fourier transform. The solid lines represent the results from the standard resummation formalism. The renormalization (μ_R) and factorization (μ_F) scales are set equal to the Higgs (m_H) and Z boson mass (m_Z) in the case of Higgs and Z boson production respectively. The lower panel shows the ration between the two prescriptions.

In addition to performing the above replacements, we need to implement the exact same procedure as in Ref. [195, 196] by suitably re-adjusting the solutions to the evolution equation and fixing the integral over ξ_p in order to recover the total inclusive cross section. In Fig. 10 we show a comparison of the Borel and Minimal prescription for both the Higgs and Z boson production. Leaving aside the interpretation of the resummed results for the time being, let us only focus on the difference between the two prescriptions. The large-$\hat{b}$ limit of the soft-improved transverse momentum resummation are represented by the dashed lines. They are produced using the Borel prescription for the inverse Fourier transform using as a value of cutoff C=2. The standard transverse momentum results are plotted as solid lines, they are produced using HQT [195] for Higgs and DYQT [242, 243] for DY. The results are shown for various logarithmic accuracies, namely LL in orange, NLL in blue, and NNLL in red. The bottom panels show the ratio between the Minimal Prescription and the Borel approach with the bands representing the errors from the numerical integration. We can clearly see that the our resummation formalism is order-by-order consistent with the standard transverse momentum resummation using the Minimal Prescription. As shown in the lower panels, the slight discrepancy between the Borel and Minimal Prescription can be attributed to numerical errors. These results confirm two things: first, that the soft-improved transverse momentum resummation indeed reproduces the standard transverse momentum resummation in the large-$\hat{b}$ limit; and second that the Borel method is an equivalent valid prescription to perform inverse Fourier-Mellin transform.

We can now move to the study of the dependence of the resummed expression in momentum space on the cutoff C. In Fig. 11, we plot the normalized cross section for

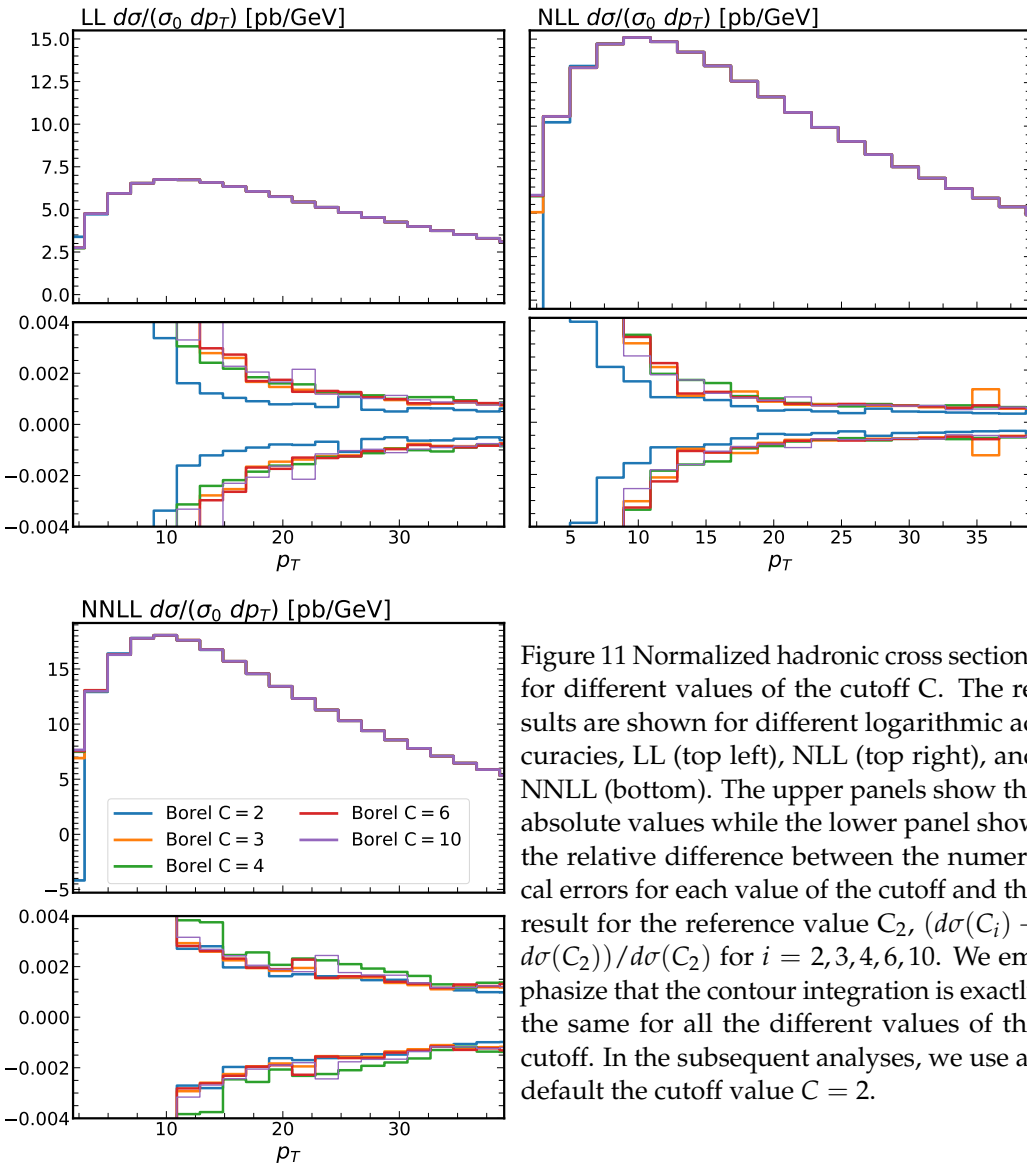


Figure 11 Normalized hadronic cross sections for different values of the cutoff C. The results are shown for different logarithmic accuracies, LL (top left), NLL (top right), and NNLL (bottom). The upper panels show the absolute values while the lower panel show the relative difference between the numerical errors for each value of the cutoff and the result for the reference value C_2, $(d\sigma(C_i) - d\sigma(C_2))/d\sigma(C_2)$ for $i = 2, 3, 4, 6, 10$. We emphasize that the contour integration is exactly the same for all the different values of the cutoff. In the subsequent analyses, we use as default the cutoff value $C = 2$.

different values of C. In the lower panel, we plot the absolute difference between the cross sections computed at various cutoff C and the reference $C = 2$. As before, the bands represent the errors from the numerical integration, in particular, we highlight the error bands for the reference values. Since we can see a clear decrease in the error as p_T gets bigger, we decided to present results for smaller values of p_T. One can see that all the variations are equivalent for all values of p_T except for the first bin in which the numerical errors are large anyway. It is very interesting to report that as the value of the cutoff increases, the errors arising from the numerical integration also increase. Similar observation can also be made as we increase in logarithmic accuracy.

3.3.2 Finite order truncation of the resummed expression

In order to provide a valid prediction, the resummed results have to be matched to the fixed-order calculations. As stressed from previous sections, whilst the small-p_T region is dominated by the resummed cross section, the large-p_T part is dominated by finite components computed from fixed-order calculations. This suggests that in the intermediate p_T region, both contributions have to be consistently matched in order to obtain uniformly sensible theoretical predictions. Such a procedure guarantees the correct behaviour from perturbative calculations up to a specified order and incorporates the large logarithms from resummation at higher-orders. In this thesis, we consider the additive matching scheme which consists on adding to the all-order results the perturbative calculations truncated at a given order in α_s and subtract to the whole the expansion of the resummed results at the same order. The matching can be performed at the level of the partonic cross section in which the matched expression can be written as follows

$$\frac{d\hat{\sigma}^{\rm M}_{ab}}{d\xi_p}(N,\chi) = \frac{d\hat{\sigma}^{\rm res}_{ab}}{d\xi_p}(N,\chi) + \left[\frac{d\hat{\sigma}^{\rm F.O}_{ab}}{d\xi_p}(N,\chi)\right]_{\mathcal{O}(\alpha_s^n)} - \left[\frac{d\hat{\sigma}^{\rm exp}_{ab}}{d\xi_p}(N,\chi)\right]_{\mathcal{O}(\alpha_s^n)}, \tag{3.3.5}$$

where the notations M and (F.O) denote the matched and the fixed-order respectively. the subscripts $\mathcal{O}(\alpha_s^n)$ indicate the order at which both the fixed-order and the expanded (exp) expressions are truncated. Here and in the sequel, the notation used to denote the fixed-order accuracy and therefore the matched results refers to the accuracy of the p_T distribution. In particular, the LO refers to the non-trivial order of the transverse momentum distribution whose integral is the NLO total cross section[1]. Therefore, LO and NLO will always refer to the accuracy of the p_T spectrum. The terms that result from the expansion of the resummed expression can be organized into contributions that are regular and singular in the large-$\hat{b}$ (or equivalently in the small-p_T limit). In Fourier-Mellin space, they generally take the following form

$$\begin{aligned}\frac{d\hat{\sigma}^{\rm exp}_{ab}}{d\xi_p}\left(N,\chi,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right) = &\sum_c \sigma^{\rm LO}_{{\rm F},c\bar{c}}\delta_{ca}\delta_{\bar{c}b} + \sum_c \sigma^{\rm LO}_{c\bar{c}} \sum_{n=1}^{\infty} \alpha_s^n(Q^2)\times \\ &\left[\Sigma^{(n)}_{ab}\left(N,\chi,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right) + \mathcal{H}^{(n)}_{ab\to c\bar{c}}\left(N,\chi,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)\right],\end{aligned} \tag{3.3.6}$$

where both the regular and singular terms depend on the resummation scale. We recall that the resummation scale has been introduced in Sec. (2.2.2). As stated previously, however, in our case we always consider the case where $Q=M$. The perturbative coefficients $\Sigma^{(n)}_{ab}$ are polynomials in the logarithm variable $\ln\chi$. They are constructed in a way that they vanish when $\ln\chi=0$ (i.e. $\hat{b}=-\bar{N}b_0$) and lead to the threshold resummation when $\hat{b}=0$ (formally at the total inclusive level). On the other hand, the function $\mathcal{H}^{(n)}_{ab\to c\bar{c}}$ contains all the terms that behave as constants in the large-$\hat{b}$ limit. Up to NLO, we only need up to

1 In literatures, different way of counting has been used. In Ref. [199], for instance, different counting of the perturbative order is adopted in which fixed-order labels NLO and NNLO are being used. Such a counting corresponds exactly to the accuracy of the total inclusive cross section.

$\Sigma^{(2)}_{ab}$ and $\mathcal{H}^{(2)}_{ab\to c\bar{c}}$ terms in the expansion. The general form of the singular contribution can be written as

$$\Sigma^{(n)}_{ab}\left(N,\chi,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=\sum_{k=1}^{2n}\Sigma^{(n,k)}_{ab}\left(N,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)\ln^k\chi, \tag{3.3.7}$$

where the functions $\Sigma^{(n,k)}_{ab}$ are purely functions of N. Their momentum space versions can be computed using the generating function in Eq. (3.1.16). That is, for a fixed value of k the computation of their Fourier inverse transform just amounts to the computation of the following limit

$$\mathcal{J}_k=\lim_{\epsilon\to 0}\frac{\partial^k}{\partial\epsilon^k}\tilde{\mathcal{M}}\left(N,\xi_p,\epsilon\right), \tag{3.3.8}$$

with $\tilde{\mathcal{M}}$ as given in Eq. (3.1.17). In the case of Higgs boson production for example, the terms that contribute up to NLO are expressed as follows

$$\mathfrak{I}_1(N,\xi_p)=-\frac{2N}{\sqrt{\xi_p}}\mathrm{K}_1(2N\sqrt{\xi_p}) \tag{3.3.9}$$

$$\mathfrak{I}_2(N,\xi_p)=-\frac{4N}{\sqrt{\xi_p}}\mathrm{K}_1^{(1)}(2N\sqrt{\xi_p})-\frac{4N}{\sqrt{\xi_p}}\left(\ln\left(\frac{N}{\sqrt{\xi_p}}\right)+\gamma_E\right)\mathrm{K}_1(2N\sqrt{\xi_p}) \tag{3.3.10}$$

$$\begin{aligned}\mathfrak{I}_3(N,\xi_p)=&-\frac{6N}{\sqrt{\xi_p}}\mathrm{K}_1^{(2)}(2N\sqrt{\xi_p})-\frac{12N}{\sqrt{\xi_p}}\left(\ln\left(\frac{N}{\sqrt{\xi_p}}\right)+\gamma_E\right)\mathrm{K}_1^{(1)}(2N\sqrt{\xi_p})\\&-\frac{6N}{\sqrt{\xi_p}}\left(\ln^2\left(\frac{N}{\sqrt{\xi_p}}\right)+2\gamma_E\ln\left(\frac{N}{\sqrt{\xi_p}}\right)-\zeta_2+\gamma_E^2\right)\mathrm{K}_1(2N\sqrt{\xi_p})\end{aligned} \tag{3.3.11}$$

$$\begin{aligned}\mathfrak{I}_4(N,\xi_p)=&-\frac{8N}{\sqrt{\xi_p}}\mathrm{K}_1^{(3)}(2N\sqrt{\xi_p})-\frac{24N}{\sqrt{\xi_p}}\left(\ln^2\left(\frac{N}{\sqrt{\xi_p}}\right)+2\gamma_E\ln\left(\frac{N}{\sqrt{\xi_p}}\right)-\zeta_2+\gamma_E^2\right)\\&\mathrm{K}_1^{(1)}(2N\sqrt{\xi_p})-\frac{24N}{\sqrt{\xi_p}}\left(\ln\left(\frac{N}{\sqrt{\xi_p}}\right)+\gamma_E\right)\mathrm{K}_1^{(2)}(2N\sqrt{\xi_p})-\frac{8N}{\sqrt{\xi_p}}\mathrm{K}_1(2N\sqrt{\xi_p})\\&\left(\ln^3\left(\frac{N}{\sqrt{\xi_p}}\right)+3\gamma_E\ln^2\left(\frac{N}{\sqrt{\xi_p}}\right)+3\left(\gamma_E^2-\zeta_2\right)\ln\left(\frac{N}{\sqrt{\xi_p}}\right)-3\gamma_E\zeta_2+\gamma_E^3+\zeta_3\right),\end{aligned} \tag{3.3.12}$$

where the $\mathrm{K}_1^{(i)}$ are the i-th derivatives of the Bessel function w.r.t. the argument. One can check that in the small-p_T limit–meaning taking $N\to\exp(-\gamma_E)$–the above expressions exactly reproduce the standard small-p_T expressions given by Eqs. (121)-(124) of Ref. [195]. The equivalence between taking $p_T\to 0$ and setting $N\to\exp(-\gamma_E)$ deserves further comment. On can check that the modified logarithmic variable $\ln(1+\hat{b}^2/b_0^2)$ of Ref. [195] is obtained from our modified $\ln(\chi(\hat{b}))$ by setting $\bar{N}=1$ which in Eqs. (3.3.9-3.3.12) translates into taking $N\to\exp(-\gamma_E)$.

3.3.3 Phenomenological results for SIPT

After checking the robustness of our resummation procedure using the Borel prescription, we are now ready to present phenomenological predictions. This section is devoted to the phenomenological studies of the impact of the soft-improved transverse momentum resummation on the production of a colourless final-system which in our case could be either the Higgs or Z-boson. The values of the central renormalization, factorization, and resummation scales is chosen to be either the Higgs mass ($m_H = 125$ GeV) or the Z-boson mass ($m_Z = 91.187$ GeV) with $G_F = 1.663 \times 10^{-5}\text{GeV}^{-2}$. The perturbative uncertainty is then computed using the seven-point method by varying the renormalization and factorization scales by a factor of two in either either direction around the central value. Whilst the small-p_T region ($p_T \ll m_{\rm F}$) of the transverse momentum spectra is dictated by the resummation, the region where $p_T \sim m_{\rm F}$ is controlled by smallness of the strong coupling in which the use of fixed-order calculations is fully justified. In this region, one therefore needs to switch on fixed-order computations through the matching procedure explained in Sec. (3.3.2). For the two processes we are interested, the QCD radiative corrections are fully known at NLO, see Sec. (2.4) and App. 2.B for details.

Higgs boson production at LHC

We begin by studying the Higgs boson produced via gluon-gluon fusion in proton-proton collision at LHC with a center of mass energy $\sqrt{s}=13$ TeV. In the following, we restrict ourselves to the region $p_T^{\rm H} \lesssim m_T$ where the HEFT description is appropriate. The fixed-order results have been obtained using HPT-MON [254] which has been benchmarked against HQT [242, 250, 255] and NNLOJET [26, 256, 257]. We first comment on results show in Fig. 12 in which the pure soft-improved transverse momentum and standard resummations are compared. The top panels show the different order of logarithmic accuracy (NLL and NNLL) along with the fixed-order results while the lower panels show the ratio of the various predictions w.r.t. the NNLL results. We see that the NLO results diverges to $(-\infty)$ as $p_T \to 0$ and it exhibits an unphysical peak at $p_T \sim 10$ GeV. This arises due the non-cancellation between negative (next-to) leading and positive subleading logarithmic contributions. On the contrary, the resummations lead to a well-behaved transverse momentum distribution that has a (Jacobian) peak at $p_T \sim 10$ GeV and vanishes in the limit $p_T \to 0$. The scale dependence of the fixed-order result is about $\pm 10\%$ at $p_T = 20$ GeV which slightly decrease as the value of p_T increases. Comparing the the NLL and NNLL of the standard transverse momentum resummation, we can highlight two main observations: first, as obviously expected, one can see that the scale uncertainties are reduced with the inclusion of NNLL terms, second, the uncertainty bands do not overlap in the small-p_T region. The latter observation suggests that the resummed series poorly converges and/or the method of scale variations significantly underestimates logarithmic contributions due to the missing higher-order corrections. Now, comparing the standard to the soft-improved transverse momentum resummation, we observe a sizeable difference in the small-p_T region. Indeed, we notice that the improved resummation displays faster perturbative convergence at small p_T. This can be understood as coming from the $N/\hat{b}$ corrections whose effect become less significant as the logarithmic accuracy increases. As a consequence, in the soft-improved transverse momentum resummation, the NLL result

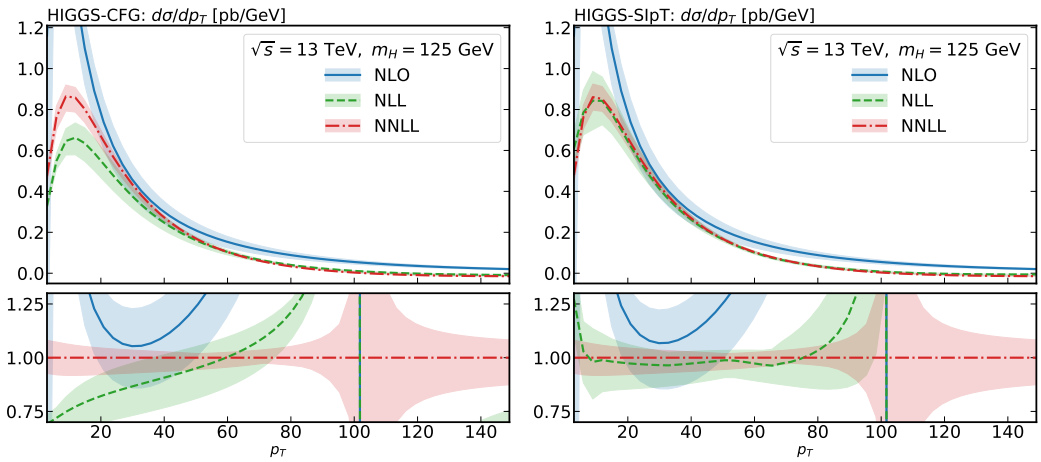


Figure 12 Higgs transverse momentum spectra from gluon fusion at $\sqrt{s} = 13$ TeV for the two types of resummation formalisms, namely the standard CFG (left) and the soft-improved transverse momentum resummation (right). The top panels show the pure resummed results at various logarithmic accuracies along with the NLO fixed-order result. The lower panels show the ratio of the various predictions w.r.t to the NNLL. The central scale is set to m_H. The uncertainty bands are computed using the seven-point scale variation, and in all cases, $Q = m_H$.

is much closer to the NNLL result that in the standard resummation formalism. As a matter of fact, the NNLL uncertainty band is mostly contained in the NLL one. But as demonstrated above, the two resummation formalisms are comparable in the large-p_T region. At NNLL, the difference between the standard and soft-improved transverse momentum resummation is almost invisible, but with the the latter having a moderately smaller uncertainty band due to the better convergence of the resummed series. Finally, one can see that the large-p_T predictions of both resummations are completely off w.r.t. the fixed order results and hence can no longer be trusted. As a side note, unlike in the standard resummation formalism, the modified resummation does not require any regulation procedure to get rid of its large-p_T behaviour. It instead relies on the fact its $\hat{b} = 0$ limit coincides with the threshold resummed expression at the inclusive level.

As just mentioned above, in order to produce reliable predictions that is valid for all values of p_T, the pure resummed results have to be matched with the fixed-order calculations according to Eq. (3.3.5). The results are shown in Fig. 13 where the results using the standard resummation formalism are shown on the left while the results from the soft-improved resummation are shown on the right. The matching of the standard resummation to the fixed-order calculations improves the agreement between the NLL+LO, NNLL+NLO, and NLO predictions, mainly in the intermediate and large-p_T region. As a matter of fact, the contribution of the LO finite component to the NLL+LO result is only about 2% in the peak and up to 50% in the large-p_T region. Similar quantitative behaviour is observed at NNLL+NLO. Now, we can see that the uncertainty bands of the NLO and NNLL+NLO in the standard formalism mostly overlap for $p_T > 20$ GeV. In the context of

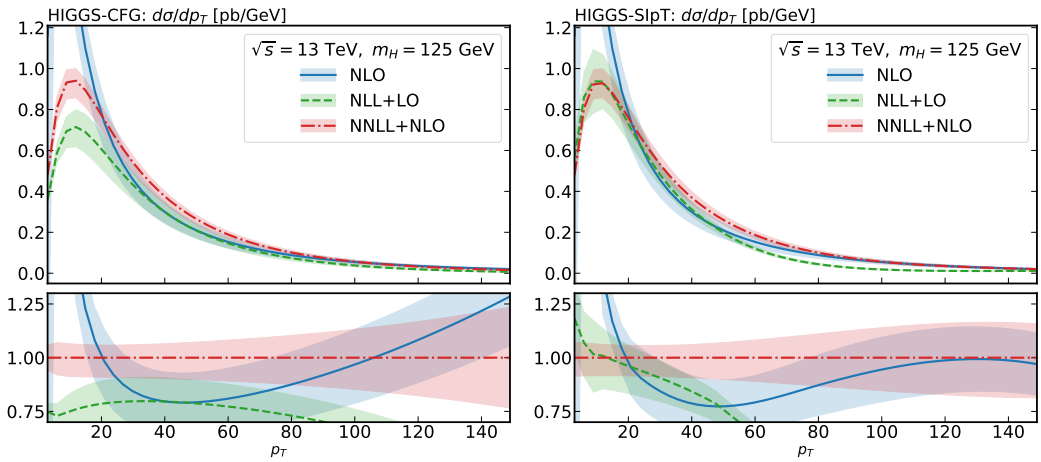


Figure 13 Higgs transverse momentum spectra from gluon fusion at $\sqrt{s} = 13$ TeV for the two types of resummation formalisms, namely the standard CFG (left) and the soft-improved transverse momentum resummation (right), when matched to fixed-order predictions. The top panels show the matched results at various logarithmic accuracies (NLO+LO and NNLL+NLO) along with the NLO fixed-order result. The lower panels show the ratio of the various predictions w.r.t to the NNLL+NLO result. The central scale is set to m_H. The uncertainty bands are computed using the seven-point scale variation, and in all cases, $Q = m_H$.

soft-improved transverse momentum resummation, the contribution of the LO and NLO finite components affect the pure resummed version as in the standard case. The difference being that the even in the small-p_T limit, the NNLL+NLO uncertainty band is contained inside the NLL+LO band. At NNLL+NLO, the same observation as for the standard case still holds since both results are almost indistinguishable at a such logarithmic accuracy. Finally, it is worth emphasizing that in the context of soft-improved resummation, the matched results have a broadly similar features to the purely resummed ones. That is, in the small-p_T region, the matching of the improved resummation changes the result less than the matching of the standard resummation.

Z boson production at the Tevatron Run II

Let us know turn our attention to the application of the improved transverse momentum resummation to the case of Z boson production decaying into a pair of leptons. In particular, we consider the production of a Z boson with mass $m_z = 91.187$ GeV a the Tevatron [258–260] with a center of mass energy $\sqrt{s} = 1.96$ TeV. Such a choice is motivated by the fact that soft contributions for DY processes are expected to be more pronounced at colliders with lower center of mass energy. In all the subsequent DY analyses, the fixed-order results are obtained using DYQT [242, 243]. Fig. 14 shows the pure resummed results in which the soft-improved transverse momentum and standard resummations are compared. Analogous to Fig. 12, the top panels show the results for the different order of

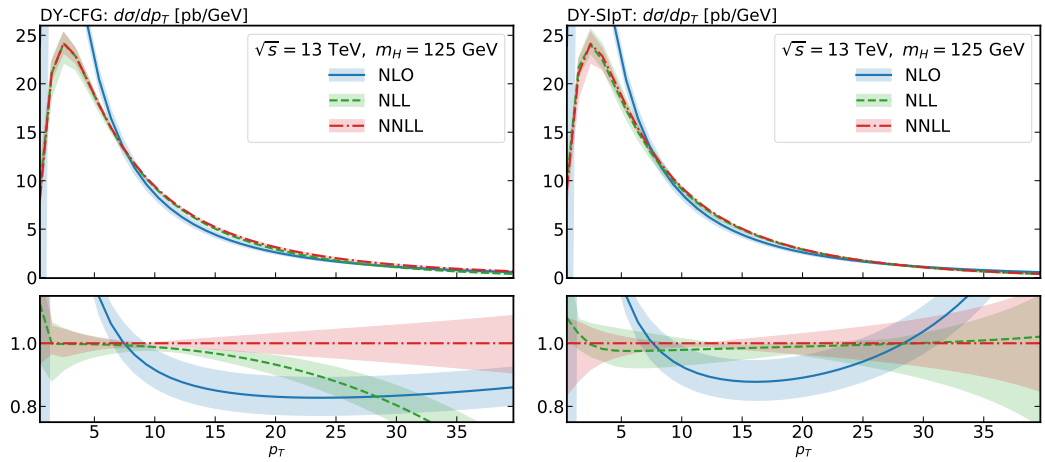


Figure 14 Transverse momentum spectra of the Z boson production at Tevatron run II with $\sqrt{s}$=1.96 TeV for the two types of resummation formalisms, namely the standard CFG (left) and the soft-improved transverse momentum resummation (right). The top panels show the pure resummed results at various logarithmic accuracies along with the NLO fixed-order result. The lower panels show the ratio of the various predictions w.r.t to the NNLL. The central scale is set to m_Z. The uncertainty bands are computed using the seven-point scale variation method, and in all cases, the resummation scale is fixed to $Q = m_Z$.

logarithmic accuracy while the lower panels show the ratio of the various predictions w.r.t the respective NNLL results. In both types of resummation, one sees similar features as for the Higgs boson production, namely the fact that the resummation yield a well-behaved transverse momentum distribution in the small-p_T region and that the distribution vanishes when $p_T \to 0$. The distribution then peak at about $p_T \sim 2$ GeV before vanishing again at large-p_T. Contrary to the case of the Higgs boson production, the resummed series, even for the standard approach, seem to converge fast: the NNLL uncertainty bands is almost entirely contained inside the NLL. Both at NLL and NNLL, the soft-improved transverse momentum resummation is analogous to the standard transverse momentum resummation at low-p_T; noticeable difference occur at values of large-p_T where the improved resummation displays better convergence. It can indeed be seen that not only the agreement between the NNLL+NLO and fixed order results has improved, but also the NLL+LO and NNLL+NLO bands overlap although the latter is not entirely contained in the former. In addition, one can also notice a slight decrease in uncertainties when going from the standard to the improved transverse momentum resummation.

The results of the resummed predictions when matched to fixed-order calculations are shown in Fig. 15. As in all the previous cases, the results using the standard resummation formalism are shown on the left while the results from the soft-improved standard resummation are shown on the right. Due to the nice convergence of the resummed series, there is a good agreement between the NLL+LO, NNLL+NLO, and NLO results. Notice that this can be seen in the two resummation formalisms. As a matter of fact, not only

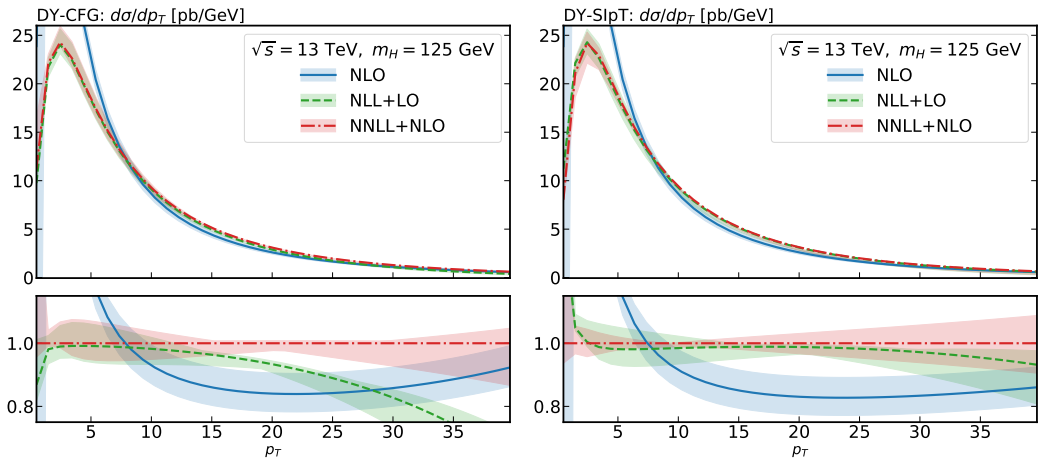


Figure 15 Transverse momentum spectra of the Z boson production at Tevatron run II with $\sqrt{s}=1.96$ TeV for the two types of resummation formalisms, namely the standard CFG (left) and the soft-improved transverse momentum resummation (right), when matched to fixed-order predictions. The top panels show the matched results at various logarithmic accuracies (NLO+LO and NNLL+NLO) along with the NLO fixed-order result. The lower panels show the ratio of the various predictions w.r.t to the NNLL+NLO result. The central scale is set to m_Z. The uncertainty bands are computed using the seven-point scale variation, and in all cases, $Q=m_Z$.

the NNLL+NLO bands are smaller compared to the NLL+LO, but also the latter is also contained in the former. For both resummation formalisms, the contributions of the finite components (LO & NLO) to the resummed results are less than 5% in the small-p_T region and between 10-20% in the region where p_T is large. The improved transverse momentum resummation only differs from the standard approach in two ways. First, the uncertainty bands in the improved resummation is seen to be smaller compared to the standard resummation. This is at least true for all values of p_T shown on the plots. Second, while both resummations clearly yield inaccurate predictions in the large-p_T regions due to the presence of non-justified logarithms, the soft-improved resummation seems to converge to the fixed-order result where the standard resummation appears to deviate from it.

3.3.4 Phenomenological results for the combined resummation

The soft-improved transverse momentum resummation clearly does not account for all the soft logarithms that appear in the threshold resummation. These are soft gluons emitted at large angle. Thus, in order to take these soft logarithms into account, one needs to consistently incorporate the pure threshold resummation into the soft-improved transverse momentum resummation. While observations suggest that this might be possible, the non-commutativity of the Mellin and Fourier integrals makes it difficult to construct such an expression. In the following passage, we briefly expand on the soft contributions

present in the improved resummation while explicitly referring to the threshold resummed expressions.

The threshold resummation at fixed ξ_p consists of four factors: three ξ_p independent and one ξ_p dependent. Two of the ξ_p independent terms originate from the incoming legs, while the third originates from the system that recoils against the large-ξ_p colourless final state. The ξ_p dependent term resums both a subclass of soft and recoil contributions.

Let us first start with the soft contributions that are not at all present in the soft-improved transverse momentum. Soft radiations originating from the outgoing legs that are soft but not collinear are only present at fixed-ξ_p and are therefore missing in the improved resummation. Such contributions can be simply added to the improved resummation with the logarithmic terms in ξ_p subtracted from the $\sigma^{\rm F}_{c\bar{c}} C^{(1)}_{abc}$ prefactor. This is because threshold resummation starts one order higher than the (soft-improved) transverse momentum resummation.

Next, Let us discuss the interference term which simultaneously resums subclasses of soft and collinear contributions. In Δ_{abc}, soft logarithms are resummed to all order while logarithms of ξ_p are only resummed at NLL. That is, only terms of the form $\alpha_s^n \ln^m N \ln \xi_p$ appears in the interference term while for example term such as $\alpha_s^2 \ln N \ln^2 \xi_p$ do not. These terms, upon expansion of the full threshold resummed expression, will be multiplied by positive powers of ξ_p which from the point of the view of the improved resummation are power suppressed. Based on the renormalization group arguments, such contributions can be made manifest in the the soft-improved transverse momentum resummation by replacing q^2 with q^2/N in the argument of α_s in the perturbative function A_c of Eq. (2.4.49).

Lastly, let us comment on the two radiative contributions which originate from the incoming legs. Subclasses of these contributions are already resummed in the standard transverse momentum resummation. Indeed, the standard transverse momentum resummation has N independent coefficients provided that the PDF evolutions are evaluated at the scale $1/\hat{b}$ in which case particular classes of soft logarithms are resummed in the anomalous dimensions. However, at the inclusive level, the total cross section also contains double soft collinear logarithms which arise when one integrates over the whole range of ξ_p. These contributions are not included in the standard resummation approach but present in the improved formulation of transverse momentum resummation. Indeed, due to the property that soft-improved transverse momentum leads to the total inclusive cross section when $\hat{b}=0$ (equivalent to integrating in ξ_p), these logarithms which appear at the total cross section level are taken into account. In addition to these, thanks to the new factorization, the soft-improved transverse momentum resummation (as in the threshold resummation) exhibits the $1/\sqrt{N}$ behaviour in the large-N limit. This, however, is only true at the level of the generating function in Eq. (2.4.6). Upon performing first the Fourier and then the Mellin integrals, this behaviour is lost. One can indeed check in Eq. (2.4.40) that such a behaviour is no longer present. Thus, in order to consistently account for these missing soft logarithms, one needs (at least) to construct an expression of the generating function where this behaviour in the large-N limit is preserved. Such an expression is not available at present.

For the sake of phenomenological studies, it suffices to combine the pure threshold and soft-improved transverse momentum resummations using some suitable matching function. The matching function must be chosen such that the the combined result reproduces the soft-improved transverse momentum and threshold resummation at small-

ξ_p and large-x respectively. One possible expression for the matching function has been proposed and studied in Refs. [165, 235] where the combined resummed expression is defined as follows

$$\frac{d\hat{\sigma}^{\text{cmb}}_{ab}}{d\xi_p}\left(N,\xi_p\right) = \left(1 - T\left(N,\xi_p\right)\right)\frac{d\hat{\sigma}^{\text{tr}}_{ab}}{d\xi_p}\left(N,\xi_p\right) + T\left(N,\xi_p\right)\frac{d\hat{\sigma}^{\text{th}}_{ab}}{d\xi_p}\left(N,\xi_p\right). \tag{3.3.13}$$

The matching function is defined as $T = N^k\xi_p^m / \left(1 + N^k\xi_p^m\right)$ where the values of k and m can be chosen arbitrarily provided that $m<k$. This is because the combined resummation results that differ from the soft-improved transverse momentum resummation by $\mathcal{O}(\xi_p^m)$ corrections when $\xi_p \to 0$, and from the threshold resummation by $\mathcal{O}(N^{-1})$ corrections when $N \to \infty$. Such a combined resummed expression resums small-ξ_p logarithms up to NNLL accuracy while threshold ones are resummed up to NNLL* according to the counting in Table 1 of Ref. [261].

For the subsequent phenomenological studies of the impact of adding the pure threshold contribution to the soft-improved transverse momentum resummation, we consider the exact same settings as in Sec. (3.3.3). In addition to defining a proper value of the cutoff entering in the Borel prescription, one also has to define the free parameters entering the matching function in Eq. (3.3.13). The following results are produced by setting $k=3$ and $m=2$. There is some arbitrariness in the definition of k and m, however, we did explicitly check that results do not change provided that k is chosen from 2 to 5 and $m<k$. The variation of these parameters can therefore be used to assess the ambiguity of the combined resummed expression.

Higgs boson production at LHC

Analogous to the previous phenomenological studies, let us first study the impact of the combined resummation formalism to the Higgs boson production. The results are shown in Fig. 16 with the pure resummed results on the left and the matched to fixed-order results on the right. Since the contribution of the pure threshold resummation is only effective when N is large and $\hat{b}$ is fixed, soft-improved transverse momentum and combined resummation are exactly similar up to scales of at least 40 GeV. This is true for both unmatched and matched results. Comparing the pure combined results with the pure soft-improved results in Fig. 12, we notice significant difference in the large-p_T region. The inclusion of the threshold contributions leads to a surprising agreement between the NNLL and the fixed-order result, although noticeable difference persists between the NLL and the NLO results for larger values of p_T. Perhaps the most interesting feature of the combined resummation is the fact that the even in the absence of a matching procedure, the resummation computation seems to capture the behaviour of the fixed-order results in the regions where where standard transverse momentum resummation fails to give accurate predictions. At the matched level, the combined results are different from the soft-improved resummation in that the former yield better agreement with the fixed-order results. When going from the pure combined resummed results to the matched results, we see some increase in the uncertainty in the intermediate and large-p_T regions which suggests that the uncertainty is underestimated in the case matching is absent. The upshot here is that the matching of the combined resummation changes the result less than the

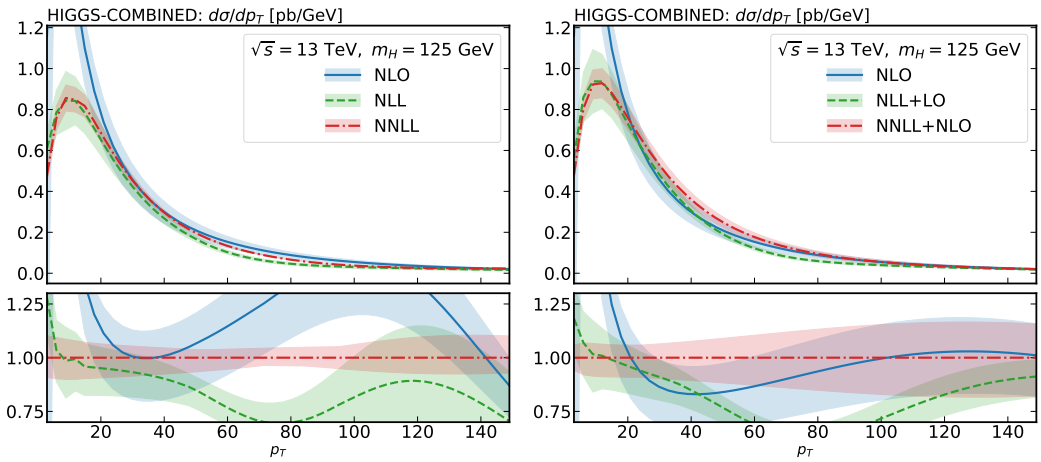


Figure 16 Higgs transverse momentum spectra from gluon fusion at $\sqrt{s}=13$ TeV for the combined resummation formalisms. The pure resummed results are shown on the left when the results matched to fixed-order calculations are shown on the right. The top panels show the absolute results for the various logarithmic orders along with the NLO fixed-order result. The lower panels show the ratio of the various predictions w.r.t to the NNLL (+NLO) result. The central scale is set to m_H. The uncertainty bands are computed using the seven-point scale variation, and in all cases, the resummation scale is set to $Q=m_H$.

matching of the soft-improved resummation, and the latter less than the matching of the standard transverse momentum resummation.

Z boson production at the Tevatron Run II

Let us finally close the section on the improved and combined resummations by presenting the numerical results for the Z boson production at Tevatron run II. The settings are exactly similar to the results presented in previous sections. The results are shown in Fig. 17 with the pure combined resummed results on the left and the results matched to fixed-order calculations on the right. By comparing the pure combined resummed results with the pure soft-improved results in Fig. 14, we see similar features as for the Higgs boson production, namely that the combined resummation yield surprising agreement with the fixed-order results in the large-p_T region. This is even more pronounced in the case of Z boson production as the the three bands (namely NLO, NLL, and NNLL) now overlap. This behaviour remains when we match the pure resummed calculations to the fixed-order results. In particular, we can notice impressive agreement between the NLL+LO, NNLL+NLO, and NLO results. Not only is the NNLL+NLO band smaller compared to the NLL+LO but the latter is also contained in the former suggesting a good convergence of the resummed perturbative expansion. These results suggest that for the transverse momentum distribution of the Z boson where x is often far from threshold (unity), the effect of the threshold resummation is less pronounced.

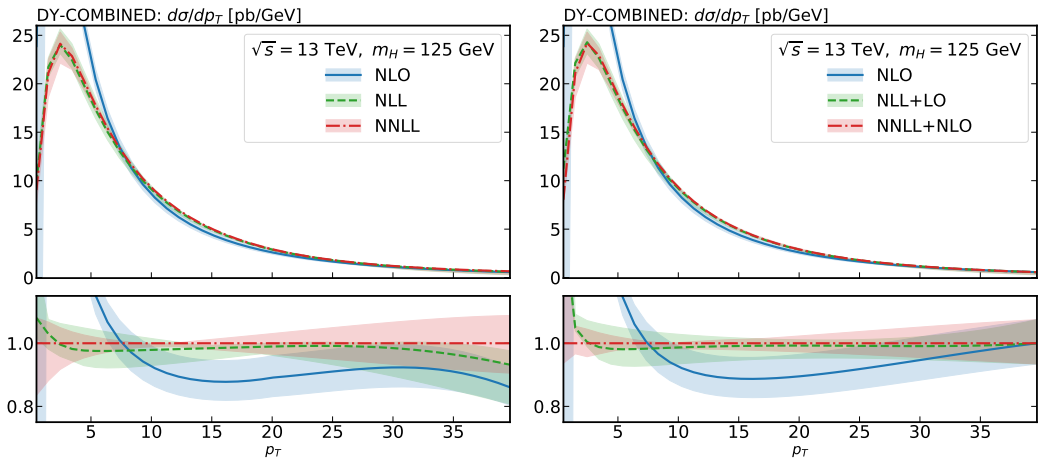


Figure 17 Transverse momentum spectra of the Z boson production at Tevatron run II with $\sqrt{s}=1.96$ TeV for the combined resummation formalisms. The pure resummed results are shown on the left when the results matched to fixed-order calculations are shown on the right. The top panels show the absolute results for the various logarithmic orders along with the NLO fixed-order result. The lower panels show the ratio of the various predictions w.r.t to the NNLL (+NLO) result. The central scale is set to m_Z. The uncertainty bands are computed using the seven-point scale variation, and in all cases, the resummation scale is set to $Q=m_Z$.

3.4 Approximating NNLO Higgs p_T distributions using resummations

To the present day, the inclusive and transverse momentum distributions of the Higgs boson production are known to very high precision. The total inclusive cross section was originally computed at next-to-leading order (NLO) in the heavy-top-quark limit in Refs. [182, 183, 262] and later on with finite top mass in Ref. [263]. These results were extended to NNLO in the infinite top mass limit in Refs. [9,69,264] and in Refs. [69,265–270] for the finite top mass. The N3LO result including the finite top quark mass effect was first available as an approximation using the knowledge of the singularity structure in Mellin space of the all-order computations [133]. Recently, several calculations at N3LO with finite-top quark mass have been available [10,271–274]. As introduced in previous passages, the differential distributions for the Higgs boson production is known up to NNLO [23–26]. Since these results are often obtained numerically, reading off the coefficients that are relevant for the comparison to all-order computations is not practically feasible.

In the following sections, we present a formalism for the construction of the Higgs transverse momentum distributions beyond NNLO using all-order computations. The approach is inspired by Ref. [133] in which it was argued at the inclusive level that an approximation to a N3LO expression can be derived from the knowledge of the singularity structure of the all-order computations in Mellin space. It was suggested that an approximation to a partonic cross section can be constructed by simply combining the small and large-N behaviours which can respectively be predicted from the high-energy

and threshold resummation. Indeed, as we have highlighted many times, threshold resummation embodies to all-order in α_s logarithms of the form $\ln N$ that drives the transverse momentum spectra in the limit $N \to \infty$, while the behaviour $(N-1)^n$ for $N \to 1$ is fully determined to all-order by the small-x resummation. Therefore, for a partonic cross section known up to N^{n}LO (i.e. $\mathcal{O}(\alpha_s^n)$), the approximate expression is constructed as a combination between fixed-order calculations and an expansion from resummations

$$\left[\frac{d\hat{\sigma}_{ab}}{d\xi_p}\right]^{\mathrm{N}^{n+1}\mathrm{LO}}(N,\xi_p) = \left[\frac{d\hat{\sigma}_{ab}}{d\xi_p}\right]^{\mathrm{N}^{n}\mathrm{LO}}(N,\xi_p) + \frac{d\hat{\sigma}_{ab}^{(n+1)}}{d\xi_p}(N,\xi_p), \tag{3.4.1}$$

where the last term on the right-hand side is the approximate expression computed by expanding the all-order computations and retaining only the $\alpha_s^{(n+1)}$-term. Its expression is given as a contribution from the different small-N and large-N behaviours

$$\frac{d\hat{\sigma}_{ab}^{(n)}}{d\xi_p}(N,\xi_p) = \frac{d\hat{\sigma}_{ab}^{\mathrm{he},(n)}}{d\xi_p}(N,\xi_p) + \frac{d\hat{\sigma}_{ab}^{\mathrm{th},(n)}}{d\xi_p}(N,\xi_p). \tag{3.4.2}$$

Notice that no matching function was introduced here. This means that Eq. (3.4.2) is only valid if the small-N behaviour controlled by the high-energy contribution is not spoiled by the threshold component and vice-versa. While in Ref [133], it was shown that better approximation can be constructed by modifying the pure threshold contribution in such a way that the singularity structure at small-N is preserved, here we first present results without modifying the singularity structure of any of the contribution. We stress that the results presented here are just meant to be a proof of concept and more deeper analyses will be the subject of forthcoming studies in the context of DY. One might argue that the construction of such an approximation is doomed to fail since the Higgs production at the LHC is far from threshold. However, by analyses of the position of the saddle point in Mellin space [275] it was shown at the level of the inclusive cross section that the bulk of the contribution to the cross section are controlled by logarithmically enhanced terms. We expect this statement to hold at the level of transverse momentum distributions.

As a proof of concept, let us mainly focus on the gg-channel. This is justified since the main bulk of the cross section comes from the gg-channel while the contribution from the other channels are small. In Fig. 18 we show the contribution of the various (sub) channels ($gg, gq, q\bar{q}$, notice that gq and qg are the same) to the full hadronic cross section. We can clearly see that both at LO and NLO, the gg-channel is by far the most dominant channel that contributes to the Higgs boson production from gluon fusion. As a matter of fact, at leading-order the gg-channel contributes to about 70% of the full cross section while at next-to-leading order it contributes to about 85%. This is even more evident when looking at the relative difference in the lower panel of Fig. 18 where we see that for all the values of p_T considered, the uncertainty bands of the gg-channel and the full result overlap. Similar behaviour can be observed by comparing the partonic channels in Mellin space. This is shown in Fig. 19 where we plot the Mellin space partonic cross section for the various partonic channels from which we can see that the gg-channel provides the most dominant contribution.

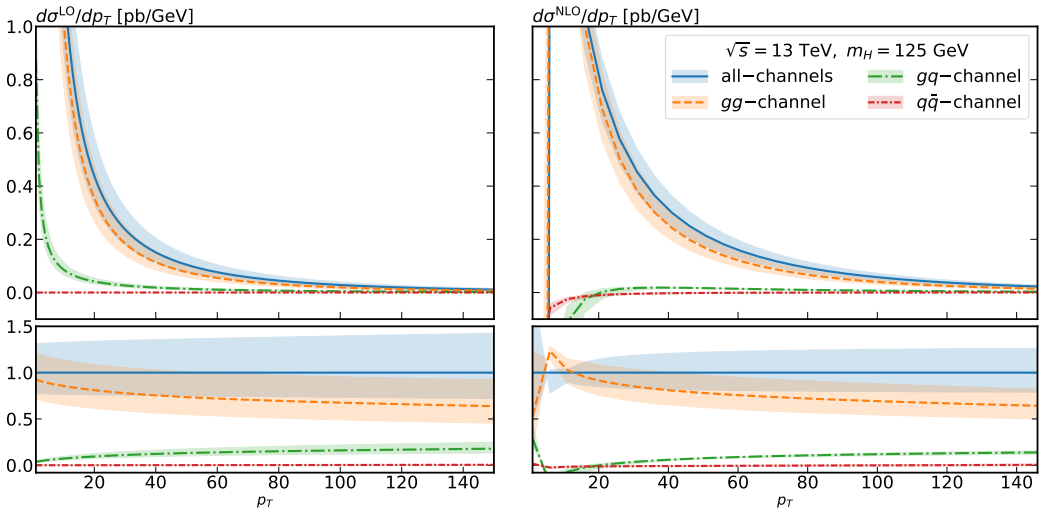


Figure 18 Contribution of the various (sub) channels $(gg, gq, q\bar{q})$ to the full hadronic cross section at the center of mass energy $\sqrt{s} = 13$ GeV. The full hadronic result (blue) is also for reference. The results are shown for both the LO (left) and NLO (right). The upper panels show the absolute results while the lower panels show the ratio between the various channels and the full hadronic result. the uncertainty bands are computed using the seven-point scale variation method depicted in Fig. 3. In all cases the mass of the Higgs is set to $m_H = 125$ GeV.

3.4.1 The large-N (large-x) region

Let us first start by describing the large-N approximation of the coefficient function (or partonic cross section) in Eq. (3.4.1). The threshold resummed expression is given by Eq. (2.3.20) which can be written in a much simpler form as

$$\frac{d\hat{\sigma}^{\text{th}}_{ab}}{d\xi_p}\left(N, \xi_p, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) = \tilde{\mathcal{C}}_{ab}(N, \xi_p) \exp\left\{\sum_{n=1}^{\infty} \alpha_s^{n-2} g_n\left(N, \xi_p, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right)\right\} \tag{3.4.3}$$

where $\tilde{\mathcal{C}}$ collects all the non-logarithmic dependence and all the enhanced logarithmic parts are resummed in the exponent. Notice that $\tilde{\mathcal{C}}$ is also a perturbative function in α_s where $\alpha_s \equiv \alpha_s(\mu_R^2)$. In our derivation, we showed that the g_n functions are known up to g_3. This will allow us to construct an approximation of the NNLO term of the transverse momentum distribution. Before doing so, let us first analyse Eq. (3.4.3). Its expansion as a series in the strong coupling can be expressed as

$$\frac{d\hat{\sigma}^{\text{th}}_{ab}}{d\xi_p}\left(N, \xi_p, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) = \tilde{\mathcal{C}}_{ab}(N, \xi_p) \sum_{n=1}^{\infty} \alpha_s^n \sum_{k=0}^{2n} \tilde{g}_{n,k}\left(N, \xi_p, \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right) (\ln N + \gamma_E)^k \tag{3.4.4}$$

where the coefficients $\tilde{g}_{n,k}$ can be determined to all-order in the strong coupling. One can immediately see that the appearance of the N-enhanced term on the right-hand side

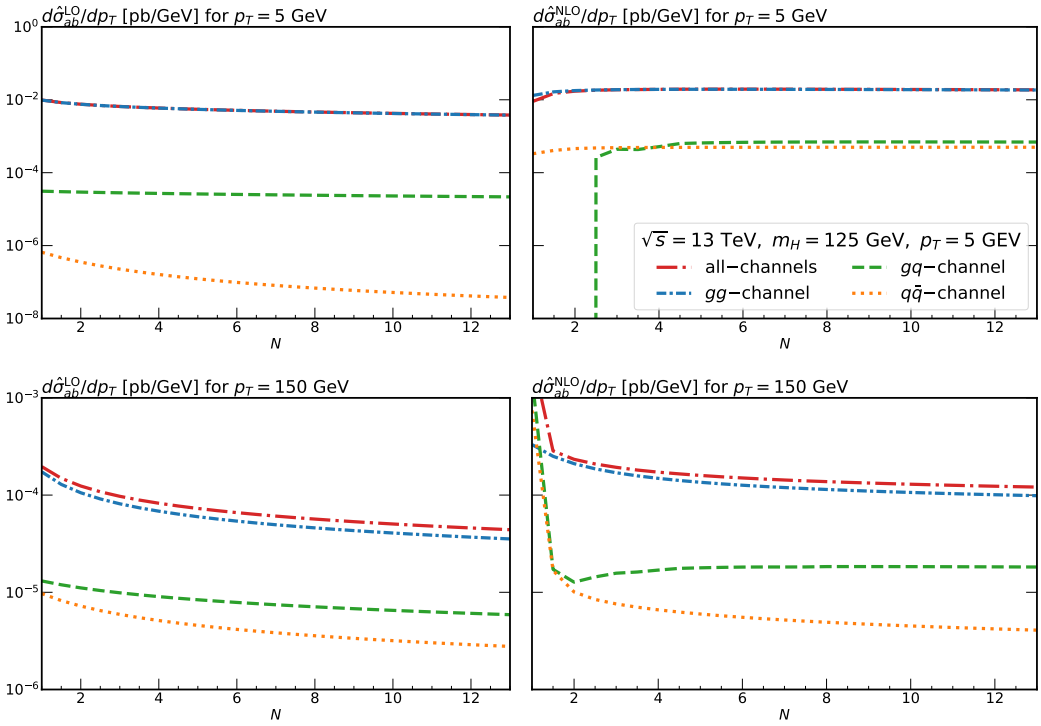


Figure 19 Contribution of the various partonic (sub) channels $(gg, gq, q\bar{q})$ to the partonic cross section in Mellin space. The results are shown for the Higgs boson production at the LHC with $\sqrt{s}=13$ TeV for different values of the transverse momentum, namely $p_T=5$ GeV (top) and $p_T=150$ GeV (bottom). The results have been computed at the central scale, i.e. the normalization, factorization, and resummation scales have been set to the Higgs boson mass.

of Eq. (3.4.4) does not exactly correspond to the Mellin transform of the x-space expressions. Indeed, the Mellin transform of the plus distributions (for $k=1$) appearing in x-space takes the following form

$$\int_0^1 dx\, x^{N-1} \left(\frac{\ln(1-x)}{1-x}\right)_+ = \frac{1}{2}\left(\gamma_E^2 + \frac{\pi^2}{6}\right) + \gamma_E \psi_0(N) + \frac{1}{2}\left(\psi_0^2(N) - \psi_1(N)\right), \quad (3.4.5)$$

where ψ_n denotes the Polygamma functions. The logarithms of N then arise upon taking the large-N limit of the digamma function ψ_0. Reciprocally, the inverse Mellin transform of $\ln N$ in terms of a plus distribution is given by

$$\mathcal{M}^{-1}(\ln N) \propto \left(\frac{\ln(\ln 1/x)}{\ln 1/x}\right)_+ . \quad (3.4.6)$$

In addition, not only the expanded expression in Eq. (3.4.4) does not exactly correspond to the ones that appear in the fixed-order calculations but also Eq. (3.4.4) displays an unphysical singularity. By inspecting the right-hand side of Eq. (3.4.4), one observes that

the expression exhibits a logarithmic branch cut at $N=0$. This is not in accordance with the exact fixed-order results whose singularity is represented by poles at small-N, not cuts. It is, however, crucial to emphasize that despite this mismatch–which only appear at small N–the asymptotic behaviour of the fixed-order calculations in the large-N limit is correctly reproduced by the expansion in Eq. (3.4.4). The correct singularity structure in the small-N region of the resummed expression can be restored back by carefully exploiting the analyticity of the Mellin space cross section, as described in Ref. [133]. Here, we resort to an approach in which the logarithms of N are simply replaced by the digamma functions [216]. This yields the following expanded resummed expression

$$\frac{d\hat{\sigma}_{ab}^{\text{th}}}{d\xi_p}\left(N,\xi_p,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)=\tilde{\mathcal{C}}_{ab}(N,\xi_p)\sum_{n=1}^{\infty}\alpha_s^n\sum_{k=0}^{2n}\tilde{g}_{n,k}\left(N,\xi_p,\frac{Q^2}{\mu_R^2},\frac{Q^2}{\mu_F^2}\right)\left(\psi_0(N)+\gamma_E\right)^k \tag{3.4.7}$$

The distinction between Eq. (3.4.4) and Eq. (3.4.7) is often referred to as N-soft versus ψ-soft. To better understand the difference between the two soft expressions, we compare them in Fig. 20 along with the partonic fixed-order result for the gg-channel for various values of p_T. On the y-axis is plotted the coefficient $d\hat{\sigma}_{ab}^{\text{tr},(2)}/d\xi_p$ which contains both the strong coupling α_s and the Born level cross section σ_0. Notice that our expressions are shifted such that the pole is now at $N=0$ instead of $N=1$. Since the transverse momentum distribution of the Higgs gluon production via gluon fusion is largely dominated by large N terms down to moderately small values of N, we decided to present the results for N between zero and one in order to clearly see what is happening close to the poles. This can be seen from the plots where the three curves start to converge from $N\sim 1.5$. We can see that the difference between the N-soft and ψ-soft schemes only lies in the small-N region. Henceforth, we choose to adopt the ψ-soft approach for the construction of the approximated expression.

3.4.2 The small-N (small-x) region

The leading logarithmic (LLx) high-energy (or often referred to as small-x) resummation of transverse momentum distribution has been derived in Ref. [276]. The computations were performed by keeping the initial-state gluons off their mass-shell, $p_i^2=|p_{T,i}|^2$. An impact factor h is then computed by means of a triple Mellin transform: one Mellin transform conjugate to the partonic variable x and two Mellin transforms w.r.t the offshellness of the incoming gluons $\xi_i=|p_{T,i}|^2/Q^2$. For the case of Higgs boson produced via gluon fusion in the HEFT formalism, the high-energy resummed transverse momentum distribution for the gg-channel is expressed as

$$\frac{d\hat{\sigma}_{gg}^{\text{he,H}}}{dy_H d\xi_p}\left(N,\hat{b},\xi_p\right)=h\left(0,\tilde{\gamma}^-\left(N,\hat{b}\right),\tilde{\gamma}^+\left(N,\hat{b}\right),\xi_p\right) \tag{3.4.8}$$

where the impact factor is defined according to

$$h\left(N,\tilde{\gamma}^-,\tilde{\gamma}^+,\xi_p\right)=\sigma_{\text{H},gg}^{\text{Born}}\frac{\tilde{\gamma}^+\tilde{\gamma}^-\xi_p^{\tilde{\gamma}^+}\xi_p^{\tilde{\gamma}^-}}{\xi_p\pi(1+\xi_p)^N}R\left(\tilde{\gamma}^+\right)R\left(\tilde{\gamma}^-\right) \tag{3.4.9}$$

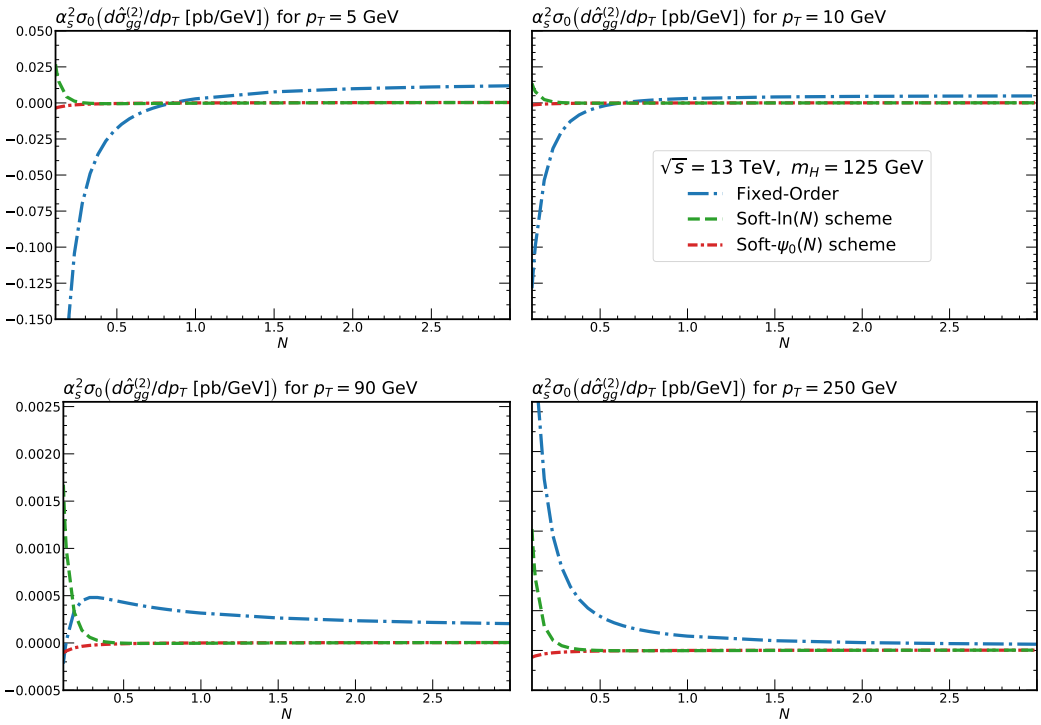


Figure 20 Second-order contributions of the Higgs transverse momentum distribution at the LHC with $\sqrt{s}=13$ TeV. The coefficient includes the strong coupling and the Born level cross section. The fixed-order result is compared to the two different types of soft approximations given in Eq. (3.4.4) and Eq. (3.4.7). The results are shown for different values of the transverse momentum, namely $p_T=5,10,90,250$ GeV. The results have been computed at the central scale, i.e. $m_H^2=\mu_F^2=\mu_R^2$.

$$\int_0^\infty \frac{d\xi_1}{\xi_1^2}\,\xi_1^{\tilde{\gamma}^+} \int_{(1-\sqrt{\xi_1})^2}^{(1+\sqrt{\xi_1})^2} \frac{d\xi_2}{\xi_2^2}\,\xi_2^{\tilde{\gamma}^-} \frac{(1-\xi_1-\xi_2)^2}{\sqrt{2\xi_1\xi_2+2\xi_1+2\xi_2-1-\xi_1^2-\xi_2^2}}.$$

As described in Ref. [276], the integrals in Eq. (3.4.9) can be computed by introducing a new variable u defined as $\xi_2 = 1+\xi_1-2\sqrt{\xi_1}u$ transforming the (ξ_1,ξ_2)-integral into an integration in ξ_1 and u. The final result can consequently be obtained in a closed form by first performing an expansion in powers of u, then performing the integration term-by-term, and finally summing back the series. This leads to the following expression

$$h\left(N,\tilde{\gamma}^-,\tilde{\gamma}^+,\xi_p\right) = \sigma_{\mathrm{H},gg}^{\mathrm{Born}} \frac{\xi_p^{\tilde{\gamma}^+}\xi_p^{\tilde{\gamma}^-}}{\xi_p(1+\xi_p)^N} R\left(\tilde{\gamma}^+\right) R\left(\tilde{\gamma}^-\right) \times \left[\left(1+\frac{2\tilde{\gamma}^+\tilde{\gamma}^-}{1-\tilde{\gamma}^+-\tilde{\gamma}^-}\right) \frac{\Gamma\left(1+\tilde{\gamma}^+\right)\Gamma\left(1+\tilde{\gamma}^-\right)\Gamma\left(2-\tilde{\gamma}^+-\tilde{\gamma}^-\right)}{\Gamma\left(2-\tilde{\gamma}^+\right)\Gamma\left(2-\tilde{\gamma}^-\right)\Gamma\left(\tilde{\gamma}^++\tilde{\gamma}^-\right)}\right]. \tag{3.4.10}$$

The perturbative functions R accounts for the factorization scheme. In $\overline{\mathrm{MS}}$ scheme, the first non-vanishing orders are given by $R(\tilde{\gamma})=1+8\zeta_3\tilde{\gamma}^3/3+\mathcal{O}(\tilde{\gamma}^4)$. The resummed

anomalous dimensions $\tilde{\gamma}^{\pm}$ are also perturbative functions and therefore can be written as a series in the coupling $\tilde{\gamma} = \sum_{n=1}^{\infty} \tilde{\gamma}^{(n)}$. It has been shown [277,278] that these anomalous dimension entirely determine the nature of the small-N singularity of the (differential) cross sections at the the resummed level. They are expressed in terms of the standard anomalous dimensions via the following relation

$$\tilde{\gamma}^{\pm}(N, \hat{b}) = \gamma\left(N \pm i\frac{\hat{b}}{2}\right). \tag{3.4.11}$$

For the construction of the approximate differential distribution, we are are interested in the finite order expansion of the resummed expression given in Eq. (3.4.8). Thus, one can expand the resummed expression in Eq. (3.4.8) as a series in the strong coupling and perform the integration over the rapidity which is equivalent to setting $\hat{b}=0$. The NLO high-energy resummed expression is then written as

$$\frac{d\hat{\sigma}^{\text{he,H}}_{gg}}{d\xi_p}(N, \xi_p) = \sigma^{\text{Born}}_{\text{H},gg}\left[\left(\frac{\alpha_s}{\pi}\right)\frac{C_A}{N}\frac{2}{\xi_p} + \left(\frac{\alpha_s}{\pi}\right)^2\left(\frac{2C_A}{N}\right)^2\frac{\ln \xi_p}{\xi_p}\right] + \mathcal{O}(\alpha_s^3). \tag{3.4.12}$$

Higher-order coefficients up to $\mathcal{O}(\alpha_s^4)$ are derived by taking the Mellin transforms of Eqs. (4.12) of Ref. [276]. Eq. (3.4.12) is what we are going to use for the construction of the approximate transverse momentum distribution.

In order to assess the goodness of the high-energy expressions as an approximate to the fixed-order result in the small-N (equivalently small-x) regions, we compare the the high-energy approximation to the known second order coefficient from the fixed-order computations. The results are shown in Fig. 21 for various values of p_T. Since we are mainly interested in the region where the bulk of the contribution is dominated by small-x physics, we show the plots for moderately small values of N. In view of combining the high-energy approximation to the threshold approximation–which provides the large-N behaviour–one has to make sure that the small-N contributions vanish at moderately large value of N. In Ref. [133], this is systematically enforced by subtracting from the high-energy expressions contributions that have the same small-N singularities but vanish as $N \to \infty$. Such a restriction is not required in our case as the distribution clearly vanishes when $N \to \infty$. Eq. (3.4.12) indeed shows that as long as the transverse momentum does not vanish, the transverse momentum distribution always vanish in the large-N limit. Interestingly, the vanishing point is located in the vicinity of $N=1$ which coincides with what has been observed at the inclusive level [133] considering that the singularity in the plots of Fig. 21 is shifted to the left by one unit w.r.t. the aforementioned reference. Fig. 21 shows that the high-energy expressions provide a good approximate to the fixed-order results in the small-N region while not spoiling the large-N singularities. Only for $p_T = 90$ GeV small discrepancies between the two results can be noticed.

3.4.3 Partonic level results at NLO

Before presenting results for the approximate NNLO transverse momentum distribution, which will be shown in the subsequent section at the hadronic level, let us first compare the approximate predictions to known exact results, namely NLO. The results are shown

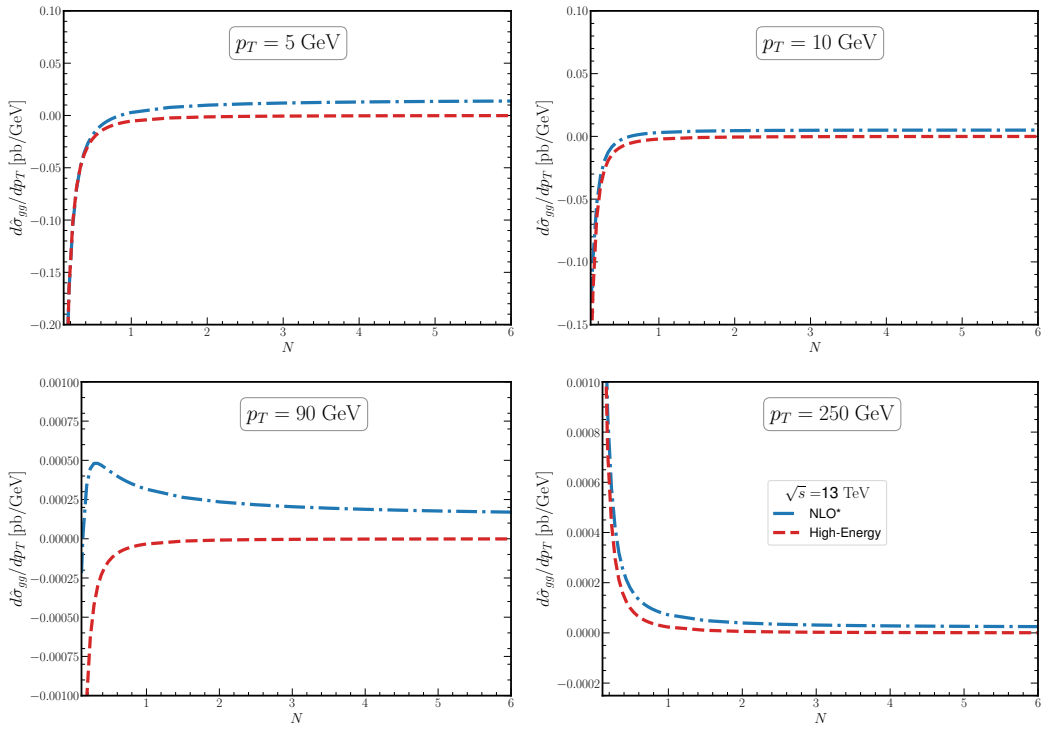


Figure 21 Second-order contributions of the Higgs transverse momentum distribution at the LHC with $\sqrt{s}=13$ TeV. The coefficient includes the strong coupling and the Born level cross section. The fixed-order result is compared to the high-energy approximation given in Eq. (3.4.12) and considering only the second-order terms. The results are shown for different values of the transverse momentum, namely $p_T=5, 10, 90, 250$ GeV. The results have been computed at the central scale.

in Fig. 22 for different values of p_T; similar results but zoomed in the small-N region are shown in Fig. 23. Each plot compares the exact NLO coefficient functions to the pure threshold and high-energy approximations, and to the combined small and large-N approximation. While the absolute results are shown on the top panels, the ratio of each approximations to the exact result is shown on the bottom panels. The uncertainty bands are obtained as the envelope from the variation of the renormalization and factorization scales according to the seven-point method.

Let us first focus our attention at the small-N region ($N<1$). For all the values of p_T, it is apparent that the combined approximate results reproduce fairly well the exact fixed-order results. Not only the two uncertainty bands of two results overlap, but in all the cases except for $p_T=5$ GeV, the uncertainty band of the exact results are contained within the approximation. This, at least, suggests that the approximations do not underestimate uncertainties associated to missing higher-order corrections. Moving to the region where $N>1$, one notices that discrepancies appear between the exact and approximate results, as if the latter is the same as the former upon addition or subtraction of a constant term.

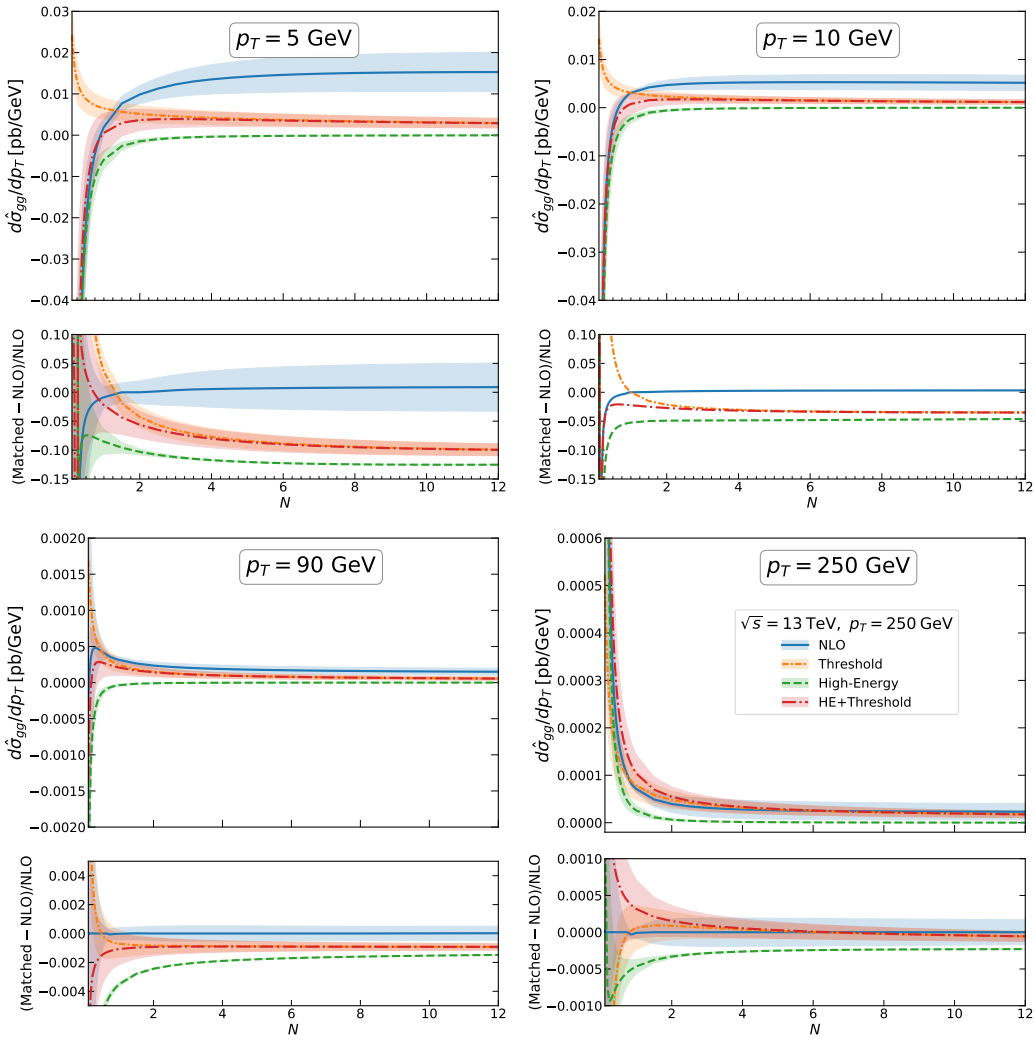


Figure 22 Comparison of the second order coefficient of the Higgs transverse momentum distribution (blue) to the various approximations resulting from the expansion of the threshold (orange), high-energy (green), and combined threshold and high-energy resummations (red). The Higgs boson is produced via gluon fusion from a collision of protons at $\sqrt{s}=13$ GeV. The top panels show the absolute results while the bottom panels show the ratio of the various predictions to the the exact NLO. The results are shown for different values of the transverse momentum, namely $p_T=5, 10, 90, 250$ GeV.

This could be due to the fact that, as opposed to Ref. [133], we have not modified the subleading terms that contribute to the threshold expressions. This difference, however, reduces as the value of the transverse momentum increases. Indeed, while the combined small and large-N behaviour provide a reasonable good approximation to the exact result when $p_T=90$ GeV, the uncertainty band of the approximate result is entirely contained in

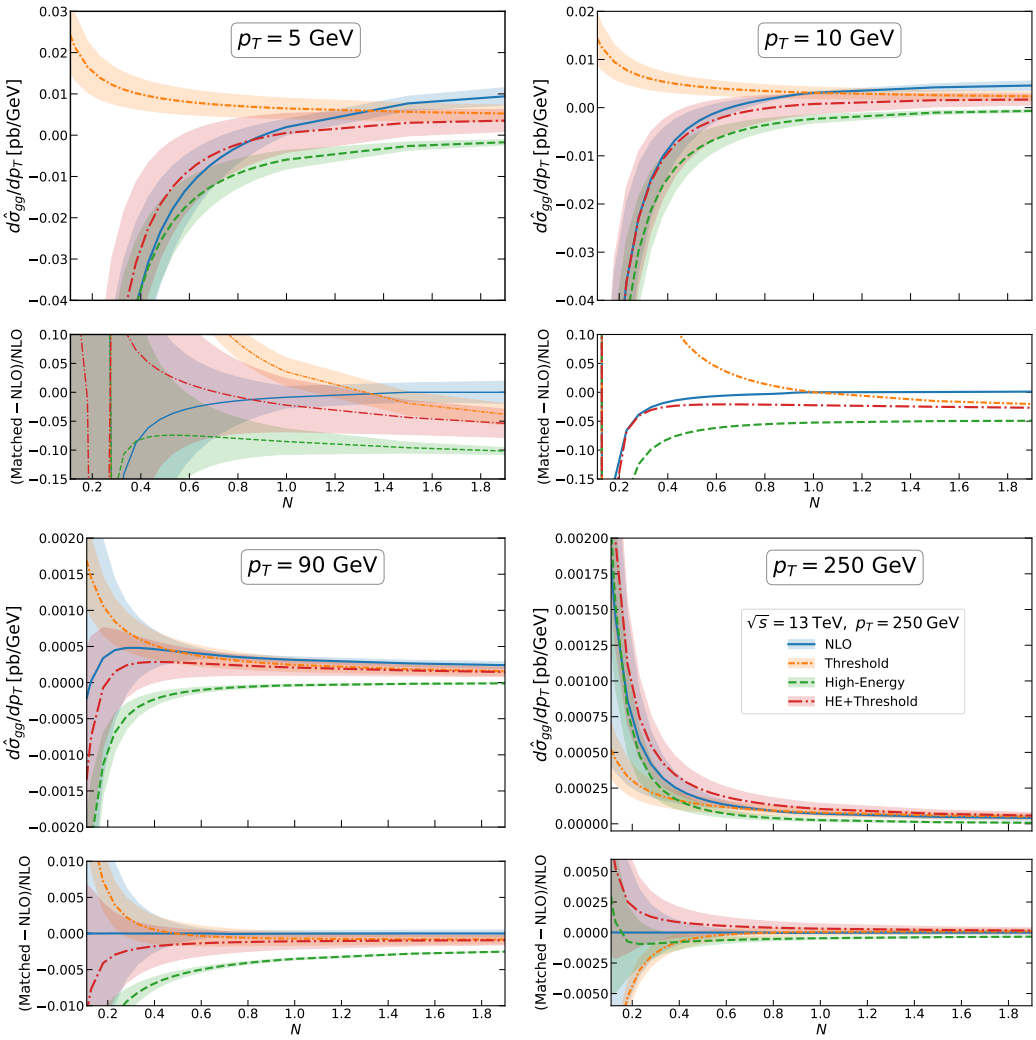


Figure 23 Same as Fig. 22 but zoomed in the small-N region.

the uncertainty band of the fixed-order exact result when $p_T = 250$ GeV. This suggests that the large-N approximation could be improved by supplementing to it a contribution that vanishes as $p_T \to \infty$. These results support the statement made earlier that the impact of the threshold resummation is more significant for large values of p_T.

3.4.4 Hadronic level results

Let us finally turn our attention to the approximate NNLO transverse momentum distribution. Similar to what was done at the partonic level, for the approximation, we supplement the known NLO transverse momentum spectra with the α_s^3-terms from the expansion of the threshold and high energy resummations. The inverse Mellin transform is then

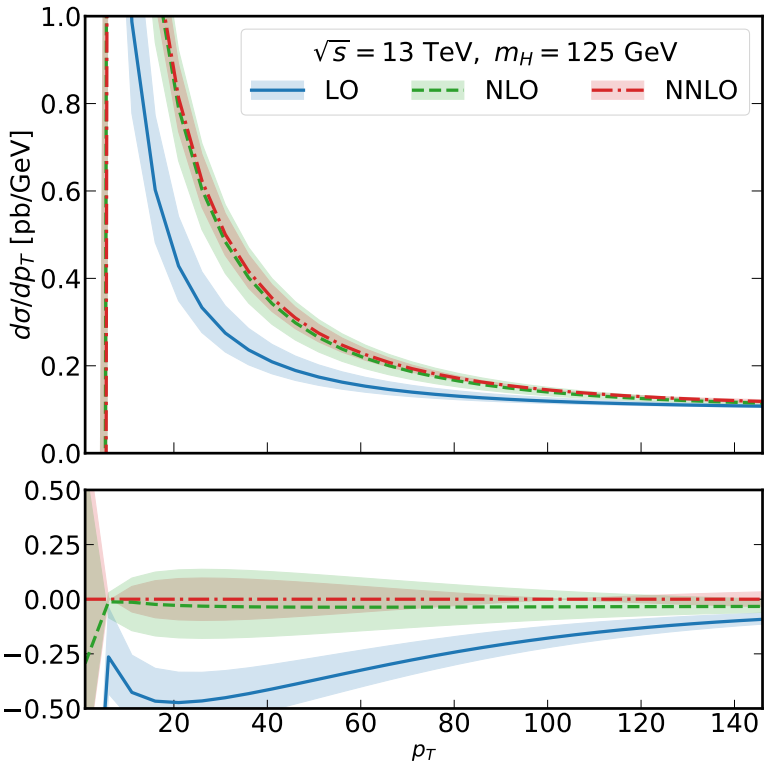


Figure 24 Approximate NNLO Higgs transverse momentum distribution for the gg-channel by combining threshold and high energy resummations. The NNLO prediction is represented by the *red* curve. For comparison, the LO (*blue*) and NLO (*green*) are also shown. The top panel represents the absolute results while the bottom panel represents the relative difference between the various predictions and the NNLO. As usual, the uncertainty bands are computed using the seven-point scale variation. In all the plots, the resummation scale Q is set to the Higgs mass m_H.

computed using the same contour deformation used to define the Minimal Prescription as described in Sec. (3.1.1). In order to alleviate the need of presenting plots with different values of p_T, we present directly the results at the hadronic level. The results are shown in Fig. 24. For comparison, both the LO and NLO results are also included. As discussed in the previous section, the results are only shown for the gg channel. As expected, the uncertainty band of the NNLO prediction is slightly smaller compared to the one from NLO. In addition, the uncertainty bands of the NLO and NNLO predictions mostly overlap in the range of p_T values considered in which the latter is mostly contained in the former.

3.5 Summary

This chapter focused on the phenomenological studies of the transverse momentum distributions of two colour singlet objects, namely the Higgs boson and the Z-boson

decaying into a lepton pair via DY mechanism. In doing so, we discussed the issues of performing the inverse Fourier and Mellin transforms in the context of soft-improved transverse momentum resummation. To circumvent such issues, we resorted to the Borel prescription that has proved to not only yield better control on the subleading terms but also yield good numerical stability. This allowed us to (separately) assess the impacts of the soft-improved transverse momentum and combined resummation. It was shown that while the modified transverse momentum resummation yields perturbative convergence in the small-p_T region, the effect of including the pure threshold contributions improves the agreement between the fixed-order calculations in the medium and large-p_T regions. In the last part, we explored the idea of combining resummations to approximate unknown orders. In particular, we used the expansions from the threshold and high energy resummations to approximate the large and small N respectively. Such an approximation seem to yield reasonable predictions as attested by the partonic and hadronic results. However, further works are required in order to fully understand the large N singularities and improve the predictions in the region where threshold behaviours dominate. As mentioned before, the main objective will be to eventually apply such procedures to DY processes that can be used in PDF fit.

3.A Chebyshev polynomials

In this section, we briefly recall the definition of Chebyshev polynomials that we used in Sec. (3.1.2) to Mellin transform parton density functions. The Chebyshev polynomials are defined as a series in a parameter $x \in [-1, 1]$ whose degree depends on the truncation of the series. They are defined as follows

$$T_i(x) = \sum_{n=k}^{n} T_{nk} x^k. \tag{A.1}$$

The first three coefficients of the above series are given by the following

$$T_0(x) = 1, \qquad T_1(x) = x, \qquad T_2(x) = 2x^2 - 1, \tag{A.2}$$

while the higher-order coefficients are defined recursively as

$$T_n(x) = 2x\, T_{n-1}(x) - T_{n-2}(x). \tag{A.3}$$

Given the definition above, a generic function $f(z)$ can be approximated in the range $[z_{\min}, z_{\max}]$ by expanding it on the basis of the Chebyshev polynomials,

$$f(z) = \sum_{k=0}^{n} \tilde{c}_k \left(Az + B\right)^k \tag{A.4}$$

where n is a positive integer and represents the degree at which the expansion is performed. The parameters A, B, and $\tilde{c}_k$ are respectively defined as

$$A = \frac{2}{z_{\max} - z_{\min}}, \qquad B = \frac{z_{\max} + z_{\min}}{z_{\min} - z_{\max}}, \qquad \tilde{c}_k = -\frac{c_0}{2}\delta_{k0} + \sum_{i=m}^{n} c_i T_{ik}. \tag{A.5}$$

The expression given in Eq. (A.4) is analytic in z and its Mellin transform can be easily computed analytically or numerically. The coefficients c_i are best determined using libraries for scientific calculations such as the GNU Scientific Library.

3.B Mellin space FO Higgs production as implemented in HPT-MON

In this Appendix, we describe the procedure to perform the numerical Mellin transform of the momentum space expressions of the transverse momentum distribution of the Higgs boson produced via gluon fusion. The following procedure is implemented in HPT-MON that can produced predictions both in momentum and Mellin space either as a single differential in p_T or as a double differential in p_T and rapidity y_H.

The double differential distribution at the hadronic level is expressed as follows

$$\frac{d\sigma}{dp_T^2 dy_H}\left(\frac{p_T^2}{Q_T^2}\right) = \sum_{a,b}\int_0^1 dx_a dx_b\, f_a(x_a,\mu_F^2) f_b(x_b,\mu_F^2)\frac{d\hat{\sigma}_{ab}}{dp_T^2 dy_H}\left(\frac{p_T^2}{Q_T^2}\right), \tag{B.1}$$

where $Q_T^2 = Q^2 + p_T^2$ with Q being the momentum of the final state partons that balance the Higgs boson. Notice that for simplicity, we have omitted the renormalization scale (μ_R^2) dependence for simplicity.In some instances, we drop the argument of the cross sections for brevity. The partonic cross section contains contributions up to next-to-leading order (NLO). For the reason that will become obvious later, we organize the terms that contribute to the partonic component in the following

$$\frac{d\hat{\sigma}_{ab}}{dp_T^2 dy_H}\left(\frac{p_T^2}{Q_T^2}\right) = \frac{\sigma_{gg}^{\rm LO}}{\hat{s}}\left[\Sigma_{ab}^{\delta(Q^2)}\left(\frac{p_T^2}{Q_T^2}\right) + \Sigma_{ab}\left(\frac{p_T^2}{Q_T^2}\right)\right], \tag{B.2}$$

where $\sigma_{gg}^{\rm LO}$ and $\sqrt{\hat{s}}$ are the usual Born level cross section and the partonic center of mass energy respectively. The first term $\Sigma_{ab}^{\delta(Q^2)}$ contains all the expressions that are proportional to the $\delta(Q^2)$ function. It therefore contains the full LO contribution and some part of NLO. Notice that such terms are singular in the small-p_T limit. They are proportional to logarithms of the form $(\alpha_s^m/p_T)\ln^n(p_T^2/m_H^2)$. The second term Σ_{ab} embodies contributions that are free of $\delta(Q^2)$. It contains both singular and regular terms when $p_T \to 0$. The perturbative expansion of the delta-dependent part can be written as

$$\Sigma_{ab}^{\delta(Q^2)}\left(\frac{p_T^2}{Q_T^2}\right) = \left(\frac{\alpha_s}{\pi}\right)\Sigma_{ab}^{\delta(Q^2),(1)}\left(\frac{p_T^2}{Q_T^2}\right) + \left(\frac{\alpha_s}{\pi}\right)^2\Sigma_{ab}^{\delta(Q^2),(2)}\left(\frac{p_T^2}{Q_T^2}\right). \tag{B.3}$$

The leading-order coefficient corresponds exactly to the LO contribution whose expression is simply given by $\Sigma_{ab}^{\delta(Q^2),(1)} = g_{ij}\delta(Q^2)$ where the sub-partonic channels are defined as follows [279]

$$g_{gq} = C_F\left(\frac{\hat{t}^2+\hat{s}^2}{-\hat{u}}\right) \tag{B.4}$$

$$g_{q\bar{q}} = 2C_F^2\left(\frac{\hat{t}^2+\hat{u}^2}{\hat{s}}\right) \tag{B.5}$$

$$g_{gg} = N_c\left(\frac{m_H^8+\hat{s}^4+\hat{t}^4+\hat{u}^4}{\hat{u}\hat{t}\hat{s}}\right). \tag{B.6}$$

Due to crossing symmetry, the qg-channel can be read-off from the expression of gq by a swap of $\hat{t}$ and $\hat{u}$. Recall that $\hat{t}$ and $\hat{u}$ are the Mandelstam variables which in terms of the rapidity can be expressed as

$$\hat{t} = m_H^2 - \sqrt{s}\, x_b m_\perp e^{y_H} \tag{B.7}$$

$$\hat{u} = m_H^2 - \sqrt{s}\, x_a m_\perp e^{-y_H} \tag{B.8}$$

and they satisfy the relation $Q^2 = \hat{s} + \hat{t} + \hat{u} - m_H^2$. The NLO contribution is more complicated as takes contributions from three different sources: the *real* corrections that arise from the Higgs plus two-partons at the Born level, the *virtual* corrections that arise from the interference of the Born and one-loop amplitudes, and the *Altarelli-Parisi* corrections that arise from the definition of the parton densities at NLO in $\overline{\text{MS}}$ scheme. Isolating the terms that are proportional to $\delta(Q^2)$, we have

$$\Sigma_{gg}^{\delta(Q^2),(2)} = \delta\left(Q^2\right)\left(\Delta + \delta + N_c U\right) g_{gg} + \left(N_c - n_f\right) \times \tag{B.9}$$
$$\times \frac{N_c}{3}\delta\left(Q^2\right)\left[\left(\frac{m_H^4}{\hat{s}}\right) + \left(\frac{m_H^4}{\hat{t}}\right) + \left(\frac{m_H^4}{\hat{u}}\right) + m_H^2\right]$$

$$\Sigma_{gq}^{\delta(Q^2),(2)} = \delta\left(Q^2\right)\left(\Delta + N_c V_1 + C_F V_2 + V_3\right) g_{gq} \tag{B.10}$$
$$+ \left(N_c - C_F\right)\left[\frac{\hat{s}^2 + \hat{t}^2 + \hat{u}^2 - \hat{u} m_H^2}{-\hat{u}}\right] C_F \delta\left(Q^2\right)$$

$$\Sigma_{q\bar{q}}^{\delta(Q^2),(2)} = \delta\left(Q^2\right)\delta\left(Q^2\right)\left(\Delta + N_c W_1 + C_F W_2 + W_3\right) g_{q\bar{q}} \tag{B.11}$$
$$+ \delta\left(Q^2\right)\left[\frac{\hat{t}^2 + \hat{u}^2 + \hat{s}^2 - \hat{s} m_H^2}{\hat{s}}\right]\left(N_c - C_F\right) 2C_F^2$$

Notice that the quark-antiquark with different flavours $\Sigma_{qq'}^{\delta(Q^2),(2)}$ do not contain terms that are proportional to $\delta(Q^2)$. The functions U, V and W do not explicitly depend on the transverse momentum. In fact, they are functions of the Mandelstam variables, the Higgs mass m_H, the number of active flavours n_f, and the QCD factor N_c. The function δ, in addition to the aforementioned variables, depends on the renormalization scale μ_R while $\Delta = 11$ in $n_f = 5$ (which is the case we are interested in). Their expressions are explicitly given in Ref. [279]. For convenience, we write their expressions here

$$\delta = \frac{3\beta_0}{2}\left(\ln\frac{\mu_R^2}{-\hat{t}} + \ln\frac{\mu_R^2}{-\hat{u}}\right) + \left(\frac{67}{18}N_c - \frac{5}{9}n_f\right) \tag{B.12}$$

$$\begin{aligned} U = {} & \frac{1}{2}\ln^2\frac{-\hat{u}}{-\hat{t}} - \ln\frac{\hat{s}}{m_H^2}\ln\frac{-\hat{t}}{m_H^2} - \ln\frac{\hat{s}}{m_H^2}\ln\frac{-\hat{u}}{m_H^2} - \ln\frac{-\hat{t}}{m_H^2}\ln\frac{-\hat{u}}{m_H^2} + \zeta_2 \\ & + \ln^2\frac{m_H^2}{\hat{s}} + \ln^2\frac{m_H^2}{m_H^2 - \hat{t}} + \ln^2\frac{m_H^2}{m_H^2 - \hat{u}} \\ & + 2\,\text{Li}_2\left(\frac{\hat{s} - m_H^2}{\hat{s}}\right) + 2\,\text{Li}_2\left(\frac{m_H^2}{m_H^2 - \hat{t}}\right) + 2\,\text{Li}_2\left(\frac{m_H^2}{m_H^2 - \hat{u}}\right) \end{aligned} \tag{B.13}$$

$$V_1 = \frac{1}{2}\ln^2\frac{-\hat{u}}{-\hat{t}} + \frac{1}{2}\ln^2\frac{\hat{s}}{-\hat{u}} - \frac{1}{2}\ln^2\frac{\hat{s}}{-\hat{t}} + \ln\frac{\hat{s}}{m_H^2}\ln\frac{-\hat{t}}{m_H^2} - \ln\frac{\hat{s}}{m_H^2}\ln\frac{-\hat{u}}{m_H^2} \tag{B.14}$$
$$- \ln\frac{-\hat{t}}{m_H^2}\ln\frac{-\hat{u}}{m_H^2} + 2\,\mathrm{Li}_2\left(\frac{m_H^2}{m_H^2-\hat{u}}\right) + \ln^2\frac{m_H^2}{m_H^2-\hat{u}} + 3\zeta_2$$

$$V_2 = \ln^2\frac{m_H^2}{\hat{s}} + \ln^2\frac{m_H^2}{m_H^2-\hat{t}} - 2\ln\frac{\hat{s}}{m_H^2}\ln\frac{-\hat{t}}{m_H^2} - \frac{7}{2} \tag{B.15}$$
$$+ 2\,\mathrm{Li}_2\left(\frac{\hat{s}-m_H^2}{\hat{s}}\right) + 2\,\mathrm{Li}_2\left(\frac{m_H^2}{m_H^2-\hat{t}}\right) - 2\zeta_2$$

$$W_1 = \ln\frac{-\hat{u}}{m_H^2}\ln\frac{-\hat{t}}{m_H^2} - \ln\frac{\hat{s}}{m_H^2}\ln\frac{-\hat{u}}{m_H^2} - \ln\frac{\hat{s}}{m_H^2}\ln\frac{-\hat{t}}{m_H^2} \tag{B.16}$$
$$+ 2\,\mathrm{Li}_2\left(\frac{\hat{s}-m_H^2}{\hat{s}}\right) + \ln^2\frac{m_H^2}{\hat{s}} - \frac{1}{2}\ln^2\frac{-\hat{u}}{-\hat{t}} - 5\zeta_2$$

$$W_2 = \frac{3}{2}\left(\ln\frac{\hat{s}}{-\hat{t}} + \ln\frac{\hat{s}}{-\hat{u}}\right) + \ln^2\frac{-\hat{u}}{-\hat{t}} - 2\ln\frac{-\hat{u}}{m_H^2}\ln\frac{-\hat{t}}{m_H^2} + \ln^2\frac{m_H^2}{m_H^2-\hat{u}} \tag{B.17}$$
$$+ \ln^2\frac{m_H^2}{m_H^2-\hat{t}} + 2\,\mathrm{Li}_2\left(\frac{m_H^2}{m_H^2-\hat{u}}\right) + 2\,\mathrm{Li}_2\left(\frac{m_H^2}{m_H^2-\hat{t}}\right) - 7 + 6\zeta_2$$

$$W_3 = \frac{\beta_0}{2}\left(4\ln\frac{\mu_R^2}{\hat{s}} + \ln\frac{\mu_R^2}{-\hat{u}} + \ln\frac{\mu_R^2}{-\hat{t}}\right) + \left(\frac{67}{6}N_c - \frac{5}{3}n_f\right) \tag{B.18}$$

Let us now turn to the $\delta(Q^2)$-independent part. The LO order term does not contain terms proportional to $\delta(Q^2)$, therefore only next-to-leading terms contribute to the expression of Σ_{ab}. Its expression can be decomposed into a singular and regular contribution. Therefore, we can write Σ_{ab} as

$$\Sigma_{ab}\left(\frac{p_T^2}{Q_T^2}\right) = \Sigma_{ab}^{\mathrm{S}}\left(\frac{p_T^2}{Q_T^2}\right) + \Sigma_{ab}^{\mathrm{R}}\left(\frac{p_T^2}{Q_T^2}\right). \tag{B.19}$$

The singular contribution can in turn be organized in terms of the classes of plus distributions. We explicitly give here the expressions for all the various (sub) partonic channels

$$\Sigma_{gg}^{\mathrm{S}}\left(\frac{p_T^2}{Q_T^2}\right) = f_0\left(\mathcal{M}, m_H^2, Q^2, \frac{p_T^2}{Q_T^2}\right) + \sum_{\mathcal{P}}\left\{\left(\frac{\ln 1-z_a}{1-z_a}\right)_+\left[\frac{p_{gg}(z_a)}{-\hat{t}}g_{gg}(z_a x_a, x_b)\right.\right.$$
$$\left. + \left(\frac{z_a}{-\hat{t}}\right)f_1\left(\mathcal{M}, m_H^2, Q^2\right)\right] + \left(\frac{1}{1-z_a}\right)_+\left[\frac{1}{\hat{t}}p_{gg}(z_a)g_{gg}(z_a x_a, x_b)\ln\left(\frac{z_a\mu_F^2}{-\hat{t}}\right)\right.$$
$$\left.\left. - \left(\frac{z_a}{-\hat{t}}\right)\ln\left(\frac{z_a Q_T^2}{-\hat{t}}\right)f_1\left(\mathcal{M}, m_H^2, Q^2\right) + \left(\frac{z_a}{\hat{t}}\right)f_2\left(\mathcal{M}, m_H^2, Q^2\right)\right] \right. \tag{B.20}$$
$$\left.\times\left(\frac{2n_f}{\hat{t}}\right)g_{qg}(z_a x_a, x_b) - \left(\frac{2n_f}{\hat{t}}\right)p_{qg}(z_a)g_{qg}(z_a x_a, x_b)\ln\left(\frac{\mu_F^2}{Q^2}\right) + C_{qg}(z_a)\right\},$$

$$\Sigma_{gq}^{\mathrm{S}}\left(\frac{p_T^2}{Q_T^2}\right) = g_0\left(\mathcal{M}, m_H^2, Q^2, \frac{p_T^2}{Q_T^2}\right) + \left(\frac{z_a}{-\hat{t}}\right)\left(\left(\frac{\ln 1-z_a}{1-z_a}\right)_+ - \frac{\ln\left(Q_T^2 z_a/(-\hat{t})\right)}{(1-z_a)_+}\right)$$

$$
\begin{aligned}
&\times g_1\left(\mathcal{M}, Q^2\right) + \left(\frac{1}{-\hat{u}}\right)\left[-P_{qq}\left(z_b\right)\ln\frac{\mu_F^2 z_b}{(-\hat{u})} + p_{qq}\left(z_b\right)\left(\frac{\ln 1-z_b}{1-z_b}\right)_+ + C^{\epsilon}_{qq}\left(z_b\right)\right]\\
&\times g_{gq}\left(x_a, z_b x_b\right) + \left(\frac{1}{-\hat{t}}\right)\left[-P_{gg}\left(z_a\right)\ln\frac{\mu_F^2 z_a}{(-\hat{t})} + p_{gg}\left(z_a\right)\left(\frac{\ln 1-z_a}{1-z_a}\right)_+\right] g_{gq}\left(z_a x_a, z_b\right)\\
&+ \left(\frac{1}{-\hat{u}}\right)\left[-P_{gq}\left(z_b\right)\ln\frac{\mu_F^2}{Q^2} + C^{\epsilon}_{gq}\left(z_b\right)\right] g_{gg}\left(z_a, z_b x_b\right) + g_{q\bar{q}}\left(z_a x_a, z_b\right) \qquad \text{(B.21)}\\
&\times\left(\frac{1}{-\hat{t}}\right)\left[-P_{qg}\left(z_a\right)\ln\frac{\mu_F^2}{Q^2} + C^{\epsilon}_{qg}\left(z_a\right)\right] - \left(\frac{z_b}{-\hat{u}}\right)\left(\frac{1}{1-z_b}\right)_+ \frac{3}{2}C_F^2\left(\frac{\hat{t}^2+\hat{s}^2}{-\hat{u}}\right),
\end{aligned}
$$

$$
\begin{aligned}
\Sigma^{\mathrm{S}}_{q\bar{q}}\left(\frac{p_T^2}{Q_T^2}\right) &= h_0\left(\mathcal{M}, Q^2, \frac{p_T^2}{Q_T^2}\right) + \sum_{\mathcal{P}}\left\{\left(\frac{z_a}{-\hat{t}}\right)\left(\left(\frac{\ln 1-z_a}{1-z_a}\right)_+ - \frac{\ln\left(Q_T^2 z_a/(-\hat{t})\right)}{(1-z_a)_+}\right)\right.\\
&\times h_1\left(\mathcal{M}\right) + \left(\frac{1}{-\hat{t}}\right)\left[-P_{qq}\left(z_a\right)\ln\frac{\mu_F^2 z_a}{(-\hat{t})} + p_{qq}\left(z_a\right)\left(\frac{\ln 1-z_a}{1-z_a}\right)_+ + C^{\epsilon}_{qq}\left(z_a\right)\right] \qquad \text{(B.22)}\\
&\left.\times\, g_{q\bar{q}}\left(z_a x_a, z_b\right) + \left(\frac{1}{\hat{t}}\right)\left[P_{gq}\left(z_a\right)\ln\frac{\mu_F^2}{Q^2} - C^{\epsilon}_{gq}\left(z_a\right)\right] - \left(\frac{z_a}{-\hat{t}}\right)\left(\frac{\hat{t}^2+\hat{u}^2}{\hat{s}}\right)\left(\frac{\beta_0 C_F^2}{1-z_a}\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
\Sigma^{\mathrm{S}}_{q\bar{q}'}\left(\frac{p_T^2}{Q_T^2}\right) &= \frac{2C_F^2}{p_T^2}\ln\frac{p_T^2}{Q_T^2}\left[\frac{\left(\hat{s}-Q^2\right)^2+\left(\hat{u}+\hat{t}-2Q^2\right)^2}{\hat{s}}\right] + \sum_{\mathcal{P}}\left\{\left(\frac{1}{-\hat{t}}\right) - \frac{p_{gq}(z_a)}{(1-z_a)_+}\right.\\
&\left.\times\ln\frac{\mu_F^2}{Q^2} + \left(\frac{1}{-\hat{t}}\right)C^{\epsilon}_{gq}(z_a) g_{qg}(z_a x_a, x_b)\right\}, \qquad \text{(B.23)}
\end{aligned}
$$

where $\mathcal{P} = \left((\hat{t}, q) \leftrightarrow (\hat{u}, b)\right)$ and $\mathcal{M}$ denotes the set of Mandelstam variables. The dimensionless variable z_a is defined in terms of the Mandelstam variables as

$$
z_a = \frac{-\hat{t}}{Q^2-\hat{t}}, \qquad z_b = \frac{-\hat{u}}{Q^2-\hat{u}}. \qquad \text{(B.24)}
$$

The functions p_{ab} instead are defined in terms of the splitting functions, whose expressions are given in App. 1.D, through the following relations

$$
p_{gg}(z_a) = (1-z_a)P_{gg}(z_a) \qquad \text{(B.25)}
$$

$$
p_{qq}(z_a) = (1-z_a)P_{qq}(z_a) \qquad \text{(B.26)}
$$

with the other splitting terms being the same, i.e $p_{ab} = P_{ab}$ (for $a \neq b$). The functions $g_{ab}(z_a x_a, x_b)$ are defined in terms of the g_{ab} functions, that were introduced at the leading-order, through a shift in the argument in the Mandelstam variables. They are defined in terms of the Mandelstam variables as follows

$$
g_{gq}(z_a x_a, x_b) = C_F\left(\frac{\hat{t}^3 + z_a^2\hat{s}^2\hat{t}}{-z_a\hat{s}p_T^2}\right) \qquad \text{(B.27)}
$$

$$
g_{q\bar{q}}(z_a x_a, x_b) = 2C_F^2\left(\frac{\hat{t}^4 + z_a^2\hat{s}^2 p_T^4}{z_a\hat{s}\hat{t}^2}\right) \qquad \text{(B.28)}
$$

$$g_{gg}(z_a x_a, x_b) = N_c \left(\frac{m_H^8 + z_a^4 \hat{s}^4 + \hat{t}^4 + (z_a \hat{s}/p_T^2)^4}{(z_a^2/p_T^2)\hat{s}^2 \hat{t}} \right). \tag{B.29}$$

The functions $g_{ab}(x_a, z_b x_b)$ that arise from the permutation can be derived in an analogous way. Finally, we list below the functions that enter into our definition of $\Sigma_{ab}^{\rm S}$. The functions entering into $\Sigma_{gg}^{\rm S}$ are given by:

$$\bullet\, f_0\left(\mathcal{M}, m_H^2, Q^2, \frac{p_T^2}{Q_T^2}\right) = \frac{N_c^2}{p_T^2} \ln \frac{p_T^2}{Q_T^2} \left(f_0^{(1)}\left(\mathcal{M}, m_H^2\right) + f_0^{(2)}\left(\mathcal{M}, m_H^2, \frac{p_T^2}{Q_T^2}\right)\right) \tag{B.30}$$

with

$$f_0^{(1)}\left(\mathcal{M}, m_H^2\right) = \frac{2m_H^4}{\hat{s}\hat{u}\hat{t}} \left(\frac{\left(m_H^2 - \hat{t}\right)^4 + \left(m_H^2 - \hat{u}\right)^4 + \hat{u}^4 + \hat{t}^4}{\left(m_H^2 - \hat{u}\right)\left(m_H^2 - \hat{t}\right)} \right) \tag{B.31}$$

$$f_0^{(2)}\left(\mathcal{M}, m_H^2, \frac{p_T^2}{Q_T^2}\right) = \left(Q^2 + Q_T^2\right) \left(\frac{m_H^8 + \hat{s}^4 + Q^8 + (\hat{u}/z_b)^4 + \left(\hat{t}/z_a\right)^4}{\hat{s}^2 Q^2 Q_T^2} \right) \tag{B.32}$$

$$\bullet\, f_1\left(\mathcal{M}, m_H^2, Q^2\right) = \frac{N_c^2}{2} \left(f_1^{(1)}\left(\mathcal{M}, m_H^2, Q^2\right) + f_1^{(2)}\left(\mathcal{M}, m_H^2, Q^2\right)\right) \tag{B.33}$$

with

$$f_1^{(1)}\left(\mathcal{M}, m_H^2, Q^2\right) = \frac{\left(m_H^8 + \hat{s}^4 + Q^8 + \hat{u}^4 + \hat{t}^4\right)}{\hat{s}\hat{u}\hat{t}} \tag{B.34}$$

$$f_1^{(2)}\left(\mathcal{M}, m_H^2, Q^2\right) = z_a z_b \left(\frac{m_H^8 + \hat{s}^4 + Q^8 + (\hat{u}/z_b)^4 + \left(\hat{t}/z_a\right)^4}{\hat{s}\hat{u}\hat{t}} \right) \tag{B.35}$$

$$\bullet\, f_2\left(\mathcal{M}, m_H^2, Q^2\right) = N_c \frac{\beta_0}{2} \left(\frac{m_H^8 + \hat{s}^4 + z_a z_b \left((\hat{u}/z_b)^4 + \left(\hat{t}/z_a\right)^4 \right)}{\hat{s}\hat{u}\hat{t}} \right) \tag{B.36}$$

For the functions that enter into the definition of $\Sigma_{gq}^{\rm S}$, they are expressed as follows:

$$\bullet\, g_0\left(\mathcal{M}, m_H^2, Q^2, \frac{p_T^2}{Q_T^2}\right) = N_c \frac{C_F}{p_T^2} \ln \frac{p_T^2}{Q_T^2} \left(g_0^{(1)}\left(\mathcal{M}, m_H^2, Q^2\right) + g_0^{(2)}\left(\mathcal{M}, Q^2\right)\right) \tag{B.37}$$

with

$$g_0^{(1)}\left(\mathcal{M}, m_H^2, Q^2\right) = -\frac{2m_H^4}{\hat{s}\hat{u}} \left(\frac{\left(m_H^2 - \hat{t}\right)^2}{\left(m_H^2 - \hat{u}\right)} + \frac{\hat{t}^2}{\left(m_H^2 - \hat{u}\right)} \right) \tag{B.38}$$

$$g_0^{(2)}\left(\mathcal{M}, Q^2\right) = \left(Q^2 + Q_T^2\right) \left(\frac{-\hat{s}^3 \left(\hat{t} z_a\right) - \hat{s} \left(\hat{t}/z_a\right)^3 - Q^6 \left(\hat{u}/z_b\right) - Q^2 \left(\hat{u}/z_b\right)^3}{\hat{s}^2 Q^2 Q_T^2} \right) \tag{B.39}$$

$$\bullet\, g_1\left(\mathcal{M}, Q^2\right) = N_c C_F \left(g_1^{(1)}\left(\{\mathcal{M}\}, Q^2\right) + g_1^{(2)}\left(\mathcal{M}, Q^2\right)\right) \tag{B.40}$$

with

$$g_1^{(1)}\left(\mathcal{M},Q^2\right)=\frac{-\hat{s}^3\hat{t}-\hat{s}\hat{t}^3+Q^6\hat{t}+Q^2\hat{t}^3}{\hat{s}\hat{u}\hat{t}} \tag{B.41}$$

$$g_1^{(2)}\left(\mathcal{M},Q^2\right)=z_a z_b\left(\frac{-\hat{s}^3\left(\hat{t}/z_a\right)-\hat{s}\left(\hat{t}/z_a\right)^3-Q^6\left(\hat{u}/z_b\right)-Q^2\left(\hat{u}/z_b\right)^3}{\hat{s}\hat{u}\hat{t}}\right) \tag{B.42}$$

Finally, the functions that enter into the definition of $\Sigma^{\mathrm{S}}_{q\bar{q}}$ are expressed as follows:

$$\bullet\, h_0\left(\mathcal{M},Q^2,\frac{p_T^2}{Q_T^2}\right)=2\frac{C_F^2}{p_T^2}\ln\frac{p_T^2}{Q_T^2}\left(\frac{\left(\hat{s}-Q^2\right)^2+\left(\hat{u}+\hat{t}-2Q^2\right)^2}{\hat{s}}\right) \tag{B.43}$$

$$\bullet\, h_1\left(\mathcal{M},Q^2\right)=\left(2C_F-N_c\right)C_F^2\left(\frac{\hat{t}^2+\hat{u}^2+\left(\hat{t}/z_a\right)^2+\left(\hat{u}/z_b\right)^2}{\hat{s}}\right) \tag{B.44}$$

The regular terms Σ^{R}_{ab} embody contributions that are not logarithmically enhanced in the small-p_T limit plus $\mathcal{O}(\epsilon)$ corrections from the collinear singularity (ϵ comes from the dimensional regularization). For the various partonic channels, their expressions are exactly the same as those appearing in Ref. [279] and are given by

$$\begin{aligned}\Sigma^{\mathrm{R}}_{gg}=\frac{1}{\hat{s}^2p_T^2Q^2}\Big(&N_c^2\Big(A_0+A_{(1234)}+A_{(2341)}+A_{(3412)}+A_{(4123)}+A_{(1324)}\\&+A_{(2413)}+A_{(3241)}+A_\epsilon\Big)+n_fC_F\Big(B_{1(+-)}+B_{1(++)}\Big)+n_fN_c\Big(B_{2(+-)}+B_{2(++)}\Big)\end{aligned} \tag{B.45}$$

$$\begin{aligned}\Sigma^{\mathrm{R}}_{gq}=\frac{1}{\hat{s}^2p_T^2Q^2}\Big(&C_F^2\Big(C_{1(+-)}+C_{1(-+)}+C_{1(++)}+C_{1(--)}\Big)\\&+N_cC_F\Big(C_{2(+-)}+C_{2(-+)}+C_{2(++)}+C_{2(--)}+C_{2\epsilon}\Big)\Big)\end{aligned} \tag{B.46}$$

$$\Sigma^{\mathrm{R}}_{q\bar{q}}=\frac{1}{\hat{s}^2p_T^2Q^2}\Big(2C_F^3\Big(D_{1(+-)}+D_{1(++)}\Big)+2N_cC_F^2\Big(D_{2(+-)}+D_{2(++)}\Big)+E_Q\Big) \tag{B.47}$$

$$\Sigma^{\mathrm{R}}_{q\bar{q}'}=\frac{C_F^2E_2}{\hat{s}^2p_T^2Q^2},\qquad \Sigma^{\mathrm{R}}_{qq}=\frac{1}{\hat{s}^2p_T^2Q^2}\left(C_F^2E_2+\frac{C_F^2}{N_c}E_4\right) \tag{B.48}$$

where the the $A_{(ijkl)}$ terms come from the $(gg\to gg)$ colour-ordered amplitudes squared in which the projection of the spins are take for outgoing momenta while the B, C and D terms come from the $(q\bar{q}\to gg)$ amplitudes squared. The E_1, E_2 and E_3 fully depend on Born kinematics, they respectively arise from the s, t and u channel diagrams. All of these expressions are explicitly given in Appendix **A** of Ref. [279].

Let us recall that the expression of the partonic cross section we are dealing with, as expressed in Eq. (B.2), is doubly differential in transverse momentum and in rapidity. In order to derive an expression of a transverse momentum distribution in Mellin space, we need to integrate the differential distribution over the rapidity y_H. It is then clear that based on the way we organized the cross section in Eq. (B.2), there are two main classes of integration that we have to deal with, namely the terms that are proportional to $\delta(Q^2)$ and the terms that δ-independent. In the subsequent sections, we describe how to deal with the δ-term in the integration and the plus prescriptions.

Integration variable and $\delta(Q^2)$ terms

The integral we are interested to solve is the following

$$\frac{d\hat{\sigma}_{ab}}{dp_T^2}\left(\frac{p_T^2}{Q_T^2}\right) = \int_{y_{\min}}^{y_{\max}} dy_H \, \frac{d\hat{\sigma}_{ab}}{dp_T^2 dy_H}\left(\frac{p_T^2}{Q_T^2}\right), \tag{B.49}$$

where $(y_{\min}, y_{\max})$ defines the rapidity range and usually taken to be $y_{\max} = -y_{\min}$. In order to deal with the $\delta(Q^2)$ terms, one option would be to keep y_H as the integration variable and change the variable that is the argument of the delta functions. As suggested in [280], a more convenient variable for integration is the momentum Q^2. The relation between y_H and Q is given by

$$\sinh y_H = \pm \frac{\sqrt{\left(\hat{s} + m_H^2 - Q^2\right)^2 - 4\hat{s}\left(p_T^2 + m_H^2\right)^2}}{2\sqrt{\hat{s}}\left(p_T^2 + m_H^2\right)}, \tag{B.50}$$

and the Jacobian corresponding to this transformation writes as

$$\mathfrak{J}(Q^2 \leftarrow y_H) = \left(-2m_T\sqrt{\hat{s}}\sinh\left(y_H\right)\right)^{-1} = \frac{1}{\sqrt{\left(\hat{s} + m_H^2 - Q^2\right)^2 - 4\hat{s}\left(p_T^2 + m_H^2\right)}}. \tag{B.51}$$

For numerical purposes, it is convenient can re-scale the integration boundaries to be between zero and one. This can be done by introducing a new variable $q = Q^2/Q^2_{\max}$ where $Q^2_{\max}$ corresponds to the momentum of the partons at $y_H = 0$. Solving Eq. (B.50) for Q^2 and subsequently setting $y_H = 0$ leads to the following

$$Q^2_{\max} = m_H^2 + \hat{s} - 2\sqrt{\hat{s}\left(p_T^2 + m_H^2\right)}. \tag{B.52}$$

Using q as the integration variable introduces the following Jacobian

$$\mathfrak{J}(q \leftarrow y_H) = \frac{Q^2_{\max}}{\sqrt{\left(\hat{s} + m_H^2 - Q^2\right)^2 - 4\hat{s}\left(p_T^2 + m_H^2\right)}}. \tag{B.53}$$

Using the above changes of variables, the transverse momentum distribution can now be written in the following form

$$\frac{d\hat{\sigma}_{ab}}{dp_T^2}\left(\frac{p_T^2}{Q_T^2}\right) = \int_0^1 dq \, \mathfrak{J}(q \leftarrow y_H)\left[\frac{d\hat{\sigma}_{ab}}{dp_T^2 dy_H}(\hat{u}, \hat{t}) + \frac{d\hat{\sigma}_{ab}}{dp_T^2 dy_H}(\hat{t}, \hat{u})\right] \tag{B.54}$$

where the permutation of $\hat{u}$ and $\hat{t}$ inside the brackets is due to the fact that there are tow possible choices for the rapidity as illustrated in Eq. (B.50). The above transformation also implies that the Mandelstam variables $\hat{u}$ and $\hat{t}$ also need to be expressed in terms of the new integration variable. Doing so yields to the following expression

$$\hat{t}_{\pm} = \frac{1}{2}\left[Q^2 + m_H^2 - \hat{s} \pm \sqrt{\left(\hat{s} + m_H^2 - Q^2\right)^2 - 4\hat{s}\left(p_T^2 + m_H^2\right)}\right] \tag{B.55}$$

$$\hat{u}_{\pm} = \frac{1}{2}\left[Q^2 + m_H^2 - \hat{s} \mp \sqrt{\left(\hat{s} + m_H^2 - Q^2\right)^2 - 4\hat{s}\left(p_T^2 + m_H^2\right)}\right] \tag{B.56}$$

where again the $(\pm)$ reflects the fact that there are two possible choices for the rapidity. Putting aside the plus distributions, the integration of all terms proportional to $\delta(Q^2)$, namely the $\Sigma_{ab}^{\delta(Q^2)}$-functions is now straightforward. As a matter of fact, there is no need at all to perform the integration over y_H since fully transverse momentum distribution can be computed by simply setting $q = Q^2 = 0$,

$$\frac{d\hat{\sigma}_{ab}^{\delta(Q^2)}}{dp_T^2}\left(\frac{p_T^2}{Q_T^2}\right) = \left\{\mathfrak{J}(q \leftarrow y_H)\left[\frac{d\hat{\sigma}_{ab}^{\delta(Q^2)}}{dp_T^2 dy_H}(\hat{u},\hat{t}) + \frac{d\hat{\sigma}_{ab}^{\delta(Q^2)}}{dp_T^2 dy_H}(\hat{t},\hat{u})\right]\right\}_{Q=0}. \tag{B.57}$$

Change of variables in Plus distributions

We have seen how a convenient choice of variable simplifies the integration of the terms that are proportional to $\delta(Q^2)$ but we have not commented yet on how this impact the plus distributions. Large logarithmic terms appear in the form of plus distribution

$$\left(f(z_a)\right)_+ = \left(\frac{\ln^k(1-z_a)}{1-z_a}\right)_+ \tag{B.58}$$

where $k = 0, 1$ for the next-to-leading contributions we are considering. The above plus distributions can be expressed in terms of the integration variable q

$$\frac{z_a}{-\hat{t}}\left(\frac{1}{1-z_a}\right)_+ = \frac{1}{Q^2_{\max}}\left\{\left[\frac{1}{q}\right]_+ + \delta(q)\ln\frac{Q^2_{\max}}{-\hat{t}}\right\} \tag{B.59}$$

$$\frac{z_a}{-\hat{t}}\left(\frac{\ln(1-z_a)}{1-z_a}\right)_+ = \frac{1}{Q^2_{\max}}\left\{\left[\frac{\ln(q)}{q}\right]_+ + \ln\frac{Q^2_{\max} z_a}{-\hat{t}}\left[\frac{1}{q}\right]_+ + \frac{\delta(q)}{2}\ln^2\frac{Q^2_{\max}}{-\hat{t}}\right\}, \tag{B.60}$$

where the square brackets indicate that the pole is at $q = 0$ as opposed to the pole being at $z_a = 1$. Analogous expressions can be derive for $\left(f(z_b)\right)_+$ by just replacing $\hat{t}$ with $\hat{u}$ in Eqs. (B.59, B.60). Therefore, The integration of the terms in which plus distributions appear is now straightforward as they are given by

$$\int_0^1 dq\ \left[f(q)\right]_+ g(q) = \int_0^1 dq\, f(q)\left(g(q) - g(1)\right) \tag{B.61}$$

where g is just an analytic function multiplying the plus distributions.

Mellin transform of the full FO predictions

Once the integration over the rapidity has been performed, one can compute the Mellin transform of the ensuing expression. Whilst the exact analytical computation of a few sets of terms appearing in the cross section is feasible both for the rapidity integration and

the Mellin transform, most of the terms have to be integrated numerically. As usual, the Mellin transform of the cross section is computed via

$$\frac{d\hat{\sigma}_{ab}}{dp_T^2}\left(N, \frac{p_T^2}{Q_T^2}\right) = \int_0^1 dx\, x^{N-1} \frac{d\hat{\sigma}_{ab}}{dp_T^2}\left(x, \frac{p_T^2}{Q_T^2}\right) \tag{B.62}$$

where the Mellin transform is taken w.r.t the partonic variable $x = Q^2/\hat{s}$. The procedure described above is what is implemented in HPT-MON [254]. This allowed us to perform checks on the approximate Higgs NNLO distribution in Mellin space.

Part III

MACHINE LEARNING & PDFS

4

GENERATING NEW FEATURES WITH DEEP LEARNING MODELS

In recent years, the field of deep generative modellings have received considerable attention due to the explosion of numerous machine learning techniques that provide various degree of generation capability. If trained successfully, deep generative models can be used to estimate the likelihood of a given data and to create new samples that follow the same underlying probability distribution. One of the most popular classes of deep generative models is known as *Generative Adversarial Networks* (or in short GANs). Applications of deep generative models, especially GANs, have reached a wide range of area such as computer visions [281–286], language and speech synthesis [287, 288], cryptography [289, 290], stenography [291–293], et cetera.

Owing to the complexity of the simulations (required for the event construction and data analysis), the field of high energy physics has not been spared. As a matter of fact, some of the shortcomings arising in standard modelling have been overcome by deep generative models, and to the present day, variants of deep generative models have been applied to various areas of LHC searches such as parton showers, detector simulation, and searches for physics beyond the standard model [66, 294–333].

The purpose of this chapter is not to present a detailed review of deep generative models (to which a vast amount of literatures already exists), nor an extensive comparative analyses of the different variants, but rather to illustrate broadly the concept of deep generative modelings and their applications. Particular attention will be given to generative adversarial neural networks which Chapter 5 will rely on.

4.1 Basics of Deep Learning

Neural network, also referred to as *Artificial Neural Network* (ANNs), generally assumes feed-forward structure in which the neurons are stacked into layers, with each of the neurons receiving inputs from the previous layers. One example of a widely used architecture is known as *Multilayer Perceptron* that is comprised of an input layer, one single hidden layer, and an output layer, with pairs of successive layer being fully connected (as illustrated in Fig. 25). If we denote by $\vec{X}$ the vector which contains as elements the values of the perceptrons in the input layer, then the hidden and output layer compute the following

$$\vec{X}_{\mathcal{H}} = f_1(W_1 \vec{X}_{\mathcal{P}} + \vec{B}_1) \tag{4.1.1}$$

$$\vec{X}_{\mathcal{O}} = f_2(W_2 \vec{X}_{\mathcal{H}} + \vec{B}_2) \tag{4.1.2}$$

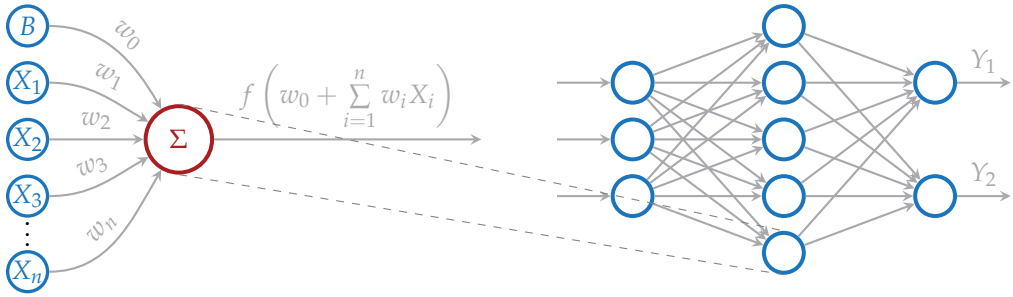


Figure 25 An example of a multilayer perceptron model. The right network represents how the input data are fed to the perceptrons, traverse through one single hidden layer and results into a prediction $\vec{Y}$. The left network describes how a perceptron is *activated* with an input vector $\vec{X}$, weights $\vec{w}$, bias w_0, and an activation function f.

where $W_{1,2}$ and $\vec{B}_{1,2}$ represent the weight and bias vectors of the multilayer perceptron layer respectively, and $f_{1,2}$ represent some activation functions. If the neural network contains more than one single hidden layer, then they are called *Deep Neural Networks* (DNNs). Whilst the *universal approximation theorem* [334] states that a Neural Network with one single hidden layer is capable of approximating any Borel-measurable function, in practice, deep neural networks can substantially improve the training. Indeed, due to their modularity, deep neural networks are more capable of learning complex features by *hierarchically* composing (simple) non-linear functions.

In order to train (deep) neural networks, two key elements are required: initialization and optimization. In turn, the latter is composed of several components which consists in computing the loss (or cost) function and updating each parameter of the networks using the gradients according to an optimization algorithm. For the sake of pedagogical clarity, let us briefly expand on these elements.

Initialization

The initialization procedure is one of the most crucial steps when training neural networks as it can significantly impact the model's performance. A properly (or conversely poorly) initialized neural network can lead to a substantially good (or conversely bad) performance. In order to illustrate this, consider a deep neural network where all the bias are zero and all the activation function are linear, i.e. $\vec{X}_n = W_n W_{n-1} \cdots W_1 \vec{X}_{\mathcal{P}}$. It can be easily argued that a *too-large* initialization of the weights leads to *exploding gradients*. As an illustration, consider the case where every weight is initialized with the identity matrix times some positive constant $\varepsilon > 1$ that makes the values on the diagonal larger than unity, i.e. $W_1 = W_2 \cdots = W_n = \varepsilon \mathbb{I}$. The output layer is then given by $\vec{X}_{\mathcal{O}} = (\varepsilon \mathbb{I})^n \vec{X}_{\mathcal{P}}$ suggesting that $\vec{X}_\ell$ increases exponentially with ℓ where ℓ denotes the ℓ-th layer. When propagated backward, the the gradients of the loss function w.r.t the parameters of the deep neural network are too big, leading to an *exploding gradient problem*. Conversely, a *too-small* initialization of the deep neural network yields *vanishing gradients*. Following the same line of thought as before, but instead considering $\tilde{\varepsilon} > 1$ such that the values of the diagonal

weights are smaller than unity. In this scenario, the output layer is given by $\vec{X}_{\mathcal{O}} = (\tilde{\varepsilon}\mathbb{I})^n \vec{X}_{\mathcal{P}}$ suggesting that $\vec{X}_\ell$ decreases exponentially with ℓ. As a result, the gradients of the (deep) neural network are too small leading to a *vanishing gradient* problem.

In order to prevent the gradients of the loss functions from vanishing or exploding, it is suitable to keep the mean of the activation (here also noted $\vec{X}_\ell$ due to the linearity of the activation function) around zero and with variance staying the same across all the layers. In a generic case where no assumption about the (deep) neural network is made, this can be translated into the following constraints:

$$\mathbb{E}\left[W_n \vec{X}_{n-1} + \vec{B}_n\right] = \mathbb{E}\left[W_{n-1}\vec{X}_{n-2} + \vec{B}_{n-1}\right] \tag{4.1.3}$$

$$\mathrm{Var}\left(W_n \vec{X}_{n-1} + \vec{B}_n\right) = \mathrm{Var}\left(W_{n-1}\vec{X}_{n-2} + \vec{B}_{n-1}\right). \tag{4.1.4}$$

This suggests that the weights should be sampled from a carefully chosen probability distributions. In practice, the most common used distributions are uniform $(W_\ell \sim \mathcal{U}(-x, x))$ and normal $(W_\ell \sim \mathcal{N}(0, \sigma))$. In particular, it has been shown that variants of the *Xavier initializer* generally lead to better performance [335].

Optimization

Once the weights of the (deep) neural network are properly initialized, samples from the training set are fed into the neural networks and forward-propagated through all the layers, resulting in some predictions $\vec{Y} = \mathrm{DNN}(\vec{X}, \theta)$. The parameters of the neural networks are then fine-tuned by comparing the predictions to the ground truths $\vec{Y}_p$. This is usually done by computing a differentiable loss function $\mathcal{L}\left(\vec{Y}_{\mathcal{O}}, \vec{Y}_p\right)$. The choice of loss function depends on the problem at hand, for simple regression problems for instance, one may use the *mean squared error* $\mathcal{L}_{\mathrm{MSE}}\left(\vec{Y}_{\mathcal{O}}, \vec{Y}_{\mathcal{P}}\right) = 1/N \sum_{n=1}^{N} |Y_{\mathcal{O},i} - Y_{\mathcal{P},i}|^2$.

In the case where the training set is large, which is often the case for deep learning studies, the loss function $\mathcal{L}$ is most appropriately optimized using *Mini-Batch Stochastic Gradient Descent* (SGD) or its variants in which only a small batch m of the training samples is considered at each iteration. The gradient of the objective loss functions w.r.t. the neural network's parameters θ are then computed using a *backpropagation algorithm*. The backpropagation enables the gradients to be propagated backwards from the output to the input layers. Such computations involve repeatedly applying the chain rule of partial derivatives. For each step of gradient descent, the parameters θ is updated according to the following transformation

$$\theta \longleftarrow \theta - \eta \nabla_\theta \mathcal{L}_m \left(g\left(\vec{X}, \theta\right), \vec{Y}\right), \tag{4.1.5}$$

where η denotes the learning rate and $\nabla_\theta \mathcal{L}_m$ represents the gradient w.r.t. the parameter θ over the mini-batch m. The choice of a learning rate is vital to the stability and convergence of the (deep) neural network. For instance, due to the small changes brought to the weights during the update, smaller learning rate may require a large number of epochs in order to reach convergence. Whilst large value of learning rate could lead to a fewer training epochs, it can cause the neural network to converge quickly to a suboptimal solution. The issue associated to the choice of learning rate can be overcome by resorting to an adaptive

optimizer where the learning rate is updated throughout the training. Such a dynamically adjusted learning rate is implemented for example in the *Adaptive Momentum Stochastic Gradient Descent*, also known as *Adam*. The fine-tune of the neural network's parameters in the Adam optimization is performed using the following expression

$$\theta \longleftarrow \theta - \gamma \frac{\hat{m}_1}{\sqrt{\hat{m}_2 + \epsilon}}, \tag{4.1.6}$$

where $\gamma \equiv 10^{-3}$ is the magnitude of the stepsize that is progressively adjusted during the training by the first and second moments $\hat{m}_1$ and $\hat{m}_2$. At the beginning of the training, the moments $\hat{m}_1$ and $\hat{m}_2$ are set to zero, and therefore in order for the gradients to not blow up, a small parameter $\epsilon = 10^{-8}$ is introduced. Throughout the training, both moments are updated according to the gradient of the loss function. Mathematically, this reads as follows

$$\hat{m}_1 \longleftarrow (1-\beta_1)\, \nabla_\theta \mathcal{L}_m \left(g\left(\vec{X}, \theta\right), \vec{Y} \right) + \beta_1 \hat{m}_1 \tag{4.1.7}$$

$$\hat{m}_2 \longleftarrow (1-\beta_2)\, \nabla_\theta \mathcal{L}_m \left(g\left(\vec{X}, \theta\right), \vec{Y} \right) + \beta_2 \hat{m}_2. \tag{4.1.8}$$

The constants β_1 and β_2 are called exponential decay rates for the moment estimates. Their default values, as proposed by Ref. [336], are given by $\beta_1 = 0.9$ and $\beta_2 = 0.999$. At every update, before entering in Eq. (4.1.6), the moments are normalized according to

$$\hat{m}_1 = \frac{\hat{m}_1}{1-\beta_1^t}, \qquad \hat{m}_2 = \frac{\hat{m}_2}{1-\beta_2^t}, \tag{4.1.9}$$

where t represents the timestep. Eq. (4.1.9) is known as *bias correction* and prevents the stepsize from becoming too large. Two optimization methods that are closely related to Adam are *RMSprop* [337] and *AdaGrad* [338].

Overfitting

Due to the fact that deep neural networks are typically overparametrized, they are more prone to *overfitting*. However, there exists *regularization procedures* that can prevent a model from learning noises. In short, regularization provides constraints to the space of parameters accessible by the deep neural networks. From the efficiency point of view, it is actually preferable to have a large model that may require regularization during the training than a model with too little generalization capacity to learn the problem at hand. In most deep learning applications, the former approach has indeed proved to lead to faster optimization resulting in an overall better performance.

- $\mathbb{L}_2$ **Regularization.** Also known as *Ridge Regression*, the $\mathbb{L}_2$ regularization is one of the most standard ways to avoid overfitting. This regularization relies on the assumption that the weight of the deep neural networks should not deviate far from zero. Such a constraint in enforced by adding a regularization term (based on $\mathbb{L}_2$ norm) to the loss function. That is,

$$\mathcal{L} \longleftarrow \mathcal{L} + \frac{\lambda}{2} ||W||^2 \tag{4.1.10}$$

where λ is a constant that penalizes the weights. This parameter has to be chosen carefully because choosing λ to be large, for instance, will lead to an *underfitting*. In the particular case where $\lambda = 0$, we get back the original definition of the loss function.

- **Batch normalization.** What can render the training of deep neural networks difficult is a phenomenon known as *internal covariate shifts* where the distribution of each layers input (activations) changes during the training. In order to improve the training stability, one must seek to reduce the internal covariate shift through a procedure known as *batch normalization*. This can be done by normalizing the output batches. For the sake of clarity, let us focus only at a particular activation $X_{\ell,k}$ in a hidden layer ℓ (where here we over-use the notation $X_{\ell,k}$ to also represent a given activation) and Consider a mini-batch $\mathcal{B}_m$ of size m. We then have m values of the activation $X_{\ell,k}$ in the mini-batch $\mathcal{B}_m = \{X_{\ell,k}^{(1)}, \cdots, X_{\ell,k}^{(m)}\}$. The *batch normalization* algorithms [339] normalizes the activation as follows

$$\mu_{\mathcal{B}} \longleftarrow \frac{1}{m}\sum_{i=1}^{m} X_{\ell,k}^{(i)}, \qquad \sigma_{\mathcal{B}}^2 \longleftarrow \frac{1}{m}\sum_{i=1}^{m}\left(X_{\ell,k}^{(i)} - \mu_{\mathcal{B}}\right), \tag{4.1.11}$$

$$\hat{X}_{\ell,k}^{(i)} \longleftarrow \frac{X_{\ell,k}^{(i)} - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}, \qquad X_{\ell+1,k}^{(i)} \longleftarrow \gamma \hat{X}_{\ell,k}^{(i)} + \beta \tag{4.1.12}$$

where ϵ is a small parameter and γ, β are learnable parameters. Several studies have shown that the batch normalization approach can circumvent the training instability due to poor initialization [339, 340].

4.2 Supervised vs. Unsupervised learning

Within the field of machine learning and artificial intelligence, there are two main tasks, namely the *supervised* and *unsupervised* learning. In the following section, we briefly describe the aims of each task and the differences between them.

Supervised learning:

It belongs to the class of machine learning that uses *labelled data sets*. As a matter of fact, this is mainly what sets supervised learning apart from unsupervised learning. The purpose of labelling the data sets is to train in a supervised manner an algorithm into accurately classifying data or predicting outcomes. This indeed allows the algorithm (model) to measure its accuracy and learn over time. Problems that require supervised learning can be categorized into *classification* and *regression*. Classification problems arise when one wants to map input to output labels, while regression problems arise when one wants to map input into continuous outputs. In both classification and regression problems, the main objective is to find specific structure in the input data that allows one to effectively produce correct output data. To dates, the most common algorithms on supervised learning include logistic regression, support vector machines, artificial neural networks (ANNs), and random forests.

Figure 26 Given a complicated distribution $\mathcal{X}$, a deep generative model G_θ is trained to map samples from a simple distribution $\mathcal{Z}$ defined over $\mathbb{R}^m$ to a more complicated distribution $G_\theta(\mathcal{Z})$ that is similar to $\mathcal{X}$ defined over $\mathbb{R}^n$. If the deep generative model is inverible as is the case for *Invertible Neural Networks* (INNs) such as *Normalizing Flows*, an inverse mapping is also possible. Notice that this is not the case for generative adversarial models.

Unsupervised learning:

It belongs to the class of machine learning algorithms that discovers hidden patterns in a data set without the need of a supervision. As opposed to supervised learning, unsupervised learning methods are designed to problems where the data sets are not labelled. In other words, the approach of unsupervised learning is suitable for problems that require the algorithm to identify and extract the inherent feature of a given input data set. The most common tasks within unsupervised learning include: clustering, representation learning, density estimation, and (some classes of) generative modelling. Some common unsupervised learning algorithms include $\mathcal{K}$-means clustering, kernel density estimation, (variational) autoencoders, and normalizing flows. Some of these algorithms will be described in details or partially in the subsequent sections.

Semi-supervised learning:

Supervised and unsupervised learning can be combined together into a single training approach known as *semi-supervised* learning. Such an approach is suitable for problems where the target labels for all the samples in the data set are not available. One of the most well-known examples of semi-supervised learning is the *Generative Adversarial Training*. Details of such a training is amply detailed in the next Chapter.

4.3 Generative Models

In the following section, we give a pedagogical introduction to a particular class of unsupervised learning known as *Generative Modelling*. The problem deep generative models are concerned with is the following: given a target distribution that follows a probability distribution p_R, we would like to generate new samples following some probability distribution p_θ such that p_θ is very close to p_R, i.e. the Kullback-Leibler divergence

$$\mathrm{KL}(p_R||p_\theta) = \int_x dx\, p_R(x) \ln\left(\frac{p_R(x)}{p_\theta(\theta)}\right) \tag{4.3.1}$$

is minimal. Thus, the goal is to construct a function $G_\theta : z \longrightarrow x_g$ that maps a latent variable z with fixed probability distribution p_z to a sample x_g with probability p_θ. The construction of the function G_θ from the first principle is infeasible for most of the data sets of interest. The fundamental design concept of deep generative models is to define G_θ in terms of a deep neural network function. By optimizing the parameters θ (which now defines the parameters of the deep neural networks), one can modify the probability distribution p_θ to resemble p_R. There exists various machine learning techniques to define G_θ, examples of the most used techniques are briefly described below. For reasons that shall be explained in the next sections, the technique that we adopt in the present thesis is known as *generative adversarial model*, which is amply detailed in Sec. (4.4).

Variational Autoencoders (VAEs)

Variational Autoencoders [341] are one of the widely used likelihood-based generative models whose basic idea consist in setting two deep neural networks known as *encoder* and *decoder* and to learn the best encoding-decoding scheme using iterative optimization process. Roughly speaking, the encoder takes an input data and compresses it down to a smaller representation with fewer dimensions. This consists in either selecting only a subset of the initial features or by combining a subset of them into a reduced number of features. The decoder then takes the low representation of the input data and tries to reconstruct new data sample that is similar to the ground truth. In more practical terms, at each iteration, the encoder is fed with some data, then the encoded-decoded output is compared to the input data and the error is backpropagated through the architecture to update the weights and biases. In order to avoid underfitting or overfitting, VAEs implements regularization procedures that ensures that the latent space has the relevant properties to enable generative process. The loss function that is minimized in VAEs is given by the following expression

$$\mathcal{L}_{\text{VAE}} = \underset{q_\phi(z|z)}{\mathbb{E}} \left[\ln \left(p_\theta(x|z)\right)\right] - \text{KL}\left[q_\phi(z|z)|p_\theta(z)\right], \tag{4.3.2}$$

where x as usual represents the true data distribution, ϕ and θ represents the parameterized distributions for the VAE probabilistic encoder-decoder networks. The first term in the above equation represents the reconstruction likelihood while the second term represents the non-negative divergence between the true and approximate posterior. The latter ensures that the learned distribution is similar to the prior distribution. One of the main drawbacks of the VAEs is that, despite the fact that they can learn feature representation, they have intractable marginal likelihoods.

Normalizing flows

Normalizing flows belong to the class of generative models that are *invertible*. That is, the function G_θ used to model the posterior probability distribution is a *bijective* function. They are constructed as a series of invertible functions which map complex data to a simple Gaussian distribution, or vice-versa. A stark difference between VAEs and normalizing flow models is that the latter converge faster than the former. However, samples generated using normalizing flow are not that rich in features as other models.

4.4 Generative Neural Networks (GANs)

In the following section, we describe how distributions can be modelled using generative adversarial neural networks by implicitly learning the representation of the probability density. We first present a methodological review by following the original description [342] (often referred to as *Vanilla-GAN*). In Sec. (4.4.2), we briefly expand on the challenges in training generative adversarial models, and provide in Sec. (4.4.3) (in addition to the ones mentioned in Sec. (4.1)) regularization procedures and GAN architecture variants to overcome those challenges in order to improve the training stability.

4.4.1 Theoretical framework

As mentioned in Sec. (4.3), the main idea that underpins all deep generative models is the construction of a function G_θ that generates new samples following a particular probability distribution. In the case of generative adversarial models, the function G_θ is a deep neural network function whose training over time is controlled by differential deep neural networks. The two neural networks are commonly known as *Generator* and *Discriminator*. The generator G_θ is a differentiable function represented in our case by a multilayer perceptron whose job is to deterministically generate samples x_g from a latent variable z. The discriminator D_ϕ is also a multilayer perceptron whose main job is to distinguish samples from real and synthetic PDF replicas.

The generator G_θ and discriminator D_ϕ are then trained in an adversarial way: G_θ tries to capture the probability distribution of a given data set and generates new samples following the same probability distribution (therefore minimizes the objective such that $p_R = p_\theta$), and D_ϕ tries to distinguish whether the sample came from an input data set rather than from the generator (therefore maximizes the objective $p_R \neq p_\theta$). This adversarial training allows both models to improve to the point where the generator is able to create synthetic data such that the discriminator can no longer distinguish between synthetic and original sets. This is a sort of min-max game where the GAN objective function is given by following expression [342]:

$$\min_\theta \max_\phi V\left(G_\theta, D_\phi\right) = \mathop{\mathbb{E}}_{x\sim p_R}\left[\ln D_\phi(x)\right] + \mathop{\mathbb{E}}_{z\sim p_z}\left[\ln\left(1 - D_\phi\left(G_\theta(z)\right)\right)\right]. \tag{4.4.1}$$

For a fixed generator G_θ, the discriminator performs a binary classification by maximizing V w.r.t. ϕ while assigning the probability *one* to samples from the prior PDF replicas $(x\sim p_R)$, and assigning probability *zero* to the synthetic samples $(x\sim p_\theta)$. Therefore, the discriminator is at its optimal efficiency when:

$$D^\star(x) = \frac{p_R(x)}{p_R(x) + p_\theta(x)}. \tag{4.4.2}$$

If we now assume that the discriminator is at its best $(D(x) = D^\star(x))$, then the objective function that the generator is trying to minimize can be expressed in terms of the Jensen-Shannon divergence $\mathrm{JSD}(p_R, p_\theta)$ as follows:

$$V\left(G_\theta, D_\phi^*\right) = 2\,\mathrm{JSD}(p_R, p_\theta) - 2\ln 2, \tag{4.4.3}$$

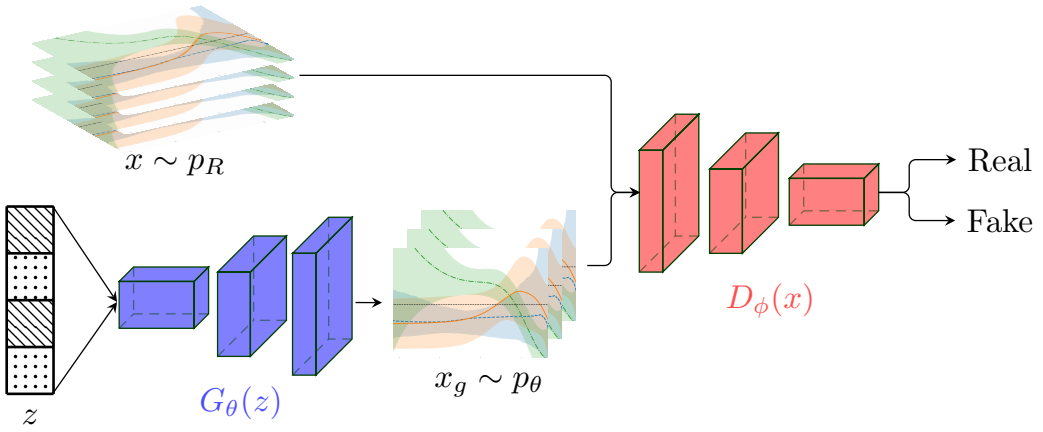


Figure 27 Diagrammatic structure of a generative adversarial neural network. Both neural networks are represented in terms of deep convolutional neural networks (DCNNs). The discriminator is expressed by D_ϕ while the generator is expressed by G_θ. Both the true input data and the generated samples are fed into the discriminator which evaluates whether or not the generated samples are similar to the true data.

where the Jensen-Shannon divergence satisfies all the properties of the Kullback-Leibler divergence and has the additional constraint that $\text{JSD}(p_R, p_\theta) = \text{JSD}(p_\theta, p_R)$. We then see from Eqs. (4.4.1-4.4.3) that the best objective value we can achieve with optimal generator and discriminator is $(-2\ln 2)$.

The generator and discriminator are generally trained in the following way: the generator generates a batch of synthetic samples that along with the samples from the input data sets are provided to the discriminator [1] (see Algorithm 1). Algorithm 1 is a typical formulation of a generative-based adversarial strategy. The adversarial training of both the discriminator and generator neural networks are described in Fig. 27. The training goes as follows. On the one hand, the generator takes as input a vector of random noises and outputs fake data which then get fed into the discriminator. During the training of the Generator, it is important that the discriminator is freezed, otherwise equilibrium might never be reached. On the other hand, the discriminator model takes as input both the true and fake data. The model then outputs predictions as to whether the samples are real with labels *ones* or fake with labels *zeros*. However, as we will discuss in the next section, working with such an implementation in practice is very challenging and often leads to poor results.

4.4.2 Challenges in training GANs

Training generative adversarial models can be very challenging due to the lack of stopping criteria that estimates when exactly GANs have finished training. Therefore, there is no guarantee that the equilibrium is reached. In addition, GANs have common failure

1 It is important that during the training of the generator, the discriminator is not trained. As it will be explained later, having an over-optimized discriminator leads to instabilities.

Algorithm 1: Algorithm describing the stochastic gradient descent training of a standard generative adversarial networks [342].

for *epochs* $1, \cdots, N$ **do**
  for *Discriminator steps* $1, \cdots, k$ **do**
    - Sample minibatch of size m from the real sample: $\{x_r^{(1)}, \cdots, x_r^{(m)}\}$
    - Sample minibatch of size m from the latent space: $\{z^{(1)}, \cdots, z^{(m)}\}$
    - Perform gradient **ascent** on discriminator:

$$\nabla_\phi V\left(G_\theta, D_\phi\right) = \frac{1}{m}\nabla_\phi \sum_{i=1}^{m} \ln D_\phi\left(x_r^{(i)}\right) + \frac{1}{m}\nabla_\phi \sum_{i=1}^{m} \ln\left(1 - D_\phi\left(G_\theta\left(z^{(i)}\right)\right)\right)$$

  end
  for *Generator steps* $1, \cdots, \ell$ **do**
    - Sample minibatch of size m from the latent space: $\{z^{(1)}, \cdots, z^{(m)}\}$
    - Perform gradient **descent** on generator:

$$\nabla_\theta V\left(G_\theta, D_\phi\right) = \frac{1}{m}\nabla_\theta \sum_{i=1}^{m} \ln\left(1 - D_\phi\left(G_\theta\left(z^{(i)}\right)\right)\right)$$

  end
end

modes due to inappropriate choices of network architecture, loss function or optimization algorithm. Several solutions have been proposed to address these issues which is still a topic subject to active research. For a review, we refer the reader to Refs. [343–345]. In this section, we briefly describe these challenges before providing some solutions in the next section. The most encountered limitations of GANs in our case are: non-convergence, vanishing gradients, and mode collapse.

It is often the case that during optimization, the losses of generator and discriminator continue to oscillate without converging to a clear stopping value. Although the existence of an equilibrium (*Nash Equilibrium*) has been proved in the original GAN paper [342], there is no guarantee that such an equilibrium will be reached in practice. This is mainly due to the fact that the generator and discriminator are modelled in terms of neural networks, thus restricting the optimization procedures to the parameter space of the networks rather than learning directly from the probability distributions [344]. On the other hand, this non-convergence could happen even if the model has ventured near the equilibrium point. In this case, when the generated samples are far from the target distribution, the discriminator pushes the generator towards the true data distribution while at the same time increasing its slope. When the generated samples approaches the target distribution, the discriminator's slope is the highest, pushing away the generator from the true data distribution [346]. As a result, depending on the stopping criteria, the GAN optimization might not necessarily converge to a Nash equilibrium [342] in which the performance of the discriminator and generator cannot be improved further (i.e. optimal conditions). In such a scenario, the generated samples are just collections of random noise.

The vanishing gradients occur when the discriminator is over-optimized and does not provide enough information to the generator to make substantial progress [343]. During backpropagation, the gradient of the generator flows backward from the last layer to the first, getting smaller at each iteration. This raises some complications since, in practice, the objective function of a discriminator close to optimal is not given by $2\,\mathrm{JSD}(p_R, p_\theta) - 2\ln 2$ as stated in Eq. (4.4.3), but rather close to zero. This then pushes the loss function to zero, providing little or no feedback to the generator. In such a case, the gradient does not change the values of the weights in the initial layers, altering the training of the subsequent layers.

Finally, the mode collapse occurs when the generator outputs samples of low diversity [347], i.e, the generator maps multiple distinct input to the same output. This is the case where the generated samples from the GAN capture only a few of the modes of the input data set. Mode collapse is a common pathology to all generative adversarial models as the cause is deeply rooted in the concept of GANs.

The challenges mentioned above contribute to the instabilities that arise in training GAN models. What often complicates the situation is that these obstacles are linked with each other and one might require a procedure that tackles all of them at the same time. As we will briefly describe in the next section, there exist various regularization techniques, some specific to the problem in question, that could influence the stability during the training.

4.4.3 Training stability & Regularization

Apart from the regularization procedures introduced in Sec. (4.1), there exists alternative procedures and techniques more suitable for generative adversarial models that can

alleviate the challenges described in Sec. (4.4.2). Whilst some of the procedures can be imposed on already defined GAN architecture, some are defined in terms of different model variants. Here, we just list some of the most used and widespread techniques.

- **Gradient Penalty:**

  One of the first line of defence against instabilities in training GANs is a technique known as *gradient penalty*. This consists in imposing a penalty on the gradient of the discriminator $D_\phi(x)$ in order to enhance its sensitivity in the regions $D_\phi(x) \to 0$ and $D_\phi(x) \to 1$. The gradient penalty is applied to the logit of $D_\phi(x)$ as follows

  $$\tilde{D}_\phi(x) = \ln\left(\frac{D_\phi(x)}{1-D_\phi(x)}\right) \quad \longrightarrow \quad \nabla\tilde{D}_\phi(x) = \frac{\nabla D_\phi(x)}{D_\phi(x)\left(1-D_\phi(x)\right)}. \tag{4.4.4}$$

  The penalty on the discriminator is applied in the regions where the predictions are wrong, i.e. the discriminator assigning labels *zero* to true data, or conversely assigning labels *one* to fake data. The gradient penalty loss is therefore defined by

  $$\mathcal{L}_{\text{GP}} = \mathop{\mathbb{E}}_{x\sim p_R}\left[\left(1-D_\phi(x)\right)^2 |\nabla\tilde{D}_\phi(x)|^2\right] + \mathop{\mathbb{E}}_{z\sim p_z}\left[D_\phi\left(G_\theta(z)\right)^2 |\nabla\tilde{D}_\phi(x)|^2\right]. \tag{4.4.5}$$

  In the case where the discriminator is perfectly well-trained the prefactors $\left(1-D_\phi(x)\right)$ and $D_\phi^2(x)$ vanish. The gradient penalty loss given in Eq. (4.4.5) can then be added to the discriminator's loss function $\tilde{\mathcal{L}}_D = \mathcal{L}_D + \lambda\mathcal{L}_{\text{GP}}$ with λ a properly chosen variable.

- **Architecture variants:**

  In order to alleviate the challenges associated with the training of GANs, various modifications on the architecture (including loss function) have been proposed. For instance, it has been studied in Ref. [348,349] that the generative capabilities of the Generator is inherently bounded by the size of the discriminator's architecture. This has motivated the use of large discriminator networks. On the other hand, the use of overparametrized discriminator may saturates the generator and hence ruin the training. However, this can be easily circumvented by restricting the weight updates of the discriminator using for example the gradient penalty described above. These eventually led to the use of *deep convolutional neural network* to represent both the discriminator and generator's. There exists thus far various GAN variants, examples are DCGAN [350,351], INFOGAN [352], BIGGAN [353], et cetera. For the study of PDFs, that will be presented in the next Chapter, we use a variant called WGAN [354].

- **Hyperparameter Optimization:**

  One of the most robust ways to come up with stable networks that do not suffer from the issues mentioned before is the hyperparameter tuning in which a space of parameter is scanned in order to find the optimal values given the metrics of interest. The choice of model (or variants) could even be regarded as a hyperparameter. Details on such a procedure are described in the next Chapter.

4.5 Summary

In this chapter, we reviewed the fundamentals of deep learning, explaining in details the various pieces that enter into the training of a deep neural network. We then provided a

brief description of what deep generative models are and how efficient they are in terms of modelling posterior distributions. In particular, we focused on generative adversarial neural networks, which differ from the other families of generative models in that GANs are likelihood-free. The challenges related to GANs were then discussed in details, as well as the standard regularization procedures used to tackle them. Two possible applications of the GANs to parton densities are presented in the next chapter.

5

EFFICIENT COMPRESSION OF PDF SETS USING GANS

It has been strongly emphasized in Chapter 1 that Parton Distribution Functions (PDFs) are crucial ingredients for all predictions of physical observables that are of interest at hadron colliders such as LHC, and efforts to make their determination accurate with smaller uncertainties have become increasingly relevant. In current phenomenology studies, apart from the theoretical uncertainties arising from missing higher-order corrections, they are one of the dominant sources of uncertainty in precision measurement. A striking example of such a limitation is the role that PDFs play in the extraction of the Higgs coupling from data [132]. It is therefore clear that for the sake of doing precision physics (an important path toward a deep understanding of the physics beyond the Standard Model), PDFs need to be accurate at a percent-level [4]. The road toward reaching such an accuracy manifests at two distinct levels: (i) the determination of the PDFs through a fitting procedure, and (ii) the a posteriori treatment of the fitted results such as the reduction of the number of Monte Carlo (MC) replicas (or number of eigenvectors in the case of Hessian PDFs). Recently, spectacular advancement has been seen at the level of the PDF determination thanks to the inclusion of new (LHC) datasets and more importantly to the improvement in methodology, see Ref. [4] for ample details. In the following chapter, what we are interested is the latter, specifically the compression of a Monte Carlo PDF replicas into a smaller set that replicates all the statistical properties of the starting PDF set (henceforth referred to as prior).

Originally, the compression of Monte Carlo PDFs has been introduced in the context of *Combined PDF Sets*. The idea consists in combining various PDF sets determined by different PDF fitter groups that implement different methodologies. As mentioned before, to the present date, there exists various PDF fitter groups that implement different methodology and provide different estimation of the PDF errors [34–68, 68–73, 355, 356]. As a consequence, even when benchmarking the different methodologies with the same input datasets, it has been shown that the resulting PDF sets still differ with a reasonable agreement [357]. Therefore, from the practical point of view, it seems reasonable to combine the different PDF sets and evaluate the combined uncertainty in order to obtain estimates of the PDF error that take into account the various methodologies.

Apart from its relevance for the combination of Monte Carlo PDF sets, the compression methodology is also useful for native Monte Carlo PDF. In the NNPDF fitting methodology where the PDFs are represented in terms of Monte Carlo replicas, a large number of such replicas are required in order to give an accurate representation of the underlying probability distribution. Indeed, there has been considerable evidence that one might require $\mathcal{O}(10^3)$ (or even $\mathcal{O}(10^4)$) Monte Carlo replicas in order to converge to the asymptotic result [358, 359]. Putting aside the practical limits that arise with the fitting of such

a large ensemble of Monte Carlo replicas, dealing with a large PDF set when producing phenomenological studies or Monte Carlo simulations is not ideal. In order to circumvent such issues, one could compress down the large sample of Monte Carlo replicas into a smaller set. However, when going from $\mathcal{O}(10^3)$ to say $\mathcal{O}(100)$, one has to make sure that all the various statistical properties such as correlations between pairs of flavours are preserved. This is where the compression approach comes into play as a mean to select a subset of Monte Carlo replicas that provides an accurate representation of the prior distribution.

A compression methodology has been introduced in Ref. [358]. The basic conceptual idea of the compression lies on the search of a subset of replicas that reproduces best the statistical features of the original Monte Carlo PDF set. It has been shown at the PDF level using NNPDF3.0 sets that a compressed set with $N_c = 100$ replicas is reasonably equivalent to a prior with $N_p = 1000$ replicas. We would like to propose a new compression strategy [360] that aims at providing a compressed set with an even smaller number of replicas when compared to the standard approach of Ref. [358], while providing a similar representation of the original probability distribution. The framework we would like to propose is based on the generative adversarial neural networks that was introduced in Sec. (4.4). The idea is to enhanced the statistics of a given input Monte Carlo PDF set by generating what we call *synthetic replicas*. These synthetic replicas are *new* replicas generated from the GANs and to some extent follow the same probability distribution as the prior. The motivation behind this is based on the following observation: large samples of Monte Carlo replicas contains fluctuations that average out to the an asymptotic limit; in the standard compression approach, the is idea is just to select samples that present less fluctuations and contain the bulk of the statistical information present in the prior. This suggests that it should be possible to generate samples that contain less fluctuations and once combined with the samples from the prior can lead to a more efficient compressed representations of the full result.

This chapter is organized as follows. In Sec. (5.1), we review the mathematical framework of the compression methodology while at the same time introducing a new implementation that is faster and most importantly can incorporate the GANs framework. The GANPDFS framework is introduced in Sec. (5.2.1) in which we present the GAN architecture used in our implementation. We also comment on some regularization procedures that are specific to our problem in order to circumvent the issues related to the training stability described in Sec. (4.4.2). The concept of hyperparameter optimization is then introduced in order to fit the methodology by scanning the hyperparameter space. The efficiency of the new framework is then assessed in Sec. (5.3) by comparing it against the standard approach based on some statistical estimators. In Sec. (5.4), we further validate our methodology by producing phenomenological results for LHC observables both at the total inclusive and differential levels. In Sec. (5.5), we give an outlook on the potential of using the GANPDFS framework to tackle the issue of the finite size effect by by-passing the need of fitting Monte Carlo replicas of $\mathcal{O}(10^4)$, or even beyond. Finally, in Sec. (5.6), we discuss why the GAN-based compression might not be relevant for NNPDF4.0 PDF sets and why that can be interpreted as an indication of the robustness of the NNPDF4.0 methodology.

5.1 Compression: methodological review

Let us briefly review the compression methodology for Monte Carlo PDF sets introduced in Ref. [358]. The following closely follows the description provided in the aforementioned reference. One can summarizes the main idea behind the compression of (combined) Monte Carlo PDFs as finding a subset of the original PDF set such that the statistical distance between the original and the compressed is minimal. This ensures that the loss of statistical information is as low as possible. In this sense, the compression methodology has to rely on two main ingredients: a proper definition of the distance metric that quantify the distinguishability between the prior and the compressed distributions, and an appropriate minimization algorithm that explores the space of possible combinations of PDF replicas that leads to such a minima.

It was originally proposed in Ref. [358] that a suitable figure of merit to measure the difference between two parton distribution functions is the *error functio*

$$\mathrm{ERF} = \frac{1}{N_{\mathrm{EST}}} \sum_k \frac{1}{N_k} \sum_i \left(\frac{C^k(x_i) - P^k(x_i)}{P^k(x_i)} \right)^2 , \tag{5.1.1}$$

where k runs over a set of statistical estimators with N_k the associated normalization factor, $P^k(x_i)$ is the value of the estimator k computed at a given point i in the x-grid for the prior set, and $C^k(x_i)$ is the corresponding value of the same estimator but for the compressed distribution. Finally, N_{EST} denotes the total number of statistical estimators involved in the minimization. The scale dependence is not explicit shown in Eq. (5.1.1) since the scale at which the the PDFs are evaluated is the same for all estimators. As in the original compression methodology, we include as statistical estimators lower moments (such as mean and standard deviation) and standardized moments (such as Skewness and Kurtosis). In addition, in order to preserve higher-moments and PDF-induced correlations, which are crucial for phenomenology studies, we also include the Kolmogorov-Smirnov distance and the correlation between multiple PDF flavours. For each statistical estimator, the normalization factor has to be defined properly as it compensates for the various orders of magnitude in different regions of (x, Q)-space that is mainly apparent in higher moments. Here, we rewrite their definitions:

Lower and higher moments:

Since the statistical properties of a PDF is mainly dominated by lower moments, they should be reproduced with higher accuracy in the compressed set. Let us denote by $P_i^r(x_j)$ and $C_i^r(x_j)$ respectively the prior and compressed sets of replicas r for a flavour i at the j-th position in the x-grid. The expression of the central value, standard deviation, Skewness, and Kurtosis for the prior are defined as

$$P_i^{\mathrm{CV}}(x_j, Q_0) = \frac{1}{N_R} \sum_{r=1}^{N_R} P_i^r(x_j, Q_0) \tag{5.1.2}$$

$$P_i^{\mathrm{STD}}(x_j, Q_0) = \sqrt{\frac{1}{N_R - 1} \sum_{r=1}^{N_R} \left(P_i^r(x_j, Q_0) - P_i^{\mathrm{CV}}(x_j, Q_0) \right)^2} \tag{5.1.3}$$

$$P_i^{\rm SKE}(x_j,Q_0) = \frac{1}{N_R}\sum_{r=1}^{N_R}\left(\frac{P_i^r(x_j,Q_0)-P_i^{\rm CV}(x_j,Q_0)}{P_i^{\rm STD}(x_j,Q_0)}\right)^3 \tag{5.1.4}$$

$$P_i^{\rm KUR}(x_j,Q_0) = \frac{1}{N_R}\sum_{r=1}^{N_R}\left(\frac{P_i^r(x_j,Q_0)-P_i^{\rm CV}(x_j,Q_0)}{P_i^{\rm STD}(x_j,Q_0)}\right)^4 \tag{5.1.5}$$

where Q_0 denotes the initial scale at which the prior PDF set is computed and N_R represents the total number of replicas. The expressions for the compressed set can be computed analogously by replacing P with C and changing the total number of replicas to size of the compressed set. Notice that in the summation the central replicas $r=0$ is not considered. From these definitions, one can therefore define the error function for the moment estimators. For each estimator k, the error function is defined as

$$\text{ERF}_k = \frac{1}{N_k}\sum_{i=-n_f}^{n_f}\sum_{j=1}^{N_x}\left(\frac{C_i^k(x_j,Q_0)-P_i^k(x_j,Q_0)}{P_i^k(x_j,Q_0)}\right)^2 \tag{5.1.6}$$

for $k = \text{CV},\text{STD},\text{SKE},\text{KUR}$. The number of active flavour is denoted by n_f while the size of the x-grid is denoted by N_x. The above equation is only valid if the denominator is non-zero, hence we only consider the case where $P_i^k(x_j,Q_0)\neq 0$. The normalization factor N_k for a statistical estimator k is defined as the lower 68% confidence-level of $N_{\rm rd}$ random selections. This can be expressed as

$$N_k = \frac{1}{N_{\rm rd}}\left[\sum_{t=1}^{N_{\rm rd}}\sum_{i=-n_f}^{n_f}\sum_{j=1}^{N_x}\left(\frac{R_i^{t,k}(x_j,Q_0)-P_i^k(x_j,Q_0)}{P_i^k(x_j,Q_0)}\right)^2\right]_{68\%} \tag{5.1.7}$$

where $R_i^t(x_j,Q_0)$ represent the t-th random selection extracted from the prior and contains the same number of replicas as the expected compressed set. This ensures that the value of the error function ERF_k for an estimator k is below 1.

Kolmogorov-Smirnov distance:

Since PDFs are continuous probability distributions, despite the fact that they are only available as an interpolation on the grid of x-points, the Kolmogorov-Smirnov test [361] (or its refinement, the Anderson-Darling test [362]) is a sensible estimator. In the context of PDF compression, the Kolmogorov-Smirnov distance is computed as follows [358]: for the prior set (resp. compressed), we count the number of replicas that falls within a given region and then normalize by the total number of replicas in the prior (resp. compressed). As in the original compression, we consider six different regions which are defined in terms of the standard deviation

$$\left[-\infty,-2P_i^{\rm STD}(x_j),-P_i^{\rm STD}(x_j),0,P_i^{\rm STD}(x_j),2P_i^{\rm STD}(x_j),+\infty\right]. \tag{5.1.8}$$

The contribution to the error function from the Kolmogorov-Smirnov can therefore be defined in the following way

$$\text{ERF}_{\rm KS} = \frac{1}{N_k}\sum_{z=1}^{6}\sum_{i=-n_f}^{n_f}\sum_{j=1}^{N_x}\left(\frac{C_i^z(x_j,Q_0)-P_i^z(x_j,Q_0)}{P_i^z(x_j,Q_0)}\right)^2, \tag{5.1.9}$$

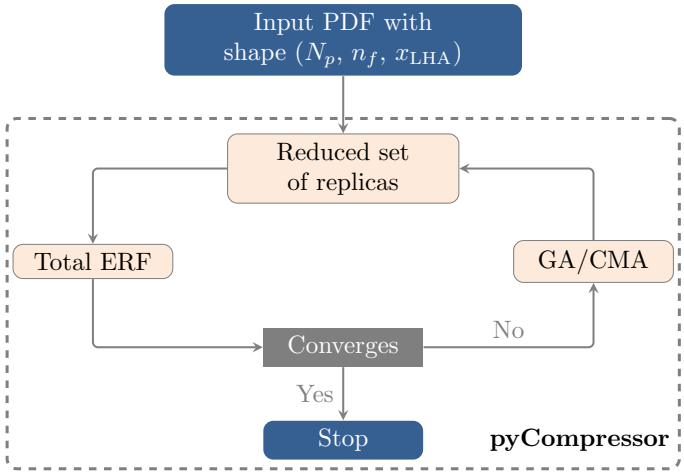


Figure 28 Flowchart describing the methodology for the compression of Monte Carlo PDF replicas. It takes as input a prior PDF set constructed as a grid in (N_R, n_f, x)-space and additional parameters such as the size of the compressed set and the choice of minimization algorithm. The output is a Monte Carlo replicas with a reduced size that follows the LHAPDF [5] grid format.

with z running through the six regions defined in Eq. (5.1.8). The normalization factor is then defined as

$$N_{\rm KS} = \frac{1}{N_{\rm rd}} \sum_{z=1}^{6} \sum_{t=1}^{N_{\rm rd}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{R_i^{t,z}(x_j, Q_0) - P_i^z(x_j, Q_0)}{P_i^z(x_j, Q_0)} \right)^2 , \tag{5.1.10}$$

where all the remaining symbols carry the same meaning as for the moments.

Correlations between PDFs:

As mentioned previously, it is important that the compressed set preserve PDF-induced correlations in physical cross sections. They play an important role in phenomenological studies. As a matter of fact, one of the main differences between a fit with 100 and 1000 replicas is that correlations between pairs of flavours are reproduced more accurately in the latter. In order to define the error function for the correlation, we introduce the correlation matrix $\mathcal{M}$ whose elements are defined as

$$\mathcal{M}_{ij} = \frac{N_R}{N_R - 1} \frac{\langle ij \rangle - \langle i \rangle \langle j \rangle}{\Sigma_i \Sigma_j}. \tag{5.1.11}$$

The brackets in the above equation are defined as follows

$$\langle i \rangle = \frac{1}{N_R} \sum_{r=1}^{N_R} f_i^r(x_i, Q_0), \qquad \langle ij \rangle = \frac{1}{N_R} \sum_{r=1}^{N_R} f_i^r(x_i, Q_0) f_j^r(x_j, Q_0) \tag{5.1.12}$$

where $f_i^r(x_i, Q_0)$ denotes the replicas r from either the prior or compressed set with flavour i and computed at the j-th position in the x-grid. The Σ_i represents the

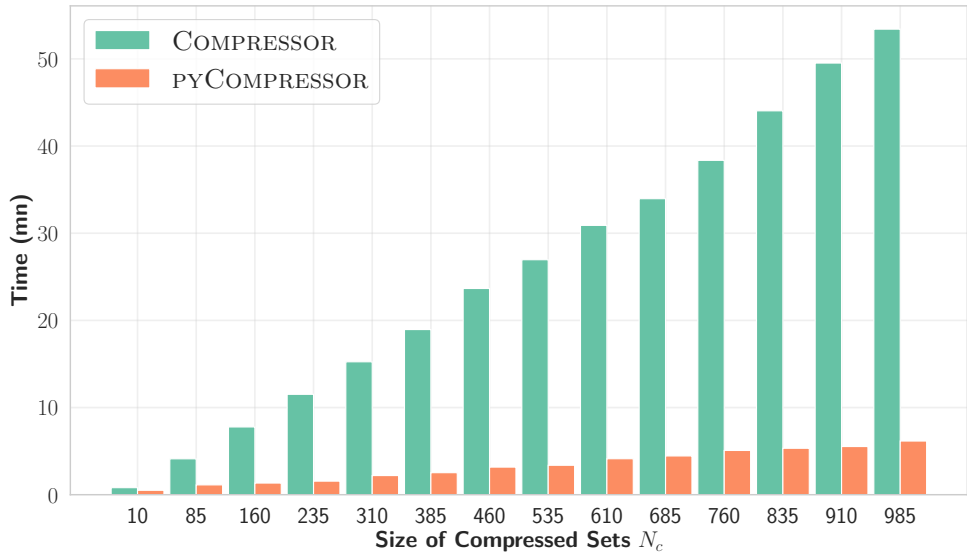


Figure 29 Speed benchmark comparing the old and new compression codes using the GA as the minimizer. The prior set is a NNPDF3.1 set with $N_p = 1000$ replicas. On the y-axis is shown the time (in minutes) that it takes to perform a compression while the x-axis represents the various size of compressed sets. For the purpose of these benchmarks, the parameters entering the Genetic Algorithm (GA) are chosen to be exactly the same across both implementations.

standard deviation and its expression is exactly similar to Eq. (5.1.3). We can now define the contribution of the correlation to the error function as

$$\text{ERF}_{\text{CR}} = \frac{1}{N_{\text{CR}}} \left(\frac{c-p}{p} \right)^2 \tag{5.1.13}$$

with the prior and compressed components respectively defined as

$$p = \text{Tr}\left(\mathcal{P} \cdot \mathcal{P}^{-1} \right), \qquad c = \text{Tr}\left(\mathcal{C} \cdot \mathcal{P}^{-1} \right), \tag{5.1.14}$$

where $\mathcal{P}$ and $\mathcal{C}$ respectively represent the correlation matrix of the prior and compressed set computed according to Eq. (5.1.11) and $\mathcal{P}^{-1}$ the inverse of $\mathcal{P}$. The normalization factor N_{CR} is computed analogously to the previous definitions, i.e. computed from a certain number of random selection.

Once the set of statistical estimators and the target size of the compressed set are defined, the compression algorithm can start the minimization procedure by searching for the combination of replicas that leads to the minimal value of the total error function defined in Eq. (5.1.1). In order to select the replicas entering into the compressed set, various minimization algorithms can be implemented. However, due to the discrete nature of the compression problem, it is convenient to use minimization based on Evolution Algorithm (EA) [363, 364] such as the Covariance Matrix Adaptation (CMA) strategy [365, 366] or

the Genetica Algorithm (GA) [367–369]. It is important to emphasize that no matter the choice of minimization algorithm is, one can never run into the risk of overfitting since the absolute minimum always exists. In Fig. 28, we show a diagrammatic representation of the compression workflow. It goes as follows: we first compute a grid of the prior PDF set for all replicas, flavours, and all points of the x-grid; then we choose a random subset from the prior in order to compute the value of the total error function; this process is repeated for a certain number of iterations making sure that the minimizer selects a better sample at every next iteration. The x-grid is constructed such that the points are distributed in the region where experimental data are available. That is, one choose x to lie within the $\left[10^{-5}, 0.9\right]$-region such that the points are logarithmically spaced between $\left[10^{-5}, 10^{-1}\right]$ and linearly spaced in the $\left[10^{-1}, 0.9\right]$-region.

The methodology described above was first implemented in a C++ code [358] and later revised in Ref. [360, 370] using state-of-the-art python development. The new implementation was required in order to accommodate for the generative models as will be discussed in the next section. Here and henceforth, we referred to the new implementation as PYCOMPRESSOR. An important benefit of the new implementation is also the gain in performance. Indeed, we observed that the new implementation is significantly faster as compared to the old one. In order to quantify such a gain, we compress a Monte Carlo PDF set with $N_p = 1000$ replicas into various sets with smaller sizes using the GA minimizer. The tests were run on a consumer-grade CPU[1] with 16 GB of memory. The results are presented in Fig. 29: on the x-axis are shown the various sizes of compressed sets while the required time for both compression codes to complete the same compression is shown on the y-axis. We can see that the new implementation is up to six times faster than the old one. One can also notice that the speed gain depends on the size of the compressed set. Not surprisingly, the difference between the two implementations increases as the the size of the compressed set gets bigger. This is due to the fact that PYCOMPRESSOR takes advantage of the multicore capabilities of modern computers[2].

5.2 Compressing PDF sets with GANs

This section is devoted to the description of our generative adversarial model, henceforth referred to as GANPDFS, and its incorporation into the compression methodology. In particular, we address the issue of instability when generating synthetic replicas. Sec. (4.4.2) introduced the general complexities of building a generative adversarial model in which we also discussed about various regularization procedures that can be used to alleviate such issues. Here, we specifically describe regularization procedures that are specific to our problem.

5.2.1 The GANPDFS methodology

Let us first restate again what we are trying to achieve. The problem we are concerned with is the following: suppose that the prior Monte Carlo PDF replicas follow a probability

1 AMD Ryzen 5 2600 with 12 threads boosted at 3.4 GHz.

2 Notice that the gain in speed due to the parallelization comes at the expense of a slightly higher memory usage. Although this is a very minor issue, this can become significant if the both he sizes of the prior and compressed sets are very large.

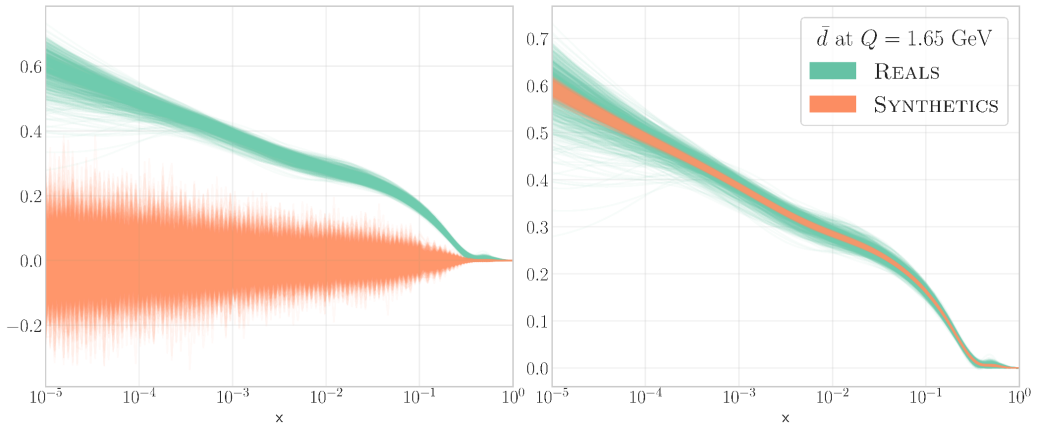


Figure 30 Illustration of non-convergence (left) and mode collapse (right) when generating Monte Carlo PDF replica. Here, we show the PDF for the $\bar{d}$-quark. The priors are represented by the green curves while the synthetic replicas are represented by the orange curves. The two plots we generated using the vanilla implementation of the generative neural network.

distribution p_R, we would like to generate synthetic replicas following a probability distribution p_θ such that p_θ is as close as possible to p_R. One way to achieve this is by defining a latent variable z with a fixed probability distribution p_z. The latent variable z can then be passed as an input to a function $G_\theta : z \to x_g$ which is generally defined to be a neural network in order to generate samples that follow a probability p_θ. By optimizing the parameter θ, one can modify the distribution p_θ to approach the target distribution p_R. Here, we propose to model the function G_θ in terms of a generative neural network and assess the goodness of the output using a discriminative neural network, hence the idea of generative adversarial neural networks (GANs).

The GANPDFS *framework & Regularization*

By adopting the standard variation (Vanilla) GANs introduced in the previous section, it is very likely that one ends up with the problem of instabilities described in Sec. (4.4.2). For the generation of synthetic replicas, the most likely scenarios are the non-convergence and mode collapse (see Fig. 30 for illustrations). Whilst the non-converging result must definitely be disregarded, the mode-collapse could still provide replicas that can be used for compression. Ideally, one would like the samples of synthetic replicas to reproduce the overall distribution of the prior, not only some part of it. It was argued in Sec. (4.4.2) that one of the main causes leading to unstable optimization during the training of a GAN is the low dimensional support of generated and target distribution. This means that the generator can only generate scarce samples from the input latent variable z and does not allow for any improvement at all. As suggested in Ref. [346], this could be overcome by providing to the latent variable z additional information. That is, one needs to supply the input latent variable with some relevant features from the real sample. In our case, this could be done by taking a linear combination of the input PDF replicas, adding on

top of it Gaussian noises, and using it as the input latent variable. This has been proved to improve significantly the stability of the GAN training and addresses the problem of non-convergence.

On the other hand, it was also discussed in Sec. (4.4.2) based on Refs. [371,372] that the objective function that the standard (original) GAN minimizes is not continuous w.r.t the generator's parameters. One can immediately see the shortcoming of this. Indeed, this can lead to both vanishing gradients and mode collapse. Such shortcomings were already perceived in the original GAN paper [342] where it was shown that the Jensen-Shannon divergence under idealized conditions contributes to oscillating behaviours. Several research papers have devoted to finding a suitable objective function in order to circumvent these particular problems (See Refs. [343,344,346] for a review). In our implementation, we resort to a loss function that was briefly introduced in Sec. (4.4.2): the Wasserstein or Earth's Mover (EM) distance. This objective function was subsequently implemented in Wasserstein GAN (WGAN) [354,373,374] as application to image generation and has proved to not only yield more stability but also more variety in the generated sample. The EM loss function is defined as follows

$$\min_{\theta}\max_{\phi} V = \mathop{\mathrm{E}}_{x\sim p_R}\left[D_\phi(x)\right] - \mathop{\mathrm{E}}_{x\sim p_\theta}\left[D_\phi(x_g)\right], \tag{5.2.1}$$

where we recall that $x_g = G_\theta(z)$ with z following a probability distribution p_z. The EM is effective in solving vanishing gradients and model collapse as it is continuously differentiable w.r.t the generator and discriminator's parameters. That is, GAN-based Wasserstein models result in a discriminator function whose gradient w.r.t its inputs is better behaved than in the standard GAN approach. This means that the discriminator can be trained until optimality without worrying about vanishing gradients.

Despite the fact that the Wasserstein distance measure yields non-zero gradients everywhere for the discriminator, the resulting GAN architecture can still be unstable when the gradients of the loss function are large. This can straightforwardly be addressed by clipping the weights of the discriminator to lie within a compact space defined by $[-c, c]$ where c is just some finite not-so-large number. Based on all of this, the default GANPDFs architecture is a modified version of a WGAN and is described in Algorithm 2. If the needs arise, the GANPDFs codes allow the users to straightforwardly define a new architecture from an input file.

The hyperparameter optimization

We have just described a possible architecture for our generative models that can alleviate the issues related to stability during the training, however, there is still some arbitrariness in choosing the parameter that each neural network (generator or discriminator) can take. Indeed, one of the main factors that influence the training of a GAN is the architecture of the neural networks. The choice of (hyper) parameters could scale up significantly the GAN's performance. However, coming up with values for hyperparameters such as the size of the network or the number of nodes in a given layer is particularly challenging as the parameter space is very large and obviously trying blindly various hyperparameters does not appear to be a suitable solution. A heuristic approach to tackle this problem–which is widely regarded as the best option within the machine learning community–is known as *hyperparameter scan* or *hyperparameter optimization*. This approach consists in searching

Algorithm 2: Default stochastic gradient descent training algorithm implemented in GANPDFS for the generation of synthetic replicas.

for *epochs* $1, \cdots, N$ **do**
  for *discriminator steps* $1, \cdots, k$ **do**
    - Sample minibatch of size m from the original input sample: $\{x_r^{(1)}, \cdots, x_r^{(m)}\}$
    - Sample minibatch of size m from the **custom** latent space: $\{z^{(1)}, \cdots, z^{(m)}\}$
    - Perform gradient **ascent** on discriminator:

$$\nabla_\phi V\left(G_\theta, D_\phi\right) = \frac{1}{m}\nabla_\phi \sum_{i=1}^{m}\left(D_\phi\left(x_r^{(i)}\right) - D_\phi\left(x_g^{(i)}\right)\right)$$

    - $\phi = \phi + \mathrm{RMSProp}(\phi, \nabla_\phi V)$
    - Clip weights within $[-c, c]$
  end
  for *generator steps* $1, \cdots, l$ **do**
    - Sample minibatch of size m from the **custom** latent space: $\{z^{(1)}, \cdots, z^{(m)}\}$
    - Perform gradient **descent** on generator:

$$\nabla_\theta V\left(G_\theta, D_\phi\right) = \frac{1}{m}\nabla_\theta \sum_{i=1}^{m} D_\phi\left(x_g^{(i)}\right)$$

    - $\theta = \theta - \mathrm{RMSProp}(\theta, \nabla_\theta V)$
  end
end

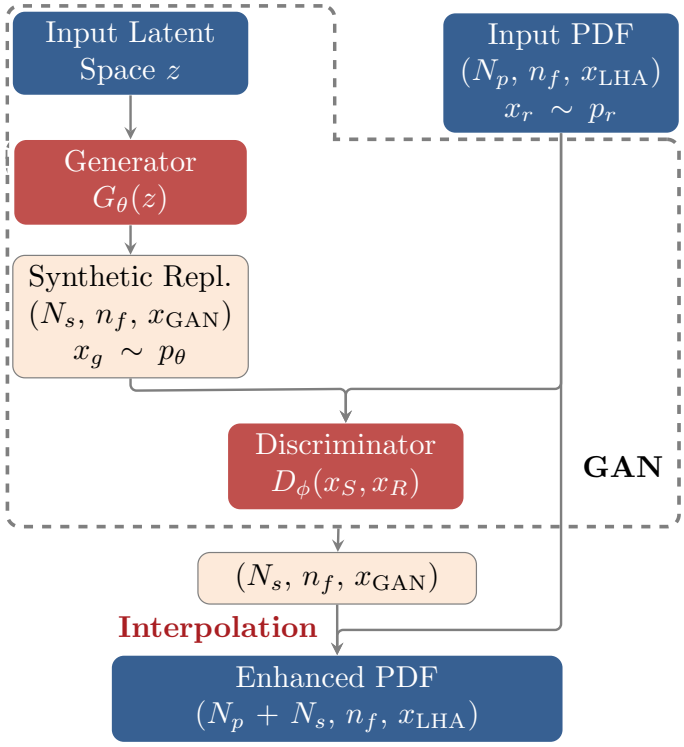


Figure 31 Flowchart describing the GANPDFS framework. The discriminator receives as input the prior PDF set computed as a grid in $(N_p, n_f, x_{\rm LHA})$-points while the generator receives as input the latent variable constructed as Gaussian noise of linear combination of the prior set. In addition to the real sample, the output of the generator also goes through the discriminator. The simoultaneous training of both neural networks goes on until a certain number of epochs is reached. The final output is an enhanced PDF following the LHAPDF grid format.

the best set of hyperparameters through an iterative search of the parameter space. In our implementation, we rely on the *Tree-structured Parzen Estimator* (TPE) [375] as an optimization algorithm. The TPE consists in building a probability model of the objective function and use it to select the most promising hyperparameters to evaluate the true objective function. Such a method outperforms by a significant amount any random search. The main challenge in performing such a hyperparameter optimization, however, is the definition of a figure of merit to hyper-optimize on. Indeed, there is no established figure of merit for generative models that can be used to assess the faithfulness of the generated samples. One possibility is to use one of the statistical estimators described in Sec. (5.1), but in doing so we restrict ourselves to only a particular feature of the distribution. In our generative model, we decide to use the Fréchet Inception Distance (FID) [376] as the figure

of merit to hyperoptimize on. For a target distribution with mean μ_r and covariance Σ_r and a synthetic distribution with mean μ_s and covariance Σ_s, the FID is defined as:

$$\text{FID} = \frac{1}{2n_f+1}\sum_{i=-n_f}^{n_f}\left\|\mu_r^{(i)}-\mu_s^{(i)}\right\|^2 + \text{Tr}\left(\Sigma_r^{(i)}+\Sigma_s^{(i)}-2\sqrt{\Sigma_r^{(i)}\Sigma_s^{(i)}}\right) \tag{5.2.2}$$

where the index i refers to the parton flavours. This suggests that the smaller the value of the FID is, the closer the generated samples are to the target distributions. For the implementation of the TPE hyperoptimization procedure in our GAN framework, we rely on a third party library called HYPEROPT [377].

We can now proceed to the description of our GANPDFS workflow. In Fig. 31 we show a diagrammatic summary of the training. Let us first mention that our implementation is based on the machine learning framework TENSORFLOW [378] using KERAS as a backend [379, 380]. Focusing now on Fig. 31, we see that in order to train, the GANPDFS receives as input a grid of shape $(N_p, n_f, x_{\text{LHA}})$ where N_p represents the number of input replicas and x_{LHA} denotes the size of the x-grid which corresponds exactly to the internal LHAPDF-grid. The input grid is then passed to the discriminator and will be used to assess the generated replicas from the generator. Separately, the generator takes as input the latent variable z and outputs a grid of shape $(N_s, n_f, x_{\text{GAN}})$. Notice that x_{GAN} does not have to be the same as x_{LHA}. This is because it has been observed that the generator trains better when provided with a large size of x-grid. In such a case, an interpolation is required in order to produce a LHAPDF-like grid. By default, x_{GAN} is a grid of $N_x = 500$ points logarithmically spaced in the small-x region $\left[10^{-9}, 10^{-1}\right]$ and linearly space in the large-x region $\left[10^{-1}, 1\right]$. The output of the GANPDFS is then a LHAPDF grid with $(N_p + N_s)$-replicas at a starting scale $Q_0 = 1.65$ GeV which can then be evolved using APFEL [381]. Notice that physical constraints such as positivity of the PDFs or sum rules are not enforced but rather inferred from the underlying distribution.

5.2.2 The GAN-enhanced compression methodology

Having separately introduced the concepts of compression and generative adversarial model for Monte Carlo PDF replicas, we now provide a description on how these two components can be incorporated into one single framework. In Fig. 32, we present a schematic diagram describing how the two frameworks are interfaced together. The workflow goes as follows: the input PDF grid is computed for a given Monte Carlo PDF set containing N_p replicas at fixed Q_0 (which by default is set to $Q_0 = 1.65$ GeV) and at some value of the Bjorken x. If enhancing the statistics of the prior with the GAN is not required, then the reduction strategy follows the standard compression methodology introduced in Sec. (5.1). If, on the other hand, the GAN-enhancement is required, the GANPDFS is used to to generate N_s synthetic replicas. The synthetic replicas combined with the prior, which henceforth we call *enhanced set*, has now a total size $N_E = N_p + N_s$. As mentioned before, since the x-grid of the output of the GAN does not necessarily have to be the same as for the prior, an interpolation might be required in order to combine the two. The enhanced (or combined) set is then passed to the PYCOMPRESSOR . It is crucial to emphasize that in the context of GAN-enhanced compression, the samples of replicas that will end up in the compressed set are drawn from the enhanced set rather than from

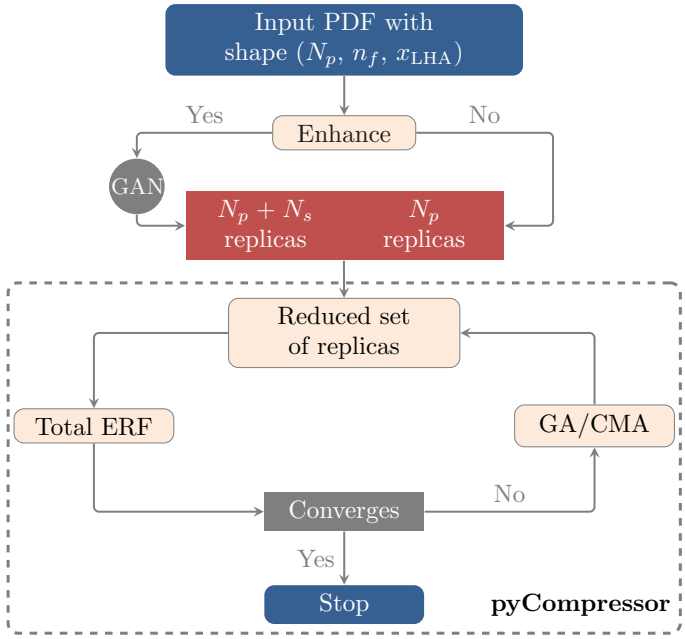


Figure 32 Flowchart describing the combined GANPDFS-PYCOMPRESSOR framework. The workflow starts with the computation of the prior MC PDF as a grid of $(N_p, n_f, x_{\rm LHA})$. If GAN-enhanced is required, the prior replicas are supplemented with synthetic replicas. If not, the compression follows the standard methodology. The output is always a compressed set containing smaller numbers of replicas.

the prior only. However, since the main goal is still to construct a compressed set that reproduces the probability distribution of the prior, the minimization has to be performed w.r.t the input Monte Carlo PDF replicas. In this sense, the expression of the total error function in Eq. (5.1.1) has to be modified such that the value of the estimator C^k for the compressed set has to be computed using samples from the enhanced distribution. We stress that the expression of the normalization factors does not change, i.e. the random sets of replicas have to be extracted from the prior.

Adiabatic minimization

Performing a compression from an enhanced set can be very challenging due to the large number of replicas to be considered. Indeed, the factorial growth of the number of replicas can probe the limit of the minimization algorithm and may potentially spoil the minimization procedure. In other words, the vast nature of the space of replica that could be explored increases the odd of the compression to be stuck in some local minima. However, if results from the standard compression are already provided, the compression code provides a more efficient compression procedure from the enhanced set with the means of an adiabatic minimization. The adiabatic minimization for the enhanced compression consists on taking as a starting point the space of replicas where the best from the standard compression was generated and start from there. Such a minimization not

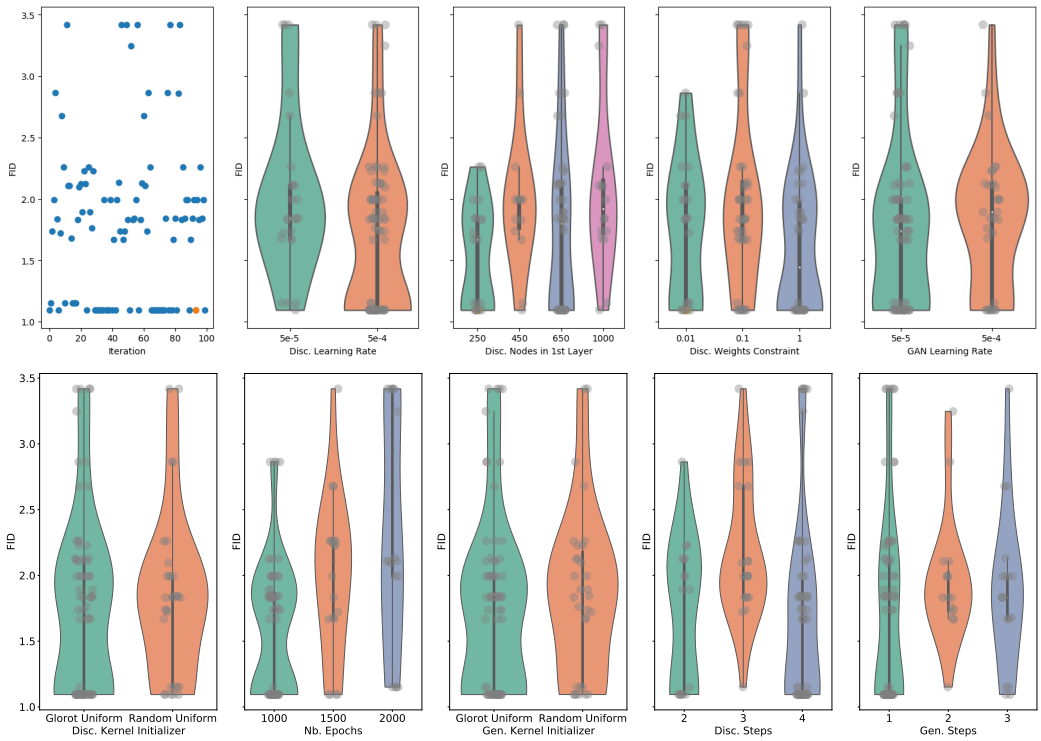


Figure 33 Graphical representation of an hyperparameter scan for a few selected parameters. The results were produced with a certain number of trial searches using the TPE algorithm. On the y-axis is represented the different values of the FID while the various hyperparameters are represented on the x-axis. The violin plots represent the behaviour of the training for a given hyperparameter and violins with denser tails are considered better choices due to their stability.

only yields faster convergence but also prevent the minimization algorithm to be trapped in some local minimum in case the generated replicas from the GAN contains fluctuations.

Hyperoptimized architecture

In this section, we present the GANPDFs architecture that has come out as the result of an hyperparameter optimization. The setup in which we performed the hyperparameter optimization is as follows: we considered as a prior a Monte Carlo PDF replicas with $N_p = 1000$ replicas fitted using the NNPDF3.1 methodology [49]. We then use the GANPDFs to generate $N_s = 2000$ synthetic replicas for a total of $N_E = 3000$ enhanced replicas. This procedure is repeated for a certain number of iteration through the hyperparameter scan by minimizing the FID. In Fig. 33, we plot an example of a hyperparameter scan in which a few selected parameters are shown. These are results after performing a certain number of trials. For each hyperparameter, the values of the FID are plotted as a function of the different parameters. For the initialization of the discriminator's weights, for instance, there are two choices, namely the Glorot Uniform and the Random Uniform. The values

	Generator	Discriminator
Network depth	3	2
kernel initializer	glorot uniform	glorot uniform
weight clipping	-	0.01
optimizer	None	RMSprop
activation	leakyrelu	leakyrelu
steps	1	4
batch size	70%	
number of epochs	1000	

Table 2 Parameters on which the hyper-parameter scan was performed are shown in the first column. The resulting best values are shown for both the generator and discriminator in the second and third column respectively.

of the FIDs are represented by the gray points and the violin shapes represent the density of gray points within a particular region. In this sense, the violin plots represent a visual illustration of how a given parameter behave during the training. That is, violins with denser tails are considered better choices as they yield more stable training. For instance, we can see that 1000 epochs lead to a slightly more stable results as opposed to 1500 or 2000. For a complete summary, the list of hyperparameters with the corresponding values are shown in Table 2.

5.3 Efficiency of the new GAN-enhanced compression

In the following sections, we quantify the performance of the GAN-enhanced framework over the standard approach by subjecting both methodologies on various statistical tests. As for the hyperparameter scan, all the analyses presented here were produced with the following setup: the prior contains $N_p = 1000$ replicas fitted with the NNPDF3.1 methodology which then gets enhanced with the GANPDFS framework in order to generate $N_E = 3000$ replicas (i.e. $N_s = 2000$ replicas). The compression of the native prior and GAN-enhanced replicas are both handled by the PYCOMPRESSOR.

5.3.1 Validation of the GAN methodology

First, we would like to see, for a given compression from the enhanced set, how many replicas are selected from the synthetic samples. In order to do so, we perform two types of compression: one with the GAN-enhanced methodology and one with the standard approach. In Fig. 34, we show the disparity between the standard and GAN-enhanced compressed sets, i.e. the number of replicas that are present in the GAN-enhanced compressed sets (including synthetic replicas) but not in the standard sets. The results are shown for different sizes of the compressed sets. The green histograms represent the number of replicas that are not present in the final set of the standard approach but present in the GAN-enhanced compressed sets. The orange histograms represent the number of synthetic replicas out of that disparity. We can see that for smaller compressed sizes (smaller than $N_c = 200$), the percentage of synthetic replicas that compose the compressed

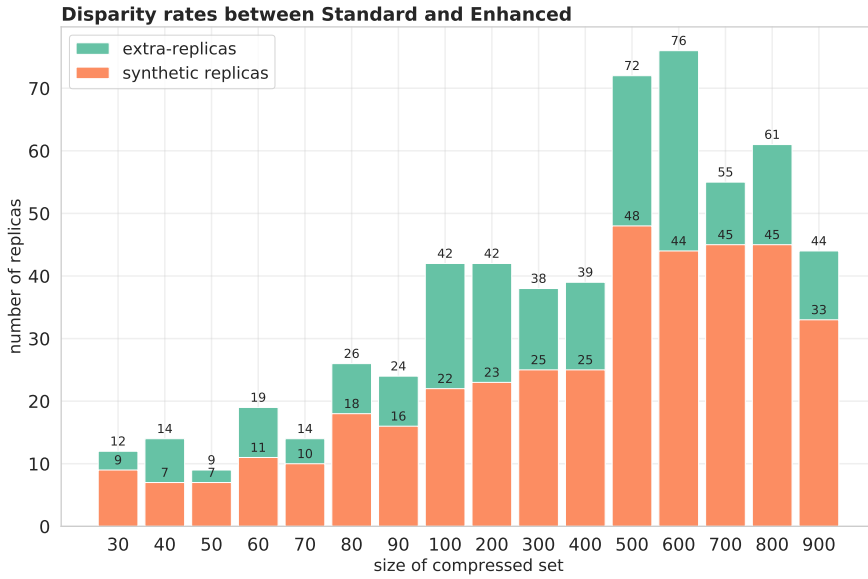


Figure 34 Histograms representing the number of disjoint replicas that are present in the GAN-enhanced compressed set but not in the standard. That is, the green histograms are constructed by counting the number of replicas from the GAN-enhanced compression that are not seen after the standard compression. Out of these numbers, the green histograms represent the number of replicas that are coming from the synthetic replicas. The results are shown as a function of the size of the compressed set.

set exceeds 10% and this percentage decreases as the size of the compressed set increases. This is explained by the fact that as the size of the targeted compressed set approaches the size of the prior Monte Carlo PDF replicas, the probability distribution of the reduced real samples get closer to the prior such that fewer synthetics are required.

Now, we turn to the validation of the GAN-enhanced compression methodology by first looking at the PDF central values and luminosities. In Fig. 35 we show the absolute central values resulting from the GAN-enhanced compression at an energy scale $Q = 1.65$ GeV. The results are shown for various PDF flavours $(g, s, \bar{d}, \bar{s})$ and for different size of compressed sets ($N_c = 50, 70, 100$). The results are normalized to the the central value of the prior Monte Carlo PDF replicas. In order to qualitatively assess whether or not compressed sets with size $N_c = 50, 70, 100$ are good representations of the prior probability distribution, we also plot the 68% and 1-sigma bands. We see that, for all the three sets, the PDF uncertainties are much larger than the fluctuations of the central values, indicating that a compressed set with $N_c = 50$ captures the main statistical properties of the prior. In Fig. 36, we plots the luminosities for the g-g, d-$\bar{u}$, and d-$\bar{u}$ combinations as a function of the invariant mass of the parton pair M_x for two different sizes of compressed sets, namely $N_c = 50, 70, 100$. As previously, the hatched bands represent the 68% confidence interval while the the 1-sigma bands are represented by the dashed lines. The compressed sets $N_c = 50, 70$ slightly deviate from the underlying probability distributions at lower values of M_x in which PDFs are known to be non-Gaussian. However, one can notice that the deviations are

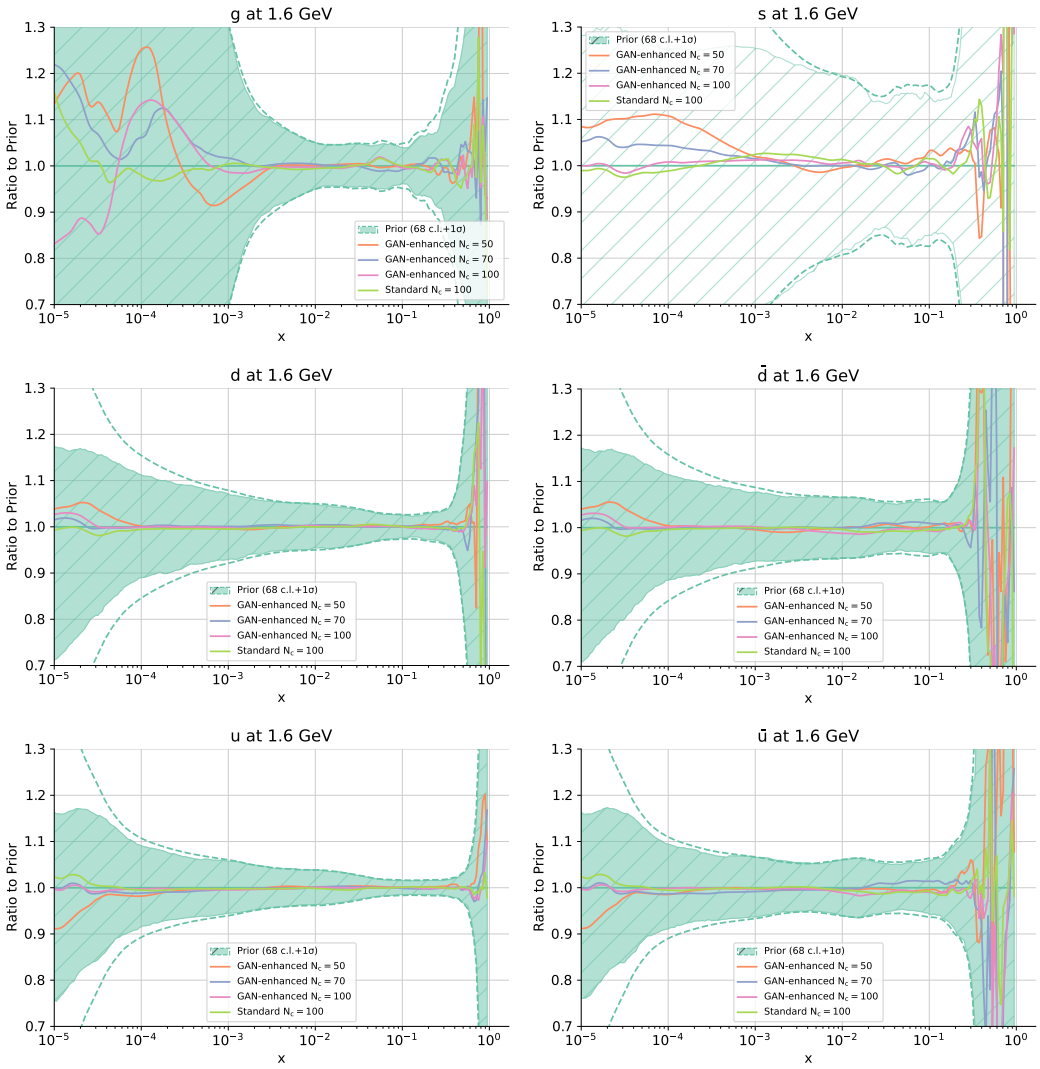


Figure 35 Comparison of the PDF central values resulting from the compression using the GAN-enhanced methodology for different values of $N_c = 50, 70, 100$ and different PDF flavours $(g, s, d, u, \bar{d})$. Results are normalized to the central value of the prior set. The 68% confidence interval is represented by the green hatched band whole the 1-sigma band is represented by the dashed green lines. The PDFs are computed at $Q = 1.65$ GeV. Plots produced using REPORTENGINE-based VALIDPHYS [6] suite.

very small compared to the uncertainty bands. For $N_c = 100$, we see very good agreement between the prior and the GAN-enhanced compressed set. These plots further strengthen the observations made at the level of central values. The above results confirm that compressed sets constructed from the GANPDFS-PYCOMPRESSOR methodology fully preserve the PDF central values and luminosities between pairs of PDFs. In particular, we

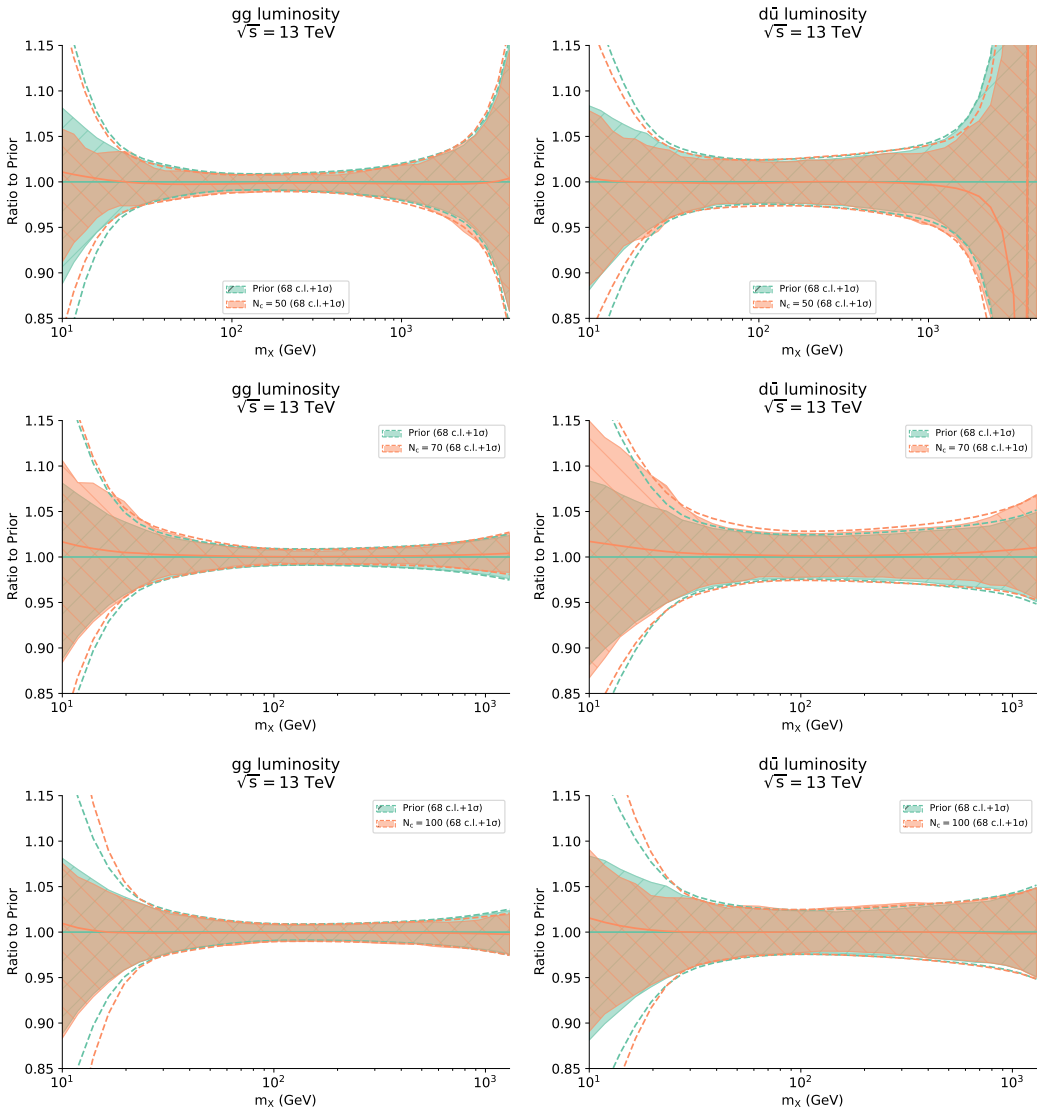


Figure 36 Comparison of the PDF luminosities between the prior and the GAN-enhanced compressed sets at LHC with $\sqrt{s}=13$ TeV. The results are shown for different sizes of compressed sets, namely $(N_c=50,70,100)$ and for different PDF luminosities (g-g, d-$\bar{u}$, and d-$\bar{u}$). The hatched error bands and the region envelopped with the dashed lines represent the 68% confidence level and 1-sigma deviation respectively. Plots produced using REPORTENGINE-based VALIDPHYS [6].

conclude that about $N_c=50$ replicas sufficient enough to reproduce the main statistical properties of a prior Monte Carlo PDF with $N_p=1000$ replicas. Next, we quantify how efficient is the generative-based compression methodology compared to the standard approach of Ref. [358].

5.3.2 Performance of the generative-based compressor

In order to qualitatively quantify the performance of the new generative-based compression methodology, we evaluate the compressed sets resulting from the new and standard methodologies on various statistical estimators. We consider the same settings as in the previous sections in which the results of the GAN-enhanced compression are compared to the results from the standard approach where the samples of replicas are selected directly from the prior Monte Carlo PDF.

Let us start by comparing the two methodologies at the level of the error functions. In Fig. 37, for each compressed set, we show the contribution of each of the statistical estimator (see in Sec. (5.1) for details) that contribute to the total value of the ERF using the standard (green) and GAN-enhanced (orange) approach as a function of the size of the compressed set. Notice that the ERFs shown in the plot are non-normalized. This is indeed irrelevant when comparing the two methodologies since the normalizations are the same in both approaches. For reference, we also show the mean (purple) and median (light blue) computed by taking the average ERF values from $N_R = 1000$ random selections. To provide an estimate of how representative of the prior distribution a random compressed set is, the confidence intervals $(50\%, 68\%, 90\%)$ computed from the random selections are also shown as error bars of varying colours.

First of all, as expected, we see that as the size of the compressed set increases, the ERF values of all the estimators tend to zero. The fact that the correlation present some fluctuations support our claim that as opposed to the lower and higher moments, correlations are difficult to preserve. On the other hand, it is clear that both compression methodologies outperform quite significantly any random selection of replicas. But in addition, by comparing the results from the standard and GAN-enhanced approach, we observe that the estimators for the GAN-enhanced compression, are in all cases except for a very few, below those of the standard compression. Indeed, it could happen that the error function for a given estimator for the standard approach is lower than the GAN while the total value of the error function is still lower for the latter. Overall, this suggests that the GAN-enhanced approach will result in a total value of the ERF that is much smaller than the one from the standard compression methodology.

In terms of efficiency, these results imply that the new generative-based methodology outperforms the standard compression approach by providing a more adequate representation of the probability distribution of the prior thanks to the synthetic replicas. This is particularly illustrated in Fig. 38 in which we plot in solid black line the total value of the error function for the standard compression as a function of the size of the compressed set. The vertical dashed lines represent the size of the compressed set $N_c = 70, 90, 100$ while the horizontal solid lines represent the respective ERF values for the GAN-enhanced compressed sets. The intersection between vertical dashed lines and horizontal solid lines being below the black line strongly indicate that the generative-based compression outperforms the standard approach. For instance, we see that $N_c = 70$ from the enhanced compressed set is equivalent to about $N_c = 110$ from the standard approach, and $N_c = 90$ from the enhanced set provides slightly more statistics that $N_c = 150$ from the standard compression methodology.

In addition to the above checks, one can also verify that correlations between PDFs are well preserved after the compression. It is important to emphasize that one of the

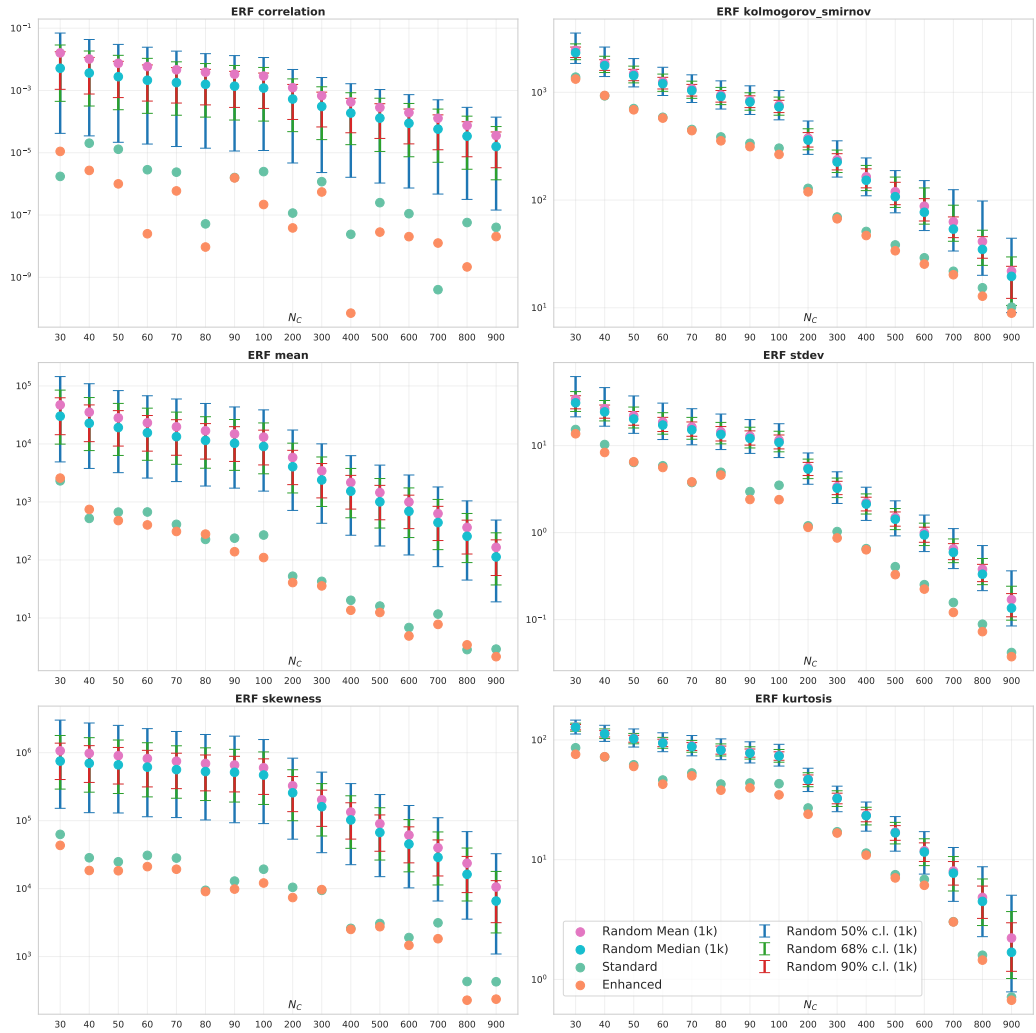


Figure 37 Comparison of the best ERF values for the compression of a Monte Carlo PDF set with $N_p = 1000$ replicas for various sizes of the compressed sets. For each compressed set, we show the contribution of the statistical estimators (see Sec. (5.1)) that contribute to the total error function using the standard compression (green) and the GAN-enhanced compression (orange) methodology. Notice that the ERFs on the plot are non-normalized. For illustration purposes, the mean (purple) and median (light blue) resulting from the average of $N_R = 1000$ random selections are shown. The resulting confidence intervals from the random selections are represented by the blue (50%), green (68%), and red (90%) error bars.

main differences between a fit with 100 and 1000 Monte Carlo replicas is that correlations are reproduced more accurately in the latter [358]. This is one of the main reasons why the compression methodology is important. Here, we show that for the same size of compressed set, the resulting compression from the GAN-enhanced methodology also

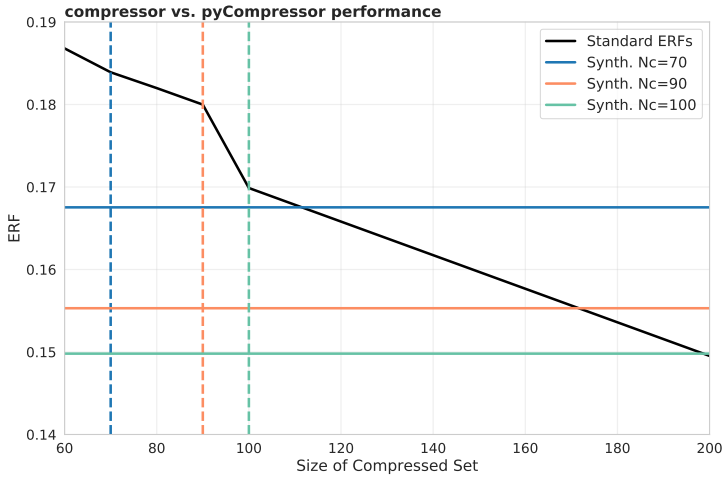


Figure 38 Comparison of the performance of the new generative-based compression algorithm (GANPDFS) and the previous methodology (standard). The normalized ERF values of the standard compression are plotted as a function of the size N_c of the compressed set (solid black line). The dashed blue, orange, and green lines represent $N_c = 70, 90, 100$ respectively. The solid lines represent the corresponding ERF values of the enhanced compression.

reproduces more accurately the correlations from the prior than the standard compression. One way of checking this is to plot the correlations between two given PDFs as a function of the Bjorken variable x. In Fig. 39, we show the correlation between a few selected pairs of PDFs (g-u, g-$\bar{u}$, and d-$\bar{u}$) for $N_c = 50, 100$ at an energy scale $Q = 100$ GeV. The results from the GAN-enhanced compression (orange) are compared to the ones from the standard approach (green). For illustration purposes, we also show PDF correlations from sets of randomly chosen replicas (dashed blue lines). We see that both compression methodologies capture very well the PDF correlations of the prior distribution. Specifically, in the case $N_c = 100$, we see small but noticeable differences between the old and new approach, with the new approach approximating best the original Monte Carlo PDF replicas. Although moderate, such a difference is also noticeable for compressed sets with $N_c = 50$ replicas.

An analogous way to verify that the compressed sets resulting from the GAN enhanced methodology reproduce more accurately the correlations of the prior PDF replicas is to compute the difference in correlation matrices. That is, compute the correlation matrix for each set (prior, standard, enhanced) and then compute the difference between the correlation matrix of the prior and the standard (or enhanced respectively). Such studies are shown in Fig. 40 where the matrices are defined in a logarithmic x grid with size $N_x = 70$ points for each of the $n_f = 8$ light partons. The first column represents the difference between the correlation matrix of the prior and the standard results while the second column represents the difference between the correlation matrix of the prior and the GAN-enhanced compressed results. The first, second, and third row shows the result for $N_c = 50$, $N_c = 70$, and $N_c = 100$ respectively. As we go from the top to bottom, we see that

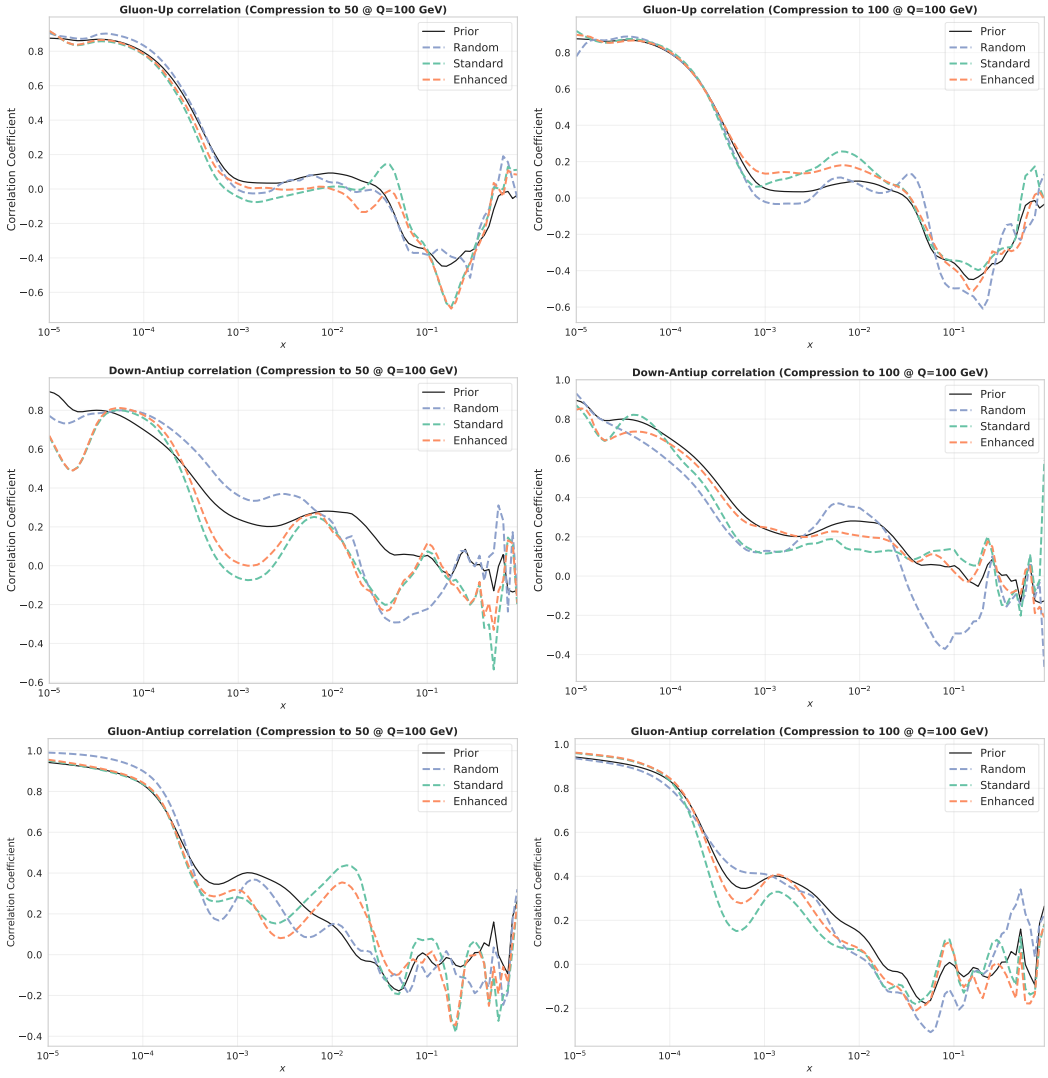


Figure 39 Comparison of the correlations between various pairs of flavours. The results are shown for different size of the compressed sets, namely $N_c = 50, 100$. The energy scale Q has been chosen to be $Q = 100$ GeV. The correlation extracted from the results of the GAN-compressor (orange) is compared to the results from the standard compressor (green). For a comparison, the results from the random selection are also shown in purple. Plots produced using REPORTENGINE-based VALIDPHYS [6].

the correlation matrix is becoming lighter, indicating an increase in similarity between the PDF correlations of the prior and the compressed sets. This feature is seen on both compression methodologies. However, as we look from left to right, we can also see that the correlation matrices on the right sides are lighter than the ones on the left. Although this is barely seen in the case $N_c = 50$, a minor difference can be seen at $N_c = 70$ while the

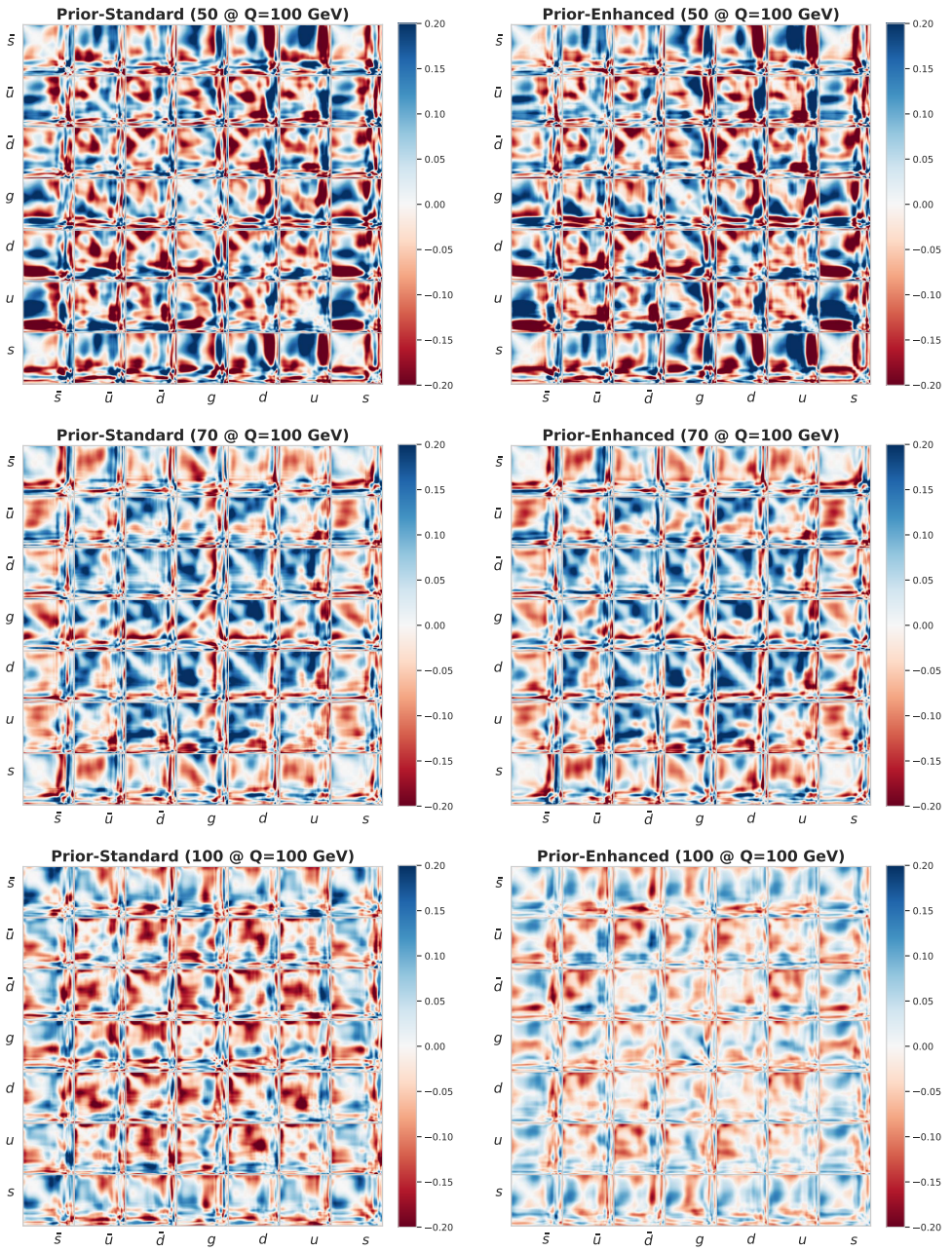


Figure 40 Difference between the correlation matrices of the prior and the compressed set resulting from the standard (first row) or enhanced (second row) compression. The correlation matrices are shown for different sizes of the compressed set.

difference is clearly significant for $N_c = 100$. These confirm the results shown in Fig. 37 in

which it was shown that the values of the error function for the correlation are smaller in the case of a GAN-enhanced compression.

5.4 Phenomenological implications

We have argued in the previous section based on various statistical estimators at the PDF level that the GAN-enhanced compression methodology outperforms the standard compression approach. In the following section, we provide the implications of such a performance for hadron collider phenomenology both at the level of the integrated and differential cross sections. Specifically, we consider the production of the Higgs boson production through the VBF (Vector Boson Fusion), top-quark pair production, and DY production decaying into leptons. The enhanced Monte Carlo PDF replicas are generated using the exact same setting as in the previous sections. In particular, here, we primarily focus on compressed sets with size $N_c = 50$ and $N_c = 70$.

Description of the theory predictions

All the theory predictions presented in the following sections have been computed for proton-proton collision at a center-of-mass energy $\sqrt{s} = 13$ TeV using next-to-leading order theory generated using Monte Carlo Generators [120, 121, 123, 382] interfaced to APPLGRID [126]. The APPLGRID grids are converted into a PINEAPPL [7] grid that provides more functionalities and flexibility. The theory predictions are then folded with the parton densities (prior or compressed set) using again PINEAPPL. For completeness, we list below the values of the input parameters:

$$\begin{aligned} m_{\rm W} &= 80.352\text{GeV}, \qquad \Gamma_{\rm W} = 2.084\text{GeV}, \qquad m_t = 172.5\text{GeV} \\ m_{\rm Z} &= 91.1535\text{GeV}, \qquad \Gamma_{\rm Z} = 2.4943\text{GeV}, \qquad \Gamma_t = 1.37758\text{GeV} \\ m_{\rm H} &= 125.0\text{GeV}, \qquad \Gamma_{\rm H} = 4.07468 \times 10^{-3}\text{GeV} \end{aligned} \tag{5.4.1}$$

For the aforementioned processes, we first present results for the total cross-sections. As will be shown later, one observes that at the level of integrated cross sections, the difference between a prior Monte Carlo PDF replicas and a compressed set (either be from the standard or GAN-enhanced approach) is extremely tiny. This is because, integrated cross sections are dominated by a specific region of the phase space where the bulk of the contribution is coming from, making them sensitive to PDFs in a very specific range of x and Q. For this reason, we also present phenomenological results for differential distributions. Specifically, we consider distributions that are differential in rapidity (namely, rapidity of the Higgs boson, rapidity of the W boson, and rapidity of the top-quark). The different distributions used in the following sections are listed in Table 3 along with the range spanned by each distribution with the number of bins. Notice that no acceptance cuts are imposed for all distributions.

Integrated cross sections

Using the theoretical predictions constructed according to the procedure described above, we show in Fig. 41 the results of the integrated cross sections. The plots describe the total

Process	N_{bin}	Rapidity Range
H-VBF	10	$\eta^{\text{H}}_{\text{min}} = -2.5, \quad \eta^{\text{H}}_{\text{max}} = 2.5$
W	10	$\eta^{\text{W}}_{\text{min}} = -4.0, \quad \eta^{\text{W}}_{\text{max}} = 4.0$
Z	10	$\eta^{\text{Z}}_{\text{min}} = -4.0, \quad \eta^{\text{Z}}_{\text{max}} = 4.0$
$t\bar{t}$	10	$\eta^{\text{t}}_{\text{min}} = -2.5, \quad \eta^{\text{t}}_{\text{max}} = 2.5$

Table 3 List of differential distributions for which the theory predictions are computed. The distributions are all differential in the rapidity of the final-state system. For each distribution, the rapidity range and the number of bins are also shown. All the processes have been computed at the LHC energy, i.e. at a center-of-mass energy $\sqrt{s}=13$ TeV.

LHC cross section at $\sqrt{s}=13$ TeV for the W and Z boson production via DY mechanism (1st and 2nd row respectively), top-quark pair production (3rd row), and Higgs boson produced via VBF (4th row). The results are shown for two compressed sets $N_c=50$ (left column) and $N_c=70$ (right column). For each process and each size of the compressed set, we compare the integrated cross section resulting from the prior PDF (with $N_p=1000$ replicas) with the standard (CMC) and GAN-enhanced (GAN) compression methodologies. For comparisons, we also plot the 1-sigma (dark blue) and 2-sigma uncertainty bands of the prior. As a general comment, we observe that the compressed sets resulting from the two different methodologies yield excellent compatibility with the prior.

For the W production, we observe very good agreement between the three different PDF sets at the 1-sigma level with the central value of the GAN-enhanced compressed set being closer to the prior than the standard. This can be observed both for $N_c=50$ and $N_c=70$. Going from $N_c=50$ to $N_c=70$, we notice that there is significant decrease in uncertainty for all three PDF sets. Whilst similar observation can be made for the Z boson production, no reduction in the error bars are seen when going from $N_c=50$ to $N_c=70$. For top quark pair production, there is basically no change between the results with $N_c=50$ and $N_c=70$. Notice that this pattern is the same across the three types of PDF sets. Finally, we turn to the case of the Higgs boson produced via the VBF mode. The same conclusions as for the previous processes still hold, namely, the compressed set from the standard and GAN-enhanced methodology are both equally compatible with the prior. This is valid both at $N_c=50$ and $N_c=70$ where no obvious distinctions can be observed due to the perfect agreement.

While in some cases, compressed sets resulting from the GAN-enhanced methodology are closer to the prior than the compressed sets resulting from the standard approach, these results suggest that integrated cross sections are not suitable quantities to assess the phenomenological implications of the new methodology. Indeed, as mentioned previously, the sensitivity of the integrated cross sections to the parton densities are only restricted to a narrow range of x and Q. In order to assess the impact of the PDFs in the phase space region that does not provide the bulk of the result, in the next section, we look into the differential distributions shown in Table 3.

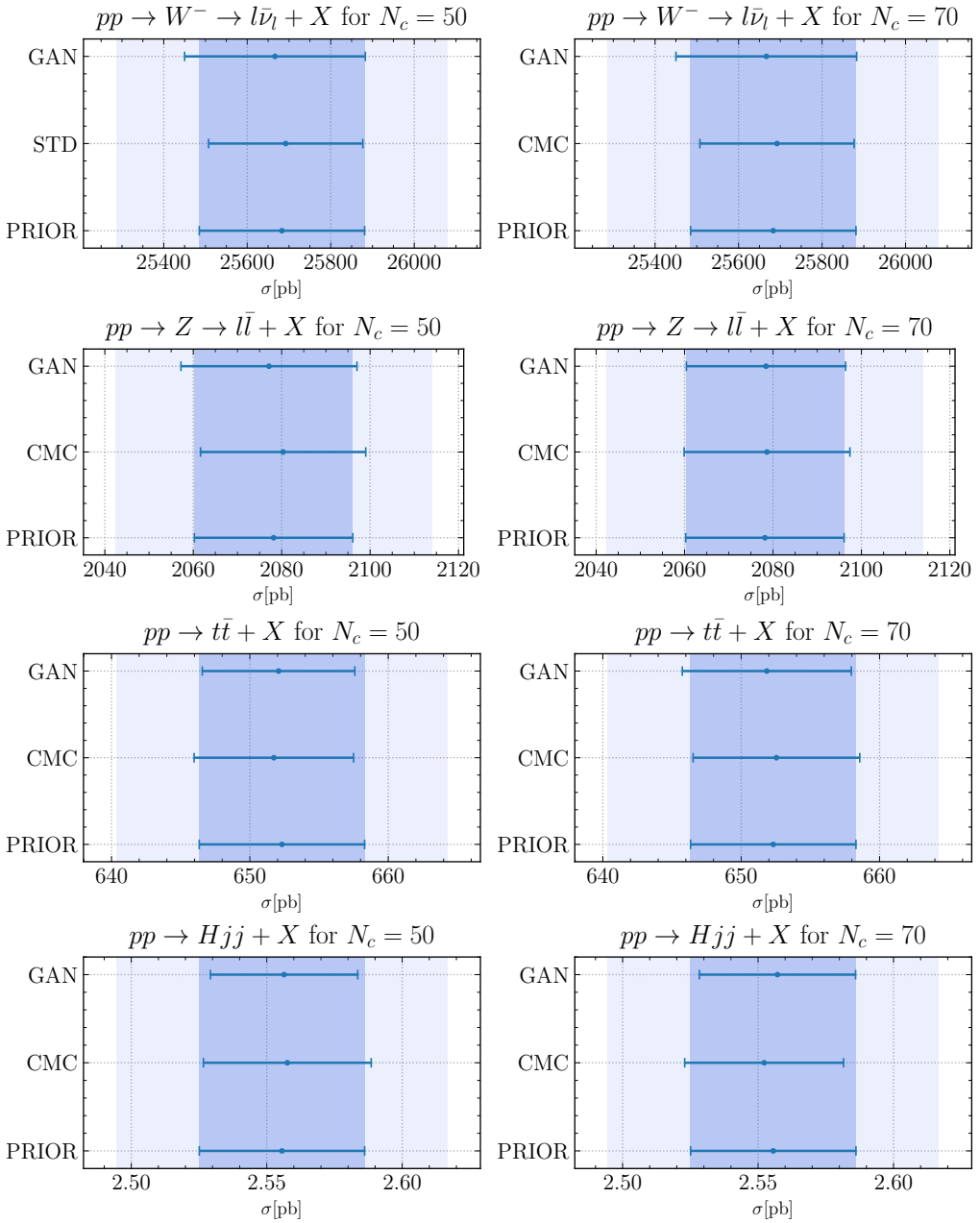


Figure 41 Integrated LHC cross sections at $\sqrt{s}$=13 TeV for W production (1st row), Z production (2nd row), top pair production (3rd row), and Higgs via VBF (4th row). The left column compares the prior with the N_c=50 compressed sets while the right column compares the prior with the N_c=70 compressed sets. Plots produced using PINEAPPL [7] scripts.

Differential distributions

Let us turn to the differential cross sections of the processes discussed in the previous section. For each process, we consider the NLO rapidity distribution with the rapidity range defined as in Table 3. In Fig. 42, we show the results for the W and Z boson production while results for the top pair and Higgs boson production are shown in Fig. 43. For each individual process, we show the results for a compressed set with size $N_c = 50$ (left column) and $N_c = 70$ (right column) using the standard (red line labelled CMC) and GAN-enhanced (orange line labelled GAN) methodology. Common to all plots, the top panel show the absolute differential cross sections with the uncertainty bands computed using seven-point scale variation. The middle panel compares the relative PDF uncertainties between the prior and the compressed Monte Carlo sets. Finally, the bottom panel shows the pull in units of the PDF uncertainty between the result found using the prior Monte Carlo PDF set and the compressed sets generated using the old and new methodology. The pull is defined in terms of theory predictions as [4]

$$\mathcal{P}\left(\Sigma_{c,i}, \Sigma_{p,i}\right) = \frac{\Sigma_{c,i}^{(0)} - \Sigma_{p,i}^{(0)}}{\sqrt{\left(\delta\Sigma_{c,i}\right)^2 + \left(\delta\Sigma_{p,i}\right)^2}} \tag{5.4.2}$$

where $\Sigma_{p,i}$ and $\Sigma_{c,i}$ represent the central values of the theory predictions using the prior and compressed set respectively at the i-th bin (for $i = 1, \cdots, N_{\rm bin}$) and $\delta\Sigma_{p,i}, \delta\Sigma_{c,i}$ the corresponding PDF uncertainties. Let us now inspect each result starting with the W production with a compressed set of size $N_c = 50$ replicas (top-left plot in Fig. 42). Whilst it is difficult to distinguish between the various PDF sets from the absolute plots, noticeable differences can be seen at the relative PDF uncertainties. Indeed, we can see that while discrepancy persists between the prior and the standard compressed set, good agreement can be seen between the prior and the compressed set generated using the GANPDFs methodology. This is more marked for large rapidity bins ($|\eta_W| > 2$). Looking now at the pull, we notice that pull with the GAN-enhanced compressed set is largely below zero while the pull with the standard approach is always positive. For both compression methodologies, the largest pulls are observed for small rapidity bins. Going from $N_c = 50$ to $N_c = 70$ (top-right plot), we can see that the same conclusion still holds but in a more pronounced way, i.e. compressed sets generated using the new methodology is closer to the prior than compressed sets generated using the standard approach. For the Z boson production, both at $N_c = 50$ and $N_c = 70$, the predictions from the standard compression approach is performing as good as the GAN-enhanced methodology. Similar to the W boson case, the largest pulls appear for small rapidity bins for $N_c = 50$ while for $N_c = 70$ the pulls are largely zero.

Let us now turn to the top pair production shown in the first row of Fig. 43. Looking at the relative PDF uncertainties, we observe the same pattern as before, namely, the results from the GANPDFs approach is generally closer to the prior than the results from the standard methodology. For both $N_c = 50$ and $N_c = 70$, the pulls with the standard and GAN approach exhibit inversely opposite behaviour, i.e. whilst the GAN result is closer to the prior in the large rapidity bins, the standard result is closer to the prior in the small rapidity bins. Finally, for the case of Higgs boson production via VBF mechanism, we observe for $N_c = 50$ that the prediction from the standard compression approach is performing

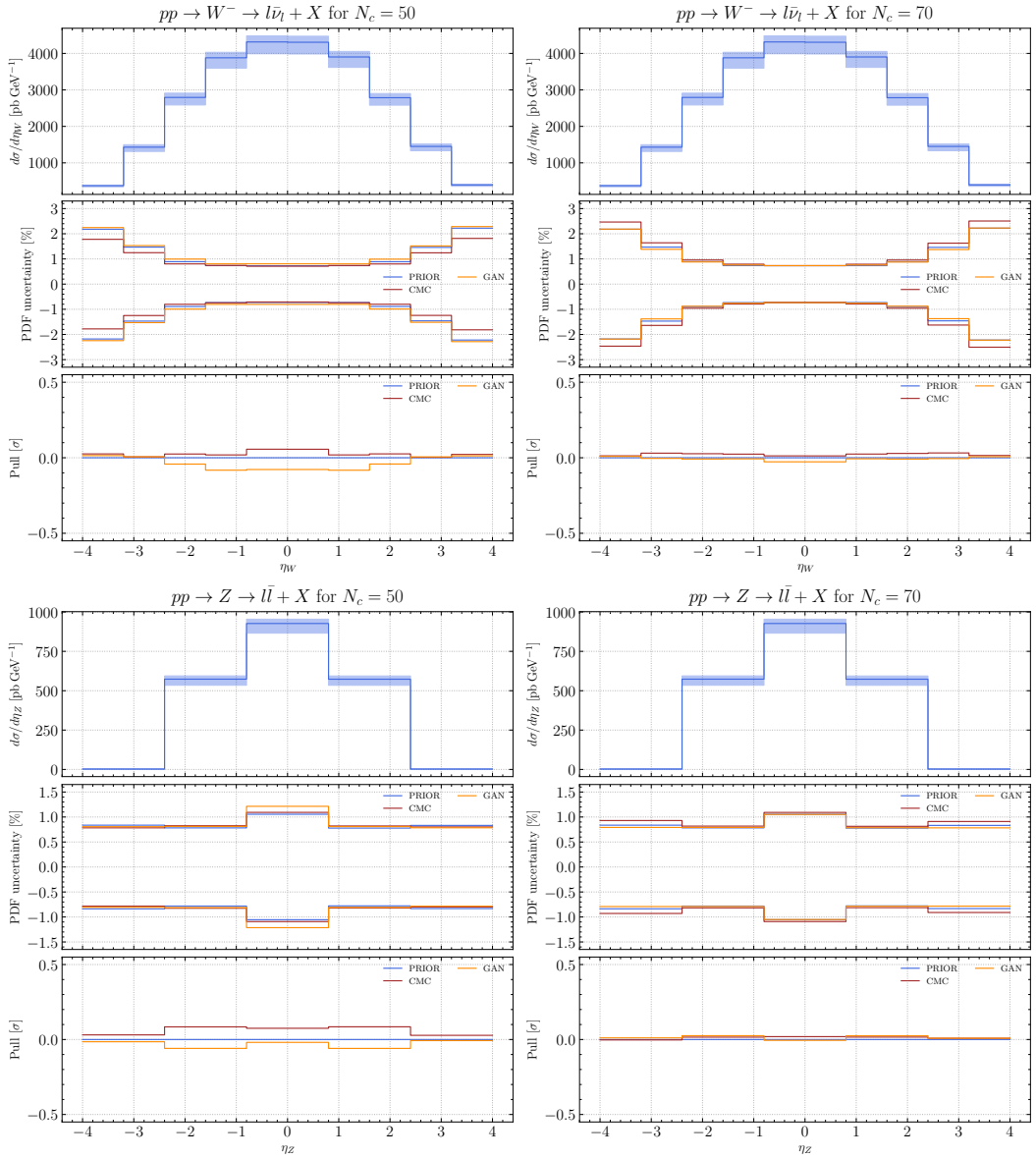


Figure 42 Differential distributions in rapidity η_X (where $X = W, Z$) during the production of a W (top) and Z bosons (bottom). The heading in each plot represents the decay mode of the corresponding weak boson. As previously, the results are shown for $N_c = 50$ (left) and $N_c = 70$ (right). The top panels show the absolute PDFs with the seven-point scale variation uncertainties, the middle panel show the relative uncertainties for all PDF sets, and the bottom panels show the pull defined in Eq. (5.4.2) between the prior and the compressed sets generated using the standard (red) and GAN-enhanced approach (orange). Plots produced using PINEAPPL [7] scripts.

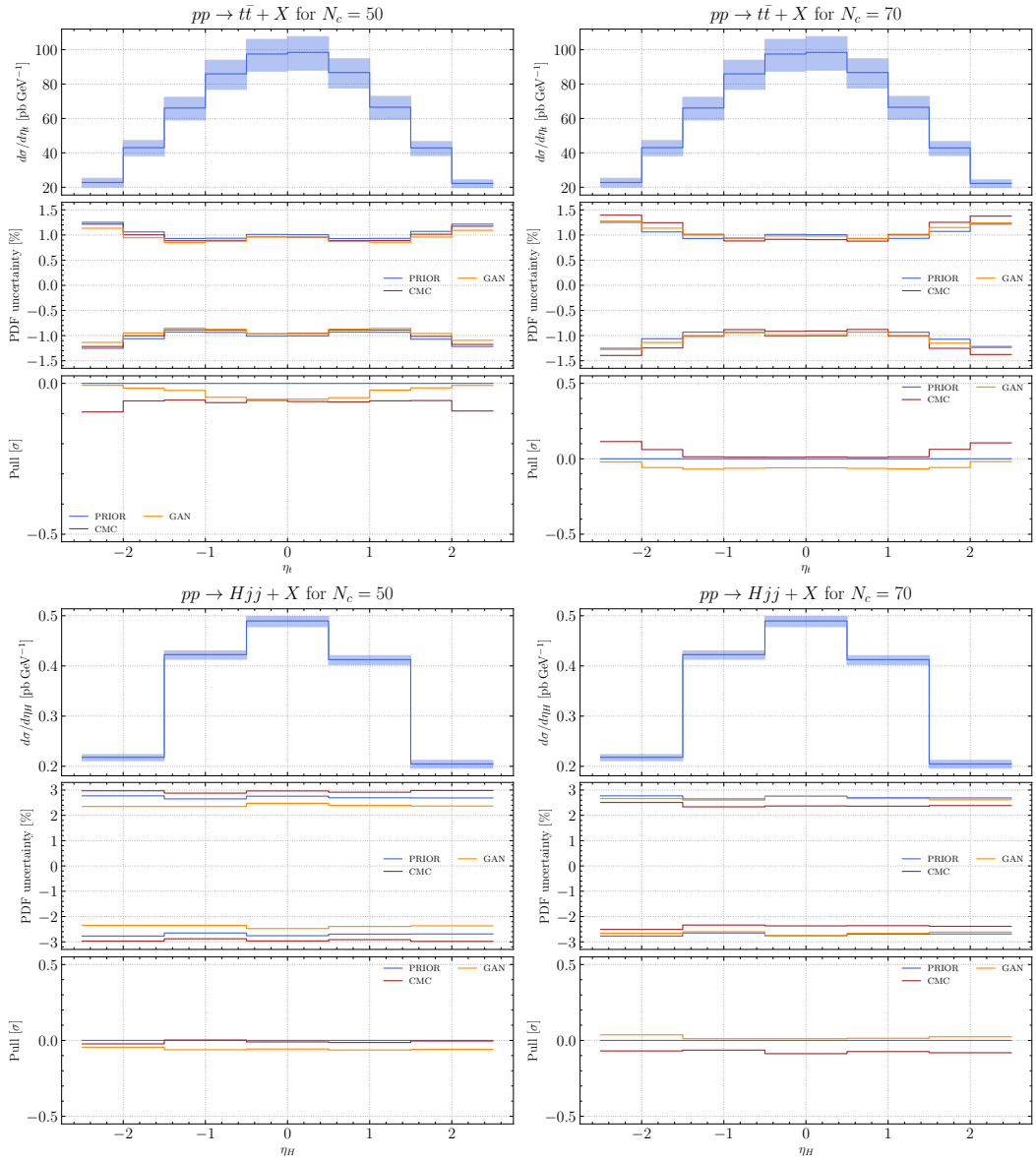


Figure 43 Same as Fig. 42 but for top-quark pair production (top) and Higgs boson production via VBF (bottom). Similar as before, the results are shown for $N_c = 50$ (left) and $N_c = 70$ (right)Plots produced using PINEAPPL [7] scripts.

as good as the prediction from the GANPDFS approach both at the level of the pull and relative PDF uncertainties. For $N_c = 70$, we see slightly better prediction from the new methodology, especially in the small rapidity bins.

Gaussianity vs. Accuracy of the predictions

One of the advantages of representing PDF errors as an ensemble of Monte Carlo replicas is the fact that one does not have to assume Gaussian approximation. Non-Gaussian features are important when experimental measurements are limited and PDF uncertainties are driven by theoretical constraints. However, in the case where experimental measurements are abundant and the distribution is dominated by Gaussian properties (such as mean and central value), it is convenient to use the Hessian representation of errors. In the particular context of compression of Monte Carlo replicas, situations might happen where the Gaussianity deteriorate in favour of the accuracy of the compression, or reciprocally, the accuracy of the compression is prioritized over the Gaussian behaviour. In the following, we quantify the Gaussianity of the new GAN-enhanced compressed results when compared to the prior and the results from the standard compression approach. Here, we follow the approach introduced in Ref. [383] that consists of two steps: first, transform the Monte Carlo samples into continuous probability distributions using the Kernel Density Estimate (KDE), and second compare the probability distribution of the prior with the Gaussian approximation and with the compressed sets resulting from the two methodologies. The former is done by computing the Kullback-Leibler (KL) divergence. For two continuous probability distributions, the KL-divergence is defined as

$$\mathrm{KL}(P|C) = \int_{-\infty}^{\infty} dx \left(P(x) \frac{\ln P(x)}{\ln C(x)} \right), \tag{5.4.3}$$

where $P(x)$ and $C(x)$ respectively represents the probability distribution of the prior and compressed set. We emphasize that here we use the same prior and compressed sets investigated in the previous sections. For the computation of the cross sections, in addition to the integrated and rapidity distributions studied earlier, we also include azimuthal, missing energy, invariant mass, and transverse momentum distributions. Summary of all the observables included in the following studies are shown in Table 4. For the computation of the cross sections, we use the SMPDF [384] package.

In Fig. 44, we show, for each class of observable, the KL distance between the prior Monte Carlo PDF with $N_p = 1000$ replicas and (a) Gaussian sets that have the same central value and standard deviation as the prior (b) compressed sets from the old (green) and new (orange) methodology for $N_c = 50$ replicas. For reference, the Hessian representation of the prior (MCH) with $N = 50$ eigenvectors is also shown in blue. We can see on the x-axis the KL divergence between the prior and its Gaussian approximation while the y-axis represents the KL distance between the prior and the reduced sets (Hessian and MC compressed sets). Apart from the Higgs boson production, all the points are clustered on the left-hand side of the plots, exhibiting Gaussian behaviours. It indeed appears that a significant fraction of the Higgs boson production cross sections are more sensitive to non-Gaussian behaviour. By inspection of the Hessian results, it is clear that the loss of accuracy in the reduction is much bigger than the non-Gaussianity. This is more apparent in the case of W and Z boson production where the KL divergence fluctuates a lot. This suggests that a Hessian representation with $N = 50$ eigenvectors is not ideal enough to capture the underlying probability distribution of the prior when the distribution is not Gaussian. In order to get a more accurate representation of the underlying probability distribution, a moderate number of eigenvectors are required.

Process (via pp collision)	**Diff. In**	N_{bin}	**Range**
$W^- + X \to \ell\bar{\nu}_\ell + X$	η^W	10	$\eta^W_{\text{min}} = -4.0, \quad \eta^W_{\text{max}} = 4.0$
	η^ℓ	10	$\eta^\ell_{\text{min}} = -2.5, \quad \eta^\ell_{\text{max}} = 2.5$
	ϕ	10	$\phi_{\text{min}} = -1.0, \quad \phi_{\text{max}} = 1.0$
	$p^{T,W}$	10	$p^{T,W}_{\text{min}} = 0, \quad p^{T,W}_{\text{max}} = 200$ (GeV)
	$p^{T,\ell}$	10	$p^{T,\ell}_{\text{min}} = 0, \quad p^{T,\ell}_{\text{max}} = 200$ (GeV)
	M^T	10	$M^T_{\text{min}} = 0, \quad M^T_{\text{max}} = 200$ (GeV)
	E^{miss}	10	$E^{\text{miss}}_{\text{min}} = 0, \quad E^{\text{miss}}_{\text{max}} = 200$ (GeV)
$Z + X \to \ell\bar{\ell} + X$	η^Z	5	$\eta^{\text{Z}}_{\text{min}} = -4.0, \quad \eta^{\text{W}}_{\text{max}} = 4.0$
	η^ℓ	10	$\eta^\ell_{\text{min}} = -2.5, \quad \eta^\ell_{\text{max}} = 2.5$
	$\eta^{\bar{\ell}}$	10	$\eta^{\bar{\ell}}_{\text{min}} = -2.5, \quad \eta^{\bar{\ell}}_{\text{max}} = 2.5$
	$p^{T,\ell}$	10	$p^{T,\ell}_{\text{min}} = 0, \quad p^{T,\ell}_{\text{max}} = 200$ (GeV)
	$p^{T,\bar{\ell}}$	10	$p^{T,\bar{\ell}}_{\text{min}} = 0, \quad p^{T,\bar{\ell}}_{\text{max}} = 200$ (GeV)
	$p^{T,Z}$	10	$p^{T,Z}_{\text{min}} = 0, \quad p^{T,Z}_{\text{max}} = 200$ (GeV)
	$p^{T,\ell\bar{\ell}}$	10	$p^{T,\ell\bar{\ell}}_{\text{min}} = 0, \quad p^{T,\ell\bar{\ell}}_{\text{max}} = 200$ (GeV)
	$M^{\ell\bar{\ell}}$	10	$M^{\ell\bar{\ell}}_{\text{min}} = 0, \quad M^{\ell\bar{\ell}}_{\text{max}} = 130$ (GeV)
$t\bar{t} + X$	η^t	10	$\eta^t_{\text{min}} = -2.5, \quad \eta^t_{\text{max}} = 2.5$
	$\eta^{\bar{t}}$	10	$\eta^{\bar{t}}_{\text{min}} = -2.5, \quad \eta^{\bar{t}}_{\text{max}} = 2.5$
	$\eta^{t\bar{t}}$	12	$\eta^{t\bar{t}}_{\text{min}} = -3.0, \quad \eta^{t\bar{t}}_{\text{max}} = 3.0$
	$p^{T,t}$	40	$p^{T,t}_{\text{min}} = 40, \quad p^{T,t}_{\text{max}} = 400$ (GeV)
	$p^{T,\bar{t}}$	10	$p^{T,\bar{t}}_{\text{min}} = 40, \quad p^{T,\bar{t}}_{\text{max}} = 400$ (GeV)
	$p^{T,t\bar{t}}$	10	$p^{T,t\bar{t}}_{\text{min}} = 20, \quad p^{T,t\bar{t}}_{\text{max}} = 200$ (GeV)
	$M^{T,t\bar{t}}$	10	$M^{T,t\bar{t}}_{\text{min}} = 300, \quad M^{T,t\bar{t}}_{\text{max}} = 1000$ (GeV)
$gg + X \to \text{H} + X$	η^{H}	10	$\eta^{\text{H}}_{\text{min}} = -2.5, \quad \eta^{\text{H}}_{\text{max}} = 2.5$
	$p^{T,\text{H}}$	10	$p^{T,\text{H}}_{\text{min}} = 0, \quad p^{T,\text{H}}_{\text{max}} = 200$ (GeV)
$\text{H}t\bar{t} + X$	η^{H}	10	$\eta^{\text{H}}_{\text{min}} = -2.5, \quad \eta^{\text{H}}_{\text{max}} = 2.5$
	$p^{T,\text{H}}$	10	$p^{T,\text{H}}_{\text{min}} = 0, \quad p^{T,\text{H}}_{\text{max}} = 200$ (GeV)
VBF: $\text{H}jj + X$	η^{H}	10	$\eta^{\text{H}}_{\text{min}} = -2.5, \quad \eta^{\text{H}}_{\text{max}} = 2.5$
	$p^{T,\text{H}}$	10	$p^{T,\text{H}}_{\text{min}} = 0, \quad p^{T,\text{H}}_{\text{max}} = 200$ (GeV)
$\text{H}W + X \to \text{H}\ell\bar{\nu}_\ell + X$	η^{H}	10	$\eta^{\text{H}}_{\text{min}} = -2.5, \quad \eta^{\text{H}}_{\text{max}} = 2.5$
	$p^{T,\text{H}}$	10	$p^{T,\text{H}}_{\text{min}} = 0, \quad p^{T,\text{H}}_{\text{max}} = 200$ (GeV)
$\text{H}Z + X \to \text{H}\ell\bar{\ell} + X$	η^{H}	10	$\eta^{\text{H}}_{\text{min}} = -2.5, \quad \eta^{\text{H}}_{\text{max}} = 2.5$
	$p^{T,\text{H}}$	10	$p^{T,\text{H}}_{\text{min}} = 0, \quad p^{T,\text{H}}_{\text{max}} = 200$ (GeV)

Table 4 List of differential distributions for which the theory predictions for the Gaussian studies are computed. The distributions are all differential in the azimuthal angle, invariant mass, missing energy, and transverse momentum.

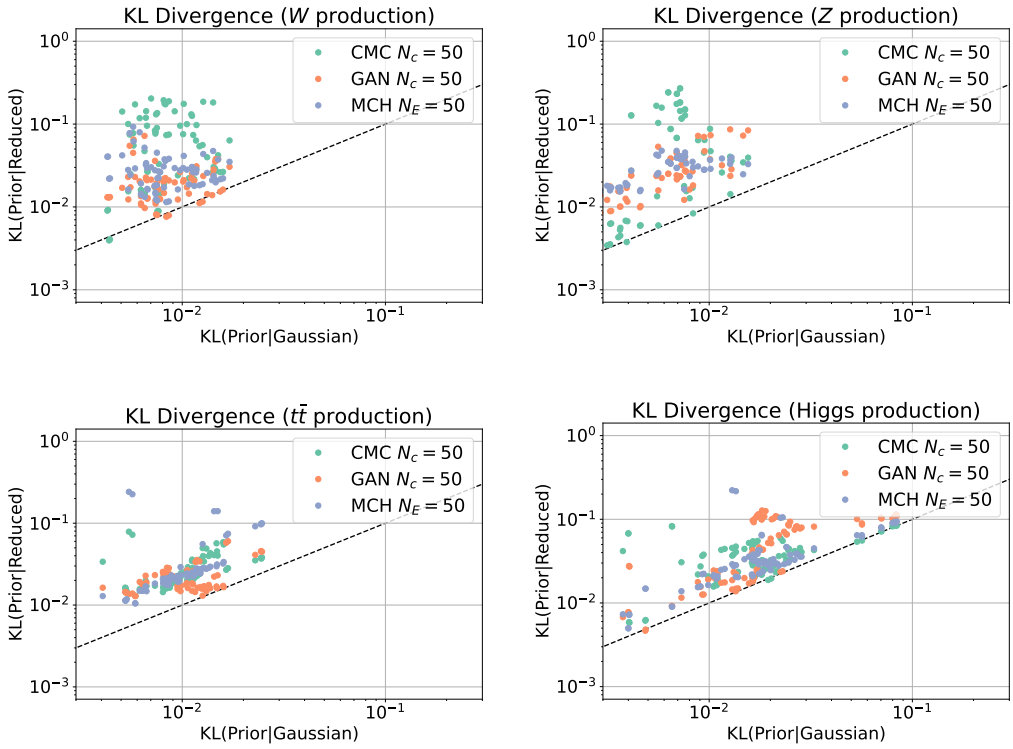


Figure 44 The KL distance as expressed in Eq. (5.4.3) between the prior Monte Carlo PDF replicas and the compressed sets resulting from the standard (green) and GAN-enhanced (orange) approach. For reference, the KL distance between the prior and its Hessian representation with $N=50$ eigenvectors (blue) is also shown. For each class of observables, the various production modes are detailed in Table 4.

Let us now compare the KL distances w.r.t the prior of the compressed sets using the standard and GANPDFS methodology. While the results from the two methodologies exhibit the same degree of Gaussianity, one can clearly see that the KL divergences between the prior and the GAN-enhanced compressed set are generally lower. This can be seen in all the processes except for the Z boson production where a few green points lie below the orange ones. From this observation, we conclude that for all the observables we are considering, it is always more advantageous to use the GAN-enhanced compressed set.

In Fig. 45 where we show results for $N_c=70$ replicas, we can immediately see that the fluctuations mainly present in the W and Z boson productions have disappeared. Now, the points are mainly clustered on the diagonal. For the particular case of the Hessian conversion, we can see that while the Gaussian features are still present, the accuracy of the reduced set has improved significantly. By comparing the compressed sets from the two methodologies, the same conclusion as for $N_c=50$ still holds, namely, the compressed set resulting from the GAN-enhanced methodology outperforms the standard approach both in terms of preserving the Gaussianity and the accuracy during the compression.

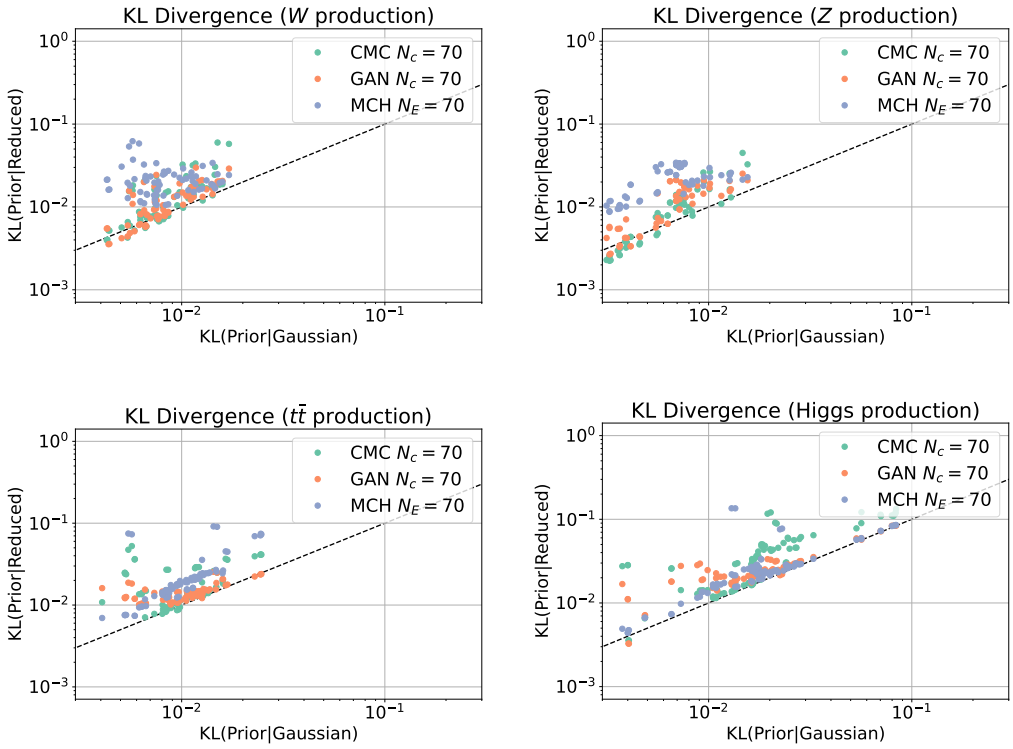


Figure 45 Same as Fig. 44 but for compressed sets with $N_c = 70$ replicas.

These results confirm at the cross section levels (both for integrated and differential cross sections) the conclusions that were drawn at the level of the parton densities described in Sec. (5.3). Indeed, these results provide strong indications that a compression methodology based on generative models (here in particular GAN) outperforms the standard compression approach.

5.5 GANs for finite size effects

In the previous sections, we only focused on the roles that generative-based adversarial models play in the improvement of the compression methodology. It was shown that, for all the statistical estimators we considered (including lower and higher moments, various distance metrics, and correlations), the new generative-based approach outperforms the standard compression. However, one main question persists as to whether the generative based adversarial PDF model can be used to bypass the fitting of a large number of Monte Carlo PDF replicas. As mentioned before, in order to converge to the asymptotic the PDF is tending to, a large number of Monte Carlo replicas is required. Despite the fact that the new N3FIT methodology [385] greatly improves over the previous NNFIT, one cannot generate as many PDF replicas as one wants. As a matter of fact, even with the

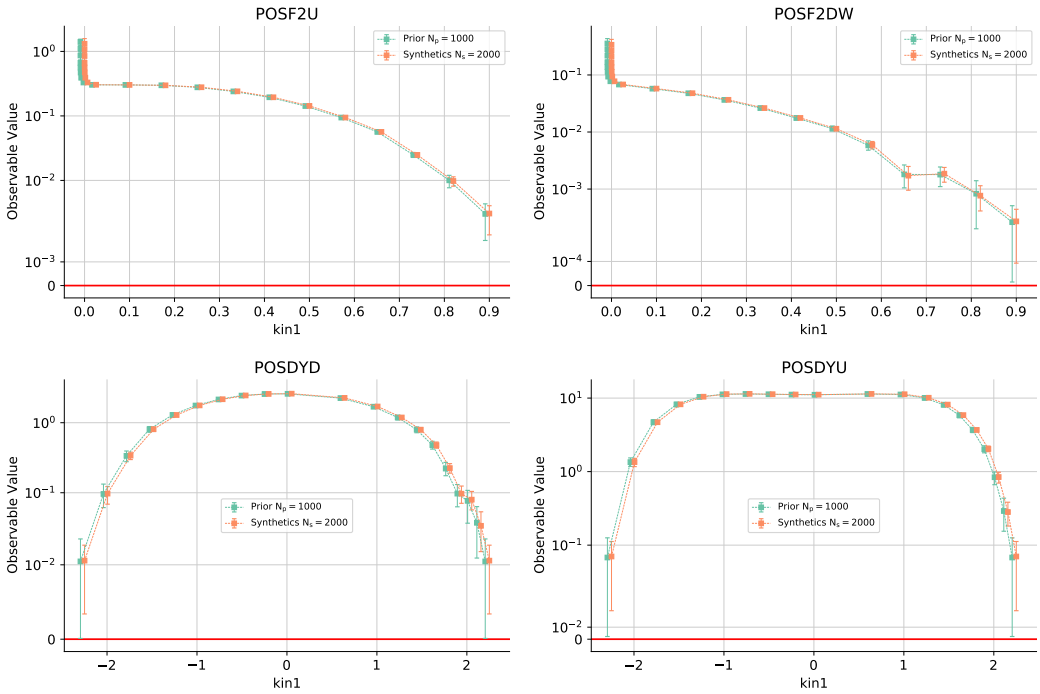


Figure 46 Positivity constraints for the prior (green) and synthetic (orange) Monte Carlo PDF replicas. The constraints correspond to the positivity of a few selected observables from Eq. (14) of Ref. [8], namely $F_2^u(x,Q^2)$, $F_2^d(x,Q^2)$, $d\sigma^2_{d\bar{d}}/dM^2dy(x_1,x_2,Q^2)$, and $d\sigma^2_{u\bar{u}}/dM^2dy(x_1,x_2,Q^2)$.

current N3FIT methodology, generating more than 1000 replicas is still computationally very expensive. If the answer to the previous section turns out to be true, then one can generate a fit with $N_0=1000$ and use the GANPDFS to generate a fit with say $N_E=5000$ synthetic replicas that contains the same statistical properties as a real fit.

Before proceeding to address the problem at hand, let us first have a closer look at the synthetic replicas generated from the GAN. First, we verify that despite the fact that physical constraints such as sum rules and positivity of the PDFs are not enforced when constructing the synthetic set, they are automatically inferred during the generation. In Fig. 46, we compare the positivity of the real and synthetic samples for four observables. Here, the prior is the same as before (containing $N_p=1000$ replicas) while the synthetic set contains only GAN-generated sample with $N_s=2000$ replicas. We see from Fig. 46 that not only the positivity of the observables are satisfied but also the the results are closely similar to the prior. Similar observation can be made by looking at the results of the sum rules shown in Table 5.

Now, in an attempt to address the posed question above, we assess the goodness of the generated samples by comparing the results to a real fit. Such an evaluation could be done by considering: two disjoint sets of N fitted replicas (say S_1 and S_2), and a set of synthetic replicas (S_3) with the same size but determined from GANPDFS using a

Prior	mean	standard deviation
momentum	0.9968	7.315×10^{-4}
u_v	1.985	3.122×10^{-2}
d_v	0.9888	3.764×10^{-2}
s_v	3.249×10^{-3}	3.547×10^{-2}
Synthetic	mean	standard deviation
momentum	0.9954	1.907×10^{-3}
u_v	1.992	3.788×10^{-2}
d_v	0.9956	3.796×10^{-2}
s_v	2.073×10^{-4}	4.833×10^{-2}

Table 5 Table comparing the values of the sum rules between the real Monte Carlo replicas (prior) with $N_p = 1000$ and the synthetic replicas with $N_s = 2000$ generated using the GANPDFS . For each replicas sample, the results are shown for the central value and the standard deviation for the various valence quarks.

starting set N_0 (with $N_0 \ll N$) of fitted replicas. Then, using various statistical estimators and distance metrics, one can measure the distance between S_1 and S_3 and compare the result with the measure of the distance between S_1 and S_2. Ideally, the uncertainty from such a measurement should be computed by repeating the exercise several time, each time picking different sets of replicas. However, not only this is computationally very expensive but also time consuming. As an alternative approach for the estimation of the uncertainty from such a study, we opt for two resampling strategies: the *(delete-1) jackknifing* [386–389] and the logically similar (non-parametric) *bootstrapping* [390–392]. Both methods provide reliable estimates of the dispersion when the statistical model is not adequately known. More complete description of these techniques may be found in the following references [393–397], here we summarize the methods in the context of parton density functions:

Delete-1 Jackknife

As the name suggests, the delete-1 Jackknife resampling strategy consists in deleting one replicas from the entire PDF set. The remaining $(N-1)$-sized samples are called (delete-1) Jackknife samples. Hence, for a prior PDF with N-sized replicas, the i-th Jackknife sample is given by the following set:

$$\mathrm{S}_k^{[i]} = \left\{ \mathrm{S}_k^{(1)}, \cdots, \mathrm{S}_k^{(i-1)}, \mathrm{S}_k^{(i+1)}, \cdots, \mathrm{S}_k^{(N)} \right\}. \tag{5.5.1}$$

We can then define the Jackknife replicate $\mathrm{D}^{(i)}$ as the distance between two Jackknife samples $\mathrm{S}_1^{[i]}$ and $\mathrm{S}_2^{[i]}$ (or $\mathrm{S}_3^{[i]}$ for the synthetics). The delete-1 Jackknife standard error can then be written as:

$$\mathrm{SE}_{\text{Jack-1}} = \sqrt{\frac{N-1}{N}\sum_{i=1}^{N}\left(\mathrm{D}^{(i)} - \mathrm{D}^{(\bullet)}\right)^2}, \tag{5.5.2}$$

where $\mathrm{D}^{(\bullet)} = \sum_{i=1}^{N} \mathrm{D}^{(i)}/N$ denotes the empirical average of the Jackknife replicates. Notice that in Eq. (5.5.2), there are N unique $\mathrm{D}^{(i)}$ Jackknife samples.

Bootstrap

The bootstrap method also belongs to a class of non-parametric resampling strategies for estimating distribution of an estimator. As opposed to the delete-1 Jackknife method, the bootstrap resampling is done through independently sampling the original sample for a certain number of times with replacement. In terms of resources, this results in a much more intensive computation.

As just mentioned above, the bootstrapping strategy consists in randomly sampling with replacement the Monte Carlo PDF replicas, with each sample set being the identical size as the original (in our case, N). The larger is the size of the random sampling or the *bootstrap size*, the more accurate our uncertainties are. For a given measurement of the distance D, the bootstrap statistical error is given by the following expression:

$$\mathrm{SE}_{\mathrm{B}} = \sqrt{\frac{1}{N_B - 1}\sum_{i=1}^{N_B}\left(\mathrm{D}^{(i)} - \mathrm{D}^{(\bullet)}\right)^2}, \tag{5.5.3}$$

where N_B and $\mathrm{D}^{(i)}$ respectively denote the bootstrap size and the distance between the i-th resampled S_1 and S_2 (or S_3 for the synthetics). The empirical average over the bootstrap sampling is given by $\mathrm{D}^{(\bullet)} = \sum_{i=1}^{N_B} \mathrm{D}^{(i)}/N_B$.

In the following studies, we consider two disjoint real fits generated using the NNFIT methodology, each with size $N=500$ replicas. The synthetic PDF replicas, that is of the same size as the real fits, were generated from a prior of $N_0=100$ replicas using the GAN methodology. The results from the two resampling methodologies are shown in Fig. 47. The histograms are constructed by simply measuring the distance between the two real fits S_1-S_2 (green) and the real and synthetic fits S_1-S_3 (orange) based on the statistical estimators we have introduced in Sec. (5.1). The error bars are computed either using the Jackknife resampling (see Eq. (5.5.2)) or the bootstrap method (see Eq. (5.5.3)). By inspecting the Jackknife results (left), we first observe that for some of the statistical estimators we considered, the synthetic PDF replicas yield lower values of the distances. Second, we see that all the error bars for the real and synthetic fits overlap except for the standard deviation. The results from the bootstrap resampling confirm the observations from the delete-1 Jackknife method, i.e. the error bars for the real and synthetic fits overlap in all but a few statistical estimators such as the standard deviation and Kurtosis.

The above results suggest that the probability distributions of the two sets of PDFs S_2 and S_3 are closely similar, indicating that the GAN could potentially be used to bypass the

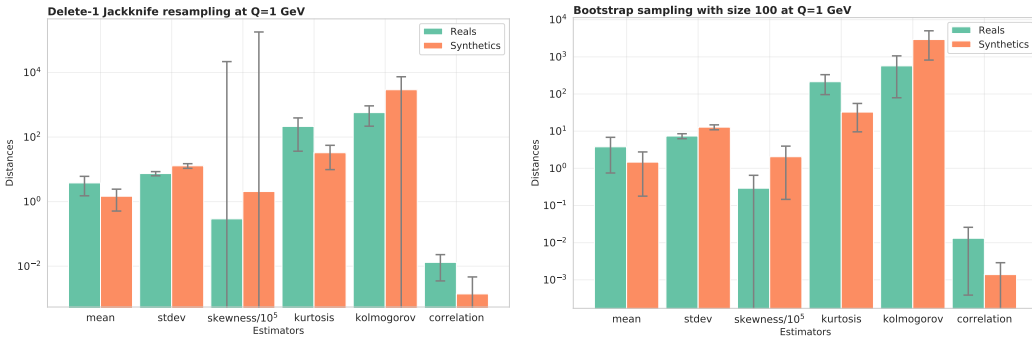


Figure 47 Comparison of the real and synthetic fits for various statistical estimators. The histograms are constructed by computing the distance between the two subsets of real replicas S_1-S_2 (green) and the distance between the synthetic replicas and one of the original subsets S_1-S_3 (orange). The error bars are computed by performing the delete-1 Jackknife (left) and the bootstrap (right). In the bootstrap resampling, the results have been evaluated by performing $N_B = 100$.

need of fitting a large sample of Monte Carlo PDF replicas. This is at least the case for a smaller sample of Monte Carlo replicas (here $N = 500$). We therefore hypothesize that this would also be the case for larger sets. If true, the present framework could be used in any of the ways proposed at the beginning of this section.

5.6 GANPDFS for NNPDF4.0 sets

It was shown in the previous section that the generative-based compression methodology significantly outperforms the standard approach when tested on a PDF set generated using the NNPDF3.1 methodology which besides of exhibiting non-Gaussian behaviour contains fluctuations. In the following section, we test the GANPDFS approach on the new fitting machinery implemented in NNPDF4.0 [385]. Hereinafter, the new settings are as follows: we consider as prior a Monte Carlo PDF with $N_p = 1000$ replicas generated using the NNPDF4.0 methodology, which is then enhanced using the GANPDFS to generate $N_s = 2000$ leading to an enhanced set with a total of $N_E = 3000$ replicas. Notice that the following analyses were performed before the official releases of the NNPDF4.0 sets, hence slight differences are present w.r.t to the PDF sets available on LHAPDF [5]. As before, both the compression using the standard and the GAN-enhanced approach are handled by the PYCOMPRESSOR [360, 370] package.

In Fig. 48 we show the same error function plots as in Fig. 37. Similar to the previous case, one can see that both the old and new methodologies still outperform by a significant amount any random selection. Whilst the value of the total error function of the GAN-enhanced methodology is always smaller than the standard approach, looking at the individual contributions in Fig. 48 there are some instances in which the standard yield smaller errors. This is, for example, the case for the skewness estimator. In addition, it seems that the difference between the orange and green points are no longer that significant. By inspection of the correlation plots, one can also observe similar patterns.

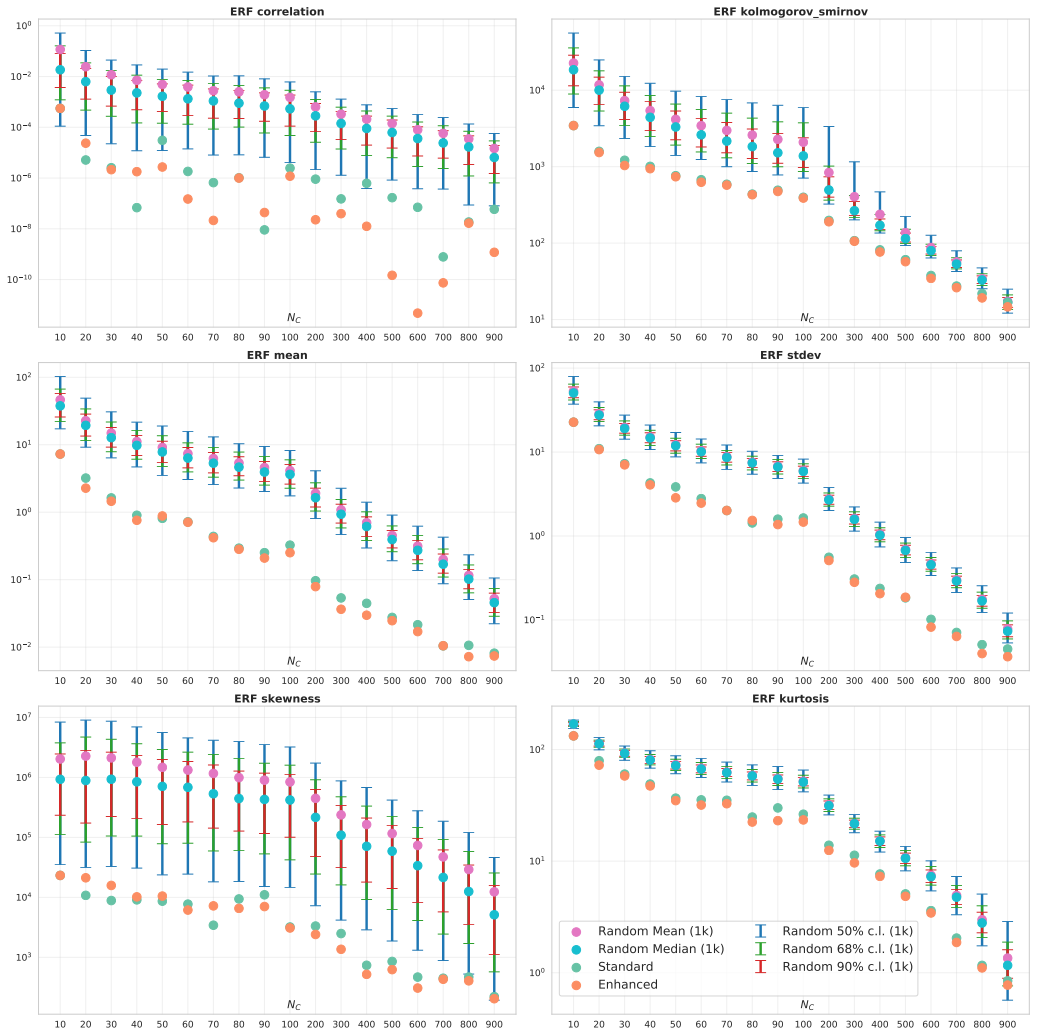


Figure 48 Same as Fig. 37 but using as a prior a Monte Carlo PDF set generated using the NNPDF4.0 methodology.

The differences in correlation matrix are shown in Fig. 49 where, as before, the results are shown for various sizes of compressed set (from top to bottom: $N_c=50$, $N_c=70$, and $N_c=100$) using the standard (left) and GAN-enhanced approach (right). While it is clear that the differences in correlation matrix between the prior and the GAN-enhanced compressed sets are smaller (lighter) than the differences in correlation matrix between the prior and the standard compressed sets, it is also evident that the difference between the left and right plots are no longer that significant.

In order to add more clarification, we can look at the equivalence and disparity plots similar to Fig. 38 and Fig. 34 respectively. The results where the prior used was generated using the NNPDF4.0 are shown in Fig. 50. On the left plot, we compare the performance

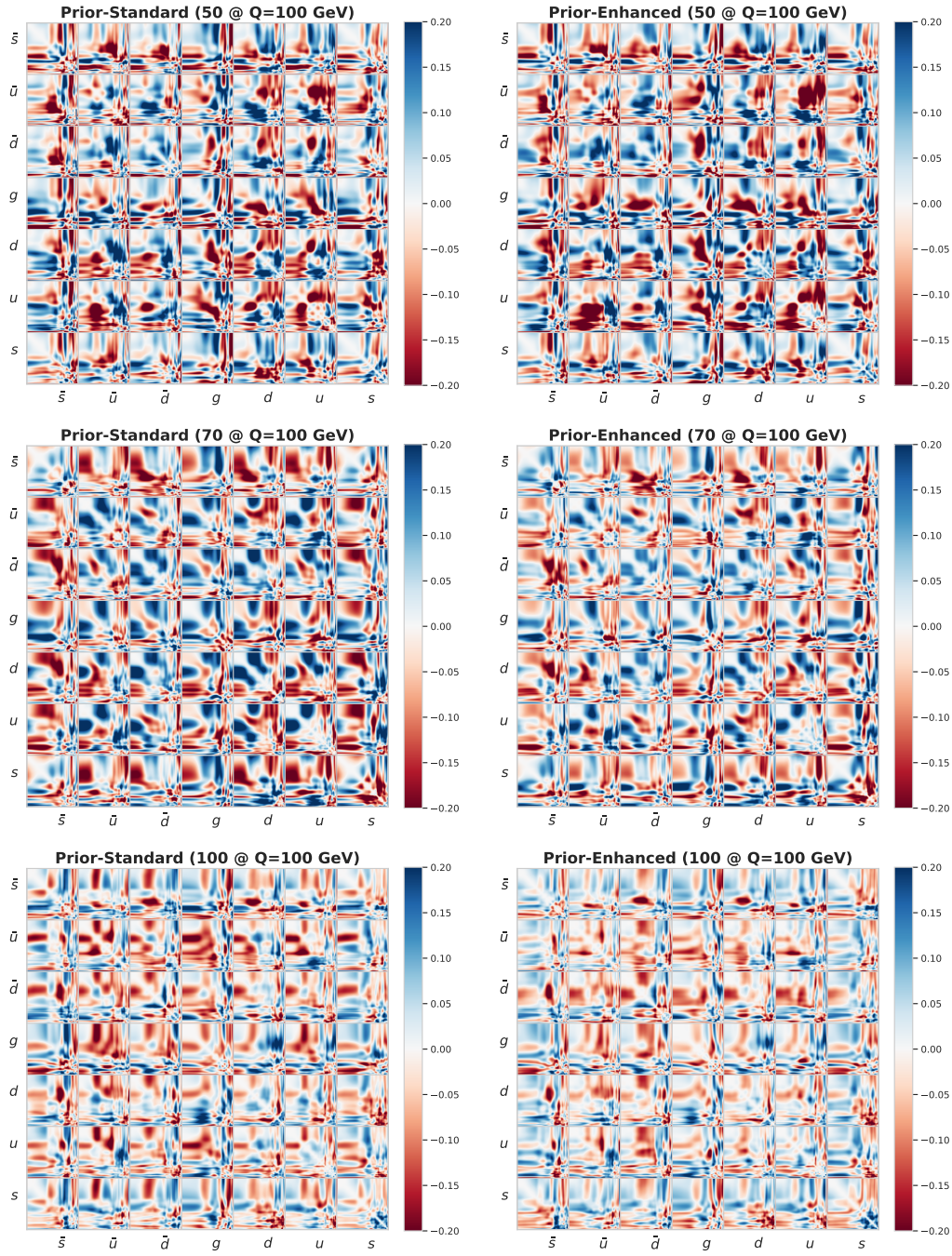


Figure 49 Same as Fig. 40 but using as a prior a Monte Carlo PDF set generated using the NNPDF4.0 methodology.

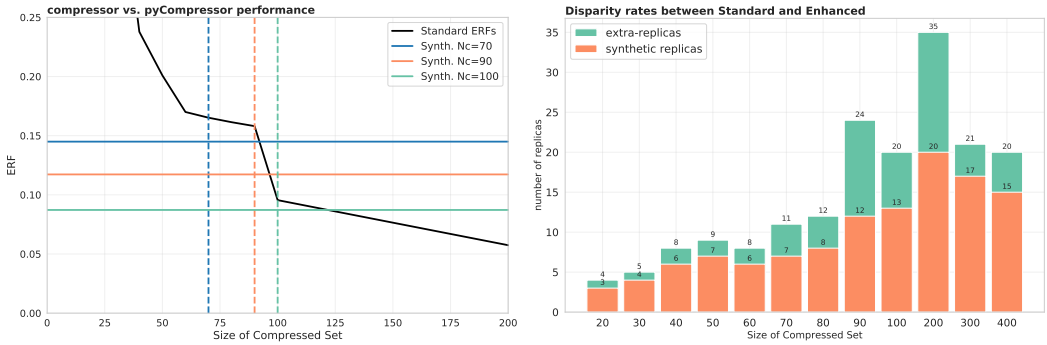


Figure 50 The left plot compares the performance of the generative-based compression algorithm and the standard approach. The plot on the right represents the number of disjoint replicas that are present in the GAN-enhanced set but not in the standard. These plots are resepectively similar to Fig. 38 and Fig. 34 with the difference that the prior was generated using the NNPDF4.0 methodology.

of the generative-based compression methodology with the standard approach for various sizes of the compressed set. The description of each line is understood from Fig. 38. In contrast to the previous results, now a GAN-enhanced compressed set with $N_c = 90$ is equivalent to the standard compressed set with just under $N_c = 100$; similarly, a GAN-enhanced compressed set with $N_c = 100$ is only equivalent to a standard compressed set with $N_c \sim 125$ replicas. These comparisons clearly demonstrate that, in the case of NNPDF4.0 set, the results from the generative-based compression model are not that notably different from the results from the standard compression. This is consistent with the disparity plot where both the number of disjoint replicas (defined to be the number of replicas that are present in the GAN-enhanced but not in standard) and the number of synthetic replicas ending up in the final compressed set are small.

The fact that the generative-based compression is not significantly impressive when compared to the standard methodology for NNPDF4.0 Monte Carlo PDF set can be attributed to the quality of the fit. Indeed, the more stable and Gaussian are the Monte Carlo PDF replicas, the less synthetic replicas are needed. The degree of stability and Gaussianity can be studied further by comparing the central value and the standard deviation of each probability distribution at the level of the cross sections with respectively the median and the minimum 68% confidence interval [359]. The deviation of minimum 68% confidence interval from the standard deviation is a measure of the presence of outliers while the deviation of the median from the central value is a measure of the asymmetry of the distribution. The deviations from the central value and standard deviation are defined according to Ref. [359] as follows

$$\Delta_\mu \equiv \frac{\text{Median} - \mu}{\Sigma}, \qquad \Delta_\Sigma \equiv \frac{\mathcal{R} - \Sigma}{\Sigma}, \tag{5.6.1}$$

where μ and Σ respectively denotes the central value and standard deviation respectively, and the minimum 68% confidence interval $\mathcal{R}$ is defined as

$$\mathcal{R} \equiv \frac{1}{2}\min\left[x_{\min}, x_{\max}\right] \qquad \text{where} \qquad \int_{x_{\min}}^{x_{\max}} dx\, \mathcal{P}(x) \equiv 0.683. \tag{5.6.2}$$

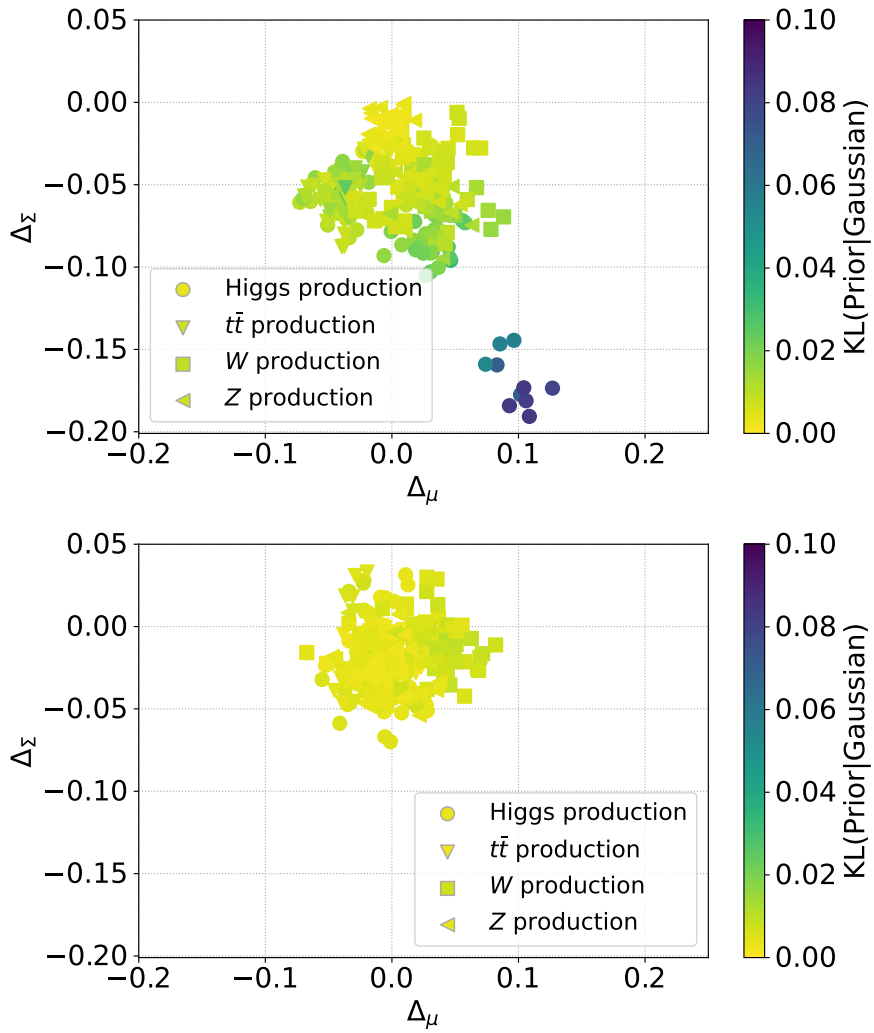


Figure 51 Measure of the degree of Gaussianity for all the differential distributions listed in Table 4. The x-axis represents the shifts between the central value and the median as represented in Eq. (5.6.1) while the y-axis represents the shift between the standard deviation and the 68% interval as represented in Eq. (5.6.1). The color map represents the KL divergence between the priors (NNPDF3.1 top and NNPDF4.0 below) and their respective Gaussian approximations.

It is therefore clear from Eq. (5.6.1) that in the case of a Gaussian in the limit of infinite sample size, both deviations vanish. In the presence of outliers, instead, $R \leq \Sigma$ and Δ_Σ is negative. Fig. 51 plots the two deviations for a NNPDF3.1 (top) and NNPDF4.0 (bottom) PDF set for all the differential distributions listed in Table 4. The shifts between the central value and the median are plotted on the x-axis while the shifts between the standard deviation and the minimum of the 68% confidence interval are plotted on the

y-axis. The density of colors represent the different values of KL divergence between the priors (either NNPDF3.1 or NNPDF4.0) and their Gaussian approximations. Looking at the results using the NNPDF3.1 PDF set, one can see that both the shifts in median and standard deviation are strongly correlated to the degree of non-Gaussianity. Indeed, for a few fractions of the observables, the shifts and the non-Gaussianity are significantly large. This mainly happens for the production of Higgs boson, consistent with Fig. 45. Looking at the results using the NNPDF4.0 instead, we see that not only the points are mainly clustered around the center $(0, 0)$ but also the KL distances are very small. This indicates that both the shifts in median and standard deviation are small. In addition, it is clear that the shift in median is not correlated with the degree of non-Gaussianity and the shift in standard deviation, and vice-versa.

5.7 Summary

In this chapter, we applied the generative adversarial neural networks to the problem of compression of Monte Carlo replicas. Conceptually, this approach consisted in enhancing the statistics of a given PDF set prior to compression. This resulted in a compression methodology that is able to provide a compressed set with smaller number of replicas and a more adequate representation of the original probability distribution. For PDF sets of the family of NNPDF3.1, the efficiency of the new GAN-based compression has been extensively verified by comparing the results with the standard compression methodology. For Monte Carlo PDF sets generated using the NNPDF4.0 methodology, while the GANPDFS still provides better accuracies for smaller sets, the results were no longer that impressive. This is mostly explained by the fact that the replicas in NNPDF4.0 PDF sets present less fluctuations and exhibit more Gaussian features. We also entertained the idea of using the GANPDFS to address the issue of finite size effects. The analyses performed in the present thesis seem to support the idea that the GANPDFS methodology could be used to bypass the fitting of a large number of PDF replicas. However, more in-depth knowledge are required for this study to be fully conclusive.

CONCLUSIONS

The main theme of the present thesis centered around the theory of strong interactions, ranging from fixed and all-order perturbative computations of observables that are of interests at hadron colliders such as LHC, to its non-perturbative aspects such as the determination of parton density functions. In more specific terms, the unifying idea has been to improve analytical and numerical computations aiming at providing accurate predictions in QCD phenomenology. The following passage summarizes three distinct yet complementary areas studied in this thesis.

The first study focused on curing logarithmic divergences appearing in fixed-order perturbative computations using techniques from resummation. In particular, we focused on the resummation of logarithmic terms that are enhanced at small transverse momentum and at partonic threshold. The two resummations were then combined into a single formalism in which small-p_T logarithms are resummed up to NNLL while threshold ones are resummed up to NNLL*. The interest in providing results for such a combined resummation was threefold. First, the results could provide a resummation of transverse momentum that is valid for the entire range of p_T. Second, the results could be used to improve our understanding of the relation between soft and collinear divergences that drive the transverse momentum spectra. And third, the expansion of the resummed expressions in the running of the perturbative coupling α_s may be used as a tool to approximate higher-order corrections that are not accounted for. The formulation of this combined resummation relied mainly on two ingredients: first, a modified transverse momentum resummation–referred to as *soft-improved transverse momentum resummation*–which leads to the threshold resummed expression upon integration over the transverse momentum, and second, a combination of the improved transverse momentum resummation and the pure threshold expression using some profile matching function that takes into account all the logarithmic enhanced terms in the soft limit for finite p_T. Based on these formulations, one of the main results of this thesis was a detailed phenomenological studies assessing separately the impacts of the improved and combined resummation on transverse momentum distributions. As a test case, we studied the transverse momentum spectra of the Higgs boson produced in gluon fusion in the HEFT picture and the Z boson production via DY mechanism. It was shown that while the the soft-improved resummation improves the convergence in the small-p_T and reduces the effects of unjustified logarithms in the medium p_T regions, the inclusion of the pure threshold resummation improves the agreement with fixed-order calculations in the medium and large-p_T regions. Whilst these effects are more pronounced in the case of Higgs boson production, which is known to have slower perturbative convergence, small but yet noticeable improvements are seen in the DY case.

Another main result from this thesis concerns the construction of an approximate NNLO transverse momentum distribution for the Higgs boson production in the infinite top-quark mass limit. As a proof of concept, our studies focused only on the gluon partonic channel which provides the main bulk of the contribution to the differential cross section. Nevertheless, the procedure described here can be easily applied to other

(sub) partonic channels. The approximate expression relied on the combination of the threshold and high-energy resummations in which the small-x logarithms are resummed to the leading-logarithmic (LLx) accuracy. The results of such an approximation seemed promising, at the very least the procedure described here can be used as a robust method to estimate theoretical uncertainties associated with missing higher-order corrections. In the context of PDF determination, such an approximation can be applied to the study of DY processes in order to provide theoretical predictions that can be compared to experimental measurements. In this case, all the partonic channels involved in the process have to be taken into account. These are currently ongoing works.

The last part of the thesis consisted in developing a machine learning framework that can compress Monte Carlo PDF replicas with as little loss of accuracy as possible. The method we proposed here is based on generative adversarial networks (GANs). Conceptually, the idea was to enhance the statistics of a given Monte Carlo PDF by generating posterior distributions that resemble the original PDF. Only then, the enhanced Monte Carlo PDF passes through a compression procedure. This has proved to yield compressed sets with smaller number of replicas and a more adequate representation of the original probability distribution. Indeed, it was shown that with the same size of compressed set, the GAN-based methodology achieves a more accurate representation of the underlying probability distribution than the standard compression. Whilst this is clearly the case for PDF sets of the family of NNPDF3.1, the results are less impressive with PDF sets generated with the NNPDF4.0 methodology. One possible application of the GANPDFs to PDFs is the problem of finite size effect. This has been briefly touched in this thesis where we compared replicas generated from GANPDFs to replicas generated from a fit using various resampling strategies. Our studies showed that there is little difference between replicas generated by the GAN and a fit. However, due to computational limitations, the studies carried on here were done with a moderate number of Monte Carlo replicas. That is, in order to reduce statistical fluctuations, large samples of Monte Carlo replicas are required. This could be a topic for future studies. The exploration of this idea might lead to important insights into a better understanding of not only the probability distribution functions but also the generative capabilities of GANs.

BIBLIOGRAPHY

[1] R. E. Cutkosky. Singularities and discontinuities of Feynman amplitudes. *J. Math. Phys.*, 1:429–433, 1960.

[2] Gerard 't Hooft and M. J. G. Veltman. DIAGRAMMAR. *NATO Sci. Ser. B*, 4:177–322, 1974.

[3] M. J. G. Veltman. Unitarity and causality in a renormalizable field theory with unstable particles. *Physica*, 29:186–207, 1963.

[4] Richard D. Ball et al. The Path to Proton Structure at One-Percent Accuracy. 9 2021.

[5] Andy Buckley, James Ferrando, Stephen Lloyd, Karl Nordström, Ben Page, Martin Rüfenacht, Marek Schönherr, and Graeme Watt. LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J. C*, 75:132, 2015.

[6] Zahari Kassabov. Reportengine: A framework for declarative data analysis, February 2019.

[7] S. Carrazza, E. R. Nocera, C. Schwan, and M. Zaro. PineAPPL: combining EW and QCD corrections for fast evaluation of LHC processes. *JHEP*, 12:108, 2020.

[8] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[9] Charalampos Anastasiou and Kirill Melnikov. Higgs boson production at hadron colliders in NNLO QCD. *Nucl. Phys. B*, 646:220–256, 2002.

[10] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Elisabetta Furlan, Thomas Gehrmann, Franz Herzog, Achilleas Lazopoulos, and Bernhard Mistlberger. High precision determination of the gluon fusion Higgs boson cross-section at the LHC. *JHEP*, 05:058, 2016.

[11] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Franz Herzog, and Bernhard Mistlberger. Higgs Boson Gluon-Fusion Production in QCD at Three Loops. *Phys. Rev. Lett.*, 114:212001, 2015.

[12] Claude Duhr, Falko Dulat, Valentin Hirschi, and Bernhard Mistlberger. Higgs production in bottom quark fusion: matching the 4- and 5-flavour schemes to third order in the strong coupling. *JHEP*, 08(08):017, 2020.

[13] Andrea Banfi, Fabrizio Caola, Frédéric A. Dreyer, Pier F. Monni, Gavin P. Salam, Giulia Zanderighi, and Falko Dulat. Jet-vetoed Higgs cross section in gluon fusion at N^3LO+NNLL with small-R resummation. *JHEP*, 04:049, 2016.

[14] Johannes M. Henn and Bernhard Mistlberger. Four-Gluon Scattering at Three Loops, Infrared Structure, and the Regge Limit. *Phys. Rev. Lett.*, 117(17):171601, 2016.

[15] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Elisabetta Furlan, Thomas Gehrmann, Franz Herzog, and Bernhard Mistlberger. Higgs Boson GluonFfusion Production Beyond Threshold in N^3LO QCD. *JHEP*, 03:091, 2015.

[16] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Elisabetta Furlan, Thomas Gehrmann, Franz Herzog, and Bernhard Mistlberger. Higgs boson gluon–fusion production at threshold in N^3LO QCD. *Phys. Lett. B*, 737:325–328, 2014.

[17] Frédéric A. Dreyer and Alexander Karlberg. Vector-Boson Fusion Higgs Pair Production at N^3LO. *Phys. Rev. D*, 98(11):114016, 2018.

[18] R. Gauld, A. Gehrmann-De Ridder, E. W. N. Glover, A. Huss, and I. Majer. Precise predictions for WH+jet production at the LHC. *Phys. Lett. B*, 817:136335, 2021.

[19] Claude Duhr, Falko Dulat, and Bernhard Mistlberger. Charged current Drell-Yan production at N^3LO. *JHEP*, 11:143, 2020.

[20] Claude Duhr, Falko Dulat, and Bernhard Mistlberger. Drell-Yan Cross Section to Third Order in the Strong Coupling Constant. *Phys. Rev. Lett.*, 125(17):172001, 2020.

[21] R. Gauld, A. Gehrmann-De Ridder, E. W. N. Glover, A. Huss, and I. Majer. Predictions for Z -Boson Production in Association with a b-Jet at $\mathcal{O}(\alpha_s^3)$. *Phys. Rev. Lett.*, 125(22):222002, 2020.

[22] Marco Bonvini, Simone Marzani, Claudio Muselli, and Luca Rottoli. On the Higgs cross section at N^3LO+N^3LL and its uncertainty. *JHEP*, 08:105, 2016.

[23] Radja Boughezal, Christfried Focke, Walter Giele, Xiaohui Liu, and Frank Petriello. Higgs boson production in association with a jet at NNLO using jettiness subtraction. *Phys. Lett. B*, 748:5–8, 2015.

[24] Radja Boughezal, Fabrizio Caola, Kirill Melnikov, Frank Petriello, and Markus Schulze. Higgs boson production in association with a jet at next-to-next-to-leading order. *Phys. Rev. Lett.*, 115(8):082003, 2015.

[25] Fabrizio Caola, Kirill Melnikov, and Markus Schulze. Fiducial cross sections for Higgs boson production in association with a jet at next-to-next-to-leading order in QCD. *Phys. Rev. D*, 92(7):074032, 2015.

[26] X. Chen, J. Cruz-Martinez, T. Gehrmann, E. W. N. Glover, and M. Jaquier. NNLO QCD corrections to Higgs boson production at large transverse momentum. *JHEP*, 10:066, 2016.

[27] Charalampos Anastasiou and Caterina Specchia. One-loop QCD contributions to differential cross-sections for Higgs production at N^3LO. *JHEP*, 05:080, 2019.

[28] Xuan Chen, Thomas Gehrmann, E. W. Nigel Glover, Alexander Huss, Ye Li, Duff Neill, Markus Schulze, Iain W. Stewart, and Hua Xing Zhu. Precise QCD Description of the Higgs Boson Transverse Momentum Spectrum. *Phys. Lett. B*, 788:425–430, 2019.

[29] Falko Dulat, Bernhard Mistlberger, and Andrea Pelloni. Precision predictions at N^3LO for the Higgs boson rapidity distribution at the LHC. *Phys. Rev. D*, 99(3):034004, 2019.

[30] X. Chen, T. Gehrmann, E. W. N. Glover, A. Huss, B. Mistlberger, and A. Pelloni. Fully Differential Higgs Boson Production to Third Order in QCD. *Phys. Rev. Lett.*, 127(7):072002, 2021.

[31] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, and J. Pires. Triple Differential Dijet Cross Section at the LHC. *Phys. Rev. Lett.*, 123(10):102001, 2019.

[32] Wojciech Bizon, Aude Gehrmann-De Ridder, Thomas Gehrmann, Nigel Glover, Alexander Huss, Pier Francesco Monni, Emanuele Re, Luca Rottoli, and Duncan M. Walker. The transverse momentum spectrum of weak gauge bosons at N ^{3} LL + NNLO. *Eur. Phys. J. C*, 79(10):868, 2019.

[33] Xuan Chen, Thomas Gehrmann, Nigel Glover, Alexander Huss, Tong-Zhi Yang, and Hua Xing Zhu. Di-lepton Rapidity Distribution in Drell-Yan Production to Third Order in QCD. 7 2021.

[34] A. M. Cooper-Sarkar. PDF Fits at HERA. *PoS*, EPS-HEP2011:320, 2011.

[35] Jan Kretzschmar. Measurement of F(2) at Medium Q**2 and the PDF Determination using H1 HERA I Data. In *17th International Workshop on Deep-Inelastic Scattering and Related Subjects*, 9 2009.

[36] Ada Solano. Diffractive PDF fits and factorisation tests at HERA. *AIP Conf. Proc.*, 1350(1):43–46, 2011.

[37] Zhiqing Zhang. HERA Inclusive Neutral and Charged Current Cross Sections and a New PDF Fit, HERAPDF 2.0. *Acta Phys. Polon. Supp.*, 8:957, 2015.

[38] Radek Zlebcik. Diffractive PDF determination from HERA inclusive and jet data at NNLO QCD. *PoS*, DIS2019:059, 2019.

[39] Radek Zlebcik. Diffractive PDF determination from HERA inclusive and jet data at NNLO QCD. *PoS*, EPS-HEP2019:500, 2020.

[40] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[41] Joan Rojo, Luigi Del Debbio, Stefano Forte, Jose I. Latorre, and Andrea Piccione. The Neural network approach to parton fitting. *AIP Conf. Proc.*, 792(1):376–379, 2005.

[42] Luigi Del Debbio, Stefano Forte, Jose I. Latorre, Andrea Piccione, and Joan Rojo. Neural network determination of parton distributions: The Nonsinglet case. *JHEP*, 03:039, 2007.

[43] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Andrea Piccione, Juan Rojo, and Maria Ubiali. A Determination of parton distributions with faithful uncertainty estimation. *Nucl. Phys. B*, 809:1–63, 2009. [Erratum: Nucl.Phys.B 816, 293 (2009)].

[44] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties. *JHEP*, 05:075, 2010.

[45] Richard D. Ball, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys. B*, 855:153–221, 2012.

[46] Richard D. Ball, Stefano Forte, Alberto Guffanti, Emanuele R. Nocera, Giovanni Ridolfi, and Juan Rojo. Unbiased determination of polarized parton distributions and their uncertainties. *Nucl. Phys. B*, 874:36–84, 2013.

[47] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, and Juan Rojo. Parton distributions with QED corrections. *Nucl. Phys. B*, 877:290–320, 2013.

[48] Emanuele R. Nocera, Richard D. Ball, Stefano Forte, Giovanni Ridolfi, and Juan Rojo. A first unbiased global determination of polarized PDFs and their uncertainties. *Nucl. Phys. B*, 887:276–308, 2014.

[49] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.

[50] Valerio Bertone, Stefano Carrazza, Nathan P. Hartland, and Juan Rojo. Illuminating the photon content of the proton within a global PDF analysis. *SciPost Phys.*, 5(1):008, 2018.

[51] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.

[52] Rabah Abdul Khalek et al. A first determination of parton distributions with theoretical uncertainties. *Eur. Phys. J.*, C:79:838, 2019.

[53] Rabah Abdul Khalek et al. Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies. *Eur. Phys. J. C*, 79(11):931, 2019.

[54] S. Alekhin, J. Blumlein, and S. Moch. The ABM parton distributions tuned to LHC data. *Phys. Rev. D*, 89(5):054028, 2014.

[55] S. Alekhin, J. Blumlein, and S. Moch. PDF fit in the fixed-flavor-number scheme. *Nucl. Phys. B Proc. Suppl.*, 222-224:41–51, 2012.

[56] Sergey Alekhin, Johannes Blümlein, Kristin Lohwasser, Lea Michaela Caminada, Katerina Lipka, Ringaile Placakyte, Sven-Olaf Moch, and Roberto Petti. Nucleon PDF separation with the collider and fixed-target data. *Nucl. Part. Phys. Proc.*, 273-275:1961–1966, 2016.

[57] A. Accardi et al. A Critical Appraisal and Evaluation of Modern PDFs. *Eur. Phys. J. C*, 76(8):471, 2016.

[58] Sergey Alekhin, Johannes Bluemlein, Sven-Olaf Moch, and Rengaile Placakyte. The new ABMP16 PDF. *PoS*, DIS2016:016, 2016.

[59] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C. P. Yuan. CT10 next-to-next-to-leading order global analysis of QCD. *Phys. Rev. D*, 89(3):033009, 2014.

[60] Pavel Nadolsky, Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Jon Pumplin, Dan Stump, and C. P. Yuan. Progress in CTEQ-TEA PDF Analysis. In *20th International Workshop on Deep-Inelastic Scattering and Related Subjects*, 6 2012.

[61] Moh'd Hussein, Joshua Isaacson, and Joey Huston. A Study of the Role of the PDF Uncertainty on the LHC *W*-Boson Mass Measurement. *J. Phys. G*, 46(9):095002, 2019.

[62] Orkash Amat et al. Impact of LHC top-quark pair measurements to CTEQ-TEA PDF analysis. 8 2019.

[63] Marco Guzzi, Tie-Jiun Hou, Sayipjamal Dulat, Jun Gao, Joey Huston, Pavel Nadolsky, Carl Schmidt, Jan Winter, Keping Xie, and C. P. Yuan. CTEQ-TEA parton distribution functions with intrinsic charm. *EPJ Web Conf.*, 192:00003, 2018.

[64] Tie-Jiun Hou, Sayipjamal Dulat, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Carl Schmidt, Jan Winter, Keping Xie, and C. P. Yuan. CT14 Intrinsic Charm Parton Distribution Functions from CTEQ-TEA Global Analysis. *JHEP*, 02:059, 2018.

[65] Richard D. Ball et al. Parton Distribution Benchmarking with LHC Data. *JHEP*, 04:125, 2013.

[66] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[67] Tie-Jiun Hou et al. New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC. *Phys. Rev. D*, 103(1):014013, 2021.

[68] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3):033006, 2016.

[69] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C*, 75(5):204, 2015.

[70] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections. *Eur. Phys. J. C*, 64:653–680, 2009.

[71] R. S. Thorne and G. Watt. PDF dependence of Higgs cross sections at the Tevatron and LHC: Response to recent criticism. *JHEP*, 08:100, 2011.

[72] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Uncertainties on α_S in the MMHT2014 global PDF analysis and implications for SM predictions. *Eur. Phys. J. C*, 75(9):435, 2015.

[73] L. A. Harland-Lang, A. D. Martin, R. Nathvani, and R. S. Thorne. Ad Lucem: QED Parton Distribution Functions in the MMHT Framework. *Eur. Phys. J. C*, 79(10):811, 2019.

[74] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs. *Eur. Phys. J. C*, 81(4):341, 2021.

[75] T. Regge. Bound states, shadow states and Mandelstam representation. *Nuovo Cim.*, 18:947–956, 1960.

[76] Walter T. Giele, Stephane A. Keller, and David A. Kosower. Parton Distribution Function Uncertainties. 4 2001.

[77] Richard D. Ball et al. An open-source machine learning framework for global analyses of parton distributions. 9 2021.

[78] Felix Hekhorn, Giacomo Magni, Juan M. Cruz-Martinez, Alessandro Candido, and Stefano Carrazza. N3pdf/eko: Nnlo, matching, backward evolution, 2021.

[79] Alessandro Candido, Felix Hekhorn, and Giacomo Magni. N3pdf/yadism: Intrinsic charm, 2021.

[80] Matteo Cacciari and Nicolas Houdeau. Meaningful characterisation of perturbative theoretical uncertainties. *JHEP*, 09:039, 2011.

[81] Emanuele Bagnaschi, Matteo Cacciari, Alberto Guffanti, and Laura Jenniches. An extensive survey of the estimation of uncertainties from missing higher orders in perturbative calculations. *JHEP*, 02:133, 2015.

[82] Marco Bonvini. Probabilistic definition of the perturbative theoretical uncertainty from missing higher orders. *Eur. Phys. J. C*, 80(10):989, 2020.

[83] Rutger Boels. Covariant representation theory of the Poincare algebra and some of its extensions. *JHEP*, 01:010, 2010.

[84] Nima Arkani-Hamed, Tzu-Chen Huang, and Yu-tin Huang. Scattering Amplitudes For All Masses and Spins. 9 2017.

[85] Johannes M. Henn and Jan C. Plefka. *Scattering Amplitudes in Gauge Theories*, volume 883. Springer, Berlin, 2014.

[86] Freddy Cachazo and David Skinner. On the structure of scattering amplitudes in N=4 super Yang-Mills and N=8 supergravity. 1 2008.

[87] Henrik Johansson and Alexander Ochirov. Color-Kinematics Duality for QCD Amplitudes. *JHEP*, 01:170, 2016.

[88] Rutger H. Boels, Reinke Sven Isermann, Ricardo Monteiro, and Donal O'Connell. Colour-Kinematics Duality for One-Loop Rational Amplitudes. *JHEP*, 04:107, 2013.

[89] Marco Chiodaroli, Qingjun Jin, and Radu Roiban. Color/kinematics duality for general abelian orbifolds of N=4 super Yang-Mills theory. *JHEP*, 01:152, 2014.

[90] Zvi Bern, Scott Davies, Tristan Dennen, Yu-tin Huang, and Josh Nohle. Color-Kinematics Duality for Pure Yang-Mills and Gravity at One and Two Loops. *Phys. Rev. D*, 92(4):045041, 2015.

[91] Josh Nohle. Color-Kinematics Duality in One-Loop Four-Gluon Amplitudes with Matter. *Phys. Rev. D*, 90(2):025020, 2014.

[92] Lance J. Dixon. Calculating scattering amplitudes efficiently. In *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 95): QCD and Beyond*, 1 1996.

[93] S. Moch, J. A. M. Vermaseren, and A. Vogt. The Three loop splitting functions in QCD: The Nonsinglet case. *Nucl. Phys. B*, 688:101–134, 2004.

[94] A. Vogt, S. Moch, and J. A. M. Vermaseren. The Three-loop splitting functions in QCD: The Singlet case. *Nucl. Phys. B*, 691:129–181, 2004.

[95] J. Davies, A. Vogt, B. Ruijl, T. Ueda, and J. A. M. Vermaseren. Large-n_f contributions to the four-loop splitting functions in QCD. *Nucl. Phys. B*, 915:335–362, 2017.

[96] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt. Four-Loop Non-Singlet Splitting Functions in the Planar Limit and Beyond. *JHEP*, 10:041, 2017.

[97] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider. The three-loop splitting functions $P_{qg}^{(2)}$ and $P_{gg}^{(2,N_F)}$. *Nucl. Phys. B*, 922:1–40, 2017.

[98] Marco Bonvini and Simone Marzani. Four-loop splitting functions at small x. *JHEP*, 06:145, 2018.

[99] Toichiro Kinoshita and Alberto Sirlin. Radiative corrections to Fermi interactions. *Phys. Rev.*, 113:1652–1660, 1959.

[100] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.

[101] T. D. Lee and M. Nauenberg. Degenerate Systems and Mass Singularities. *Phys. Rev.*, 133:B1549–B1562, 1964.

[102] George F. Sterman and Steven Weinberg. Jets from Quantum Chromodynamics. *Phys. Rev. Lett.*, 39:1436, 1977.

[103] S. D. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann. Jets in hadron-hadron collisions. *Prog. Part. Nucl. Phys.*, 60:484–551, 2008.

[104] Gerald C. Blazey et al. Run II jet physics. In *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting*, pages 47–77, 5 2000.

[105] Gavin P. Salam. Towards Jetography. *Eur. Phys. J. C*, 67:637–686, 2010.

[106] G. Ballocchi and R. Odorico. The String Effect and Independent Fragmentation Models: Lore and Facts. *Nucl. Phys. B*, 345:173–185, 1990.

[107] X. Artru and G. Mennessier. String model and multiproduction. *Nucl. Phys. B*, 70:93–115, 1974.

[108] Torbjorn Sjostrand. The Lund Monte Carlo for Jet Fragmentation. *Comput. Phys. Commun.*, 27:243, 1982.

[109] Torbjorn Sjostrand. The Lund Monte Carlo for Jet Fragmentation and e+ e- Physics: Jetset Version 6.2. *Comput. Phys. Commun.*, 39:347–407, 1986.

[110] M. G. Bowler. e+ e- Production of Heavy Quarks in the String Model. *Z. Phys. C*, 11:169, 1981.

[111] Richard D. Field and Stephen Wolfram. A QCD Model for e+ e- Annihilation. *Nucl. Phys. B*, 213:65–84, 1983.

[112] B. R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys. B*, 238:492–528, 1984.

[113] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.

[114] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[115] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[116] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.

[117] A. Vogt. Efficient evolution of unpolarized and polarized parton distributions with QCD-PEGASUS. *Comput. Phys. Commun.*, 170:65–92, 2005.

[118] R. Keith Ellis, Z. Kunszt, and E. M. Levin. The Evolution of parton distributions at small x. *Nucl. Phys. B*, 420:517–549, 1994. [Erratum: Nucl.Phys.B 433, 498–498 (1995)].

[119] Johannes Blumlein and Andreas Vogt. The Evolution of unpolarized singlet structure functions at small x. *Phys. Rev. D*, 58:014020, 1998.

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

[121] Zoltan Nagy. Three jet cross-sections in hadron hadron collisions at next-to-leading order. *Phys. Rev. Lett.*, 88:122003, 2002.

[122] John M. Campbell, R. Keith Ellis, and Walter T. Giele. A Multi-Threaded Version of MCFM. *Eur. Phys. J. C*, 75(6):246, 2015.

[123] Valerio Bertone, Rikkert Frederix, Stefano Frixione, Juan Rojo, and Mark Sutton. aMCfast: automation of fast NLO computations for PDF fits. *JHEP*, 08:166, 2014.

[124] T. Kluge, K. Rabbertz, and M. Wobisch. FastNLO: Fast pQCD calculations for PDF fits. In *14th International Workshop on Deep Inelastic Scattering*, pages 483–486, 9 2006.

[125] M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober. Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes. 9 2011.

[126] Tancredi Carli, Dan Clements, Amanda Cooper-Sarkar, Claire Gwenlan, Gavin P. Salam, Frank Siegert, Pavel Starovoitov, and Mark Sutton. A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project. *Eur. Phys. J. C*, 66:503–524, 2010.

[127] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk. Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions. *Phys. Rev. D*, 101(7):074020, 2020.

[128] Alessandro Candido, Stefano Forte, and Felix Hekhorn. Can $\overline{\text{MS}}$ parton distributions be negative? *JHEP*, 11:129, 2020.

[129] Emanuele R. Nocera, Maria Ubiali, and Cameron Voisey. Single Top Production in PDF fits. *JHEP*, 05:067, 2020.

[130] Rabah Abdul Khalek et al. Phenomenology of NNLO jet production at the LHC and its impact on parton distributions. *Eur. Phys. J. C*, 80(8):797, 2020.

[131] Ferran Faura, Shayan Iranipour, Emanuele R. Nocera, Juan Rojo, and Maria Ubiali. The Strangest Proton? *Eur. Phys. J. C*, 80(12):1168, 2020.

[132] M. Cepeda et al. Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7:221–584, 2019.

[133] Richard D. Ball, Marco Bonvini, Stefano Forte, Simone Marzani, and Giovanni Ridolfi. Higgs production in gluon fusion beyond NNLO. *Nucl. Phys. B*, 874:746–772, 2013.

[134] Ruth Britto, Freddy Cachazo, Bo Feng, and Edward Witten. Direct proof of tree-level recursion relation in Yang-Mills theory. *Phys. Rev. Lett.*, 94:181602, 2005.

[135] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.

[136] E. G. Floratos, D. A. Ross, and Christopher T. Sachrajda. Higher Order Effects in Asymptotically Free Gauge Theories: The Anomalous Dimensions of Wilson Operators. *Nucl. Phys. B*, 129:66–88, 1977. [Erratum: Nucl.Phys.B 139, 545–546 (1978)].

[137] E. G. Floratos, D. A. Ross, and Christopher T. Sachrajda. Higher Order Effects in Asymptotically Free Gauge Theories. 2. Flavor Singlet Wilson Operators and Coefficient Functions. *Nucl. Phys. B*, 152:493–520, 1979.

[138] E. G. Floratos, R. Lacaze, and C. Kounnas. Space and Timelike Cut Vertices in QCD Beyond the Leading Order. 1. Nonsinglet Sector. *Phys. Lett. B*, 98:89–95, 1981.

[139] E. G. Floratos, R. Lacaze, and C. Kounnas. Space and Timelike Cut Vertices in QCD Beyond the Leading Order. 2. The Singlet Sector. *Phys. Lett. B*, 98:285–290, 1981.

[140] G. Curci, Mario Greco, and Y. Srivastava. QCD Jets From Coherent States. *Nucl. Phys.*, B159:451–468, 1979.

[141] G. Curci, W. Furmanski, and R. Petronzio. Evolution of Parton Densities Beyond Leading Order: The Nonsinglet Case. *Nucl. Phys. B*, 175:27–92, 1980.

[142] G. Curci and Mario Greco. Soft Corrections to the Drell-Yan Process in QCD. *Phys. Lett. B*, 102:280–284, 1981.

[143] W. Furmanski and R. Petronzio. Singlet Parton Densities Beyond Leading Order. *Phys. Lett. B*, 97:437–442, 1980.

[144] Antonio Gonzalez-Arroyo and C. Lopez. Second Order Contributions to the Structure Functions in Deep Inelastic Scattering. 3. The Singlet Case. *Nucl. Phys. B*, 166:429–459, 1980.

[145] Johannes Blumlein and Stefan Kurth. Harmonic sums and Mellin transforms up to two loop order. *Phys. Rev. D*, 60:014018, 1999.

[146] Johannes Blumlein. Analytic continuation of Mellin transforms up to two loop order. *Comput. Phys. Commun.*, 133:76–104, 2000.

[147] Johannes Blumlein. Algebraic relations between harmonic sums and associated quantities. *Comput. Phys. Commun.*, 159:19–54, 2004.

[148] J. Blumlein and V. Ravindran. O (alpha**2(s)) Timelike Wilson Coefficients for Parton-Fragmentation Functions in Mellin Space. *Nucl. Phys. B*, 749:1–24, 2006.

[149] Johannes Blumlein. Structural Relations of Harmonic Sums and Mellin Transforms up to Weight w = 5. *Comput. Phys. Commun.*, 180:2218–2249, 2009.

[150] Marco Bonvini and Simone Marzani. Double resummation for Higgs production. *Phys. Rev. Lett.*, 120(20):202003, 2018.

[151] Simone Marzani. Combining Q_T and small-x resummations. *Phys. Rev. D*, 93(5):054047, 2016.

[152] Wojciech Bizon, Pier Francesco Monni, Emanuele Re, Luca Rottoli, and Paolo Torrielli. Momentum-space resummation for transverse observables and the Higgs $p_\perp$ at $N^3LL+NNLO$. *JHEP*, 02:108, 2018.

[153] Wojciech Bizoń, Xuan Chen, Aude Gehrmann-De Ridder, Thomas Gehrmann, Nigel Glover, Alexander Huss, Pier Francesco Monni, Emanuele Re, Luca Rottoli, and Paolo Torrielli. Fiducial distributions in Higgs and Drell-Yan production at $N^3LL+NNLO$. *JHEP*, 12:132, 2018.

[154] Markus A. Ebert, Johannes K.L. Michel, Iain W. Stewart, and Frank J. Tackmann. Drell-Yan q_T Resummation of Fiducial Power Corrections at N^3LL. 6 2020.

[155] Thomas Becher and Monika Hager. Event-Based Transverse Momentum Resummation. *Eur. Phys. J. C*, 79(8):665, 2019.

[156] Ignazio Scimemi and Alexey Vladimirov. Non-perturbative structure of semi-inclusive deep-inelastic and Drell-Yan scattering at small transverse momentum. *JHEP*, 06:137, 2020.

[157] Stefan Kallweit, Emanuele Re, Luca Rottoli, and Marius Wiesemann. Accurate single- and double-differential resummation of colour-singlet processes with MATRIX+RadISH: W^+W^- production at the LHC. 4 2020.

[158] Marius Wiesemann, Luca Rottoli, and Paolo Torrielli. The Z γ transverse-momentum spectrum at NNLO+N3LL. 6 2020.

[159] Hsiang-nan Li. Unification of the k(T) and threshold resummations. *Phys. Lett.*, B454:328–334, 1999.

[160] Eric Laenen, George F. Sterman, and Werner Vogelsang. Recoil and threshold corrections in short distance cross-sections. *Phys. Rev.*, D63:114018, 2001.

[161] Anna Kulesza, George F. Sterman, and Werner Vogelsang. Joint resummation for Higgs production. *Phys. Rev.*, D69:014012, 2004.

[162] Anna Kulesza, George F. Sterman, and Werner Vogelsang. Joint resummation in electroweak boson production. *Phys. Rev. D*, 66:014011, 2002.

[163] Simone Marzani and Vincent Theeuwes. Vector boson production in joint resummation. *JHEP*, 02:127, 2017.

[164] Gillian Lustermans, Wouter J. Waalewijn, and Lisa Zeune. Joint transverse momentum and threshold resummation beyond NLL. *Phys. Lett. B*, 762:447–454, 2016.

[165] Claudio Muselli, Stefano Forte, and Giovanni Ridolfi. Combined threshold and transverse momentum resummation for inclusive observables. *JHEP*, 03:106, 2017.

[166] John C. Collins, Davison E. Soper, and George F. Sterman. Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson Production. *Nucl. Phys. B*, 250:199–224, 1985.

[167] C. T. H. Davies and W. James Stirling. Nonleading Corrections to the Drell-Yan Cross-Section at Small Transverse Momentum. *Nucl. Phys. B*, 244:337–348, 1984.

[168] John C. Collins and Davison E. Soper. Back-To-Back Jets: Fourier Transform from B to K-Transverse. *Nucl. Phys. B*, 197:446–476, 1982.

[169] John C. Collins and Davison E. Soper. Back-To-Back Jets in QCD. *Nucl. Phys. B*, 193:381, 1981. [Erratum: Nucl.Phys.B 213, 545 (1983)].

[170] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization for One Loop Corrections in the Drell-Yan Process. *Nucl. Phys. B*, 223:381–421, 1983.

[171] Geoffrey T. Bodwin. Factorization of the Drell-Yan Cross-Section in Perturbation Theory. *Phys. Rev. D*, 31:2616, 1985. [Erratum: Phys.Rev.D 34, 3932 (1986)].

[172] Guido Altarelli, R. Keith Ellis, Mario Greco, and G. Martinelli. Vector Boson Production at Colliders: A Theoretical Reappraisal. *Nucl. Phys. B*, 246:12–44, 1984.

[173] Daniel de Florian and Massimiliano Grazzini. Next-to-next-to-leading logarithmic corrections at small transverse momentum in hadronic collisions. *Phys. Rev. Lett.*, 85:4678–4681, 2000.

[174] Dominik Kara. *Higher-Order Corrections to Higgs Boson Amplitudes with Full Quark Mass Dependence in Quantum Chromodynamics*. PhD thesis, Zurich U., 2018.

[175] C. T. H. Davies, B. R. Webber, and W. James Stirling. Drell-Yan Cross-Sections at Small Transverse Momentum. *Nucl. Phys. B*, 256:413, 1985.

[176] C. Balazs, Edmond L. Berger, S. Mrenna, and C. P. Yuan. Photon pair production with soft gluon resummation in hadronic interactions. *Phys. Rev. D*, 57:6934–6947, 1998.

[177] C. Balazs and C. P. Yuan. Soft gluon effects on lepton pairs at hadron colliders. *Phys. Rev. D*, 56:5558–5583, 1997.

[178] Jiro Kodaira and Luca Trentadue. Summing Soft Emission in QCD. *Phys. Lett. B*, 112:66, 1982.

[179] Jiro Kodaira and Luca Trentadue. Single Logarithm Effects in electron-Positron Annihilation. *Phys. Lett. B*, 123:335–338, 1983.

[180] S. Catani, E. D'Emilio, and L. Trentadue. The gluon form factor to higher orders: Gluon-gluon annihilation at small qt. *Physics Letters B*, 211(3):335–342, 1988.

[181] Daniel de Florian and Massimiliano Grazzini. Next-to-next-to-leading-order logarithmic corrections at small transverse momentum in hadronic collisions. *Phys. Rev. Lett.*, 85:4678–4681, Nov 2000.

[182] S. Dawson. Radiative corrections to Higgs boson production. *Nucl. Phys. B*, 359:283–300, 1991.

[183] A. Djouadi, M. Spira, and P. M. Zerwas. Production of Higgs bosons in proton colliders: QCD corrections. *Phys. Lett. B*, 264:440–446, 1991.

[184] Ye Li and Hua Xing Zhu. Bootstrapping Rapidity Anomalous Dimensions for Transverse-Momentum Resummation. *Phys. Rev. Lett.*, 118(2):022004, 2017.

[185] F. Landry, R. Brock, Pavel M. Nadolsky, and C. P. Yuan. Tevatron Run-1 Z boson data and Collins-Soper-Sterman resummation formalism. *Phys. Rev. D*, 67:073016, 2003.

[186] Anton V. Konychev and Pavel M. Nadolsky. Universality of the Collins-Soper-Sterman nonperturbative function in gauge boson production. *Phys. Lett. B*, 633:710–714, 2006.

[187] G. A. Ladinsky and C.-P. Yuan. Nonperturbative regime in qcd resummation for gauge boson production at hadron colliders. *Phys. Rev. D*, 50:R4239–R4243, Oct 1994.

[188] Jian-wei Qiu and Xiao-fei Zhang. QCD prediction for heavy boson transverse momentum distributions. *Phys. Rev. Lett.*, 86:2724–2727, 2001.

[189] R. Keith Ellis and Sinisa Veseli. W and Z transverse momentum distributions: Resummation in q_T space. *Nucl. Phys. B*, 511:649–669, 1998.

[190] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. Problems in the resummation of soft gluon effects in the transverse momentum distributions of massive vector bosons in hadronic collisions. *Nucl. Phys. B*, 542:311–328, 1999.

[191] Anna Kulesza and W. James Stirling. On the resummation of subleading logarithms in the transverse momentum distribution of vector bosons produced at hadron colliders. *JHEP*, 01:016, 2000.

[192] Markus A. Ebert and Frank J. Tackmann. Resummation of Transverse Momentum Distributions in Distribution Space. *JHEP*, 02:110, 2017.

[193] Pier Francesco Monni, Emanuele Re, and Paolo Torrielli. Higgs Transverse-Momentum Resummation in Direct Space. *Phys. Rev. Lett.*, 116(24):242001, 2016.

[194] Marco Guzzi, Pavel M. Nadolsky, and Bowen Wang. Nonperturbative contributions to a resummed leptonic angular distribution in inclusive neutral vector boson production. *Phys. Rev. D*, 90(1):014030, 2014.

[195] Giuseppe Bozzi, Stefano Catani, Daniel de Florian, and Massimiliano Grazzini. Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC. *Nucl. Phys. B*, 737:73–120, 2006.

[196] Stefano Catani, Daniel de Florian, and Massimiliano Grazzini. Universality of nonleading logarithmic contributions in transverse momentum distributions. *Nucl. Phys. B*, 596:299–312, 2001.

[197] S. Catani and M. Grazzini. Higgs Boson Production at Hadron Colliders: Hard-Collinear Coefficients at the NNLO. *Eur. Phys. J. C*, 72:2013, 2012. [Erratum: Eur.Phys.J.C 72, 2132 (2012)].

[198] Stefano Catani, Leandro Cieri, Daniel de Florian, Giancarlo Ferrera, and Massimiliano Grazzini. Universality of transverse-momentum resummation and hard factors at the NNLO. *Nucl. Phys. B*, 881:414–443, 2014.

[199] Stefano Catani, Daniel de Florian, Giancarlo Ferrera, and Massimiliano Grazzini. Vector boson production at hadron colliders: transverse-momentum resummation and leptonic decay. *JHEP*, 12:047, 2015.

[200] G. Parisi and R. Petronzio. Small Transverse Momentum Distributions in Hard Processes. *Nucl. Phys. B*, 154:427–440, 1979.

[201] Stefano Catani, Leandro Cieri, Daniel de Florian, Giancarlo Ferrera, and Massimiliano Grazzini. Vector boson production at hadron colliders: hard-collinear coefficients at the NNLO. *Eur. Phys. J. C*, 72:2195, 2012.

[202] S. Catani, G. Turnock, B. R. Webber, and L. Trentadue. Thrust distribution in e+ e- annihilation. *Phys. Lett. B*, 263:491–497, 1991.

[203] S. Catani, L. Trentadue, G. Turnock, and B. R. Webber. Resummation of large logarithms in e+ e- event shape distributions. *Nucl. Phys. B*, 407:3–42, 1993.

[204] Stefano Catani, Michelangelo L. Mangano, Paolo Nason, Carlo Oleari, and Werner Vogelsang. Sudakov resummation effects in prompt photon hadroproduction. *JHEP*, 03:025, 1999.

[205] Roberto Bonciani, Stefano Catani, Michelangelo L. Mangano, and Paolo Nason. NLL resummation of the heavy quark hadroproduction cross-section. *Nucl. Phys. B*, 529:424–450, 1998. [Erratum: Nucl.Phys.B 803, 234 (2008)].

[206] Stefano Catani and Massimiliano Grazzini. QCD transverse-momentum resummation in gluon fusion processes. *Nucl. Phys. B*, 845:297–323, 2011.

[207] M. Dasgupta and G. P. Salam. Resummation of the jet broadening in DIS. *Eur. Phys. J. C*, 24:213–236, 2002.

[208] George F. Sterman. Summation of Large Corrections to Short Distance Hadronic Cross-Sections. *Nucl. Phys. B*, 281:310–364, 1987.

[209] Roberto Bonciani, Stefano Catani, Michelangelo L. Mangano, and Paolo Nason. Sudakov resummation of multiparton QCD cross-sections. *Phys. Lett. B*, 575:268–278, 2003.

[210] Nikolaos Kidonakis and George F. Sterman. Resummation for QCD hard scattering. *Nucl. Phys. B*, 505:321–348, 1997.

[211] Stefano Catani, Daniel de Florian, Massimiliano Grazzini, and Paolo Nason. Soft gluon resummation for Higgs boson production at hadron colliders. *JHEP*, 07:028, 2003.

[212] Valentin Ahrens, Thomas Becher, Matthias Neubert, and Li Lin Yang. Renormalization-Group Improved Prediction for Higgs Production at Hadron Colliders. *Eur. Phys. J. C*, 62:333–353, 2009.

[213] Stefano Forte, Giovanni Ridolfi, and Simone Rota. Threshold resummation of transverse momentum distributions beyond next-to-leading log. 6 2021.

[214] Pulak Banerjee, Goutam Das, Prasanna K. Dhani, and V. Ravindran. Threshold resummation of the rapidity distribution for Drell-Yan production at NNLO+NNLL. *Phys. Rev. D*, 98(5):054018, 2018.

[215] Stefano Catani, Leandro Cieri, Daniel de Florian, Giancarlo Ferrera, and Massimiliano Grazzini. Threshold resummation at N^3LL accuracy and soft-virtual cross sections at N^3LO. *Nucl. Phys. B*, 888:75–91, 2014.

[216] Marco Bonvini and Simone Marzani. Resummed Higgs cross section at N^3LL. *JHEP*, 09:007, 2014.

[217] Stefano Catani, Massimiliano Grazzini, and Alessandro Torre. Soft-gluon resummation for single-particle inclusive hadroproduction at high transverse momentum. *Nucl. Phys. B*, 874:720–745, 2013.

[218] Michal Czakon, Alexander Mitov, and George F. Sterman. Threshold Resummation for Top-Pair Hadroproduction to Next-to-Next-to-Leading Log. *Phys. Rev. D*, 80:074017, 2009.

[219] Wim Beenakker, Tim Janssen, Susanne Lepoeter, Michael Krämer, Anna Kulesza, Eric Laenen, Irene Niessen, Silja Thewes, and Tom Van Daal. Towards NNLL resummation: hard matching coefficients for squark and gluino hadroproduction. *JHEP*, 10:120, 2013.

[220] G. Parisi. Summing Large Perturbative Corrections in QCD. *Phys. Lett. B*, 90:295–296, 1980.

[221] D. Amati, A. Bassetto, M. Ciafaloni, G. Marchesini, and G. Veneziano. A Treatment of Hard Processes Sensitive to the Infrared Structure of QCD. *Nucl. Phys. B*, 173:429–455, 1980.

[222] M. Ciafaloni and G. Curci. Exponentiation of large n singularities in qcd. *Physics Letters B*, 102(5):352–356, 1981.

[223] P. Chiappetta, T. Grandou, M. Le Bellac, and J. L. Meunier. The Role of Soft Gluons in Lepton Pair Production. *Nucl. Phys. B*, 207:251, 1982.

[224] Stefano Forte and Giovanni Ridolfi. Renormalization group approach to soft gluon resummation. *Nucl. Phys. B*, 650:229–270, 2003.

[225] S. Catani and L. Trentadue. Comment on QCD exponentiation at large x. *Nucl. Phys. B*, 353:183–186, 1991.

[226] Stefano Catani, Michelangelo L. Mangano, Paolo Nason, and Luca Trentadue. The Resummation of soft gluons in hadronic collisions. *Nucl. Phys. B*, 478:273–310, 1996.

[227] D. de Florian, A. Kulesza, and W. Vogelsang. Threshold resummation for high-transverse-momentum Higgs production at the LHC. *JHEP*, 02:047, 2006.

[228] S. Moch, J. A. M. Vermaseren, and A. Vogt. Higher-order corrections in threshold resummation. *Nucl. Phys. B*, 726:317–335, 2005.

[229] Marco Bonvini, Stefano Forte, Margherita Ghezzi, and Giovanni Ridolfi. Threshold Resummation in SCET vs. Perturbative QCD: An Analytic Comparison. *Nucl. Phys. B*, 861:337–360, 2012.

[230] Stefano Forte, Giovanni Ridolfi, Joan Rojo, and Maria Ubiali. Borel resummation of soft gluon radiation and higher twists. *Phys. Lett. B*, 635:313–319, 2006.

[231] Riccardo Abbate, Stefano Forte, and Giovanni Ridolfi. A New prescription for soft gluon resummation. *Phys. Lett. B*, 657:55–63, 2007.

[232] Marco Bonvini, Stefano Forte, and Giovanni Ridolfi. Borel resummation of transverse momentum distributions. *Nucl. Phys. B*, 808:347–363, 2009.

[233] Marco Bonvini, Stefano Forte, and Giovanni Ridolfi. Soft gluon resummation of Drell-Yan rapidity distributions: Theory and phenomenology. *Nucl. Phys. B*, 847:93–159, 2011.

[234] Marco Bonvini. *Resummation of soft and hard gluon radiation in perturbative QCD*. PhD thesis, Genoa U., 2012.

[235] Tanjona R. Rabemananjara. Phenomenology of combined resummation for Higgs and Drell-Yan. *JHEP*, 12:073, 2020.

[236] T. Huber and Daniel Maitre. HypExp: A Mathematica package for expanding hypergeometric functions around integer-valued parameters. *Comput. Phys. Commun.*, 175:122–144, 2006.

[237] Tobias Huber and Daniel Maitre. HypExp 2, Expanding Hypergeometric Functions about Half-Integer Parameters. *Comput. Phys. Commun.*, 178:755–776, 2008.

[238] Leonard Maximon. The dilogarithm function for complex argument. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, 459:2807–2819, 11 2003.

[239] John R. Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. A Phenomenological Profile of the Higgs Boson. *Nucl. Phys. B*, 106:292, 1976.

[240] Mikhail A. Shifman, A. I. Vainshtein, M. B. Voloshin, and Valentin I. Zakharov. Low-Energy Theorems for Higgs Boson Couplings to Photons. *Sov. J. Nucl. Phys.*, 30:711–716, 1979.

[241] Richard J. Gonsalves, Jerzy Pawlowski, and Chung-Fai Wai. QCD Radiative Corrections to Electroweak Boson Production at Large Transverse Momentum in Hadron Collisions. *Phys. Rev. D*, 40:2245, 1989.

[242] Giuseppe Bozzi, Stefano Catani, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini. Transverse-momentum resummation: A Perturbative study of Z production at the Tevatron. *Nucl. Phys. B*, 815:174–197, 2009.

[243] Giuseppe Bozzi, Stefano Catani, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini. Production of Drell-Yan lepton pairs in hadron collisions: Transverse-momentum resummation at next-to-next-to-leading logarithmic accuracy. *Phys. Lett. B*, 696:207–213, 2011.

[244] C. P. Yuan. Kinematics of the Higgs boson at hadron colliders: NLO QCD gluon resummation. *Phys. Lett. B*, 283:395–402, 1992.

[245] Yuri L. Dokshitzer, Dmitri Diakonov, and S. I. Troian. Hard Processes in Quantum Chromodynamics. *Phys. Rept.*, 58:269–395, 1980.

[246] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989.

[247] S. Catani and L. Trentadue. Resummation of the QCD Perturbative Series for Hard Processes. *Nucl. Phys.*, B327:323–352, 1989.

[248] Simon Albino and Richard D. Ball. Soft resummation of quark anomalous dimensions and coefficient functions in MS-bar factorization. *Phys. Lett.*, B513:93–102, 2001.

[249] Eric Laenen. Resummation for observables at TeV colliders. *Pramana*, 63:1225–1249, 2004.

[250] Daniel de Florian, Giancarlo Ferrera, Massimiliano Grazzini, and Damiano Tommasini. Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC. *JHEP*, 11:064, 2011.

[251] Eric Laenen, George F. Sterman, and Werner Vogelsang. Higher order QCD corrections in prompt photon production. *Phys. Rev. Lett.*, 84:4296–4299, 2000.

[252] S. Farid Khwaja and A. B. Olde Daalhuis. Uniform asymptotic expansions for hypergeometric functions with large parameters IV. *Analysis and Applications*, 12(06):667–710, October 2014.

[253] Nico M Temme. Large parameter cases of the gauss hypergeometric function. *Journal of Computational and Applied Mathematics*, 153(1-2):441–462, April 2003.

[254] Tanjona Rabemananjara and Roy Stegeman. N3pdf/hpt-mon: Hpt-mon-v1.0.1, Jun 2021.

[255] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini. The q(T) spectrum of the Higgs boson at the LHC in QCD perturbation theory. *Phys. Lett. B*, 564:65–72, 2003.

[256] X. Chen, T. Gehrmann, E. W. N. Glover, and M. Jaquier. Precise QCD predictions for the production of Higgs + jet final states. *Phys. Lett. B*, 740:147–150, 2015.

[257] A. Banfi, S. Redford, M. Vesterinen, P. Waller, and T. R. Wyatt. Optimisation of variables for studying dilepton transverse momentum distributions at hadron colliders. *Eur. Phys. J. C*, 71:1600, 2011.

[258] T. Affolder et al. The transverse momentum and total cross section of e^+e^- pairs in the Z boson region from $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.*, 84:845–850, 2000.

[259] V. M. Abazov et al. Measurement of the shape of the boson transverse momentum distribution in $p\bar{p} \to Z/\gamma^* \to e^+e^- + X$ events produced at $\sqrt{s}$=1.96-TeV. *Phys. Rev. Lett.*, 100:102002, 2008.

[260] Victor Mukhamedovich Abazov et al. Measurement of the Normalized $Z/\gamma^* - > \mu^+\mu^-$ Transverse Momentum Distribution in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Lett. B*, 693:522–530, 2010.

[261] Marco Bonvini, Stefano Forte, Giovanni Ridolfi, and Luca Rottoli. Resummation prescriptions and ambiguities in SCET vs. direct QCD: Higgs production as a case study. *JHEP*, 01:046, 2015.

[262] D. de Florian, M. Grazzini, and Z. Kunszt. Higgs production with large transverse momentum in hadronic collisions at next-to-leading order. *Phys. Rev. Lett.*, 82:5209–5212, 1999.

[263] M. Spira, A. Djouadi, D. Graudenz, and P. M. Zerwas. Higgs boson production at the LHC. *Nucl. Phys. B*, 453:17–82, 1995.

[264] V. Ravindran, J. Smith, and W. L. van Neerven. NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions. *Nucl. Phys. B*, 665:325–366, 2003.

[265] Simone Marzani, Richard D. Ball, Vittorio Del Duca, Stefano Forte, and Alessandro Vicini. Higgs production via gluon-gluon fusion with finite top mass beyond next-to-leading order. *Nucl. Phys. B*, 800:127–145, 2008.

[266] Robert V. Harlander and Kemal J. Ozeren. Top mass effects in Higgs production at next-to-next-to-leading order QCD: Virtual corrections. *Phys. Lett. B*, 679:467–472, 2009.

[267] Robert V. Harlander and Kemal J. Ozeren. Finite top mass effects for hadronic Higgs production at next-to-next-to-leading order. *JHEP*, 11:088, 2009.

[268] Robert V. Harlander, Hendrik Mantler, Simone Marzani, and Kemal J. Ozeren. Higgs production in gluon fusion at next-to-next-to-leading order QCD for finite top mass. *Eur. Phys. J. C*, 66:359–372, 2010.

[269] Alexey Pak, Mikhail Rogal, and Matthias Steinhauser. Virtual three-loop corrections to Higgs boson production in gluon fusion for finite top quark mass. *Phys. Lett. B*, 679:473–477, 2009.

[270] Alexey Pak, Mikhail Rogal, and Matthias Steinhauser. Finite top quark mass effects in NNLO Higgs boson production at LHC. *JHEP*, 02:025, 2010.

[271] Bernhard Mistlberger. Higgs boson production at hadron colliders at N^3LO in QCD. *JHEP*, 05:028, 2018.

[272] Thomas Becher and Tobias Neumann. Fiducial q_T resummation of color-singlet processes at N^3LL+NNLO. *JHEP*, 03:199, 2021.

[273] Simone Alioli, Alessandro Broggio, Alessandro Gavardi, Stefan Kallweit, Matthew A. Lim, Riccardo Nagar, Davide Napoletano, Christian W. Bauer, and Luca Rottoli. Matching NNLO to parton shower using N^3LL colour-singlet transverse momentum resummation in GENEVA. 2 2021.

[274] Emanuele Re, Luca Rottoli, and Paolo Torrielli. Fiducial Higgs and Drell-Yan distributions at N^3LL'+NNLO with RadISH. 4 2021.

[275] Marco Bonvini, Stefano Forte, and Giovanni Ridolfi. The Threshold region for Higgs production in gluon fusion. *Phys. Rev. Lett.*, 109:102002, 2012.

[276] Stefano Forte and Claudio Muselli. High energy resummation of transverse momentum distributions: Higgs in gluon fusion. *JHEP*, 03:122, 2016.

[277] Guido Altarelli, Richard D. Ball, and Stefano Forte. Small x Resummation with Quarks: Deep-Inelastic Scattering. *Nucl. Phys. B*, 799:199–240, 2008.

[278] Richard D. Ball. Resummation of Hadroproduction Cross-sections at High Energy. *Nucl. Phys. B*, 796:137–183, 2008.

[279] Christopher J. Glosser and Carl R. Schmidt. Next-to-leading corrections to the Higgs boson transverse momentum spectrum in gluon fusion. *JHEP*, 12:016, 2002.

[280] V. Ravindran, J. Smith, and W. L. Van Neerven. Next-to-leading order QCD corrections to differential distributions of Higgs boson production in hadron hadron collisions. *Nucl. Phys. B*, 634:247–290, 2002.

[281] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. *CoRR*, abs/1703.05192, 2017.

[282] Zihan Ding, Xiao-Yang Liu, Miao Yin, Wei Liu, and Linghe Kong. TGAN: deep tensor generative adversarial nets for large image generation. *CoRR*, abs/1901.09953, 2019.

[283] Jingwei Guan, Cheng Pan, Songnan Li, and Dahai Yu. SRDGAN: learning the noise prior for super resolution with dual generative adversarial networks. *CoRR*, abs/1903.11821, 2019.

[284] Jyh-Jing Hwang, Sergei Azernikov, Alexei A. Efros, and Stella X. Yu. Learning beyond human expertise with generative models for dental restorations. *CoRR*, abs/1804.00064, 2018.

[285] Matheus Gadelha, Subhransu Maji, and Rui Wang. 3d shape induction from 2d views of multiple objects. *CoRR*, abs/1612.05872, 2016.

[286] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. *CoRR*, abs/1610.07584, 2016.

[287] Kevin Lin, Dianqi Li, Xiaodong He, Zhengyou Zhang, and Ming-Ting Sun. Adversarial ranking for language generation. *CoRR*, abs/1705.11001, 2017.

[288] Chin-Cheng Hsu, Hsin-Te Hwang, Yi-Chiao Wu, Yu Tsao, and Hsin-Min Wang. Voice conversion from unaligned corpora using variational autoencoding wasserstein generative adversarial networks. *CoRR*, abs/1704.00849, 2017.

[289] Aidan N. Gomez, Sicong Huang, Ivan Zhang, Bryan M. Li, Muhammad Osama, and Lukasz Kaiser. Unsupervised cipher cracking using discrete gans. *CoRR*, abs/1801.04883, 2018.

[290] Martín Abadi and David G. Andersen. Learning to protect communications with adversarial neural cryptography. *CoRR*, abs/1610.06918, 2016.

[291] Denis Volkhonskiy, Ivan Nazarov, Boris Borisenko, and Evgeny Burnaev. Steganographic generative adversarial networks. *CoRR*, abs/1703.05502, 2017.

[292] Haichao Shi, Jing Dong, Wei Wang, Yinlong Qian, and Xiaoyu Zhang. SSGAN: secure steganography based on generative adversarial networks. *CoRR*, abs/1707.01613, 2017.

[293] Kevin Alex Zhang, Alfredo Cuesta-Infante, Lei Xu, and Kalyan Veeramachaneni. Steganogan: High capacity image steganography with gans. *CoRR*, abs/1901.03892, 2019.

[294] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. *Comput. Softw. Big Sci.*, 1(1):4, 2017.

[295] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev. D*, 97(1):014021, 2018.

[296] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters. *Phys. Rev. Lett.*, 120(4):042003, 2018.

[297] Anja Butter, Tilman Plehn, and Ramon Winterhalder. How to GAN LHC Events. *SciPost Phys.*, 7(6):075, 2019.

[298] Anja Butter, Tilman Plehn, and Ramon Winterhalder. How to GAN Event Subtraction. 12 2019.

[299] Anja Butter, Sascha Diefenbacher, Gregor Kasieczka, Benjamin Nachman, and Tilman Plehn. GANplifying event samples. *SciPost Phys.*, 10(6):139, 2021.

[300] Jesús Arjona Martínez, Thong Q. Nguyen, Maurizio Pierini, Maria Spiropulu, and Jean-Roch Vlimant. Particle Generative Adversarial Networks for full-event simulation at the LHC and their application to pileup description. *J. Phys. Conf. Ser.*, 1525(1):012081, 2020.

[301] Marco Bellagente, Anja Butter, Gregor Kasieczka, Tilman Plehn, and Ramon Winterhalder. How to GAN away Detector Effects. *SciPost Phys.*, 8(4):070, 2020.

[302] Sofia Vallecorsa, Federico Carminati, and Gulrukh Khattak. 3D convolutional GAN for fast simulation. *EPJ Web Conf.*, 214:02010, 2019.

[303] C. Ahdida et al. Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks. *JINST*, 14:P11028, 2019.

[304] Stefano Carrazza and Frédéric A. Dreyer. Lund jet images from generative and cycle-consistent adversarial networks. *Eur. Phys. J. C*, 79(11):979, 2019.

[305] Joshua Lin, Wahid Bhimji, and Benjamin Nachman. Machine Learning Templates for QCD Factorization in the Search for Physics Beyond the Standard Model. *JHEP*, 05:181, 2019.

[306] Riccardo Di Sipio, Michele Faucci Giannelli, Sana Ketabchi Haghighat, and Serena Palazzo. DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC. *JHEP*, 08:110, 2019.

[307] Bobak Hashemi, Nick Amin, Kaustuv Datta, Dominick Olivito, and Maurizio Pierini. LHC analysis-specific datasets with Generative Adversarial Networks. 1 2019.

[308] Viktoria Chekalina, Elena Orlova, Fedor Ratnikov, Dmitry Ulyanov, Andrey Ustyuzhanin, and Egor Zakharov. Generative Models for Fast Calorimeter Simulation: the LHCb case. *EPJ Web Conf.*, 214:02034, 2019.

[309] Aishik Ghosh. Deep generative models for fast shower simulation in ATLAS. *J. Phys. Conf. Ser.*, 1525(1):012077, 2020.

[310] F. Carminati, A. Gheata, G. Khattak, P. Mendez Lorenzo, S. Sharan, and S. Vallecorsa. Three dimensional Generative Adversarial Networks for fast simulation. *J. Phys. Conf. Ser.*, 1085(3):032016, 2018.

[311] S. Vallecorsa. Generative models for fast simulation. *J. Phys. Conf. Ser.*, 1085(2):022005, 2018.

[312] Kaustuv Datta, Deepak Kar, and Debarati Roy. Unfolding with Generative Adversarial Networks. 6 2018.

[313] Pasquale Musella and Francesco Pandolfi. Fast and Accurate Simulation of Particle Detectors Using Generative Adversarial Networks. *Comput. Softw. Big Sci.*, 2(1):8, 2018.

[314] Martin Erdmann, Lukas Geiger, Jonas Glombitza, and David Schmidt. Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks. *Comput. Softw. Big Sci.*, 2(1):4, 2018.

[315] Kamil Deja, Tomasz Trzcinski, and Łukasz Graczykowski. Generative models for fast cluster simulations in the TPC for the ALICE experiment. *EPJ Web Conf.*, 214:06003, 2019.

[316] Martin Erdmann, Jonas Glombitza, and Thorben Quast. Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network. *Comput. Softw. Big Sci.*, 3(1):4, 2019.

[317] Jan M. Pawlowski and Julian M. Urban. Reducing Autocorrelation Times in Lattice Simulations with Generative Adversarial Networks. *Mach. Learn. Sci. Tech.*, 1:045011, 2020.

[318] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Controlling Physical Attributes in GAN-Accelerated Simulation of Electromagnetic Calorimeters. *J. Phys. Conf. Ser.*, 1085(4):042017, 2018.

[319] Steven Farrell, Wahid Bhimji, Thorsten Kurth, Mustafa Mustafa, Deborah Bard, Zarija Lukic, Benjamin Nachman, and Harley Patton. Next Generation Generative Neural Networks for HEP. *EPJ Web Conf.*, 214:09005, 2019.

[320] Dawit Belayneh et al. Calorimetry with deep learning: particle simulation and reconstruction for collider physics. *Eur. Phys. J. C*, 80(7):688, 2020.

[321] Erik Buhmann, Sascha Diefenbacher, Engin Eren, Frank Gaede, Gregor Kasieczka, Anatolii Korol, and Katja Krüger. Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed. *Comput. Softw. Big Sci.*, 5(1):13, 2021.

[322] Y. Alanazi et al. AI-based Monte Carlo event generator for electron-proton scattering. 8 2020.

[323] Sascha Diefenbacher, Engin Eren, Gregor Kasieczka, Anatolii Korol, Benjamin Nachman, and David Shih. DCTRGAN: Improving the Precision of Generative Models with Reweighting. *JINST*, 15(11):P11004, 2020.

[324] Raghav Kansal, Javier Duarte, Breno Orzari, Thiago Tomei, Maurizio Pierini, Mary Touranakou, Jean-Roch Vlimant, and Dimitrios Gunopulos. Graph Generative Adversarial Networks for Sparse Data Generation in High Energy Physics. In *34th Conference on Neural Information Processing Systems*, 11 2020.

[325] A. Maevskiy, F. Ratnikov, A. Zinchenko, and V. Riabov. Simulating the time projection chamber responses at the MPD detector using generative adversarial networks. *Eur. Phys. J. C*, 81(7):599, 2021.

[326] Yue Shi Lai, Duff Neill, Mateusz Płoskoń, and Felix Ringer. Explainable machine learning of the underlying physics of high-energy particle collisions. 12 2020.

[327] Suyong Choi and Jae Hoon Lim. A Data-driven Event Generator for Hadron Colliders using Wasserstein Generative Adversarial Network. *J. Korean Phys. Soc.*, 78(6):482–489, 2021.

[328] Florian Rehm, Sofia Vallecorsa, Vikram Saletore, Hans Pabst, Adel Chaibi, Valeriu Codreanu, Kerstin Borras, and Dirk Krücker. Reduced Precision Strategies for Deep Learning: A High Energy Physics Generative Adversarial Network Use Case. 3 2021.

[329] Florian Rehm, Sofia Vallecorsa, Kerstin Borras, and Dirk Krücker. Physics Validation of Novel Convolutional 2D Architectures for Speeding Up High Energy Physics Simulations. *EPJ Web Conf.*, 251:03042, 2021.

[330] Thabang Lebese, Bruce Mellado, and Xifeng Ruan. The use of Generative Adversarial Networks to characterise new physics in multi-lepton final states at the LHC. 5 2021.

[331] Raghav Kansal, Javier Duarte, Hao Su, Breno Orzari, Thiago Tomei, Maurizio Pierini, Mary Touranakou, Jean-Roch Vlimant, and Dimitrios Gunopulos. Particle Cloud Generation with Message Passing Generative Adversarial Networks. 6 2021.

[332] Ramon Winterhalder, Marco Bellagente, and Benjamin Nachman. Latent Space Refinement for Deep Generative Models. 6 2021.

[333] Florian Rehm, Sofia Vallecorsa, Kerstin Borras, and Dirk Krücker. Validation of Deep Convolutional Generative Adversarial Networks for High Energy Physics Calorimeter Simulations. 3 2021.

[334] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[335] Saiprasad Koturwar and Shabbir Merchant. Weight initialization of deep neural networks(dnns) using data statistics. *CoRR*, abs/1710.10570, 2017.

[336] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[337] Amitabh Basu, Soham De, Anirbit Mukherjee, and Enayat Ullah. Convergence guarantees for rmsprop and ADAM in non-convex optimization and their comparison to nesterov acceleration on autoencoders. *CoRR*, abs/1807.06766, 2018.

[338] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12(null):2121–2159, July 2011.

[339] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[340] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[341] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. 2014.

[342] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. 2014.

[343] Divya Saxena and Jiannong Cao. Generative adversarial networks (gans): Challenges, solutions, and future directions. 2020.

[344] Maciej Wiatrak, Stefano V. Albrecht, and Andrew Nystrom. Stabilizing generative adversarial networks: A survey. 2020.

[345] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng. Recent progress on generative adversarial networks (gans): A survey. *IEEE Access*, 7:36322–36333, 2019.

[346] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for gans do actually converge? 2018.

[347] Duhyeon Bang and Hyunjung Shim. Mggan: Solving mode collapse using manifold guided training. 2018.

[348] Sanjeev Arora and Yi Zhang. Do gans actually learn the distribution? an empirical study. *CoRR*, abs/1706.08224, 2017.

[349] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang. Generalization and equilibrium in generative adversarial nets (gans). *CoRR*, abs/1703.00573, 2017.

[350] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

[351] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. 2016.

[352] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *CoRR*, abs/1606.03657, 2016.

[353] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. *CoRR*, abs/1809.11096, 2018.

[354] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. 2017.

[355] A. Accardi, W. Melnitchouk, J. F. Owens, M. E. Christy, C. E. Keppel, L. Zhu, and J. G. Morfin. Uncertainties in determining parton distributions at large x. *Phys. Rev. D*, 84:014008, 2011.

[356] Pedro Jimenez-Delgado and Ewald Reya. Delineating parton distributions and the strong coupling. *Phys. Rev. D*, 89(7):074049, 2014.

[357] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[358] Stefano Carrazza, José I. Latorre, Juan Rojo, and Graeme Watt. A compression algorithm for the combination of pdf sets, 2015.

[359] J. R. Andersen et al. Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report. In *9th Les Houches Workshop on Physics at TeV Colliders*, 5 2016.

[360] Stefano Carrazza, Juan M. Cruz-Martinez, and Tanjona R. Rabemananjara. Compressing PDF sets using generative adversarial networks. *Eur. Phys. J. C*, 81(6):530, 2021.

[361] Frank J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[362] Lloyd S. Nelson. The anderson-darling test for normality. *Journal of Quality Technology*, 30(3):298–299, 1998.

[363] Adam Slowik and Halina Kwasnicka. Evolutionary algorithms and their applications to engineering problems. *Neural Computing and Applications*, 32:12363–12379, 08 2020.

[364] Darrell Whitley, Soraya Rana, John Dzubera, and Keith E. Mathias. Evaluating evolutionary algorithms. *Artificial Intelligence*, 85(1):245–276, 1996.

[365] Miftah Bedru Jamal, Fang Ming, and Jiang Zhengang. Solving constrained optimization problems by using covariance matrix adaptation evolutionary strategy with constraint handling methods. In *Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence*, ICIAI '18, page 6–15, New York, NY, USA, 2018. Association for Computing Machinery.

[366] Christian Igel, Nikolaus Hansen, and Stefan Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28, 03 2007.

[367] Scott M. Le Grand and K. Merz. The application of the genetic algorithm to the minimization of potential energy functions. *Journal of Global Optimization*, 3:49–66, 1993.

[368] Sourabh Katoch, S. Chauhan, and Vijay Kumar. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, pages 1 – 36, 2021.

[369] John McCall. Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics*, 184(1):205–222, 2005. Special Issue on Mathematics Applied to Immunology.

[370] Tanjona R. Rabemananjara, Juan M. Cruz-Martinez, and Stefano Carrazza. N3pdf/pycompressor: pycompressor v1.1.0, 2020.

[371] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks. 2018.

[372] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. 2017.

[373] Thomas Pinetz, Daniel Soukup, and Thomas Pock. On the estimation of the wasserstein distance in generative models. 2019.

[374] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. 2017.

[375] James Bergstra, R. Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for Hyper-Parameter Optimization. In J. Shawe-Taylor, R.S. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *25th Annual Conference on Neural Information Processing Systems (NIPS 2011)*, volume 24 of *Advances in Neural Information Processing Systems*, Granada, Spain, December 2011. Neural Information Processing Systems Foundation.

[376] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. 2018.

[377] James Bergstra, Daniel Yamins, and David Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 115–123, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

[378] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[379] Francois Chollet et al. Keras, 2015.

[380] Antonio Gulli and Sujit Pal. *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[381] Valerio Bertone, Stefano Carrazza, and Juan Rojo. APFEL: A PDF Evolution Library with QED corrections. *Comput. Phys. Commun.*, 185:1647–1668, 2014.

[382] John M. Campbell, R. Keith Ellis, and Francesco Tramontano. Single top production and decay at next-to-leading order. *Phys. Rev. D*, 70:094012, 2004.

[383] J. R. Andersen et al. Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report. In *9th Les Houches Workshop on Physics at TeV Colliders*, 5 2016.

[384] Stefano Carrazza, Stefano Forte, Zahari Kassabov, and Juan Rojo. Specialized minimal PDFs for optimized LHC calculations. *Eur. Phys. J. C*, 76(4):205, 2016.

[385] Stefano Carrazza and Juan Cruz-Martinez. Towards a new generation of parton densities with deep learning models. *Eur. Phys. J. C*, 79(8):676, 2019.

[386] D.W. BRADLEY, R.E. LANDRY, and C.T. COLLINS. *The Use of Jackknife Confidence Intervals with the Richards Curve for Describing Avian Growth Patterns*. 1984.

[387] A. Ralph Gibson, Allan J. Baker, and Abdul Moeed. Morphometric Variation in Introduced Populations of the Common Myna (Acridotheres tristis): An Application of the Jackknife to Principal Component Analysis, 12 1984.

[388] Scott M. Lanyon. Molecular Perspective on Higher-Level Relationships in the Tyrannoidea (Aves), 12 1985.

[389] Wesley E. Lanyon and Scott M. Lanyon. Generic Status of Euler's Flycatcher: A Morphological and Biochemical Study, 04 1986.

[390] B. Efron. Bootstrap methods: Another look at the jackknife, 01 1979.

[391] Y. Hamamoto, S. Uchimura, and S. Tomita. A bootstrap technique for nearest neighbor classifier design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):73–79, 1997.

[392] Thomas J. Diciccio and Joseph P. Romano. A review of bootstrap confidence intervals, 1988.

[393] M. H. Quenouille. Notes on bias in estimation, 1956.

[394] Rupert G. Miller. A trustworthy jackknife, 12 1964.

[395] William C. Parr and William R. Schucany. The jackknife: A bibliography, 1980.

[396] P Diaconis and B Efron. Computer-intensive methods in statistics, 6 1983.

[397] Bradley Efron and Gail Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation, 1983.

LIST OF PUBLICATIONS

As of November 18th, 2021

Refereed publications

Compressing PDF sets using generative adversarial networks
Eur.Phys.J.C 81 (2021) 6, 530 - arXiv:2104.04535
Stefano Carrazza, **Tanjona R. Rabemananjara**, Juan Cruz-Martinez

Phenomenology of combined resummation for Higgs and Drell-Yan
Nucl.Part.Phys.Proc. 312-317 (2021) 27-31 - arXiv:2011.13842
Tanjona R. Rabemananjara

Publications in preparation

NNNPDF3.0*: Evidence for gluon shadowing*
Rabah Abdul Khalek, Rhorry Gauld, Tommaso Giani, Emanuele R. Nocera, **Tanjona R. Rabemananjara**, Juan Rojo

Publications in conference proceedings

Small-p_T and large-x regions for Higgs transverse momentum distributions
Journal of High Energy Physics 12 (2020) 073 - arXiv:2007.09164
Tanjona R. Rabemananjara

OPEN SOURCE SOFTWARES

• CHAPTER 3:

HPT-MON: github/N3PDF/HpT-MON

Stands for Higgs p_T Distribution in Momentum and N space. As its name suggests, it computes the partonic and hadronic Higgs cross sections from a gluon fusion (in pp collision) up to NLO both in the momentum x and Mellin N space.

HPT-N3LO: github/N3PDF/HpT-N3LO

Implements the expansion of the small-p_T and large-x resummation. Interfaced with HPT-MON, it approximates the NNLO higgs p_T distribution by approximating the extra-order with the consistent matching of both resummations (will be extended to DY).

• CHAPTER 5:

GANPDFS: github/N3PDF/GanPDFs

Given an input Monte Carlo PDF replicas, it can generate synthetic replicas that follow the same underlying probability distribition as the prior. It relies on machine learning techniques based on adversarial generative models.

PYCOMPRESSOR: github/N3PDF/pyCompressor

Given an input Monte Carlo PDF replicas, it extracts a smaller subsets that reprocudes best the statistics of the prior. Interfaced with GANPDFS, it can enhances the statistics of the input Monte Carlo prior to compression.

ACKNOWLEDGMENTS

> Science is a field which grows continuously with ever expanding frontiers. Furthermore, it is truly international in scope. Any particular advance has been preceded by the contributions of those from many lands who have set firm foundations for further developments ... Further, science is a collaborative effort. The combined results of several people working together is often much more effective than could be that of an individual scientist working alone.
>
> John Bardeen's speech at the Nobel Banquet in Stockholm, December 10, 1972

In these last pages, I would like to take the opportunity to thank the people that have helped me during these past three years, both at a personal and professional level. This work would not have been possible without the support of various people.

First and foremost, I would like to express my sincere gratitude to my supervisor, Stefano Forte. It was a great pleasure and an honour for me to be part of your group. Not only you gave me the opportunity to pursue a PhD in your group, but most importantly you introduced me to numerous exciting physics' topics. In particular, thanks for initiating me to the field of Machine Learning and to the vast and complicated world of all-order computations (aka resummation). I truly appreciate all the trust and freedom you granted me from the start, let alone the useful advice and guidance you gave me throughout the years that have helped me grow as an aspiring scientist. In addition to those, you have also given me opportunities to explore the world of physics through schools, workshops, and conferences. Your door was always open no matter how busy you were–I would not have wished for better supports. I also highly benefited from my co-supervisor, Stefano Carrazza. I am extremely indebted to you for the constant support, for being a guide in the academic world, and for always inspiring me to pursue various research interests.

Particular thanks goes to Petra Dell'Arme for the vital and decisive role you had played throughout this journey. In literal terms, I would not have accomplished anything if it was not for your help, which started even before I arrived to Milan. Your impressive efficiency in dealing with any kind of bureaucracy has made not only my life but I am sure also to the rest of group infinitely easier and less cumbersome.

I would also like to express my sincere gratitude to the members of the Doctoral committee for reviewing this thesis and providing vital comments and suggestions. In no particular order, those are Marco Bonvini, Riccardo Torre, Juan Rojo, and Tilman Plehn. I

hereby extend my gratitude to Juan Rojo for the opportunity he is giving me to pursue research at his group.

I am deeply indebted to all the members of the N3PDF group. First, to those who started the journey at the same time as me: Christopher Schwan, Juan M. Cruz-Martinez, and Jesus Urtasun Elizari. Christopher, I have always admired your meticulous attention to details–no wonder why you are using GENTOO as a GNU/Linux distribution. Thanks for proofreading all my papers, for all the wise and thoughtful advices–be they related to physics, coding, or life in general. Juan, you have taught me a lot about the art of programming, most of the tools I am currently using were inspired by your workflow. I have learnt a lot from the way you approach a problem and your effectiveness in completing a project. A very special thanks to both of you for playing such a vital role in my learning process. Thanks to Jesus who shared with me the ups and downs of being a PhD student in Theoretical Physics from the very beginning. Second, to the members that have joined later: Felix Hekhorn, Kirill Kudashkin, Alessandro Candido, and Roy Stegeman. Despite the fact that we have only been interacting for a brief period (thanks to COVID-19), I have learnt a lot from each one of you. Felix, your kindness and ever-willingness to help has always been a source of inspiration in the often harsh academic world. No matter the amount of work you have to do, you always manage to find time to help. I am also indebted to you for having proofread this thesis. Kirill, I have also benefited a lot from you by learning how you approach difficult problems. Alessandro, I have always admired your vast knowledge of computer science and tools for programming. I have no doubt that this will play a major role in your future endeavour. Last but not the least, Roy, with whom I spent most time with, countless hours on Zoom discussing (combined) resummation. Your calm and reassuring presence has always made a difference in the darkest moment. Thanks for sharing with me this journey, it would have definitely been much harder if you had not been part of it. Thanks to all of you for making the experience so much worthwhile, in particular, for all the extra-activities (movie sessions and late Apperitivos). I have to admit that I really enjoyed those moments and I will definitely miss your company a lot.

At the department, I have also benefited immeasurable supports–both at the professional and logistic level–from Giancarlo Ferrera and Alessandro Vicini. In particular, I would like to express my gratitude to the Giancarlo Ferrera for explaining to me various details on transverse momentum resummation. I would also like to acknowledge the people who made my daily life at the department a really exciting and enjoyable experience. Those who were present since the beginning: Narayan, Claudia, Francesco, Benedetta, Maria Giulia, Martina, Federico, Giulio, Mrunali; and those who joined later: Raoul Röntsch, Simone, Davide, Niccolò, and Andrea. The lunch time was one of the few moments in which we all come together. This was also an opportunity for me to learn more about the Italian culture (and cuisine). I am grateful to Claudia and Martina for inviting us to join their Astrophysics group that later on led to this amazing group.

I wish also to thank all the friends that I have met during winter and summer schools and all the people that I have interacted with in Milan. Particularly, I would like to express my sincere gratitude to Barbara & Davide for their hospitality and generosity. You both made me feel home away from home and made my life in Milan an unforgettable experience.

I would like to close the acknowledgments by expressing my sincere gratitude to my family, who despite the distance are always the closest: my mother, my sister Niry, my brother Hasimbola. Your support, understanding, and unconditional love have always given me strengths to move on. A special thanks goes to my brother Ando, his wife Ony, and their daughter Vahatra for always supporting and guiding me throughout my development. Everything I have achieved was possible because of your help, support, and wise advice. Finally, I would like to thank Angela for always understanding me as no one else does and supporting me at every stage of my life. You have made me the person I have become today and these few words cannot really convey my gratitude. Thank you!

> Three passions, simple but yet overwhelmingly strong, have governed my life: the longing for love, the search for knowledge, and unbearable pity for the suffering of mankind. These passions, like great winds, have blown me hither and thither, in a wayward course, over a great ocean of anguish, reaching to the very verge of despair.
>
> From the Prologue of Bertran Russel's Autobiography entitled **Three Passions**.