Université de Paris
École Doctorale STEP'UP - ED N. 560
Composante Physique de l'Univers
Laboratoire de Physique Théorique et Hautes Energies, UMR 7589, CNRS

en cotutelle avec / in cotutela con

Università degli Studi di Milano - Dipartimento di Fisica
Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata
Ciclo XXXIII
Istituto Nazionale di Fisica Nucleare - Sezione di Milano

Substructure at colliders

Thèse de Doctorat de: / Tesi di Dottorato di:
Giovanni STAGNITTO

en: / in:
Physique de l'Univers / Fisica, Astrofisica e Fisica Applicata

Dirigée par: / Supervisionata da:

Prof. Matteo CACCIARI   Université de Paris
Prof. Stefano FORTE   Università degli Studi di Milano
Dr. Gregory SOYEZ   IPhT, CEA Saclay

Date de soutenance: / Data dell'esame finale: 14/09/2020

Membres du jury: / Membri della giuria:

Prof. Simone MARZANI	Università di Genova	Examinateur (esaminatore)
Prof. Gavin SALAM	University of Oxford	Examinateur (esaminatore)
Prof. Danièle STEER	Université de Paris	Examinatrice (esaminatrice)
Dr. Jean-Philippe GUILLET	CNRS, LAPTh	Rapporteur (referee esterno)
Prof. Michael SPANNOWSKY	University of Durham	Rapporteur (referee esterno)

Cover illustration:

V. Kandinsky, *On the points (Auf Spitzen)*, 1928, Centre Pompidou - Paris

Abstract

In this PhD thesis we investigate several aspects of parton distribution functions (PDFs) and jets as applicable to the physics underpinning the Large Hadron Collider (LHC) as well as future colliders. We first discuss jet observables at the LHC, focusing on the single-jet inclusive cross section. We introduce possible alternative definitions, which weigh the individual contributions coming from each jet in the event and are thus unitary by construction. We also clarify the origin of some problematic aspects of the standard definition. Secondly, within the hadronic PDF fitting framework of the NNPDF collaboration, we investigate the inclusion of single-jet inclusive and dijet measurements into a global PDF fit, using QCD next-to-next-to-leading order predictions for jet processes. We field-test which observables lead to better perturbative stability, better PDF compatibility with other data, better fit quality, and more stringent constraints on the PDFs. Thirdly, we focus on an analytical understanding of machine learning techniques used for quark versus gluon discrimination, a hot topic in jet substructure studies. We construct a new version of the widely used N-subjettiness variable, which features a simpler theoretical behaviour than the original one, while maintaining, if not exceeding, the discriminating power. We input these new observables to the simplest possible neural network, with only one neuron, and we study analytically the network behaviour at leading logarithmic accuracy. We also compare our analytic findings to a more realistic neural network trained with Monte Carlo pseudo-data. Fourthly, we compute the unpolarised electron, positron, and photon PDFs at next-to-leading logarithmic accuracy in QED, which are crucial for high-precision predictions needed for future e^+e^- colliders. We present both numerical and analytical results. The analytical predictions, defined by means of a specific additive formula, provide a large-z analytical solution that includes all orders in the QED coupling constant α, with a small- and intermediate-z solution that includes terms up to $\mathcal{O}(\alpha^3)$. The content of this thesis is based on arXiv:1906.11850, arXiv:1911.12040, arXiv:2005.11327, and arXiv:2007.04319.

Keywords: Particle Physics Phenomenology; Standard Model; Collider Physics; QCD; QED; Parton Distribution Functions; Jets

SOUS-STRUCTURE AU COLLISIONNEUR DE PARTICULES

Résumé

Dans cette thèse de doctorat, nous étudions plusieurs aspects des fonctions de distribution de partons (PDF) et des jets applicables à la physique qui sous-tend le Grand collisionneur de hadrons (LHC) ainsi que les futurs collisionneurs. Nous discutons d'abord des jets observables au LHC, en nous concentrant sur la section transversale inclusive d'un seul jet. Nous introduisons d'autres définitions possibles, qui pèsent les contributions individuelles de chaque jet dans l'événement et sont donc unitaires par construction. Nous clarifions également l'origine de certains aspects problématiques de la définition standard. Deuxièmement, dans le cadre des ajustements de PDF hadroniques de la collaboration NNPDF, nous étudions l'inclusion de mesures de dijet et de d'un seul jet dans un ajustement global des de PDF, en utilisant des prédictions de la QCD à l'ordre sous-sous-dominant pour les processus de jet. Nous testons les observables qui conduisent à une meilleure stabilité perturbative, à une meilleure compatibilité des PDF avec d'autres données, à une meilleure qualité d'ajustement et à des contraintes plus strictes sur les PDF. Troisièmement, nous nous concentrons sur une compréhension analytique des techniques de machine learning utilisées pour la discrimination quark contre gluon, un sujet d'actualité dans les études de sous-structure des jets. Nous construisons une nouvelle version de la variable N-subjettiness, largement utilisée, qui présente un comportement théorique plus simple que l'originale, tout en conservant, voire en dépassant, le pouvoir discriminant. Nous introduisons ces nouvelles observables dans le réseau de neurones le plus simple possible, avec un seul neurone, et nous étudions analytiquement le comportement du réseau avec une précision logarithmique à l'ordre dominant. Nous comparons également nos résultats analytiques à un réseau de neurones plus réaliste formé avec des pseudo-données Monte Carlo. Quatrièmement, nous calculons les PDF d'électrons, de positons et de photons non polarisés avec une précision logarithmique sous-dominante dans la QED, ce qui est crucial pour les prévisions de haute précision nécessaires aux futurs collisionneurs e^+e^-. Nous présentons des résultats à la fois numériques et analytiques. Les prédictions analytiques, définies au moyen d'une formule additive spécifique, fournissent une solution analytique à grand z qui inclut tous les ordres de la constante de couplage de la QED α, avec une solution à petit et moyen z qui inclut des termes jusqu'à $\mathcal{O}(\alpha^3)$. Le contenu de cette thèse est basé sur arXiv:1906.11850, arXiv:1911.12040, arXiv:2005.11327, et arXiv:2007.04319.

Mots clés: Phénoménologie de la physique des particules; modèle standard; physique des collisionneurs; QCD; QED; fonctions de distribution de partons; jets

SOTTO-STRUTTURA AI COLLIDER DI PARTICELLE

Sintesi

In questa tesi di dottorato studiamo diversi aspetti delle funzioni di distribuzione partonica (PDF) e di jet adronici per applicazioni alla fisica del Large Hadron Collider (LHC) e dei futuri collisori di particelle. Per prima cosa discutiamo le osservabili di jet misurate a LHC, concentrandoci sulla sezione d'urto di singolo jet inclusiva. Introduciamo possibili definizioni alternative, che pesano i singoli contributi provenienti da ciascun jet in ogni evento e sono quindi unitarie per costruzione. Chiariamo anche l'origine di alcuni aspetti problematici della definizione standard. In secondo luogo, nell'ambito dei fit delle PDF adroniche della collaborazione NNPDF, esaminiamo l'inclusione di misure di sezioni d'urto di singolo jet e di dijet in un fit di PDF globale, utilizzando le previsioni all'ordine next-to-next-to-leading in QCD per i processi di jet. Questo ci permette di testare sul campo quali osservabili di jet portano a una migliore stabilità perturbativa, a una migliore compatibilità con altri dati sperimentali, a una migliore qualità di fit e a vincoli più rigorosi sulle PDF. In terzo luogo, ci concentriamo su una comprensione analitica delle tecniche di machine learning utilizzate per la discriminazione tra jet originati da quark o gluoni, un tema caldo negli studi sulle sotto-strutture dei jet. Costruiamo una nuova versione della variabile N-subjettiness ampiamente utilizzata, che presenta un comportamento teorico più semplice di quello originale, mantenendo, se non superando, il potere discriminante. Usiamo queste nuove variabili come input della rete neurale più semplice possibile, con un solo neurone, e studiamo analiticamente il comportamento della rete, mantenendo una precisione logaritmica all'ordine leading. In seguito confrontiamo i nostri risultati analitici con una rete neurale più complessa allenata con pseudodati Monte Carlo. In quarto luogo, calcoliamo le PDF di elettrone, positrone e fotone non polarizzati con una precisione logaritmica next-to-leading in QED, ingredienti cruciali per le previsioni ad alta precisione necessarie per i futuri collisori e^+e^-. Presentiamo risultati sia numerici che analitici. Le previsioni analitiche, definite per mezzo di una specifica formula additiva, forniscono una soluzione analitica a grande-z che include tutti gli ordini nella costante di accoppiamento di QED α, combinata con una soluzione a piccolo e intermedio z, che include termini fino a $\mathcal{O}(\alpha^3)$. Il contenuto di questa tesi si basa su arXiv:1906.11850, arXiv:1911.12040, arXiv:2005.11327, e arXiv:2007.04319.

Parole chiave: Fenomenologia della fisica della particelle; Modello standard; Fisica dei collider di particelle; QCD; QED; Funzioni di distribuzione partonica; Jets

Vedi, in questi silenzi in cui le cose
s'abbandonano e sembrano vicine
a tradire il loro ultimo segreto,
talora ci si aspetta
di scoprire uno sbaglio di Natura,
il punto morto del mondo, l'anello che non tiene,
il filo da disbrogliare che finalmente ci metta
nel mezzo di una verità.

Eugenio Montale, *I limoni*, in *Ossi di Seppia* (1925)

La peste che il tribunale della sanità aveva temuto che
potesse entrar con le bande alemanne nel milanese, c'era
entrata davvero, come è noto; ed è noto parimente che non
si fermò qui, ma invase e spopolò una buona parte d'Italia.

Alessandro Manzoni, *I promessi sposi*, Capitolo XXXI

Contents

List of publications

The content of this thesis is based on the following publications:

1. M. Cacciari, S. Forte, D. Napoletano, G. Soyez and G. Stagnitto, *Single-jet inclusive cross section and its definition*, *Phys. Rev.* **D100** (2019) 114015 [1906.11850]

2. V. Bertone, M. Cacciari, S. Frixione and G. Stagnitto, *The partonic structure of the electron at the next-to-leading logarithmic accuracy in QED*, *JHEP* **03** (2020) 135 [1911.12040]

3. R. Abdul Khalek, S. Forte, T. Gehrmann, A. Gehrmann-De Ridder, T. Giani, E. W. N. Glover, A. Huss, E. R. Nocera, J. Pires, J. Rojo and G. Stagnitto, *Phenomenology of NNLO jet production at the LHC and its impact on parton distributions*, *Eur. Phys. J. C* **80** (2020) 797 [2005.11327]

4. G. Kasieczka, S. Marzani, G. Soyez and G. Stagnitto, *Towards Machine Learning Analytics for Jet Substructure*, 2007.04319, accepted by JHEP

Introduction

The idea is very simple: smash particles together and see what comes out. Since the first electron-positron collider built in the 1950s near Rome, particle accelerators have played an essential role both in the development and the validation of the Standard Model (SM) of particle physics. Elaborating a theory of fundamental interaction with the right to be called "standard" required decades of outstanding ideas and powerful advances in technology. Each major achievement in the recent history of particle physics is linked to the construction of a new particle collider: the proton-antiproton collider SPS (1981-1991) discovered the W and Z bosons; the proton-antiproton collider Tevatron (1987-2011) discovered the top quark; the electron-positron collider LEP (1989-2000) tested the electroweak sector of the SM. The latest milestone in the history of particle physics has been the discovery of the Higgs boson in 2012, at the Large Hadron Collider (LHC), a proton-proton collider at CERN.

On the theoretical side, we know that the SM is not the ultimate theory of Nature, and several theories that account for phenomena beyond the SM have been proposed over the past decades. Some of these models rely on the discovery of new particles at the electroweak scale. The conspicuous absence of these new particles rules out a plethora of hypothetical scenarios, at least in their simplest variants, thus making it increasingly difficult to find a "beautiful" theory[1] that extends the SM. Ideally, such an extension would fit two main criteria: it should be able to predict at least as well as the SM the huge amount of experimental collider data, while also taking into account other phenomena as yet unexplained by the SM.

On the experimental side, today's physics community is pushing the boundary of the current *intensity* frontier. LHC has just finished in December 2018 its Run 2 at $\sqrt{s} = 13$ TeV with 140 fb^{-1} of accumulated luminosity. We are currently, and until summer 2021, in the middle of the Long Shutdown 2 (LS2). At the end of the next run, 2021-2025, LHC will have triplicated its total integrated luminosity. The LS3, which is expected from 2025 to 2027, will include a major revision of the detectors and notably the installation of High Luminosity LHC (HL-LHC). This upgrade will permit LHC to deliver between 5 to 7.5 times its former nominal luminosity. As for the *energy* frontier, the status is much less clear: the construction of a future linear or circular collider, with different beam configurations (e^+e^-, ep, or pp), is still under discussion. CERN has just announced in the 2020 Update of the European Strategy for Particle Physics [6] several priorities: a new e^+e^- collider operating as a "Higgs factory", possibly followed by a future hadron collider. In any case, we will most likely need to wait decades before we can see a collider with a significant improvement in the center of mass energy.

The vast amount of data that will be collected by LHC and HL-LHC in the near future will

---

[1]Maybe we just have to change our concept of beauty [5].

require developing novel approaches to look for signs of physics beyond the SM. Two possibilities arise. In the first scenario, we examine the distinctive experimental signatures of new physics signals as predicted by the proposed beyond the SM theories. In the second scenario, we run blind analyses of the data collected by LHC in order to find a possible deviation from the SM background without any theoretical bias.

Both strategies require a detailed analysis of the SM background. We already know that if the LHC data contains a deviation from the SM, it would be tiny, and possibly it would only appear with a huge amount of experimental data, as expected at the end of the HL-LHC runs. At that point, the statistical error will certainly be small, and most likely so will the systematic error, thanks to state-of-the-art detectors and improved analysis techniques. Accordingly, to perform a proper comparison, we need to increase the accuracy of the SM predictions as much as possible, thereby decreasing the associated theoretical error. Only in this way will we be able to bring to light a possible deviation from the SM.[2] The topics presented is this thesis move within this spirit.

What level of accuracy are we talking about? At hadron colliders, relative errors on the experimental measurements can be as small as a few percentage points. The plot in Fig. 1 shows a summary of SM measurements performed by the ATLAS collaboration at LHC. Starting from the top, we see that for some very inclusive cross sections the experimental error is already smaller than the theoretical one. Moving towards the bottom of the plot, the cross sections are less inclusive, and the associated experimental error is still quite large, but certainly going to decrease in the near future. As Fig. 1 shows, the theoretical predictions currently have a relative uncertainty of 5–10%. By the end of the HL-LHC runs we will need this relative uncertainty to go down to 1% in order to properly compare with the experimental data. As for e^+e^- colliders, they are strikingly different from hadron colliders: the cleaner environment favours more precise measurements. Fig. 2 displays the projected experimental error on several electroweak observables at two future e^+e^- colliders — the linear ILC and the circular FCCee — and compares these to the LEP results. The relative error at ILC will reach 0.01% and at FCCee will be possibly even smaller. Accordingly, we need the theory to also be that accurate.

More accurate predictions require calculating higher order terms in perturbative calculations. Such fixed-order predictions, though necessary, are not sufficient for a comparison with the experimental data. This is due to both practical and fundamental reasons. Practical, because fixed-order results fail to account for higher order radiative corrections, which can have a large impact in determinate regions of the phase space, thereby spoiling the convergence of the perturbative series. Fundamental, because quarks and gluons, which possibly enter the fixed-order predictions, live inside composite objects and thus are not directly observable by experiments.

In this thesis, we will focus on the study of two theoretical objects, *parton distribution functions* (PDFs) and *jets*, introduced to complement the fixed-order predictions and solve the two issues mentioned in the preceding paragraph. These objects enter in the computation of almost any observable at colliders and thus are of crucial importance. Fig. 3 depicts a typical LHC event in which two protons collide. A proper description of the initial state requires the introduction of PDFs, whereas jets come into play in the description of strongly interacting

---

[2]Moreover, a better understanding of SM phenomena is a noble purpose in and of itself. For instance, even if we believe we know the fundamental laws of strong interactions, some mechanisms are still unclear. For example, why is the spin of a composite objects such as the proton the observed value?
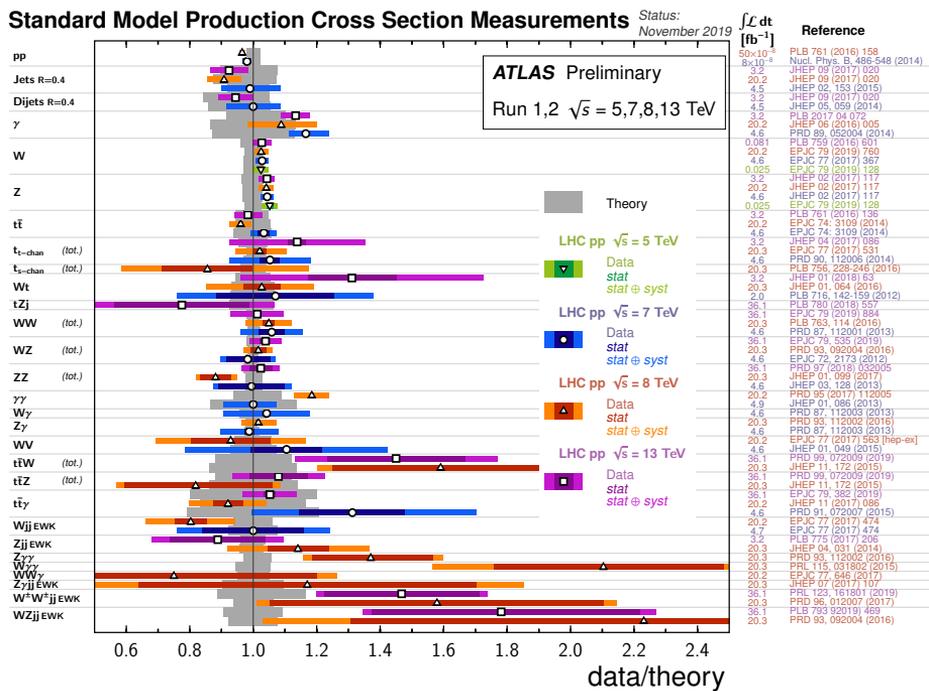
Figure 1: Summary of ratios with respect to best theory for several Standard Model production cross section measurements at the LHC, corrected for branching fractions. Taken from Ref. [7] (ATLAS collaboration).
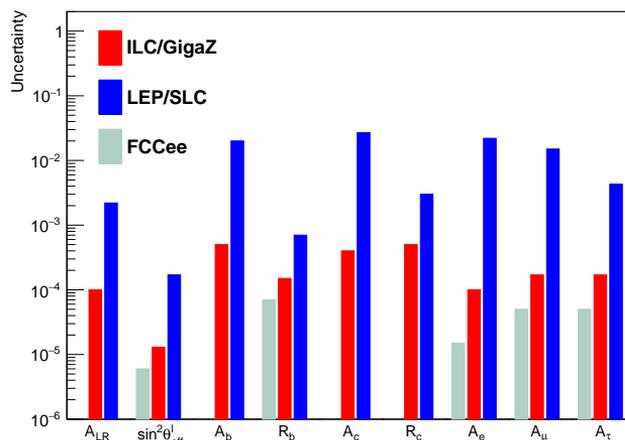
Figure 2: Estimated experimental uncertanties on several precision electroweak observables at future e^+e^- colliders, compared with the final combined results of LEP and SLC. "ILC/GigaZ" refers to the run of the International Linear Collider at the Z pole, while "FCCee" refers to the Future Circolar Collider in the e^+e^- configuration. Taken from Ref. [8].
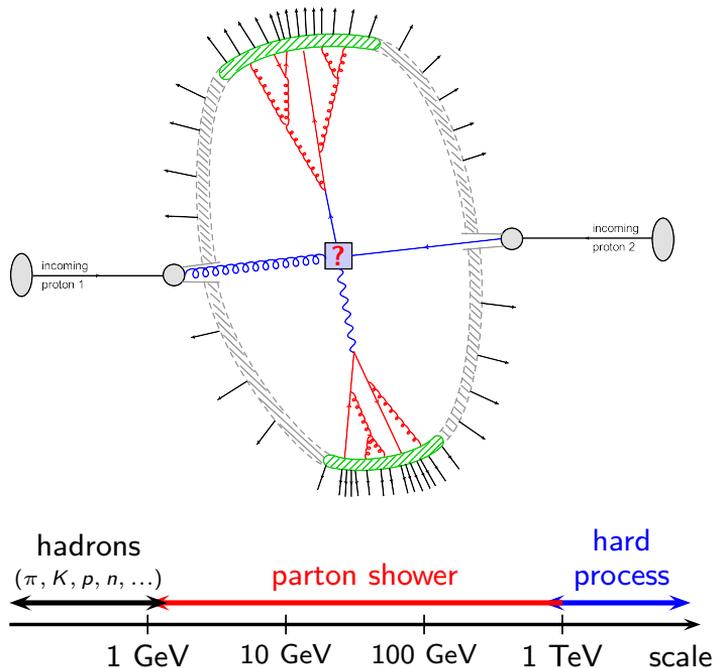
Figure 3: A pictorial representation of the theory behind a dijet collider event. Two partons are extracted from the incoming protons, hard scatter in the question mark box and then progressively loose energy by radiating particles. The colors code the typical energies of the particles involved. Taken from Ref. [9].

final states.[3] Loosely speaking, both PDFs and jets feature a *substructure*: as we increase the resolution scale, we better resolve the internal structure of these objects. Discussing what happens in Fig. 3 gives us the opportunity to talk about PDFs and jets, and comment on their substructure.

Hadronic PDFs Since the parton model proposal, we know that the hard scattering process actually occurs between two partons, quarks or gluons, extracted from the incoming protons. In Fig. 3 the partons are depicted as blue incoming lines. Each parton carries a momentum fraction x of the longitudinal momentum of the parent hadron; a parton distribution function (PDF) describes this dependence on x. PDFs are an essential input for any theory prediction at hadron colliders. However, PDFs have an intrinsic non-perturbative component and cannot be calculated using perturbation theory. Instead, PDFs must be extracted from the experimental data, and that can only be done by means of a fitting procedure.

Actually, a PDF is a function of two variables: the momentum fraction x and a scale Q^2. The scale may be thought of as the resolution of a magnifier. For example, let us suppose that, at a scale Q_0^2, we see a single quark with momentum fraction x coming out of the proton. If

---

[3]QCD particles, and thus jets, may also be present in the final state of e^+e^- collisions. Hadronic PDFs are obviously not present when we collide leptons. However, we can still adopt the same formalism in order to take into account some potentially large contributions, due to collinear photon emissions, to all orders in perturbation theory. See "Electron PDFs" below.

we increase the resolution of our magnifier (i.e. we increase our Q^2), we can start to resolve the quark emitting a gluon. Both partons now carry a momentum fraction less than x. Hence, increasing Q^2 reveals the substructure of the PDFs, and we see more partons with less energy. The Q^2-dependence of the PDFs — unlike their x-dependence — is entirely predictable within perturbative QCD.

Electron PDFs In principle, e^+e^- colliders do not require introducing the concept of PDFs, but it may be useful for practical purposes. Calculations of processes in QED have some large contributions stemming from photon collinear emissions in the initial state. These contributions appear as logarithms (possibly to some power) of some hard physical scale E over the mass of the electron m, $\log^k(E^2/m^2)$. These terms can be numerically large, preventing the perturbative series from being well behaved. Using the PDF formalism allows us to collect such large terms in a universal object, an electron (or positron) PDF, and then resum the logarithms to all order through the Q^2-evolution. In this scenario, the incoming electron is seen as a composite object and acquires a substructure. By looking at the electron with a magnifier we can resolve a photon emission or, at higher order, even a positron emission!

Jet physics In addition to PDFs, we are also interested in jets. When hard QCD particles come out of the scattering region (the blue lines in Fig. 3), they start radiating and progressively loose energy. This radiation process is depicted in red in Fig. 3, and the energy scales involved span several order of magnitude. When they reach the non-perturbative region ($\sim$ 1 GeV), quarks and gluons start recombining into hadrons in the process of hadronization, depicted as a green band in Fig. 3. The final state hadrons, the black arrows in Fig. 3, are the ones actually detected by the experiments. Given that QCD dynamics favours emissions at small angles, most of the detected particles will be distributed around the direction of the original parton that originated the shower. These collimated bunches of energetic particles are called jets and can be considered proxies for the hard scattering partons. Fig. 4 shows a typical event display from the ATLAS collaboration at CERN. In the image, we clearly see two jets emerging from the interaction point. In practice, however, in order to properly count the number of jets present, we need an unambiguous jet definition. If both theory and experiment adopt the same definition, we are in a position to systematically compare theoretical predictions with experimental data.

Jets are abundant at hadron colliders. Total cross sections for jet production at LHC ($\sqrt{s} \sim$ 7–14 TeV) are of the order of 10^7 pb, two orders of magnitude more than W or Z production and four orders of magnitude more than top pair production. In addition, observables involving only jets are the simplest ones with a purely strongly interacting final state. Because of this and because of the abundance of jets at LHC, jet cross sections are particularly well suited for precision QCD studies, such as the determination of the parton distributions and of the strong coupling constant α_s. There are, however, a number of unsettled theoretical issues related to the choice of the most appropriate setup to investigate jet events. To begin with, there are two ways to describe an event with jets in the final state. One possibility is to describe the final state kinematics by adopting individual jet variables (e.g. jet transverse momentum, jet rapidity) and measuring these variables on each jet in the event. An alternative approach is to select only the n hardest jets in the event and use variables suited for the description of the n-jets system as a whole (e.g. dijet mass, rapidity separation). Furthermore, once we have chosen a specific observable, we have to deal with the follow up choice of renormalization and factorization scales. Since jet production is naturally a multiscale process, this is a complex decision.
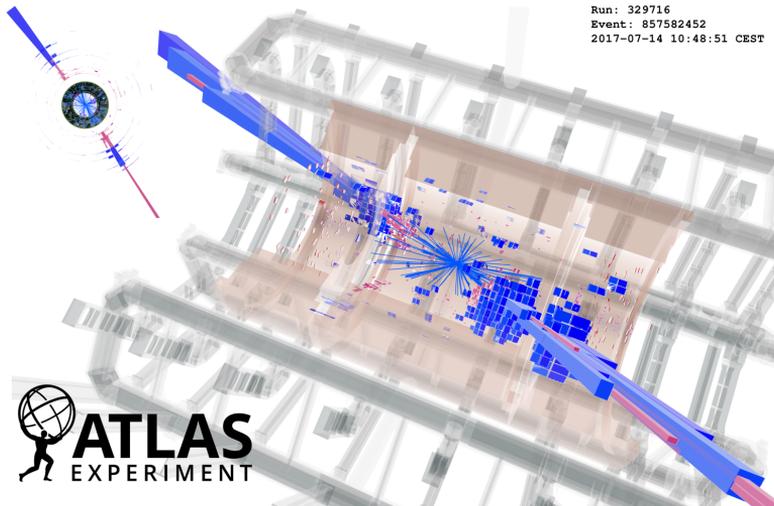
Figure 4: An experimental view of a collider dijet event, produced in pp collisions at $\sqrt{s} = 13$ TeV. The blue towers are proportional to the energy released in the detector. ATLAS Experiment © 2020 CERN

Jet substructure Like PDFs, jets are also naturally endowed with a substructure, and by studying their internal structure we can learn valuable information. For instance, the pattern of radiation around a hard particle is tightly linked to the nature of that particle; the distribution of the radiation in energy and angle is different if the particle is a quark or a gluon. Being able to discriminate between a quark-initiated jet and a gluon-initiated jet is important for a multitude of studies: for the isolation of specific production channels, for precision α_s studies, for PDF extraction, and for the search of new physics.

The large number of variables required for a full description of the radiation pattern around the hard prongs in a jet naturally calls for machine learning (ML) techniques. The field of jet substructure has recently seen an exponential growth in the adoption of these techniques. Neural networks (NN) in particular are used to attack this quark- vs. gluon-jet classification problem. Historically, studies have focused mainly on neural network performance. However, we also need to ensure that the physics we obtain as an output of the ML algorithm are the same as the physics we obtain from resummed QCD calculations.

This thesis studies several aspects of hadronic PDFs, electron PDFs, jet observables, and jet substructure mentioned above. The concept of substructure is the real common denominator among the different topics: as we increase the resolution, we can resolve more quarks and gluons inside the colliding protons, more electrons and photons inside the colliding electrons, or more constituents inside jets. A better understanding of these different kind of substructures allows us to increase the precision of our predictions (PDFs and jets are crucial ingredients in precision calculations of Standard Model cross sections) or better interpret experimental data (jet substructure techniques are important tools for analyses of hadronic final states). We shall conclude this introduction with a description of each chapter's content, highlighting where personal contributions can be found.

Chapter 1: Quantum field theories at colliders

This introductory chapter introduces the reader to the basic elements of QED and QCD relevant for this thesis. We briefly present the running of coupling constants, the problem of infrared divergences, collinear factorization and parton distribution functions, DGLAP equations, jets and their substructure, structure of QCD predictions, and dependence on renormalization and factorization scales.

Chapter 2: Theory calculations for jet processes

For the purposes of Chapter 3 and Chapter 4, we discuss the available theory predictions for jet observables at various different perturbative orders in the QCD coupling constant. We carry out explicit computations at leading order (LO). We then briefly present the available next-to-leading order (NLO) calculations and report on the recent next-to-next-to-leading order (NNLO) calculations. In particular, for single-jet inclusive cross section, we comment on the apparent perturbative instability shown by higher order corrections.

Chapter 3: On the definition of single-jet inclusive cross section (based on M. Cacciari, S. Forte, D. Napoletano, G. Soyez and G. Stagnitto, *Single-jet inclusive cross section and its definition*, *Phys. Rev.* **D100** (2019) 114015 [1906.11850])

We study possible alternative unitary definitions of the single-jet inclusive cross section. These new definitions weigh the individual contributions coming from each jet in the event. The weighted definitions are unitary by construction, whereas the usual standard definition is non-unitary as each event is counted more than once. This chapter also clarifies the origin of some problematic aspects of the standard definition.

Chapter 4: Impact of jet measurements on parton distributions (based on R. Abdul Khalek, S. Forte, T. Gehrmann, A. Gehrmann-De Ridder, T. Giani, E. W. N. Glover, A. Huss, E. R. Nocera, J. Pires, J. Rojo and G. Stagnitto, *Phenomenology of NNLO jet production at the LHC and its impact on parton distributions*, *Eur. Phys. J. C* **80** (2020) 797 [2005.11327])

Within the PDF fitting framework of the NNPDF collaboration, we investigate the inclusion of single-jet inclusive and dijet measurements into a global PDF fit, using QCD NNLO prediction for jet processes. Jet data provide unique constraints on the gluon PDF at large values of the momentum fraction. The fits performed allow us to field-test which observables and which choice of scale leads to better perturbative stability, better PDF compatibility with other data, better fit quality, and more stringent constraints on the PDFs.

Chapter 5: Towards machine learning analytics for jet substructure (based on G. Kasieczka, S. Marzani, G. Soyez and G. Stagnitto, *Towards Machine Learning Analytics for Jet Substructure*, 2007.04319, accepted by JHEP)

In the context of jet substructure studies, we introduce a variant of the N-subjettiness variable, the *primary N-subjettiness* $\mathcal{T}_N$, which is more amenable to an all-order QCD analysis. The primary N-subjettiness is such that, if we measure a set of n variables $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$, at leading logarithmic accuracy a cut on the likelihood ratio — which is the optimal single-variable discriminant — corresponds to a simple cut on $\mathcal{T}_n$. This important property allows us to determine whether a neural network fed with such n primary N-subjettiness variables, after proper training, leads to the same classifier as the one dictated by the likelihood ratio.

Chapter 6: Electron PDFs (based on V. Bertone, M. Cacciari, S. Frixione and G. Stagnitto, *The partonic structure of the electron at the next-to-leading logarithmic accuracy in QED*, *JHEP* **03** (2020) 135 [1911.12040])

Working within QED, we obtain the electron, positron, and photon PDFs of the unpolarised electron at the next-to-leading logarithmic (NLL) accuracy in the $\overline{\text{MS}}$ subtraction scheme. We present both numerical and analytical results. The analytical predictions are defined by means of a specific additive formula. This formula matches a large-z analytical solution that includes all orders in the QED coupling constant α, with a small- and intermediate-z solution that includes terms up to $\mathcal{O}(\alpha^3)$.

Finally, in the Conclusions, we summarize the main results found in the thesis and discuss possible avenues for future studies.

<div align="right">

CHAPTER 1

</div>

Quantum field theories at colliders

The reason for this chapter is twofold. First, the reader not acquainted with the topics of the thesis will find here some introduction to the main concepts covered in the manuscript. We hope that these pages can also become useful for future graduate students taking the first steps in this field of research. Second, we set up the notation used in the rest of the thesis, and in the subsequent chapters we shall often refer to equations presented here.

The beauty behind quantum field theories will hardly emerge in our discussions. We just recall the main results and comment on them, without providing (almost) any proof. For an extensive and more detailed treatment, we refer the reader to standard textbooks on quantum field theory [10–13], to books more focused on perturbative QED/QCD [14–18], and to specific lecture notes [19–23].

The outline of the chapter is as follows. In Sec. 1.1 we discuss the main similarities and differences between QED and QCD, by also showing how the coupling constant runs in both theories. In Sec. 1.2 we present the master factorization formula for processes at hadron colliders, and we introduce the parton distribution functions (PDFs). We then move to the problem of infrared divergences: in Sec. 1.3 we show how matrix elements behave in the *soft* and *collinear* limits — specifying what this nomenclature means — and in Sec. 1.4 we explain why infrared divergences should (or not) cancel in physical results. Starting with Sec. 1.5, we concentrate on PDFs. We first introduce the collinear factorization procedure for the absorption of initial state collinear divergences in the PDFs. We also show how the DGLAP equations govern the factorization scale dependence of PDFs (Sec. 1.6), and how one can practically extract the proton PDFs from experimental data (Sec. 1.7). We then turn to jets: in Sec. 1.8 we present the clustering algorithms usually adopted to define jets and in Sec. 1.9 we introduce the reader to jet substructure techniques. Finally, Sec. 1.10 discusses the structure of pertubative QCD predictions, especially in relation to the choice of the renormalization and factorization scale.

1.1 QED versus QCD

Quantum electrodynamics (QED) and quantum chromodynamics (QCD) are the gauge field theories of electromagnetic interactions and strong interactions, respectively. As quantum field theories, they are based on different symmetry groups. QED is based on the U(1) symmetry group, while QCD is based on the SU(N_c) symmetry group, with $N_c = 3$. The charge of strong interaction is called *colour*, in analogy to the three main colours: red, green and blue, which sum up to give the "colourless" white. Only colour-singlet combinations of quarks exist in Nature[1]. The Lagrangian densities of QED and QCD are reported in Appendix 1.A.

The parameter defining the strength of the interaction between coloured or charged particles is the *coupling constant*. We denote with g_S the QCD coupling constant, while e is the QED coupling constant. For historical reasons, one also defines the fine structure constant $\alpha_{\rm em}$:

$$\alpha_{\rm em} = \frac{e^2}{4\pi}\,, \tag{1.1}$$

and, by analogy, the strong coupling constant α_S:

$$\alpha_S = \frac{g_S^2}{4\pi}\,. \tag{1.2}$$

Note that it is customary to simply denote $\alpha_{\rm em}$ as α. Throughout this thesis, this will be clear in the context, and in case of ambiguities we will adopt the $\alpha_{\rm em}$ notation.

Starting with the Lagrangian densities, we can derive the Feynman rules of the theory and calculate amplitudes $\mathcal{M}$, and eventually take the product $\mathcal{M}^*\mathcal{M}$ to obtain the matrix element squared $|\mathcal{M}|^2$ which enters in physical quantities. By doing so, in QCD one finds recurrent expressions in colour algebra involving products of colour matrices t_{ab}^C or structure constants f_{ABC}, which can be simplified in the following way:

$$\sum_{a,b} t_{ab}^A t_{ab}^B = T_R \delta_{AB}\,, \quad \sum_A t_{ab}^A t_{bc}^A = C_F \delta_{ac}\,, \quad \sum_{A,B} f_{ABC} f_{ABD} = C_A \delta_{CD}\,, \tag{1.3}$$

with

$$T_R = \frac{1}{2}\,, \quad C_F = \frac{N_c^2 - 1}{2N_c} \xrightarrow{N_c=3} \frac{4}{3}\,, \quad C_A = N_c \xrightarrow{N_c=3} 3\,. \tag{1.4}$$

These constants have a nice physical interpretation: C_F is the colour factor associated to a gluon emission of a quark; C_A is the colour factor associated to a gluon emission off a gluon; finally, T_R is the colour factor associated to a gluon splitting in a quark-antiquark pair.

By comparing the Lagrangian densities, one finds a practical recipe for obtaining QED results starting with QCD expressions:

$$g_S^2 \to e^2\,, \quad T_R \to Q_i^2\,, \quad C_F \to Q_i^2\,, \quad C_A \to 0\,, \tag{1.5}$$

with Q_i the electric charge of each particle in units of the positron charge. Eq. (1.5) is motivated by the fact that Q_i is the formal analogue of t_{ab}^A and the QED structure constants are null by definition. However, this prescription should be interpreted *cum grano salis*, notably in the case

---

[1]The *confinement* of quarks inside colour-singlet composite objects, even though not rigorously proven in QCD, may be naively understood as a consequence of the behaviour of the running coupling α_S at small energies, see Sec. 1.1.2.

of sums over quark flavours: we need to make sure to proper take into account the fractional charge of quarks, and the fact that quarks come in three colours.

1.1.1 Renormalization procedure

As it is well known, once we go beyond the tree-level approximation, the *bare* Lagrangians in Appendix 1.A suffer from divergences of ultraviolet (UV) origin i.e. coming from loop integrations which involve arbitrarily large momenta. Is it customary to renormalize the theory by splitting the bare Lagrangian in two terms:

$$\mathcal{L}_{\text{bare}} = \mathcal{L}_{\text{R}} + \mathcal{L}_{\text{c.t.}} \tag{1.6}$$

where the renormalized Lagrangian $\mathcal{L}_{\text{R}}$ is formally the same as $\mathcal{L}_{\text{bare}}$, but now it contains only physical parameters and fields, while the counterterm Lagrangian $\mathcal{L}_{\text{c.t.}}$ provides additional Feynman graphs needed to cancel UV divergences.

In order to isolate the divergences, we need to regularize loop integrations. This is usually done through *dimensional regularization*, by shifting the momentum integration from 4 to $d = 4 - 2\epsilon$ dimensions, whereby the integral converges. This allows us to parametrize the degrees of divergence of the loop integration as simple poles proportional to $1/\epsilon$:

$$g^2 \int \frac{d^4 q}{\dots} \longrightarrow g^2 \widetilde{\mu}^{2\epsilon} \int \frac{\mathrm{d}^{4-2\epsilon} q}{\dots} = g^2 \widetilde{\mu}^{2\epsilon} \left[\frac{C_{-1}}{\epsilon} + C_0 + C_1 \epsilon + \mathcal{O}(\epsilon^2) \right] \tag{1.7}$$

Note that a scale $\widetilde{\mu}$ appears, as we need to maintain the correct mass dimension of the overall result. We usually refer to this scale as *regularization scale*. The final result at the end of the renormalization procedure will not depend on this scale[2].

While the cancellation of divergences between $\mathcal{L}_{\text{bare}}$ and $\mathcal{L}_{\text{c.t.}}$ fixes the divergent parts of the counterterms, the finite terms included in $\mathcal{L}_{\text{c.t.}}$ are arbitrary. Different conventions for the finite part of the counterterms are known as different *subtractions schemes*. In the *minimal subtraction* scheme (MS) the finite part of each counterterm is set to zero. In the *modified minimal subtraction* scheme ($\overline{\text{MS}}$) a $\ln 4\pi - \gamma_E$ finite term[3], which usually appears in the evaluation of integrals in d dimensions, is included in the definition of the counterterms.

The counterterms are defined at a *renormalization scale* μ. In the MS or $\overline{\text{MS}}$ scheme, the renormalization scale μ and the regularization scale $\widetilde{\mu}$ are usually made to coincide. As pointed out in Ref. [11], in dimensional regularization the scales $\widetilde{\mu}$ and μ are not in principle the same; but we can always replace μ with $\widetilde{\mu}$ by including in the counterterms appropriate logarithms such as $\ln(\widetilde{\mu}^2/\mu^2)$.

1.1.2 Running coupling

The bare coupling g_0 appearing in the bare Lagrangian $\mathcal{L}_{\text{bare}}$ must be independent of the renormalization scale μ introduced in the renormalization procedure. By imposing the condition

$$\mu \frac{d}{d\mu} g_0 = 0 \,, \tag{1.8}$$

---

[2]We could have alternatively regularized the infinities by introducing a parameter Λ to cutoff the integration region of large momenta. In such a case, the regularization scale would have been Λ.

[3]Throughout this thesis, we will denote the natural logarithm either as ln or log. In case, decimal logarithms will be explicitly written as $\log_{10}$.

one finds that the physical coupling g_R appearing in $\mathcal{L}_R$ retains a dependence on μ. The *running* of the coupling is encoded in the β-function, which is nothing but the renormalization group equation for the coupling constant. The β-function is usually written in the following form[4]:

$$\mu^2 \frac{d\alpha}{d\mu^2} = \frac{d\alpha}{d\ln\mu^2} = \beta(\alpha) = b_0\alpha^2 + b_1\alpha^3 + \dots . \tag{1.9}$$

The *one-loop* (*two-loop*) coefficient is b_0 (b_1), and in QCD they are given by:

$$b_0^{\text{QCD}} = -\frac{11\,C_A - 4\,n_F T_R}{12\pi}, \quad b_1^{\text{QCD}} = -\frac{17\,C_A^2 - n_F T_R(10\,C_A + 6\,C_F)}{24\pi^2}, \tag{1.10}$$

whereas in QED they simplify to:

$$b_0^{\text{QED}} = \frac{1}{3\pi}\left[N_c \sum_{i=1}^{n_F} Q_i^2 + n_L\right], \quad b_1^{\text{QED}} = \frac{1}{4\pi^2}\left[N_c \sum_{i=1}^{n_F} Q_i^4 + n_L\right]. \tag{1.11}$$

n_F denotes the number of quarks, while n_L denotes the number of leptons. The coefficients b_i are known up to b_4, and beyond two-loop they are scheme-dependent (see Ref. [24] for their expressions in the $\overline{\text{MS}}$ scheme). If we retain only the first coefficient, a simple exact analytic solution of Eq. (1.9) exists:

$$\alpha(\mu^2) = \frac{\alpha(\mu_0^2)}{1 - b_0\alpha(\mu_0^2)\ln(\mu^2/\mu_0^2)}, \tag{1.12}$$

with μ_0^2 and $\alpha(\mu_0^2)$ defining the initial condition for the evolution. It is also possible to write down the equivalent of Eq. (1.12) accurate up to two loops:

$$\frac{\alpha(\mu^2)}{\alpha(\mu_0^2)} = \left(1 - \alpha(\mu_0^2)b_0 L + \alpha(\mu_0^2)\frac{b_1}{b_0}\ln\frac{b_0 + \alpha(\mu_0^2)(b_1 - b_0^2 L)}{b_0 + \alpha(\mu_0^2)b_1}\right)^{-1}, \tag{1.13}$$

with $L = \ln(\mu^2/\mu_0^2)$. This expression is not unique; however it guarantees the correctness of the $\alpha(\alpha L)^k$ and $\alpha^2(\alpha L)^k$ terms to all orders, whereas Eq. (1.12) resums only and solely the former class of terms.

Let us discuss the run of the coupling constant in QED and in QCD. This discussion has significant implications on the validity of a perturbative approach.

- The QED β-function is positive, therefore as energy increases the coupling α_{em} becomes larger. However, the small value of $\alpha_{\text{em}}(m_e^2) \simeq 1/137$ at the electron mass does not significantly increase with the evolution in the energy range usually probed at colliders. A perturbative treatment is thus justified.

- In QCD, the sign of the β-function is negative, thus $\alpha_s \to 0$ for $\mu \to +\infty$. This particular behaviour is called *asymptotic freedom* [25,26] i.e. the fact that the strong coupling becomes weaker for processes involving large momentum transfer.

  On the other hand, we see that there exists a constant Λ_{QCD}, such that $\alpha_s \to +\infty$ for $\mu \to \Lambda_{\text{QCD}}^+$. The value of Λ_{QCD} (~ 200 MeV), is indicative of the energy range where the

---

[4]Note it is customary in QCD to collect a global minus sign on the r.h.s. of Eq. (1.9), leading to the alternative definition of the β-function coefficients as $\beta_i = -b_i$.

non-perturbative dynamics becomes relevant, and it is closely connected with the scale of hadron masses. If we want to pursue a perturbative calculation, we need to make sure to work far from the Λ_{QCD} pole.

In QCD, usually one adopts as initial condition for the QCD evolution the value of the strong coupling at the mass of the Z boson, M_Z:

$$\alpha_s(M_Z^2) \simeq 0.118 \,, \quad M_Z \simeq 91.188 \text{ GeV} \,. \tag{1.14}$$

When evaluating the β-function coefficients, we fix n_F as the number of quark flavours considered light, with the remaining heavier quark flavours decoupling from the theory [27]. One can then relate the coupling for a theory with $n_F + 1$ light flavours to that with n_F flavours through specific equations (see e.g. Ref. [24]).

For what concerns the QED evolution, one can adopt as initial condition the fine structure constant α_{em} at a scale equal to the electron mass, m_e^2:

$$\alpha_{\text{em}}(m_e^2) \simeq 1/137.036 \,, \quad m_e \simeq 0.511 \text{ MeV} \,. \tag{1.15}$$

Quark contributions to the running of α_{em}, usually estimated starting from e^+e^- low-energy data, are also present. One may take into account such contributions coming from the light hadronic resonances in an inclusive way. Namely, by starting from a precise determination of $\alpha = \alpha_H$ that does include low-energy contributions, and that can be associated with a scale μ_H (just) larger than the mass of the heaviest hadronic resonance, one can backward-evolve $\alpha_H = \alpha(\mu_H)$ from μ_H down to the preferred initial scale. By doing so, the possible local effects of the resonances on the evolution are still neglected, but this is not important if the final scale of the evolution is meant to be hard.

1.2 Hadron collisions

Since the parton model [28], we know that at sufficiently high energies the scattering between two protons actually takes place between point-like particles, *partons*, extracted from the two nucleons. Each parton comes with a certain fraction x of the longitudinal momentum of the parent hadron. In order to obtain the cross section for a determinate process with a hadron with momentum p in the initial state, we should write a sort of weighted mean over the possible momentum fractions of the extracted quark. This requires the introduction of a function $f_i(x)$ which gives the "probability"[5] to extract a parton of flavour i with the momentum fraction x. This function is called *parton distribution function* (PDF). With two protons in the initial state, the hadronic cross section is written as:

$$\sigma(Q, p_1, p_2) = \sum_{i,j} \int_0^1 dx_1 dx_2 \, f_i(x_1) f_j(x_2) \, \hat{\sigma}_{ij}(Q, x_1 p_1, x_2 p_2) + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}^2}{Q^2}\right) \,, \tag{1.16}$$

---

[5]Strictly speaking, the PDFs admit a probabilistic interpretation only at leading order. At higher orders, PDFs can even become negative as a consequence of the subtraction of collinear divergences. This will be clear in the following.

where $\hat{\sigma}_{ij}$ is called *partonic* or *short-distance* cross section, encoding the hard scattering between partons i and j.[6] Q denotes a physical scale of the hard scattering process e.g. the mass of the produced object. Finally, the last term on the r.h.s. of Eq. (1.16) reflects the fact that the factorization of a total hadronic cross section as a "product" of parton distrubution functions and a partonic cross section is valid up to power corrections in $\Lambda_{\mathrm{QCD}}^2/Q^2$.

This factorization is based on a separation of the long-distance non-perturbative regime from the short-distance behaviour computable in perturbative QCD. We thus need to make sure that diagrams with lines connecting the partonic part of the process directly to the non-perturbative hadronic part do not count. Factorization theorems have been proved at a reasonable level of rigor for deep inelastic scattering (just one hadron in the initial state), for the Drell-Yan process (two hadrons in the initial state, but leptons in the final state) and for other sufficiently inclusive processes. However, factorization is often assumed to be true also in a more general context, see Ref. [31] for a review.

Eq. (1.16) may be rewritten in a more suggestive way. If we denote with s the overall hadron-hadron c.m. energy squared, the fraction of s available for the parton-parton collision will be $\hat{s} = x_1 x_2 s$. By introducing $\tau = \hat{s}/s$, we can rewrite Eq. (1.16) as:

$$\sigma = \sum_{i,j} \int_{\tau_0}^1 d\tau \int_0^1 dx_1 dx_2 \, f_i(x_1) f_j(x_2) \, \hat{\sigma}_{ij}(x_1, x_2) \, \delta(\tau - x_1 x_2) \,, \tag{1.17}$$

where τ_0 is the minimum value of τ at which the hard process can occur. We define the convolution of two functions $g(z)$ and $h(z)$ as:

$$f(z) = g \otimes_z h = \int_0^1 dx \, dy \, g(x) h(y) \, \delta(z - xy) = \int_z^1 \frac{dx}{x} \, g(x) \, h\left(\frac{z}{x}\right) \,, \tag{1.18}$$

and if $\hat{\sigma}_{ij}$ depends only on $\hat{s}$, Eq. (1.17) may be arranged as:

$$\sigma = \sum_{i,j} \int_{\tau_0}^1 \frac{\mathrm{d}\tau}{\tau} \cdot \left[\frac{\tau}{\hat{s}} \frac{d\mathcal{L}_{ij}}{d\tau} \right] \cdot [\hat{s} \, \hat{\sigma}_{ij}(\hat{s})] \tag{1.19}$$

where the differential *parton luminosity*, with dimensions of a cross section, is defined as

$$\frac{\tau}{\hat{s}} \frac{d\mathcal{L}_{ij}}{d\tau} = \frac{\tau}{\hat{s}} \left(f_i \otimes_\tau f_j \right) \,. \tag{1.20}$$

Hence, the hadronic cross section is obtained after integration over the possible values of the c.m. energy fraction τ and sum over the possible partonic flavours i and j of the product between the parton luminosity $\mathcal{L}_{ij}$ and the short-distance cross section $\hat{\sigma}_{ij}$.

---

[6]In Eq. (1.16), we are neglecting that initial state partons carry transverse momentum as a consequence of their motion inside the parent nucleon. Actually, what we call as PDFs are more properly defined as *collinear* PDFs. Transverse-Momentum-Dependent PDFs (TMD PDFs) describe the distribution of partons inside the nucleon as a function of both the intrinsic partonic transverse momentum $\mathbf{k}_\perp$ and the longitudinal momentum fraction x (see e.g. [29, 30]). After integration over $\mathbf{k}_\perp$, we recover the standard collinear PDFs. At the current level of experimental precision, the adoption of TMD PDFs is required for a faithful description of the cross section in the region of total low transverse momentum of the final state system.

1.3 Infrared divergences

The renormalization procedure provides us with a systematic way to handle the divergences of UV origin. However, when computing matrix elements, we find other type of divergences, due to the vanishing masses of some particles involved. These *infrared* (IR) divergences originate from the integration over the region of low momenta.

Given a matter field of mass m and a gauge field of mass λ, suppose you want to compute the matrix element for the emission of the gauge field off the matter field. It can be shown that: if $\lambda = 0$ i.e. the gauge field is massless, the matrix element diverges when the energy of the gauge field goes to zero (*soft* limit); if $\lambda = m = 0$ i.e. the matter field is also massless, the matrix element diverges when the emission angle becomes null (*collinear* limit).

We will now explicitly show that these singular behaviours are universal i.e. the divergent part factorizes from the matrix element in the soft and/or collinear limit.

1.3.1 Preludium: classical theory of radiation

Soft divergence is an effect already present in classical electrodynamics, as photon can be emitted with arbitrary small energies. Suppose we have an electron in motion interacting with a potential. Due to the presence of the potential, it will undergo acceleration and it will emit radiation. This radiation is traditionally called *bremsstrahlung*, "braking radiation", because it was first observed with electrons stopped in a metallic target. In classical electrodynamics, the distribution in frequency and angle of the intensity of energy radiated by an accelerated charge is given by (Chapter 15 of [32]):

$$\frac{d^2 I}{d\omega \, d\Omega} = e^2 \left| \int dt \left[\frac{d}{dt} \left(\frac{\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta}(t))}{1 - \mathbf{n} \cdot \boldsymbol{\beta}(t)} \right) \right] \exp i\omega(t - \mathbf{n} \cdot \mathbf{r}(t)) \right|^2 , \tag{1.21}$$

where $\mathbf{r}$ and $\boldsymbol{\beta}$ are the position and the velocity of the radiating particle, while $\mathbf{n}$ is the direction of propagation of the radiation. In a photon description of radiation, the latter equation times $1/\omega$ gives the number of photons emitted per unit energy interval, with $E_\gamma = \omega$ (in natural units), and per unit solid angle:

$$\frac{d^2 N}{dE_\gamma \, d\Omega_\gamma} = \frac{1}{\omega} \frac{d^2 I}{d\omega \, d\Omega} . \tag{1.22}$$

For sufficiently long wavelengths of the emitted radiation — or small photon energies compared to the total energy available — the radiation pattern won't depend on the details of the trajectory in the scattering region, but only on the initial and final velocities, and on the direction in which radiation is observed. In fact, if we denote with $\boldsymbol{\beta}$ and $\boldsymbol{\beta'}$ the velocity of the electron before and after the collision, in the limit $\omega \to 0$ the exponential in Eq. (1.21) is equal to unity, and the integrand is a perfect differential. By introducing the polarization vector $\boldsymbol{\epsilon}$, lying in the plane orthogonal to $\mathbf{n}$, we obtain:

$$\lim_{E_\gamma \to 0} d^2 N = e^2 \left| \left(\frac{\boldsymbol{\epsilon} \cdot \boldsymbol{\beta'}}{1 - \mathbf{n} \cdot \boldsymbol{\beta'}} \right) - \left(\frac{\boldsymbol{\epsilon} \cdot \boldsymbol{\beta}}{1 - \mathbf{n} \cdot \boldsymbol{\beta}} \right) \right|^2 \frac{dE_\gamma}{E_\gamma} d\Omega_\gamma . \tag{1.23}$$

We find that the number of photons emitted per unit energy emitted is inversely proportional to the energy, and diverges as $E_\gamma \to 0$, leading to the *infrared divergence*. Moreover, the radiation

is strongly peaked about the directions of $\boldsymbol{\beta}$ (and $\boldsymbol{\beta'}$), as the denominators in Eq. (1.26) can be rewritten as:

$$(1 - \mathbf{n} \cdot \boldsymbol{\beta}) = (1 - |\boldsymbol{\beta}| \cos \vartheta) . \tag{1.24}$$

Note that we would also have a *collinear* divergence for $\cos \vartheta \to 0$ if it were not for the $|\boldsymbol{\beta}| = |\mathbf{p}|/E$ factor, which is different from zero for massive particles, like the electron.

Eq. (1.23) may be written in a manifestly Lorentz-invariant form by replacing the 3-dimensional products with the analogous 4-dimensional products between:

$$k = E_\gamma (1, \mathbf{n}) , \quad p = E (1, \boldsymbol{\beta}) , \quad \epsilon = (0, \boldsymbol{\epsilon}) , \tag{1.25}$$

and by multiplying the result by $[(2\pi)^3 2E_\gamma]^{-1}$ in order to introduce the usual Lorentz-invariant phase space. Thus, we obtain:

$$\lim_{E_\gamma \to 0} d^2 N = e^2 \left| \frac{\epsilon \cdot p'}{k \cdot p'} - \frac{\epsilon \cdot p}{k \cdot p} \right|^2 \frac{d^3 k}{(2\pi)^3 2E_\gamma} . \tag{1.26}$$

This is exactly the same expression one would get in a quantum field theory calculation. Of course, far from the $\omega \to 0$ limit, the classical result completely fails in describing the pattern of photon radiation. In this case we need a proper quantum description.

1.3.2   Soft limit of matrix element
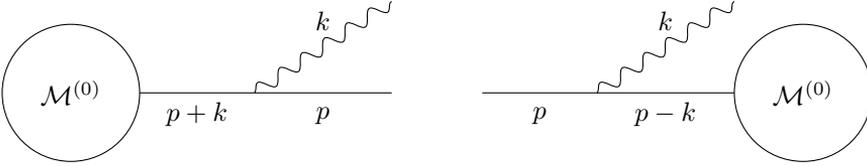
Figure 1.1: Final state (left) and initial state (right) emission of photon or gluon.

Suppose we have a diagram for a generic process, with an outgoing lepton with momentum $p + k$. This lepton later emits a photon with momentum k, and it is left with momentum p (Fig. 1.1, on the left). Feynman rules provide an expression for the complete matrix element $\mathcal{M}(p, k)$, dependent both on p and k. This matrix element contains the propagator of the fermion line, which in the limit of small k can be written as:

$$\frac{\not{p} \pm \not{k}}{(p \pm k)^2 - m^2} \xrightarrow{k \to 0} \frac{\not{p}}{\pm (p \cdot k)} . \tag{1.27}$$

It is thus straightforward to conclude that in this same limit the full matrix element factorizes as

$$\mathcal{M}(p, k) \xrightarrow{k \to 0} eQ \frac{p_\mu}{p \cdot k} \epsilon_\mu(k) \mathcal{M}^{(0)}(p) , \tag{1.28}$$

where $\mathcal{M}^{(0)}(p)$ is the matrix element with an outgoing lepton with momentum p. In a similar fashion, for a photon emission off an initial state lepton, as depicted in Fig. 1.1 on the right, we arrive at the result

$$\mathcal{M}(p - k) \xrightarrow{k \to 0} -eQ \frac{p_\mu}{p \cdot k} \epsilon_\mu(k) \mathcal{M}^{(0)}(p) . \tag{1.29}$$

The fraction $p_\mu/(p \cdot k)$ is usually called *eikonal current*. Note how in the limit $k \to 0$ both Eq. (1.28) and Eq. (1.29) diverge. As for the emissions from the internal lines of a diagram, we do not have any problems of IR divergence.

Since we are allowed to neglect k also in the momentum conservation delta function, we obtain a factorization of the phase space as well. Thus, supposing for instance a single charged particle in the final and initial state, and taking into account both diagrams of Fig. 1.1, the cross section for one photon emission in the soft limit reads

$$d\sigma^{\text{soft}}(k, p, p') = d\sigma^{(0)}(p, p') \frac{\alpha}{(2\pi)^2} \left[\frac{\epsilon \cdot p'}{k \cdot p'} - \frac{\epsilon \cdot p}{k \cdot p} \right]^2 \frac{d^3 k}{k_0} , \qquad (1.30)$$

where $d\sigma$ denotes an arbitrary fully differential cross section. Eq. (1.30) coincides with Eq. (1.26) we obtained within a classical framework. It is possible to write down a similar result for an arbitrary number of photon emissions, both from initial and final state fermions, see e.g. Ref. [18]. In the limit of infinite photon emissions, the one-photon result simply exponentiates, since emissions can be assumed as independent.

Expressions for the soft limit of matrix elements are similar in QCD, apart from a more complex colour structure, see Ref. [33]. Colour correlations among different legs and the fact that soft radiation can come also from hard gluons preclude a simple exponentiation of the single emission result, see Ref. [34] and references therein.

1.3.3 Collinear limit of matrix element

Matrix elements factorize also in the collinear limit. Strictly speaking, in presence of a fermion mass there are no collinear divergences, since the mass acts as a natural cutoff to the integration. However, it is still useful to consider the collinear limit of matrix elements even in presence of a finite fermion mass m, usually called *quasi*-collinear limit.

Consider again the final state branching depicted in Fig. 1.1 on the left, with a photon emission off an electron. A convenient parametrization of final state momenta, suitable for a precise definition of the collinear limit, is the Sudakov parametrization. Namely, we write the momentum p and k as [35]:

$$p^\nu = z\tilde{p}^\nu - k_\perp^\nu + \frac{m^2 + \mathbf{k}_\perp^2 - z^2 m^2}{z} \frac{n^\nu}{2\,\tilde{p} \cdot n} , \qquad (1.31)$$

$$k^\nu = (1-z)\tilde{p}^\nu + k_\perp^\nu + \frac{\mathbf{k}_\perp^2 - (1-z)^2 m^2}{1-z} \frac{n^\nu}{2\,\tilde{p} \cdot n} , \qquad (1.32)$$

where $\tilde{p}^\nu$ is the momentum defining the "collinear" direction, with $p^2 = \tilde{p}^2 = m^2$; n^ν is an auxiliary light-like vector ($n^2 = 0$); finally, $k_\perp^\nu$ denotes the transverse direction, with $k_\perp \cdot \tilde{p} = k_\perp \cdot n = 0$. Note that within this parametrization, the product between p and k reads:

$$2\, p \cdot k = \frac{\mathbf{k}_\perp^2 + (1-z)^2 m^2}{z(1-z)} . \qquad (1.33)$$

Given Eqs. (1.31)-(1.32), the *quasi*-collinear limit is defined as $k_\perp \to 0$ and $m \to 0$, with fixed $m^2/\mathbf{k}_\perp^2$. If $m = 0$, the collinear limit is simply defined as $k_\perp \to 0$. In this (these) limit(s), we

arrive at

$$|\mathcal{M}(p,k)|^2 \xrightarrow{p\|k} \frac{e^2}{p\cdot k}\,\hat{P}(z)\,|\mathcal{M}^{(0)}(\tilde{p})|^2\,, \tag{1.34}$$

where the function $\hat{P}(z)$ is defined as

$$\hat{P}(z) = Q_i^2\left[\frac{1+z^2}{1-z} - \frac{m^2}{p\cdot k}\right]\,. \tag{1.35}$$

with Q^2 the charge of the lepton squared. This function is called *splitting kernel*: it encodes the behaviour of the matrix element as a function of the momentum fraction z taken by one of the two outgoing particles (in this case by the lepton). Similarly, we arrive at the factorization of the phase space:

$$d\Phi(p,k) \to \frac{d\Phi(\tilde{p})}{z}d\Phi(k) = d\Phi(\tilde{p})\frac{1}{16\pi^2}\frac{d\mathbf{k}_\perp^2\,dz}{z(1-z)}\,. \tag{1.36}$$

The factor $1/z$ takes into account the replacement $p \to \tilde{p}$. By using Eq. (1.34) and Eq. (1.36), we obtain for the cross section in the collinear limit:

$$d\sigma(p,k) \to d\sigma^{(0)}(\tilde{p})\frac{\alpha}{2\pi}\frac{d\mathbf{k}_\perp^2}{\mathbf{k}_\perp^2 + (1-z)^2 m^2}\hat{P}(z)dz\,. \tag{1.37}$$

Two comments are in order. First, note that the previous equation diverges as $z \to 1$ i.e. when the photon goes soft, see Eq. (1.35). Indeed, the collinear limit entails the soft limit (but the opposite is not true). Second, if $m = 0$, the integration over the transverse momentum $\mathbf{k}_\perp^2$ diverges. This is the strictly collinear divergence, due to the vanishing fermion mass.

As in the case of the soft limit, the expressions presented here are also valid in QCD, upon replacement of couplings and colour factors e.g. in Eq. (1.35), Q^2 becomes C_F.

1.4 Cancellation of infrared divergences

In the previous section, we have examined the IR singular behaviour of matrix elements, both in QED and in QCD. We have found that in suitable limits the matrix elements diverge.

However, besides the *real* matrix elements we have studied, at the same perturbative order there are also *virtual* corrections, namely diagrams where a virtual photon or gluon is emitted and later absorbed by the fermion leg. Supposing that $|\mathcal{M}^{(0)}|^2$ is of order α^k, diagrams with a virtual emission are of order α^{k+2}, but their interference with the Born level $\mathcal{M}^{(0)}$ is of order α^{k+1}, which is the same order of a real emission diagram. Thus we have to take into account both contributions. Let us discuss what happens in QED and in QCD when combining real and virtual contributions.

QED Soft divergences cancel between real and virtual terms if the observable we are calculating is sufficiently *inclusive* i.e. we cannot distinguish a virtual event and a real event with a soft photon. For instance, this requirement is satisfied if we assume to not be able to detect photons with energy less than some experimental cutoff ΔE. In this way infinities cancel and the leftover finite terms give a perturbative correction to the tree-level cross section. Consider for example the initial state radiation corrections to the process $e^+e^- \to \mu^+\mu^-$. After summing

of real and virtual contributions, we obtain

$$\sigma_R + \sigma_V = \sigma_B \left[\frac{\alpha}{\pi} \left(\frac{\beta^2 + 1}{\beta} \ln \frac{1+\beta}{1-\beta} - 2 \right) \ln \frac{E}{\Delta E} \right] , \tag{1.38}$$

where σ_B is the Born cross section, E is some upper limit above which the soft approximation is no longer valid, and β is defined as:

$$\beta = \sqrt{1 - \frac{m^2}{s}} , \tag{1.39}$$

with m the electron mass and s the centre of mass energy squared. Eq. (1.38) is finite, and we see the appearance of a typical logarithm $\ln(E/\Delta E)$, which is the left after the cancellation of the soft divergence. A systematic proof of cancellation of soft divergences in QED has been proved in 1937 by Bloch and Nordsieck [36]. In the 1960s, Yennie, Frautschi and Suura provided the full treatment of IR divergences in QED [37], with an exponentiation of soft photon effects to all order in perturbation theory.

As for collinear divergences, they are not present in QED, as the mass of the leptons act as regulator. Indeed, the first logarithm in Eq. (1.38) is of collinear origin, and in the limit $s \gg m^2$ (as it is the case at typical collider energies) it becomes

$$\ln \frac{1+\beta}{1-\beta} \xrightarrow{s \gg m^2} \ln \frac{s}{m^2} . \tag{1.40}$$

Even though not divergent, this is a large number, which can spoil the convergence of the perturbative series. It is possible to take into account the presence of these large logarithms of collinear origin at all order in α: this will be the subject of Chapter 6.

QCD In QCD we have to deal with both soft and collinear divergences. Indeed, even if quark masses are retained in the calculations[7], gluon self-interactions are a source of intrinsic collinear divergence. However, there exists a a theorem of cancellation even in the case of both soft and collinear divergences. This result is called KLN theorem, proved by Kinoshita [38] and Lee and Nauenberg [39]. Cancellation of soft and collinear divergences is guaranteed if a sum over *initial* and *final* degenerate states is carried out. By *degenerate* states we mean states which cannot be distinguished from one another e.g. a single-quark state and a quark accompanied by an arbitrary number of collinear gluons.

In general, given a final state with $n+1$ QCD particles, in the limit in which one particle becomes soft or two particles become collinear, the observable should scale as:

$$O_{n+1}(k_1, \ldots, k_i, \ldots, k_n) \xrightarrow{k_i \to 0} O_n(k_1, \ldots, k_{i-1}, k_{i+1}, \ldots, k_n) , \tag{1.41}$$

$$O_{n+1}(k_1, \ldots, k_i, k_j, \ldots, k_n) \xrightarrow{k_i \parallel k_j} O_n(k_1, \ldots, k_i + k_j, \ldots, k_n) , \tag{1.42}$$

where O_{n+1} and O_n is the value of the observable evaluated on a set of n or $n+1$ particles respectively. This property is called *infrared and collinear* (IRC) safety. IRC safety guarantees

---

[7]It is actually possible to keep the quark mass finite in the case of *heavy* quarks only (like charm or bottom), since the masses of the lightest quarks are not well defined and certainly smaller than Λ_{QCD}, therefore a perturbative calculation is meaningless.

the cancellation of soft and collinear divergences, because in the soft and collinear limit O_{n+1} reduces to O_n with the right mapping of momenta to permit the cancellation with the virtual contributions.

However, in real situations we are not in the position to actually take the sum over *initial* degenerate states. This fact can be simply understood as follows. Consider Fig. 1.2, which is the analogue of Fig. 1.1 with a collinear splitting in the final or in the initial state.
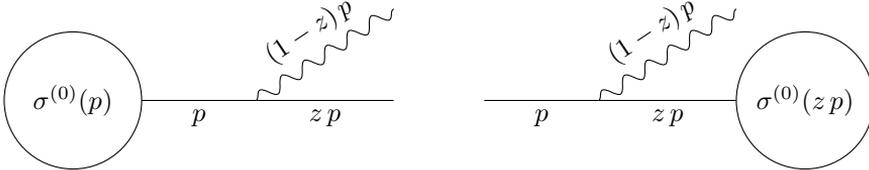

Figure 1.2: Final state (left) and initial state (right) collinear splitting.

In Fig. 1.2 on the left, the final state quark undergoes a collinear splitting, emitting a gluon with momentum fraction $1 - z$. We have already found the expression of the cross section in this limit, see Eq. (1.37) and Eq. (1.35), upon replacements $\alpha \to \alpha_s$ and $Q^2 \to C_F$, and with $m = 0$. This is the result for the real cross section. As said before, we need to also take into account virtual contributions, which in the collinear limit can be found to be given by the exact opposite of Eq. (1.37):

$$\sigma_V = -\sigma^{(0)}(p) \frac{\alpha_s}{2\pi} \frac{d\mathbf{k}_\perp^2}{\mathbf{k}_\perp^2} \hat{P}(z) dz \,. \tag{1.43}$$

This implies that, when summing real and virtual contributions, and performing the integration to get the total cross section, we obtain:

$$\sigma^{(1)} = \frac{\alpha_s C_F}{2\pi} \int \left[\sigma^{(0)}(p) - \sigma^{(0)}(p) \right] \left(\frac{1 + z^2}{1 - z} \right) \frac{d\mathbf{k}_\perp^2}{\mathbf{k}_\perp^2} dz = 0 \,. \tag{1.44}$$

Hence, the divergences cancel, because the momentum entering the Born cross section is the same in the real and in the virtual case. By now performing the same calculation in the case of initial state splitting, as depicted in Fig. 1.2 on the right, we find:

$$\sigma^{(1)} = \frac{\alpha_s C_F}{2\pi} \int \left[\sigma^{(0)}(zp) - \sigma^{(0)}(p) \right] \left(\frac{1 + z^2}{1 - z} \right) \frac{d\mathbf{k}_\perp^2}{\mathbf{k}_\perp^2} dz \,. \tag{1.45}$$

In the real emission case, the momentum entering the Born cross section is zp; instead, in the virtual case, it is simply p, as virtual matrix elements have Born kinematics. In the soft limit i.e. $z \to 1$, we get the cancellation of soft divergence, since the difference in the square brackets vanishes. Conversely, the collinear divergence does not cancel, since the integration over $\mathbf{k}_\perp^2$ is not bounded from below. The KLN theorem does not apply in this case, since we are not summing over initial state potential collinear configurations.

1.5   Collinear factorization

Initial state collinear divergences can be absorbed into universal objects through the *collinear factorization* procedure. The main idea is to absorb the initial state collinear divergences in the

parton distribution functions, similarly to what we have done in Sec. 1.1, where UV divergences were absorbed in the bare coupling. We can sketch this procedure as follows. According to Eq. (1.16), the hadronic cross section is given by the "product" of a *bare* parton luminosity $\mathcal{L}_0$ and a parton cross section $\hat{\sigma}$ (with collinear divergences):

$$\sigma = \mathcal{L}_0 * \hat{\sigma}\,. \tag{1.46}$$

As collinear divergences come from the integration over small transverse momenta, we introduce a scale μ_F, called *factorization scale*[8], to isolate this region and we write:

$$\sigma = \mathcal{L}_0 * \Gamma_{\mathrm{div}}(\mu_F^2) * \hat{\sigma}(\mu_F^2)\,, \tag{1.47}$$

where $\Gamma_{\mathrm{div}}(\mu_F^2)$ contains the singular collinear behaviour of $\hat{\sigma}$, whereas $\hat{\sigma}(\mu_F^2)$ is finite. At this point we redefine the parton luminosity by absorbing the divergent factor:

$$\mathcal{L}(\mu_F^2) = \mathcal{L}_0 * \Gamma_{\mathrm{div}}(\mu_F^2)\,, \tag{1.48}$$

and in the end we obtain:

$$\sigma = \mathcal{L}(\mu_F^2) * \hat{\sigma}(\mu_F^2)\,. \tag{1.49}$$

Both factors are now finite, even though they both depend on an unphysical scale μ_F^2. Of course, we should make sure that the divergent part actually factorizes and it is independent of the observable and of the process. This can be accomplished through the *generalized ladder expansion* of Ref. [40]. See also Ref. [41] for a more pedagogical presentation.

We are now going to provide a concrete example of Eqs. (1.46)-(1.49). For sake of simplicity, suppose we have only one proton in the initial state and just a single partonic flavour, say a quark. At leading order, the momentum entering the the partonic cross section is the same as the momentum carried by the extracted quark i.e.

$$\sigma(p) = \int_0^1 dx\, q_0(x)\, \hat{\sigma}^{(0)}(xp)\,, \tag{1.50}$$

with q_0 denoting the bare quark PDF. At next-to-leading order, we take into account the emission of a real gluon in the initial state, plus virtual corrections. This emission rescales the momentum from $xp \equiv \hat{p}$ to $z\hat{p}$. In the collinear limit, the partonic cross section is given by Eq. (1.45). In order to regularize the singularity, we retain the quark mass m in the calculation. By doing so, Eq. (1.45) becomes:

$$\frac{\alpha_S C_F}{2\pi} \int_0^1 dz \int_0^{Q^2} \frac{d\mathbf{k}_\perp^2}{\mathbf{k}_\perp^2 + (1-z)^2 m^2} \left(\frac{1+z^2}{1-z}\right) \left[\hat{\sigma}^{(0)}(z\hat{p}) - \hat{\sigma}^{(0)}(\hat{p})\right] \tag{1.51}$$

$$= \frac{\alpha_S C_F}{2\pi} \int_0^1 dz \left(\frac{1+z^2}{1-z} \ln \frac{Q^2}{(1-z)^2 m^2} - \frac{2z}{1-z}\right) \left[\hat{\sigma}^{(0)}(zp) - \hat{\sigma}^{(0)}(p)\right] + \mathcal{O}\left(\frac{m^2}{Q^2}\right) \tag{1.52}$$

where Q^2 is some hard scale of the process, setting the upper limit of integration. As we are

---

[8] As for the renormalization scale, in the $\overline{\mathrm{MS}}$ scheme the factorization scale is usually made to coincide with the regularization scale $\widetilde{\mu}$ of dimensional regularization, see Sec. 1.1.1.

interested in the terms divergent as $m \to 0$, we simply write:

$$\hat{\sigma}^{(1)}(\hat{p}) = \frac{\alpha_s C_F}{2\pi} \int_0^1 dz \left(\frac{1+z^2}{1-z} \ln \frac{Q^2}{m^2} \right) \left[\hat{\sigma}^{(0)}(z\hat{p}) - \hat{\sigma}^{(0)}(\hat{p}) \right] . \tag{1.53}$$

The latter equation may be further simplified by introducing the plus prescription notation, which is a proxy for the virtual corrections in the collinear limit:

$$\int_0^1 dz \, [f(z)]_+ \, g(z) = \int_0^1 f(z) \left[g(z) - g(1) \right] . \tag{1.54}$$

Hence Eq. (1.53) becomes

$$\hat{\sigma}^{(1)}(\hat{p}) = \frac{\alpha_s}{2\pi} \ln \frac{Q^2}{m^2} \int_0^1 dz \, P(z) \hat{\sigma}^{(0)}(z\hat{p}) , \tag{1.55}$$

with P the regularized version of the quark-quark splitting kernel:

$$P(z) = C_F \left(\frac{1+z^2}{1-z} \right)_+ . \tag{1.56}$$

At this point, according to Eq. (1.48), we split the integration over the transverse momentum in Eq. (1.51) through the introduction of the factorization scale μ_F^2. At the end this simply amounts to splitting the logarithm in Eq. (1.55) into the sum of two pieces:

$$\ln \frac{Q^2}{m^2} = \ln \frac{Q^2}{\mu_F^2} + \ln \frac{\mu_F^2}{m^2} . \tag{1.57}$$

and thus Eq. (1.55) is now given by the sum of two terms:

$$\hat{\sigma}^{(1)}(\hat{p}) = \Gamma^{(1)}(\hat{p}, \mu_F^2) + \hat{\sigma}^{(1)}(\hat{p}, \mu_F^2) . \tag{1.58}$$

In general, $\hat{\sigma}^{(1)}(\hat{p}, \mu_F^2)$ denotes the finite part of the next-to-leading order partonic cross section, *subtracted* of the initial state collinear divergence, which has been collected in $\Gamma^{(1)}(\hat{p}, \mu_F^2)$. Given Eq. (1.58), the NLO version of Eq. (1.50) reads:

$$\sigma(p) = \int_0^1 dx \, q_0(x) \left[\hat{\sigma}^{(0)}(xp) + \frac{\alpha_s}{2\pi} \ln \frac{\mu_F^2}{m^2} \int_0^1 dz \, P(z) \hat{\sigma}^{(0)}(zxp) + \hat{\sigma}^{(1)}(xp, \mu_F^2) \right]$$
$$= \int_0^1 dy \int_0^1 dx \int_0^1 dz \, q_0(x) \, \Gamma(z, \mu_F^2) \, \hat{\sigma}^{(0)}(yp) \, \delta(y - xz) + \int_0^1 dx \, q_0(x) \, \hat{\sigma}^{(1)}(xp, \mu_F^2) \tag{1.59}$$

with the distribution $\Gamma(z, \mu_F^2)$ defined as:

$$\Gamma(z, \mu_F^2) = \delta(1 - z) + \frac{\alpha_s}{2\pi} \ln \frac{\mu_F^2}{m^2} P(z) . \tag{1.60}$$

Note that we have the freedom to include in the subtraction term of Eq. (1.60) any finite term. Different choices are associated to different *subtraction schemes*.

According to Eq. (1.48), we are now ready to redefine the parton distribution as

$$q(y, \mu_F^2) \coloneqq \int_0^1 dx \int_0^1 dz \, q_0(x) \, \Gamma(z, \mu_F^2) \, \delta(y - xz) \tag{1.61}$$

$$= q_0(y) + \frac{\alpha_S}{2\pi} \ln \frac{\mu_F^2}{m^2} \int_y^1 \frac{dx}{x} \, q_0(x) \, P\left(\frac{y}{x}\right) \equiv q_0 \otimes_y \Gamma(\mu_F^2) \tag{1.62}$$

At the end we have achieved the desired factorization of Eq. (1.49):

$$\sigma(p) = \int_0^1 dy \, q(y, \mu_F^2) \left[\hat{\sigma}^{(0)}(yp) + \hat{\sigma}^{(1)}(yp, \mu_F^2) \right] \tag{1.63}$$

which is equal to the starting expression, Eq. (1.59), up to terms of order $\mathcal{O}(\alpha_S^2)$.

In this derivation we have considered just one hadron in the initial state (and therefore one PDF) and just one partonic flavour. By restoring two hadrons in the initial state and by summing over partonic flavours, we arrive at the analogue of Eq. (1.16):

$$\sigma(p_1, p_2) = \sum_{i,j} \int_0^1 dx_1 dx_2 \, f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \, \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \mu_F^2) \,. \tag{1.64}$$

Note the explicit dependence on the factorization scale μ_F^2. The subtracted short-distance cross section is free of collinear divergence, and the parton distributions have been redefined accordingly to:

$$f_i(x, \mu_F^2) = f_{i,0}(x) + \frac{\alpha_S}{2\pi} \ln \frac{\mu_F^2}{m^2} \sum_j \int_x^1 \frac{d\xi}{\xi} \, f_{j,0}(\xi) P_{ij}\left(\frac{x}{\xi}\right) \,. \tag{1.65}$$

The sum in Eq. (1.65) is over the possible parton flavours j which undergo a collinear splitting, by leaving a parton with flavour i entering the short-distance cross section. The function P_{ij} in Eq. (1.65) is the regularized Altarelli-Parisi splitting function for the branching $i \leftarrow j$. The leading order QCD splitting functions are:

$$P_{qq}(x) = C_F \left[\frac{1 + x^2}{1 - x} \right]_+ \tag{1.66}$$

$$P_{gq}(x) = C_F \left[\frac{1 + (1 - x)^2}{x} \right] \tag{1.67}$$

$$P_{qg}(x) = T_R \left[x^2 + (1 - x)^2 \right] \tag{1.68}$$

$$P_{gg}(x) = C_A \left[\frac{x}{(1 - x)_+} + \frac{1 - x}{x} + x(1 - x) \right] + \delta(1 - x) \frac{11 C_A - 4 n_F T_R}{6} \tag{1.69}$$

The splitting function P in Eq. (1.56) is the quark-quark splitting function $P \equiv P_{qq}$. Several comments are in order. First, the regular (i.e. $x < 1$) part of P_{qq} and P_{gq} swaps under $z \leftrightarrow (1-z)$, since P_{qq} and P_{gq} refer to the same $q \to qg$ vertex with different kinematics. Second, P_{qg} and P_{gg} are symmetric under $z \leftrightarrow (1-z)$, since we have only quarks or only gluons in the final state of the $g \to q\bar{q}$ and $g \to gg$ branching respectively. Third, the diagonal splitting functions P_{qq} and P_{gg} are intrinsically distributions, containing both plus prescriptions and delta functions,

in order to take into account the virtual corrections. Finally, note that Eqs. (1.66)-(1.69) are valid also in QED with leptons and photons, upon the replacements indicated in Eq. (1.5). The initial state splitting kernels with massive partons can be found in Ref. [42].

We conclude this section with a couple of final remarks. First, we point out that more sophisticated way of regularizing mass singularities exist. In Eq. (1.51), we have simply avoided the collinear divergence by adopting a finite quark mass m. In real cases, regularizing singularities in QCD by giving masses to the particles is extremely complicated, and not gauge-invariant. A more elegant approach is through dimensional regularization, already introduced in Sec. 1.1.1 in the context of UV divergences.

Secondly, in this section we have focused on initial state collinear singularities, but we mention that we could handle also final state collinear singularities in the same way. Indeed, if the observable under study is not sufficiently inclusive, i.e. we are not summing over the whole set of possible degenerate final states, the KLN theorem does not apply. This happens when taking into account single-hadron production cross sections, and when considering distributions differential in the kinematical variables of the detected particle. In such cases, one is left with final state collinear divergences, and the idea is to absorb these divergences in the final state analogue of a PDF. Namely, a *fragmentation function* (FF). In practice one writes down an equation similar to Eq. (1.50), and the same collinear factorization procedure presented in this section for the PDFs applies for the FFs as well. Of course, as the PDFs, FFs have an intrinsic non-perturbative component.

1.6   DGLAP equation

As with the renormalization scale μ_R, physical predictions have to be independent, up to higher orders, on the factorization scale μ_F introduced in the factorization procedure. This requirement leads to a renormalization group equation for the PDFs. In the context of our discussion, such an equation can be obtained by simply taking the derivative of both members of Eq. (1.65). We arrive at the important result:

$$\frac{\partial f_i(z, \mu_F^2)}{\partial \log \mu_F^2} = \frac{\alpha_s(\mu_F)}{2\pi} \left[P_{ij} \otimes f_j \right](z, \mu_F^2). \tag{1.70}$$

This is the DGLAP equation, first introduced in the 1970s by Dokshitzer [43], Gribov, Lipatov [44,45], Altarelli and Parisi [46]. The coupling has been evaluated at the scale μ_F[9], and the splitting kernels can be expanded in a perturbative series in $\alpha_s(\mu_F)$:

$$P_{ij}(x, \mu_F) = \sum_{k=0}^{\infty} \left(\frac{\alpha_s(\mu_F)}{2\pi} \right)^k P_{ij}^{[k]}(x). \tag{1.71}$$

The leading order splitting functions $P_{ij}^{[0]}$ have already been introduced in Eqs. (1.66)-(1.69). The next-to-leading order splitting functions $P_{ij}^{[1]}$ have been calculated in the 1980s by Curci, Furmanski and Petronzio [40,47]. In 2004 the next-to-next-to-leading order splitting functions appeared [48,49].

---

[9]Strictly speaking, the scale in the argument of the coupling should be μ_R, and one should evolve α_s and f_i by varying independently μ_R and μ_F. However, it is simpler to choose a single scale $\mu = \mu_R = \mu_F$ to solve the DGLAP equation, and later evolve only α_s by varying μ_R through the β-function.

1.6.1 Non-singlet and singlet-gluon evolution

The splitting kernels P_{ij} are not all independent. Exploiting flavour symmetry, we can split the quark-quark and the quark-antiquark splitting functions in the following way:

$$\begin{aligned}
P_{q_i q_j} &= \delta_{ij} P_{qq}^{\mathrm{V}} + P_{qq}^{\mathrm{S}}, \\
P_{q_i \bar{q}_j} &= \delta_{ij} P_{q\bar{q}}^{\mathrm{V}} + P_{q\bar{q}}^{\mathrm{S}},
\end{aligned} \tag{1.72}$$

where the superscripts stand for the *singlet* (S) and the *valence* (V) components of the splitting functions. Furthermore, the splitting functions associated to a quark-gluon branching are flavour independent:

$$P_{q_i g} \equiv P_{qg}, \quad P_{g q_i} \equiv P_{gq}. \tag{1.73}$$

These properties allow us to rewrite the system of $2n_F + 1$ equations in Eq. (1.70) as a system of two coupled equations plus $2n_F - 1$ linear equations. We briefly illustrate how this works. We start with introducing the *singlet* and the *non-singlet* combinations:

$$f_{\mathrm{S}} = \sum_{i=1}^{n_F} f_{q_i} + f_{\bar{q}_i}, \qquad f_{\mathrm{NS}} = \sum_{i=1}^{n_F} f_{q_i} - f_{\bar{q}_i}. \tag{1.74}$$

The difference in the non-singlet combination cancels out the gluon PDF contribution to the evolution. Indeed, we see that f_{NS} evolves according to:

$$\frac{\partial f_{\mathrm{NS}}}{\partial \log \mu^2} = \frac{\alpha_S}{2\pi} P_{\mathrm{NS}} \otimes f_{\mathrm{NS}}, \tag{1.75}$$

where P_{NS} is defined as:

$$P_{\mathrm{NS}} = P_{qq}^{\mathrm{V}} - P_{q\bar{q}}^{\mathrm{V}} + n_F (P_{qq}^{\mathrm{S}} - P_{q\bar{q}}^{\mathrm{S}}). \tag{1.76}$$

On the other hand, the singlet combination is coupled to the gluon in the evolution:

$$\frac{\partial}{\partial \log \mu^2} \begin{pmatrix} f_{\mathrm{S}} \\ f_g \end{pmatrix} = \frac{\alpha_S}{2\pi} \mathbb{P}_{\mathrm{S}} \otimes \begin{pmatrix} f_{\mathrm{S}} \\ f_g \end{pmatrix}, \tag{1.77}$$

with the singlet-gluon matrix $\mathbb{P}_{\mathrm{S}}$ defined as:

$$\mathbb{P}_{\mathrm{S}} = \begin{pmatrix} P_{\Sigma\Sigma} & P_{\Sigma g} \\ P_{g\Sigma} & P_{gg} \end{pmatrix}, \tag{1.78}$$

where the components read as follows:

$$\begin{aligned}
P_{\Sigma\Sigma} &= P_{qq}^{\mathrm{V}} + P_{q\bar{q}}^{\mathrm{V}} + n_F (P_{qq}^{\mathrm{S}} + P_{q\bar{q}}^{\mathrm{S}}), \tag{1.79} \\
P_{\Sigma g} &= 2n_F P_{qg}, \tag{1.80} \\
P_{g\Sigma} &= P_{gq}. \tag{1.81}
\end{aligned}$$

If $n_F = 1$ i.e. we only have a quark flavour q with its own antiquark $\bar{q}$, plus the gluon g, Eq. (1.75) and Eq. (1.77) are sufficient, and one recovers the solutions for f_q and $f_{\bar{q}}$ by trivially inverting Eq. (1.74). Otherwise, if $n_F = 2$, we need more independent equations. It is customary to define

other non-singlet-like combinations such as:

$$f_{\mathrm{NS},ij}^{\pm} = f_{q_i} \pm f_{\bar{q}_i} - (f_{q_j} \pm f_{\bar{q}_j}) \tag{1.82}$$

which evolve linearly as in Eq. (1.75), with the replacement:

$$P_{\mathrm{NS}} \longrightarrow P_{\mathrm{NS}}^{\pm} = P_{qq}^{\mathrm{v}} \pm P_{q\bar{q}}^{\mathrm{v}}\,. \tag{1.83}$$

1.6.2   Solving the DGLAP equations

The DGLAP equations are usually solved numerically. Because of the convolution, they are integro-differential equations, with a non-trivial dependence on the momentum fraction x and on the factorization scale μ^2. Two main strategies have been designed in the past years to attack this problem, both relying on the idea of transforming the convolution into a product. Let us discuss them in turn.

- We can start with representing the PDFs on a grid i.e. write the PDF as

$$f(x,\mu^2) = \sum_{\alpha=1}^{N_x} w_\alpha(x)\, f(x_\alpha,\mu^2) \equiv \sum_{\alpha=1}^{N_x} w_\alpha(x)\, f_\alpha(\mu^2) \tag{1.84}$$

where w are some interpolation functions, and x_α with $\alpha = 1,\ldots,N_x$ are the grid nodes[10]. In this way, the evolution equation becomes:

$$\frac{\partial f_\alpha(\mu^2)}{\partial \log \mu^2} = \frac{\alpha_s(\mu)}{2\pi} \sum_{\alpha=1}^{N_x} [P \otimes w_\beta](x_\alpha,\mu^2)\, f_\beta(\mu^2) \equiv \frac{\alpha_s(\mu)}{2\pi} \sum_{\alpha=1}^{N_x} P w_{\beta\alpha}(\mu^2)\, f_\beta(\mu^2)\,. \tag{1.85}$$

After having computed the N_x^2 convolutions between the interpolation weights and the splitting kernels, one is left with a differential equation in μ^2 which can be solved by means of standard Runge-Kutta methods. Such a method is employed in tools like HOPPET [50], QCDNUM [51] or APFEL [52]. One is able to exploit the flexibility given by the grids to efficiently reproduce any functional form for the PDFs.

- A more radical approach is based on a fundamental property of the Mellin transform. Given a function $f(z)$ whose domain is $[0,1]$, this is defined as follows:

$$M[f] \equiv f_N = \int_0^1 dz\, z^{N-1} f(z)\,. \tag{1.86}$$

If $f(z)$ is the convolution of two functions $g(z)$ and $h(z)$, then the convolution factorizes in N-space:

$$M[g \otimes h] = M[g]\, M[h] \quad \Longleftrightarrow \quad [g \otimes h]_N = g_N\, h_N\,. \tag{1.87}$$

This simple result imply that the DGLAP equation, written in Mellin space, is a pure matrix differential equation, and can be solved exactly with a closed expression, or approximately in an iterative fashion. Usually the PDFs at the initial scale are written in

---

[10]Note that it is customary to build the grid as linearly spaced in variables different from x and μ^2. For instance, variables such as $\ln(1/x)$ and $\ln \mu^2$ are more natural, given the DGLAP equation and the functional form of the PDFs.

terms of a basis of functions with known Mellin transforms; the evolution is then carried out in Mellin space, and eventually one numerically inverts the Mellin transform to obtain the evolved PDFs in x-space. This is the method adopted in PEGASUS [53] or in PARTONEVOLUTION [54].

We will give further details on the second strategy, based on the Mellin transform, in Chapter 6, where we will explicitly solve the DGLAP equations in QED.

1.7 How to determine the proton PDFs

Hadronic PDFs have an intrinsic non-perturbative component: as such, they are not calculable within perturbation theory. Unlike the Q^2-dependence, predicted by the DGLAP equation of Eq. (1.70), the x-dependence must be derived from the comparison with the experimental data, through a PDF fit. The essence of a PDF fit is encoded in Eq. (1.64): the partonic cross section $\hat{\sigma}_{ij}$ can be calculated order by order in perturbation theory, the hadronic cross section σ is measured and the PDFs f_i and f_j are extracted. The underlying assumption here is the *universality* of parton distribution functions i.e. the PDFs found when fitting the experimental data of a determinate process can be later applied in the calculation of another process. In other words, the purpose of the PDFs is to allow one to express the prediction for a process in terms of that for another process. In practice, a fitting procedure requires several elements [55]:

- a database of experimental data points (σ^{exp}) and theoretical predictions for the correspondent short-distance cross sections ($\hat{\sigma}_{ij}$).

- a parametrization of the functions f_i, through the introduction of a specific functional form with unknown free fit parameters $\vec{a}$. This parametrization is often driven by QCD insight e.g.

$$f_i(x, Q_0^2; \vec{a}) = x^{-a_1}(1-x)^{a_2} g(a_3, x) \,, \tag{1.88}$$

where Q_0^2 is a reference scale, the factor $x^{-a_1}(1-x)^{a_2}$ encodes the expected PDF behaviour at small and at large x, and g is a generic function that tends to a constant for $x \to 0$ and $x \to 1$.

- a figure of merit to minimize, usually the quadratic difference between the cross section measured, σ^{exp}, and the cross section calculated as the result of the convolution on the r.h.s. of Eq. (1.16), $\sigma^{\text{th}}(\vec{a})$, now depending on $\vec{a}$:

$$\chi^2 \propto (\sigma^{\text{exp}} - \sigma^{\text{th}}(\vec{a}))^2 \,. \tag{1.89}$$

The best fit set of parameters will be thus identified with the absolute minimum of this χ^2 function in the parameter space.

- a method of propagation of the experimental uncertainties associated to σ^{exp} on the PDFs. The final result will be then given in the form:

$$(f_i + \delta f_i)(x, Q_0^2) \,. \tag{1.90}$$

In particular, it is essential to have a proper statistical interpretation of δf_i, with an associated confidence level.

Given the wide range of options about how to actually implement each of the above points, each PDF fitting group adopts its own strategy. Just to mention some of the main collaborations: MSTW [56], CTEQ-TEA [57] and NNPDF [58] (the references point to the latest global fit of each series). We are not going to expand on this here, as Section 4.3 will introduce in more details the NNPDF collaboration fitting methodology, based on a Monte Carlo approach for the propagation of uncertainties and the adoption of neural networks as interpolating functions. This is the methodology employed in Chapter 4 for the study of the impact of jet measurements on parton distribution functions.

1.8 Jet physics

In the previous sections, we have explained in some detail what happens in the initial state of a proton-proton collision. In a similar fashion, let us discuss what happens in the final state of a collider event. This will naturally drive us towards the concept of jet.

1.8.1 Parton shower

Within a fixed-order computation, we obtain a final state with a handful of hard QCD partons. This image is quite different from what one would observe in a real experiment, where detectors around the interaction point see a great number of pions, kaons and other kind of hadrons. The process linking the description of a final state collider event as seen by a fixed-order calculation and as observed in experiments may be described as follows.

Given the set of energetic partons coming out from the hard scattering, they start progressively loosing energy, by radiating mostly soft and collinear particles, as QCD dynamics dictates. Unlike QED, emitted gluons carry colour charge, and they further emit, generating a cascade of QCD particles. This *parton shower* can be approximately calculated in perturbative QCD, as the most dominant terms are the ones associated to soft and collinear emissions, which factorize from matrix elements.

One might think to describe this sequence of radiations by iterating the basic factorization equation for one collinear emission, Eq. (1.37). For later purposes, it is useful to exploit the $z \leftrightarrow 1 - z$ symmetry of the splitting functions and move all the soft singularities to $z = 0$ [59]. Namely, we define a quark splitting function P_q for the branching of a quark into a gluon with momentum fraction z and a quark with momentum fraction $1 - z$:

$$P_q(z) = C_F \left[\frac{1 + (1-z)^2}{z} \right] , \tag{1.91}$$

and a gluon splitting function P_g for the branching of a gluon into a pair of gluons or a quark-antiquark pair:

$$P_g(z) = C_A \left[2\frac{1-z}{z} + z(1-z) + \frac{T_R n_F}{C_A}(z^2 + (1-z)^2) \right] . \tag{1.92}$$

In the soft limit these functions read:

$$P_q(z) \xrightarrow{z \to 0} \frac{2C_F}{z} , \quad P_g(z) \xrightarrow{z \to 0} \frac{2C_A}{z} , \tag{1.93}$$

with the finite part of the splitting functions describing the hard-collinear radiation. We shall use these expressions in Chapter 3, where we study the single-jet inclusive cross section at NLO in the collinear approximation.

However, this method of building approximate cross sections with a great number of legs in the final state is not efficient, as it lacks flexibility and it is difficult to implement in a computer code. An alternative way of reproducing parton showers is by means of Monte Carlo methods. The splitting function (times the particle emitted phase space) can be interpreted as a probability, and one is able to simulate parton emissions in the same way one simulates a radioactive particle decay. We are not going to expand on this: the interested reader may find more details in Ref. [60]. Thanks to Monte Carlo parton showers, starting with a handful of hard particles, we can reconstruct the complexity of a real event, even though approximately.

However, at some point in the radiation process, when the scale of energy involved is $\lesssim$ 1 GeV, we enter in the non-pertubative regime (see Fig. 3). During *hadronization*, partons start recombining and forming bound states, baryons and mesons, which are the ones actually detected. A precise description of the mechanisms behind hadronization is still unknown, and numerical simulations are based only on phenomenological models.

1.8.2 Jet definition

How is possible to link the complexity of the final state (before or after hadronization) to the picture provided by fixed-order perturbative QCD calculations, with only a few hard particles in the final state? We would like to define an observable for the description of a final state composed of QCD particles which is well defined in perturbation theory, is formulated in such a way it could be straightforwardly implemented both in theoretical predictions or in experimental analysis, and in addition does not depend too much on the hadronization process.

This leads us to the concept of *jet*. Naively, a jet may be thought of as a bunch of energetic collimated particles distributed around a determinate direction. This direction would reflect the direction of the original parton which originated the shower, since QCD dynamics encourages the production of soft and collinear particles, and thus jets are somehow proxies for the hard partons. The first jet definition dates back to 1977, with the article of Sterman and Weinberg [61]. It was based on the idea of jet as a cone containing a determinate fraction of the total energy present in the final state of the event. Over time many other definitions have been proposed, we refer the reader to Ref. [62] for a detailed discussion. Remember that an important property that observables involving jets (as any other observable) has to satisfy is IRC safety, see Eqs. (1.41)-(1.42), to guarantee cancellation of divergences and therefore calculability.

Nowadays, jets are almost always defined by means of sequential recombination *clustering algorithms*. Clustering is a form of unsupervised learning for classification, based on a proximity distance between objects. For instance, the family of gen-k_t algorithms can be formulated in the following way. For a given set of QCD partons (or hadrons) in the final state of a collider event, with each particle described by the kinematic variables (p_t, y, ϕ) discussed in Appendix 1.B, we define the following distances:

$$d_{ij} = \min\left(p_{ti}^{2n}, p_{tj}^{2n}\right) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{ti}^{2n}, \tag{1.94}$$

where the parameter R is the jet *radius*, and the angular distance ΔR_{ij}^2 is defined as:

$$\Delta R_{ij}^2 = (\Delta y_{ij})^2 + (\Delta \phi_{ij})^2 \,. \tag{1.95}$$

We first find the minimum among the whole set of distances d_{ij} and d_{iB}. If the minimum is a d_{iB} distance, the i-th particle is declared to be a jet and it is removed from the list of particles; otherwise, if the minimum is a d_{ij} distance, the i-th and the j-th particles are combined to form a new "protojet" (we usually sum the 4-momenta of the two particles to define the 4-momenta of the protojet). The procedure is iterated until there are no more particles are left. In the end, only the jets with p_t greater than some p_t^{cut} are retained.

Let us briefly comment how the gen-k_t algorithm behaves for different values for the exponent n which appears in Eq. (1.94):

- $n = 1$ (k_t algorithm [63–65]): in the collinear limit, the distance is proportional to the transverse momentum of i relative to j, and the clustering sequence emulates the QCD branching process. It tends to follows the aggregation of soft particles, thus resulting in irregular jet borders.

- $n = 0$ (Cambridge/Aachen (C/A) algorithm [66,67]): in this case, the distance is purely angular, and the clustering will follow the angular ordering among particles. This algorithm is useful to study the substructure of jets.

- $n = -1$ (anti-k_t algorithm [68]): here, the hard particles are clustered in the first steps of the algorithm. If the distance between two hard particles is greater than R, each hard particle will tend to accumulate soft particles, providing perfect conical shapes.

The anti-k_t algorithm has been adopted as the default choice by the LHC experimental collaborations. This is due to different reasons. From the experimental point of view, an algorithm which returns circular shapes in the y–ϕ plane is better for calibration purposes. In addition, when adopting algorithms of the gen-k_t family, computational geometry techniques allow to reduce the number of operations for clustering of N particles from $\mathcal{O}(N^3)$ to $\mathcal{O}(N \log N)$ [69], making the sequential recombination algorithms orders of magnitude faster than cone algorithms. The package FASTJET [70] implements the algorithms of the gen-k_t family, as well as many other tools for jet clustering and jet substructure studies.

1.9 Jet substructure

In the context of analyses involving hadronic final states, a well-behaved jet definitions is a basic requirement. However, when collecting radiation inside jets, we lose valuable information that can help us to better describe the event. This observation is even more true in the case of highly *boosted* objects i.e. objects whose transverse momentum p_t is very large compared to the invariant mass m. For instance, suppose that we have a W boson decaying hadronically in two jets. A back-of-the-envelope calculation shows that the angle θ between the decay products is proportional to m/p_t. For values of p_t much larger than m, the two jets are not individually resolved anymore, as they are clustered together is a single jet, and thus in the boosted regime the 2-pronged decay structure appears as a QCD jet with the same mass.

Another typical case scenario where one may be interested in going beyond the monolithic picture of jet is in the context of *quark- versus gluon-jet discrimination*. Namely, being able

to disentangle jets that can be thought of as originating by the fragmentation of a high-energy quark from the ones originating from a gluon. In a first approximation, the radiation potential for a gluon is greater than one for a quark, due for the different colour factor i.e. $C_A > C_F$. Hence, by quantifying the amount of radiation around a hard prong we can — in principle at least — discriminate between a quark- and a gluon-jet[11].

Jet substructure techniques have been developed in the past decade to address these topics. Exciting tools have been developed, successfully tested, and are currently used in experimental analyses. Here we limit ourselves to discuss one of the most used jet substructure variables, the N-subjettiness [73], that we will adopt in the studies of Chapter 5, and we refer the reader to a dedicated textbook [59] for a detailed discussion about other jet substructure tools.

As its name suggests, N-subjettiness aims to identify jets with an N-prong structure. It takes inspiration from the event-shape N-jettiness [74]. In order to achieve this, a set of axes $a_1, \ldots, a_N$ is introduced and the N-subjettiness is defined as

$$\tau_N^{(\beta)} = \frac{1}{p_{tJ} R_0^\beta} \sum_{i \in \text{jet}} p_{ti} \min(\Delta R_{ia_1}^\beta, \ldots, \Delta R_{ia_N}^\beta),$$

where β is a free parameter, p_{tJ} and R_0 are the jet transverse momentum and radius respectively and ΔR_{ia_j} is the distance between particle i and the axis a_j in the azimuth-rapidity plane, see Eq. (1.95). The axes a_j can be defined in several ways (see for instance Ref. [59]). One common choice is to re-cluster the jet with an algorithm of the gen-k_t family and take a_i as the N exclusive jets. In this case, if we set the exponent p of the gen-k_t algorithm to $p = 1/\beta$, we match the distance measure used for the clustering with the one used to compute N-subjettiness. This property can be used in analytic calculations. Indeed, IRC safety guarantees that N-subjettiness distributions can be meaningfully calculated in perturbative QCD. These distributions have been the subject of several theoretical investigations [75–78].

N-subjettiness has been originally introduced in the context of boosted objects phenomenology. If ones defines the N-subjettiness ratio $\tau_{N,N-1}^{(\beta)}$ as

$$\tau_{N,N-1}^{(\beta)} = \frac{\tau_N^{(\beta)}}{\tau_{N-1}^{(\beta)}},$$

one is provided with a good discriminating variable for N-prong signal jets against the QCD background. Indeed, N-subjettiness quantify the amount of radiation around N prongs, then for a structure with N prongs one expects a transition point between τ_{N-1} (small) and τ_N (large). For instance, by imposing a cut on $\tau_{21} < \tau_{\text{cut}}$ one can differentiate W jets from QCD jets.

Since gluon jets have generally larger value of N-subjettiness compared to quark jets, distributions based N-subjettiness can be applied as well in the context of quark/gluon discrimination. As an example, let us calculate what is the probability of having τ_1 smaller than a cut τ in the case of a quark or a gluon jet[12]. In other words, we would like to calculate the cumulative

---

[11]Note that the association of a single parton to a final-state jet is an intrinsically ambiguous — if not ill-defined, operation, essentially because of higher-order corrections. It is however possible to employ operational definitions that associate the value of a measurable quantity to an enriched sample of quarks or gluons [71, 72].

[12]Our discussion is along the lines of Ref. [79].

probability distribution $\Sigma(\tau_1 < \tau)$. If $\beta = 2$, τ_1 is simply given by:

$$\tau_1 = \sum_i z_i \theta_i^2 = \sum_i \rho_i = \rho \,, \tag{1.98}$$

where z_i is the momentum fraction and θ_i is the angle (in the unit of the jet radius) of each gluon emission off the quark (or gluon) hard line. Note that τ_1 is the same as the (normalised) jet mass $\rho = (m/p_t R_0)^2$, and each emission contributes with $\rho_i = z_i \theta_i^2$ to the total jet mass.

For the purpose of this calculation and of Chapter 5, we introduce the radiator $\mathcal{R}(X)$, calculated considering the contribution of a soft and collinear gluon carrying a fraction z_i of the hard parton transverse momentum p_t, emitted at an angle $\theta_i R_0$ with respect to the hard parton direction, and imposing that $\rho_i > X$. At LL accuracy it is the same for quarks and gluons and its expression reads

$$\mathcal{R}(X) = 2 \int_0^1 \frac{d\theta_i^2}{\theta_i^2} \int_0^1 \frac{dz_i}{z_i} \frac{\alpha_S(z\theta p_t R_0)}{2\pi} \Theta(z_i \theta_i^\beta > X) = \frac{2}{\beta} \int_X^1 \frac{d\rho_i}{\rho_i} \int_{\rho_i}^1 \frac{dz_i}{z_i} \frac{1}{\pi} \alpha_S \left(z_i^{\frac{\beta-1}{\beta}} \rho_i^{\frac{1}{\beta}} p_t R_0 \right). \tag{1.99}$$

For clarity, let us neglect the running of α_S. We then find:

$$\mathcal{R}(X) = \frac{2\alpha_S}{\beta\pi} \int_X^1 \frac{d\rho}{\rho} \log\left(\frac{1}{\rho}\right) = \frac{\alpha_S}{\beta\pi} \log^2\left(\frac{1}{X}\right). \tag{1.100}$$

At LL accuracy, one emission dominates the jet mass and all the others just provide tiny corrections. This is easy to understand, as the emission probability in the soft and collinear limit is

$$P(z) \, d\theta \, dz = C_R \frac{\alpha_S}{\pi} \, d\log\left(\frac{1}{\theta}\right) \, d\log\left(\frac{1}{z}\right) \tag{1.101}$$

Eq. (1.101) implies that the emissions are uniformly distributed in the $\log(1/\theta)$-$\log(1/z)$ plane, and are thus *exponentially* far apart on a linear scale. This implies that we can approximate the condition $\tau_1 < \tau$ as:

$$\tau_1 = \sum_i \rho_i \simeq \max_i(\rho_i) < \tau \,, \tag{1.102}$$

and the last inequality is equivalent to the requirement $\rho_i < \tau \, \forall i$. All the real emissions are thus independent, and we have to sum over any possible number p of real emissions. Conversely, the virtual emissions are (by definition) unconstrained, and naturally exponentiate. In the end we obtain

$$\Sigma(\tau_1 < \tau) = \sum_{p=0}^\infty \left[\frac{1}{p!} \left(C_R \frac{\alpha_S}{\pi} \int_0^\tau \frac{d\rho_i}{\rho_i} \ln\left(\frac{1}{\rho_i}\right) \right)^p \right] \left[\exp\left(-C_R \frac{\alpha_S}{\pi} \int_0^1 \frac{d\rho}{\rho} \ln\left(\frac{1}{\rho}\right) \right) \right]$$

$$= \exp\left(-C_R \frac{\alpha_S}{\pi} \int_\tau^1 \frac{d\rho}{\rho} \ln\left(\frac{1}{\rho}\right) \right) \equiv \exp\left(-C_R \mathcal{R}(\tau) \right). \tag{1.103}$$

Hence, the cumulative distribution for quark and gluon jet at LL differs only by the colour factor C_R. This result is usually called *Sudakov factor*, and it has a nice physical interpretation related to the partial cancellation of real and virtual emissions. Indeed, real and virtual emissions cancel each other, apart from the region where $\tau_1 > \tau$, where real emissions are forbidden and only
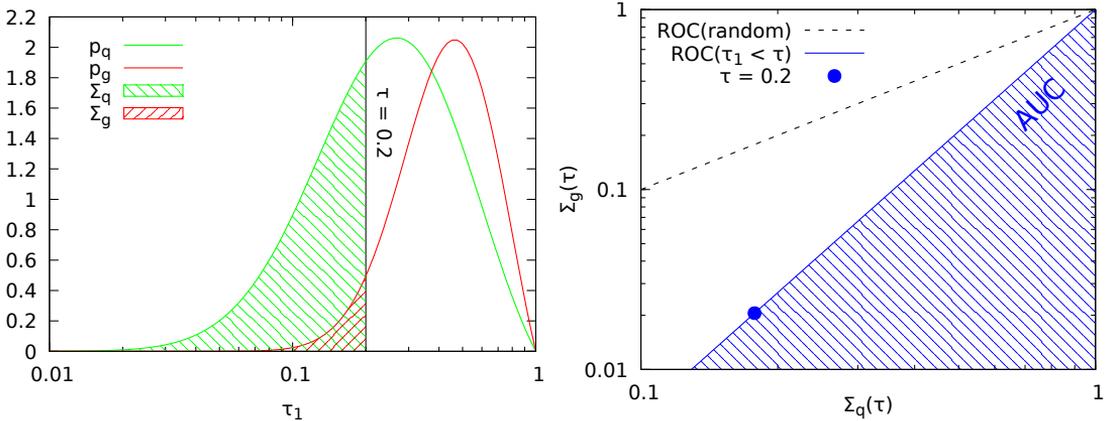
Figure 1.3: Left: differential $p_i(\tau_1)$ and cumulative $\Sigma_i(\tau_1 < \tau)$ distributions for quark and gluon jets. Right: ROC curves for a random classifier and the classifier corresponding to a simple cut on τ_1. In both plots an illustrative cut equal to $\tau = 0.2$ is shown.

virtual emissions are allowed, giving rise to an exponential factor with negative sign.

We can straightforwardly derive the differential distribution p for observing a value $\tau_1 = \tau$ by simply taking the derivative of Eq. (1.103):

$$p(\tau_1 = \tau) = C_R \frac{2\alpha_s}{\beta\pi} \frac{\log(1/\tau)}{\tau} \exp\left(-C_R \frac{\alpha_s}{\beta\pi} \log^2\left(\frac{1}{\tau}\right)\right) . \tag{1.104}$$

These differential distributions for quark and gluon jets are shown in Fig. 1.3 on the left. The cumulative distributions for an illustrative cut $\tau = 0.2$ are also shown. It is clear from Fig. 1.3 that the less the two differential distributions overlap, the more a simple cut τ on τ_1 is effective when separating quark jets from gluon jets.

In order to assess the discriminating power of an observable, Receiver Operating Characteristic (ROC) curves are often considered. These curves show the background (gluon) efficiency against the signal (quark) efficiency and are remarkably useful to directly compare the performance of different tools. In terms of the normalised cumulative distributions for quark and gluon jets, Σ_q and Σ_g respectively, the ROC curve is defined as:

$$\text{ROC}(x) = \Sigma_g\left(\Sigma_q^{-1}(x)\right) , \tag{1.105}$$

where x is the signal (quark) efficiency. If the observable is τ_1, it's trivial to find an explicit expression for the ROC curve:

$$\Sigma_g(\tau) = \left[\exp\left(-C_F \mathcal{R}(\tau)\right)\right]^{C_A/C_F} \equiv \left[\Sigma_q(\tau)\right]^{C_A/C_F} \longrightarrow \text{ROC}(x) = x^{C_A/C_F} . \tag{1.106}$$

Such a ROC curve is shown in Fig. 1.3 on the right. The area under the ROC curve (AUC) can be used as a single-number quantifier of the discrimination power. $\text{AUC} = 0$ corresponds to perfect performance, whereas $\text{AUC} = 1/2$ to a random classifier. In this latter case the ROC

curve is just the straight line $y = x$, see Fig. 1.3. The AUC for a cut on τ_1 is given by:

$$\text{AUC} = \int_0^1 dx \, x^{C_A/C_F} = \frac{C_F}{C_F + C_A} \simeq 0.308 \,. \tag{1.107}$$

1.10 Structure of QCD predictions

We conclude this Chapter by showing how a QCD perturbative result is organized, and explaining how any fixed order prediction retains a dependence on the unphysical renormalization and factorization scales.

1.10.1 Scale dependence

In perturbative QCD, we express the theoretical prediction as the sum of a series in α_s. For example if we consider a total cross section we can write:

$$\sigma = \sum_{k=0}^{\infty} c_k \alpha_s^k = c_0 + c_1 \alpha_s + c_2 \alpha_s^2 + \dots \tag{1.108}$$

If we retain only some terms of the expansion, we can claim a *fixed-order prediction*:

$$\sigma_p = \sum_{k=0}^{p} c_k \alpha_s^k \tag{1.109}$$

People usually refer to the p-th order as the N^pLO order: leading order (LO), next-to-leading order (NLO), next-to-next-to-leading order (NNLO) and so on. As we include further orders in the expansion, we can reasonably hope to see systematic improvements in the accuracy of our predictions and if we have an arbitrarily large numbers of terms we can imagine to reach the "true sum" of the series[13].

The renormalization of the theory leads α_s to depend on the renormalization scale μ_R — in hadronic collisions, we would have also a dependence on the factorization scale μ_F, but let us neglect this dependence for a moment. In order to calculate the total cross section we need to make a choice for μ_R. Suppose that we evaluate the coupling at $\mu_R = Q$, and then the coefficients refer to an expansion around $\alpha_s(Q)$:

$$\sigma = \sum_k c_k \alpha_s^k(Q) = c_0 + c_1 \alpha_s(Q) + c_2 \alpha_s^2(Q) + c_3 \alpha_s^3(Q) + \dots \tag{1.110}$$

Choosing another value for μ_R leads to a variation of α_s and it seems that this arbitrariness introduces a dependence on the unphysical scale μ_R in the physical observable σ. However it can be shown that with a different choice of μ_R, given an infinite number of terms in the α_s expansion, the coefficients c_k change their values in order to compensate for the variation of α_s and this cancellation leads the final result to be independent of the choice of μ_R. In other

---

[13]Since QCD series are usually asymptotic and not convergent, the "true sum" of the series must be defined by an appropriate prescription.

words, this means that Eq. (1.110) must be equivalent to:

$$\sigma = \sum_{k=0}^{\infty} c_k(Q, \mu_R) \alpha_s^k(\mu_R) \,. \tag{1.111}$$

However, when we deal with a fixed-order prediction, σ_p, this cancellation cannot be exact. Any fixed order prediction has an unavoidable scale dependence: $\sigma_p = \sigma_p(\mu_R)$. It can be shown that the scale dependence is always of order α_s^{p+1}, so it is greatly reduced with an increasing of the perturbation order. Thus, order by order we have the relation:

$$\sigma_p(\mu_R) = \sigma_p(Q) + \mathcal{O}(\alpha_s^{p+1}) \,. \tag{1.112}$$

This last relation allows us to find the expression of the $c_k(Q, \mu_R)$ coefficients as a function of the c_k ones. For instance, by using Eq. (1.12), we find that $c_2(Q, \mu_R)$ is constrained to be:

$$c_2(Q, \mu_R) = c_2 - 2b_0 c_1 \ln \frac{\mu_R}{Q} \,. \tag{1.113}$$

Similarly, for what concerns the factorization scale μ_F, by imposing order by order the μ_F-independence of the physical prediction, one finds relations similar to Eq. (1.113), but involving convolutions with the Altarelli-Parisi splitting kernels P_{ij}, see Sec. 1.5. Indeed, like the β-function of Eq. (1.9) controls the dependence on μ_R, the DGLAP equation of Eq. (1.70) controls the dependence on μ_F.

1.10.2 Missing higher order uncertainties and scale variation

Strictly speaking, the fixed-order prediction in Eq. (1.109) is given without uncertainty, because the coefficients c_k are perfectly determined. However, it is customary to see σ_p as an estimation of the infinite series of Eq. (1.108) and then we need to associate a missing higher order uncertainty (MHOU) to any fixed-order prediction. This uncertainty is related to the fact that we are neglecting the remainder:

$$\Delta_p = \sum_{k=p+1}^{\infty} c_k \alpha_s^k = c_{p+1}\alpha_s^{p+1} + c_{p+2}\alpha_s^{p+2} + \dots \tag{1.114}$$

In order to estimate the presumed value of Δ_p and then the uncertainty related to σ_p, one typically quotes an uncertainty interval $[\sigma_p^-, \sigma_p^+]$ around $\sigma_p(\mu_R = Q)$. One choice for such an interval is:

$$[\sigma_p^-, \sigma_p^+] = [\min\{\sigma_p(Q/r), \sigma_p(rQ)\}, \max\{\sigma_p(Q/r), \sigma_p(rQ)\}] \tag{1.115}$$

where r is a factor (usually $r = 2$ or 4) fully conventional. This approach to estimate MHOUs is referred to as *scale variation* procedure, because it relies on calculations of σ_p at renormalization scales different from the central one.

A priori there is no reason why this interval should represent a sensible estimate of the remainder Δ_p of the series, because the former is a function of the c_k coefficients for $k \leq p$, while the latter is a function of the c_k coefficients for $k \geq p$. It can be shown that the interval given by scale variation and the remainder of the series are comparable under the assumption that all the coefficient in the series share the same magnitude [80], and thus the use of scale

variation is in some sense justified, even if this procedure has no rigorous theoretical foundation.

If our predictions depend also on the factorization scale μ_F, we could still use a prescription similar to Eq. (1.115), by choosing as uncertainty interval the one obtained after a concurrent variation of μ_R and μ_F by a factor of r around the central scale chosen. However, as suggested in Ref. [81], during the variation of the scales it is preferable to avoid values of μ_R or μ_F which differ by a factor greater then r. In the case $r = 2$, this prescription reduces to the well-known *7-point scale variation rule*, see Fig. 1.4.
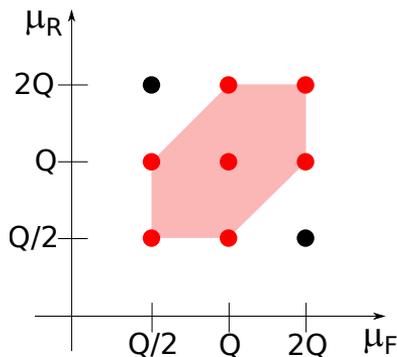

Figure 1.4: *7-point scale variation rule*: the renormalization and the factorization scales are varied around the central energy Q, but the opposite ends of the square - which imply a factor 4 of variation between the scales - are not taken into account. The uncertainty band is obtained by taking the minimum and the maximum value of the envelope.

1.10.3 What is the right scale?

We have seen how any fixed-order prediction, due to a truncation of the perturbation series, retains a scale dependence. One can then try to make a "good" choice for the scales, in order to make the truncated and the all-orders prediction as similar as possible. Renormalization scale or factorization scale μ always appears in fractions like μ/Q inside a logarithmic term, where Q is a characteristic hard scale of the process. Thus, choosing a scale μ radically different from Q generates large logarithms which deteriorates the convergence of the perturbative series. The simple choice $\mu = Q$ could be a good choice for processes in which only one characteristic energy appears e.g. usually the inclusive ones, such as $e^+e^- \to$ hadrons.[14] The situation is quite different for processes which naturally involve different physical scales, for example processes with a more exclusive final state. Jet observables we shall study in Chapters 2–4 are multi-scale processes and the choice of the proper scale is a delicate issue.

In any case, remember that, while there could be choices *better* than others[15], there is never a *best* choice, as fixing a scale does not remove the theoretical error on the prediction: for a quantity up to order $\mathcal{O}(\alpha_S^p)$, there will always be an error of order $\mathcal{O}(\alpha_S^{p+1})$.

---

[14]Actually there is no reason to have μ *exactly* equal to Q. If we choose for example $\mu = 2Q$ or $\mu = Q/2$ for a fixed order prediction at order p, there are small logarithms which appear at order $\mathcal{O}(\alpha_S^{p+1})$. But this does not improve nor deteriorate the convergence of the series, because we don't know the value of the c_{p+1} coefficient.

[15]For instance, in processes with one distinctive scale Q, the choice better than others is Q, since this choice resums the $\log(\mu_R/Q)$ terms to all orders.

1.A QED and QCD Lagrangian densities

In this appendix we report explicit expressions for the Lagrangian densities in covariant gauges in both QED and in QCD, and we compare them side by side. They are given by:

$$
\mathcal{L}_{\text{QCD}} = -\frac{1}{4}\mathcal{F}_{\mu\nu}^{A}\mathcal{F}^{A\,\mu\nu} \tag{1.116}
$$
$$
+ \sum_{q} \overline{\psi}_{q,a}(i\gamma^{\mu}\partial_{\mu} - m_q)\psi_{q,a}
$$
$$
+ \sum_{q} \overline{\psi}_{q,a}(-g_S\gamma^{\mu}t_{ab}^{C}\mathcal{A}_{\mu}^{C})\psi_{q,b}
$$
$$
+ \left[\mathcal{L}_{\text{GF}} = -\frac{1}{2\xi}\left(\partial^{\mu}\mathcal{A}_{\mu}^{A}\right)^{2}\right]
$$
$$
+ \left[\mathcal{L}_{\text{ghost}}\right]
$$

$$
\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.117}
$$
$$
+ \sum_{i} \overline{\psi}_{i}(i\gamma^{\mu}\partial_{\mu} - m_i)\psi_{i}
$$
$$
+ \sum_{i} \overline{\psi}_{i}(-e\gamma^{\mu}Q_{i}A_{\mu})\psi_{i}
$$
$$
+ \left[\mathcal{L}_{\text{GF}} = -\frac{1}{2\xi}\left(\partial^{\mu}A_{\mu}\right)^{2}\right]
$$

A sum over repeated indices is understood. Fermion and boson fields appear:

$\psi_{q,a}$ is the quark field of flavour q and mass m_q, with colour index a running from 1 to $N_c = 3$. The sum with index q runs over the different flavours of quarks. Quark fields transform under the fundamental representation of SU(3). $\mathcal{A}_{\mu}^{C}$ is the gluon field, the gauge field of SU(3), with the index C running from 1 to $N_c^2 - 1 = 8$. Gluon fields transform under the adjoint representation of SU(3).

ψ_i is a fermion field with mass m_i. The sum with index i runs over electrically charged particles (both leptons and quarks). This field transforms under U(1). A_μ is the photon field, the gauge field of U(1).

The second line of Eqs. (1.116)-(1.117) describes a kinetic propagation term for the fermion fields. The third line of both Lagrangians encodes the fermion-boson interaction vertex. Such a vertex contains:

t_{ab}^{C} are the $N_c^2 - 1 = 8$ generators of SU(3), whose 3-dimensional representation is provided by the eight Gell-Mann matrices, hermitean and traceless (see Ref. [24] for explicit expressions).

Q_i is the charge of the field ψ_i in units of the positron charge e.g. the electron has charge $Q = -1$.

The first line of Eqs. (1.116)-(1.117) contains the contraction of field strength tensors:

$\mathcal{F}_{\mu\nu}^A$ is the QCD field strength tensor, built as:

$$\mathcal{F}_{\mu\nu}^A = \partial_\mu \mathcal{A}_\nu^A - \partial_\nu \mathcal{A}_\mu^A - g_s f_{ABC} \mathcal{A}_\mu^B \mathcal{A}_\nu^C \tag{1.118}$$

where f_{ABC} are the structure constants of SU(3), appearing in the defining relation of the algebra associated to the group:

$$\left[t^A, t^B\right] = i f_{ABC} t^C \tag{1.119}$$

$F_{\mu\nu}$ is the QED field strength tensor:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.120}$$

Note the absence of the *non-abelian* term. The structure constants of U(1) are null, and the commutator in Eq. (1.119) is always zero.

By expanding Eq. (1.118), we find a kinetic term for the gluons and self-interaction terms among gauge fields, namely the triple-gluon vertex (of order g_s) and the four-gluon vertex (of order g_s^2). These vertices are present because the carriers of the colour force are themselves coloured, unlike the electrically neutral photon. In fact, there are no interactions among photons in QED, with Eq. (1.120) encoding just the propagation of the photon field.

The fourth line in both Lagrangians is related to the choice of the gauge. The class of gauges with the given expression for $\mathcal{L}_{\mathrm{GF}}$ are called covariant gauges, and the gauge with $\xi = 1$ (Feynman gauge) is particularly convenient, as the expression for the gluon and photon propagator simplifies in this gauge.

As for the term on the fifth line in Eq. (1.116), $\mathcal{L}_{\mathrm{ghost}}$, this is only present in non-abelian theories. It is related to the Faddeev-Popov ghosts, fields with spin 0 and fermionic statistic. Their presence is required in covariant gauges, in order to cancel the unphysical degrees of freedom appearing in loops.

Finally, note that both in QCD and in QED the gauge fields are massless, as there is no gauge invariant way of including a mass for them.

1.B   Collider kinematics

In this appendix we introduce the variables usually adopted to study the kinematics of final state particles at colliders. In e^+e^- collisions, one can choose a spherical coordinate system around the interaction point, with individual energies E_i and angles θ_{ij} between particles as good variables. In *pp* collisions, as we have seen in Sec. 1.5, each parton carries only a momentum fraction x of the parent proton momentum, and these fractions are different for the two partons. Thus the interacting partons usually have a boost in the direction of the beam line (a *longitudinal* boost), and spherical symmetry is broken under this kind of boost.

Hence, in hadronic collisions we adopt a cylindrical coordinate system. We recast the 3-momentum $\mathbf{p}$ of a final state particle in the following form:

$$\mathbf{p} = (p_x, p_y, p_z) = (\mathbf{p_t}, |\mathbf{p}| \cos\theta) , \tag{1.121}$$

where θ is the angle between $\mathbf{p}$ and the beam axis — supposed to lie along the z direction, while the 2-dimensional *transverse momentum* $\mathbf{p_t}$ is defined as the component of $\mathbf{p}$ in the plane

transverse to the beam axis. Explicitly we write:

$$\mathbf{p_t} = p_t(\cos\phi, \sin\phi)\,, \quad p_t = \sqrt{p_x^2 + p_y^2} = |\mathbf{p}|\sin\theta\,, \tag{1.122}$$

with ϕ an azimuthal angle. Under longitudinal boosts, $\Delta\phi$ differences are invariant; however, this is not true for $\Delta\theta$ differences. Therefore, we introduce another variable, the *rapidity y*, defined as:

$$y = \frac{1}{2}\log\left(\frac{E + p_z}{E - p_z}\right)\,. \tag{1.123}$$

often used to parametrize Lorentz transformations. Rapidity differences Δy can be proven to be invariant under longitudinal boost. Moreover, in the case of massless particles, for which $E = |\mathbf{p}|$, the rapidity coincides with another variable, the *pseudorapidity η*, defined as:

$$\eta = \frac{1}{2}\log\left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z}\right) = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{1.124}$$

which is a direct proxy for the polar angle θ. Note that in general pseudorapidity differences are *not* invariant under longitudinal boosts, unlike rapidity differences.

Finally note that, by using the transverse momentum p_t, the rapidity y and the azimuthal angle ϕ, we can write the 4-momentum of a particle with mass m as:

$$p = \left(\sqrt{p_t^2 + m^2}\,\cosh y, p_t\cos\phi, p_t\sin\phi, \sqrt{p_t^2 + m^2}\,\sinh y\right)\,. \tag{1.125}$$

For a massless particle, the latter expression reduces to:

$$p = p_t(\cosh\eta, \cos\phi, \sin\phi, \sinh\eta)\,. \tag{1.126}$$

Theory calculations for jet processes

The purpose of this chapter is to introduce the reader to the available theory calculations of jet processes at pp colliders, in increasing order of accuracy. We start in Section 2.1 by providing an explicit derivation of leading order (LO) results. Then, in the subsequent sections, we present an overview of QCD higher order corrections to jet processes. We first discuss the next-to-leading order (NLO) corrections (Sec. 2.2). We then report some details about the recent computation of next-to-next-to-leading order (NNLO) corrections (Sec. 2.3). Finally, we comment on the issues about predictions for the single-jet inclusive observable that have emerged (Sec. 2.4).

2.1 Leading order cross sections

As a warm-up, in this section we calculate predictions for jet observables at hadron colliders at the lowest perturbative order. These results date back to the 1970s [82] and can be found on standard textbooks [14] or in lecture notes [83]. However, it is useful to revise them, as they provide a good starting point for forthcoming discussions. We will first study the dynamics and the kinematics of jet production processes, then we will present explicit results both for inclusive jet and dijet differential cross sections.

2.1.1 Dynamics

At lowest order, all of the $2 \rightarrow 2$ partonic processes which involve quarks and gluons contribute. If we denote as (p_a, p_b) and (p_1, p_2) the incoming and the outgoing momenta respectively, the fully differential partonic cross section for a specific subprocess $ij \rightarrow kl$ (with these indices running over quarks, antiquarks and gluon, with all the allowed quark flavours) may be written as

$$\frac{d\hat{\sigma}_{ij \rightarrow kl}}{d\Phi_2} = \frac{g_s^4}{2(p_a + p_b)^2} \overline{\sum} |\mathcal{M}_{ij \rightarrow kl}|^2 \,, \tag{2.1}$$

where the fraction on the r.h.s. is the flux factor, $d\Phi_2$ is the two-particle Lorentz invariant phase space (containing the momentum conservation delta) and $|\mathcal{M}_{ij \rightarrow kl}|^2$ is the matrix element squared at leading order, with the strong coupling g_s factorised out.

The symbol $\overline{\sum}$ denotes an average (sum) over the initial (final) state spins and colours i.e.

$$\overline{\sum}|\mathcal{M}_{ij\to kl}|^2 = \frac{1}{1+\delta_{kl}}\frac{1}{N_i}\frac{1}{N_j}\left(\frac{1}{4}\sum_{\text{spin,colour}}|\mathcal{M}_{ij\to kl}|^2\right) \equiv \frac{M_{ij\to kl}(\hat{s},\hat{t},\hat{u})}{(1+\delta_{kl})N_iN_j}\,, \qquad (2.2)$$

with N_q or $N_{\bar{q}}$ equal to $N_c = 3$ and N_g equal to $N_c^2 - 1 = 8$, and where we have implicitly defined $M_{ij\to kl}$ for later convenience. Note that we are including also a statistical factor in the definition, necessary in the case of final state identical partons. The variables appearing in Eq. (2.2) are the usual Mandelstam invariants, defined as

$$\hat{s} = (p_a + p_b)^2\,, \quad \hat{t} = (p_1 - p_a)^2\,, \quad \hat{u} = (p_1 - p_b)^2 \qquad (2.3)$$

and satisfying $\hat{s} + \hat{t} + \hat{u} = 0$, since our particles are all massless.

We need explicit expressions for $M_{ij\to kl}$. Because of crossing symmetry, which relates the expression of a generic QCD amplitude upon exchange of an incoming particle for an outgoing antiparticle, there are only four independent building blocks, whose representative diagrams are depicted in Fig. 2.1. They can be read from e.g. [14], or computed directly with automatic tools e.g. FeynCalc [84,85]:

$$M_{qq'\to qq'} = \frac{N_c^2 - 1}{2}\left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2}\right)\,, \qquad (2.4)$$

$$M_{qq\to qq} = \frac{N_c^2 - 1}{2}\left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2} + \frac{\hat{s}^2 + \hat{t}^2}{\hat{u}^2}\right) - \frac{N_c^2 - 1}{N_c}\left(\frac{\hat{s}^2}{\hat{u}\hat{t}}\right)\,, \qquad (2.5)$$

$$M_{q\bar{q}\to gg} = \frac{(N_c^2 - 1)^2}{2N_c}\left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{t}\hat{u}}\right) - N_c(N_c^2 - 1)\left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}\right)\,, \qquad (2.6)$$

$$M_{gg\to gg} = 4N_c^2(N_c^2 - 1)\left(3 - \frac{\hat{t}\hat{u}}{\hat{s}^2} - \frac{\hat{s}\hat{u}}{\hat{t}^2} - \frac{\hat{s}\hat{t}}{\hat{u}^2}\right)\,. \qquad (2.7)$$

In the previous equations, q denotes any quark of given flavour, $\bar{q}$ the related antiquark, q' a quark of different flavour (with its own antiquark $\overline{q'}$), and g is the gluon. It is possible to write down analogous expressions for the remaining processes by switching $\hat{s} \leftrightarrow \hat{u}$ or $\hat{s} \leftrightarrow \hat{t}$. We need also to tune accordingly the colour and statistical factors in Eq. (2.2).

At level of matrix elements, Eq. (2.7) gives by large the dominant contribution[1]. The reason is twofold: on one hand, processes with gluons are more important because of the larger colour charge; on the other hand, diagrams with a gluon exchanged in the $\hat{t}$ or $\hat{u}$ channel contribute more than a propagator in the $\hat{s}$ channel, as noted in Ref. [22]. This latter fact can be understood by simply computing explicit expressions for the partonic Mandelstam variables:

$$\hat{t} = -\frac{\hat{s}}{2}(1 - \cos\hat{\theta})\,, \quad \hat{u} = -\frac{\hat{s}}{2}(1 + \cos\hat{\theta})\,, \qquad (2.8)$$

where $\hat{\theta}$ is the scattering angle in the partonic center of mass frame. We clearly see that $\min(|\hat{t}|, |\hat{u}|) \leq \hat{s}/2$, and thus terms with $\hat{t}^2$ and $\hat{u}^2$ dominate the amplitude.

---

[1]This does not hold anymore after convolution with the PDFs, especially at large p_t, because the gluon PDF is suppressed at large x.
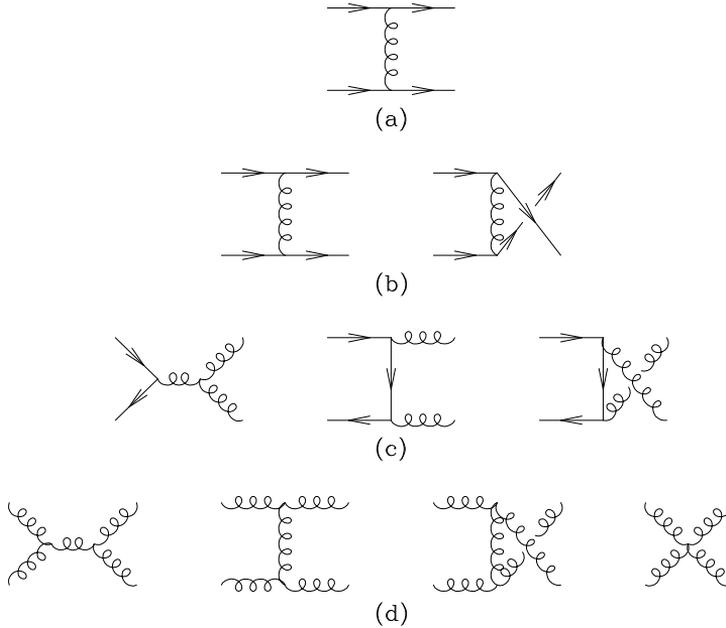
(a)

(b)

(c)

(d)

Figure 2.1: Representative Feynman diagrams for jet observables at leading order. The diagrams depicted in (a), (b), (c) and (d) are related to Eqs. (2.4), (2.5), (2.6) and (2.7) respectively. Taken from Ref. [14].

2.1.2 Kinematics

The final state phase space reads as follows:

$$
\begin{aligned}
d\Phi_2 &= \frac{d^4 p_1}{(2\pi)^4} (2\pi)\delta(p_1^2) \frac{d^4 p_2}{(2\pi)^4} (2\pi)\delta(p_2^2) (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_a - p_b) \\
&= \frac{p_t \, dp_t \, dy \, d\phi}{2(2\pi)^3} (2\pi)\delta((p_a + p_b - p_1)^2) \,.
\end{aligned}
\tag{2.9}
$$

In the second line we have first exploited the 4-dimensional delta in order to constrain the momenta of the one of two final state partons. Then, we have rephrased the phase space of the leftover parton with the usual kinematic collider variables, p_t, y and ϕ, defined in Eq. (1.126). We have chosen to work in the hadronic c.m. frame, where we have

$$
p_a = x_a \left(\frac{\sqrt{s}}{2}, 0, 0, \frac{\sqrt{s}}{2} \right), \quad p_b = x_b \left(\frac{\sqrt{s}}{2}, 0, 0, -\frac{\sqrt{s}}{2} \right),
\tag{2.10}
$$

with s the c.m. energy squared and x_a, x_b longitudinal momentum fractions, relative to the parent hadrons, of the scattering partons. The c.m. energy in the partonic frame $\hat{s}$, defined in Eq. 2.3, also equal to the invariant mass of the dijet system m_{jj}^2, is related to s as

$$
\hat{s} = m_{jj}^2 = x_a x_b s \,.
\tag{2.11}
$$

43

Transverse momentum conservation implies that the final state system will be balanced in p_t and the final state partons will be back-to-back in azimuth i.e. $p_{t2} = p_{t1}$ and $\phi_2 = \phi_1 + \pi$, see Eq. (1.122). Energy and longitudinal momentum conservation imply:

$$\cosh y_1 + \cosh y_2 = \frac{\sqrt{s}}{2p_t}(x_a + x_b)\,,$$

$$\sinh y_1 + \sinh y_2 = \frac{\sqrt{s}}{2p_t}(x_a - x_b)\,, \tag{2.12}$$

and by exploiting hyperbolic identities, we find expressions for the average sum and difference of final state rapidities:

$$Y \equiv \frac{y_1 + y_2}{2} = \frac{1}{2} \log\left(\frac{x_a}{x_b}\right)\,, \quad y^* \equiv \frac{y_1 - y_2}{2} = \cosh^{-1}\left(\frac{\sqrt{\hat{s}}}{2p_t}\right)\,, \tag{2.13}$$

and thus:

$$y_{1,2} = Y \pm y^*\,. \tag{2.14}$$

We could have derived Eq. (2.14) by means of physical considerations, as rapidity is additive under boosts along the z direction. In fact, Y is the rapidity of the partonic system in the hadronic c.m. frame, while y^* $(-y^*)$ is the rapidity of p_1 (p_2) in the partonic c.m. frame. Also note that the introduction of y^* allows one to write an alternative expression for $\hat{s}$ (and m_{jj}^2):

$$\hat{s} = m_{jj}^2 = 4p_t^2 \cosh^2(y^*)\,, \tag{2.15}$$

and by inverting Eq. (2.13) one finds that the momentum fractions are determined to be:

$$x_{a,b} = \frac{p_t}{\sqrt{s}}\left(e^{\pm y_1} + e^{\pm y_2}\right)\,, \tag{2.16}$$

or alternatively

$$x_{a,b} = \frac{2p_t}{\sqrt{s}} \cosh y^* \, e^{\pm Y} = \frac{m_{jj}}{\sqrt{s}} e^{\pm Y}\,. \tag{2.17}$$

From Eqs. (2.16)-(2.17) we can appreciate how a double differential cross section in m_{jj} and Y, or a triple differential cross section in p_t, y_1 and y_2 (or p_t, Y and y^*) is able to fix the values of the momentum fractions x_a and x_b, although higher order kinematics has the potential to change this picture.

By using Eq. (2.11) and Eq. (2.15), we see that:

$$\frac{4p_t^2}{s} \leq x_a x_b \leq 1\,. \tag{2.18}$$

For fixed p_t, the smallest and the largest accessible x values are given by:

$$x_{\min} = \frac{4p_t^2}{\sqrt{s}} \quad x_{\max} = 1\,, \tag{2.19}$$

whereas, for fixed p_t and y, we have:

$$x_{\min} = 2\frac{p_t}{\sqrt{s}}e^{-|y|}, \quad x_{\max} = 2\frac{p_t}{\sqrt{s}}e^{+|y|}, \qquad (2.20)$$

This corresponds to a configuration with $y_1 = y_2 = Y$ and $y^* = 0$ i.e. a same-side jet event, with one large and one small x value. The configuration with $y_1 = -y_2 = y^*$ and $Y = 0$ would be instead an opposite-side jet event, with the same momentum fractions $x_a = x_b$.

Finally, for later purposes, we can rearrange the phase space in Eq. (2.9) as follows:

$$\begin{aligned} d\Phi_2 &= \frac{p_t\, dp_t\, dy\, d\phi}{2(2\pi)^3}\,(2\pi)\delta((p_a + p_b - p_1)^2) = \frac{dy}{4\pi}\frac{p_t^2}{\hat{s}} \\ &= \frac{dy}{16\pi\cosh^2(y - Y)} = \frac{dy^*}{16\pi\cosh^2(y^*)}\,. \end{aligned} \qquad (2.21)$$

We have used the delta function to constrain the p_t value to be

$$p_t = \frac{\sqrt{s}x_a x_b}{x_a\exp(-y) + x_b\exp(y)} = \frac{x_a\sqrt{s}}{2}\frac{1}{\cosh(y - Y)\exp(Y)}\,. \qquad (2.22)$$

Note that, for fixed momentum fraction x_a and x_b, we have only one degree of freedom in the final state, and one can choose the rapidity of one of the two parton, y, or the rapidity difference, y^* as the independent variable.

2.1.3 Study of subprocesses

After having studied the dynamics and the kinematics of the partonic process, we can fold in the parton distributions functions, as in Eq. (1.64), and write a result for the total cross section:

$$\sigma = \sum_{ijkl}\int dx_a dx_b\, f_i(x_a)f_j(x_b)\int d\Phi_2\,\frac{d\hat{\sigma}_{ij\to kl}}{d\Phi_2}\,. \qquad (2.23)$$

If we exploit the sum over partonic flavours in Eq. (2.23), since the matrix element does not discriminate between up and down quarks, or between quarks of different generations, but only between $q,\overline{q}$, q', $\overline{q'}$ and g (as indicated after Eqs. (2.4)-(2.7)), we see that only the following seven linear combination of parton distributions appear in the calculation:

$$\begin{aligned} F_{qq}(x_a, x_b) &= D_+(x_a, x_b)\,, \\ F_{q\overline{q}}(x_a, x_b) &= D_-(x_a, x_b)\,, \\ F_{qq'}(x_a, x_b) &= Q_+(x_a)Q_+(x_b) + Q_-(x_a)Q_-(x_b) - D_+(x_a, x_b)\,, \\ F_{q\overline{q'}}(x_a, x_b) &= Q_+(x_a)Q_-(x_b) + Q_-(x_a)Q_+(x_b) - D_-(x_a, x_b)\,, \\ F_{qg}(x_a, x_b) &= (Q_+(x_a) + Q_-(x_a))\, f_g(x_b)\,, \\ F_{gq}(x_a, x_b) &= f_g(x_a)\, (Q_+(x_b) + Q_-(x_b))\,, \\ F_{gg}(x_a, x_b) &= f_g(x_a)f_g(x_b)\,, \end{aligned} \qquad (2.24)$$

where we have defined:

$$Q_\pm(x) = \sum_{i=\{q\}} f_{\pm i}(x) , \quad D_\pm(x_a, x_b) = \sum_{i=\{q,\bar{q}\}} f_i(x_a) f_{\pm i}(x_b) . \tag{2.25}$$

The first sum runs over the set of quark (or antiquark) PDFs, while the second sum runs over both quark and antiquark PDFs. In addition, if f_i is the PDF for the quark of flavour i, f_{-i} is the PDF for the antiquark of the same flavour.

Each linear combination of PDFs define a subprocess, marked by a partonic channel ij, contributing to the total cross section in Eq. (2.23), that we can recast as:

$$\sigma = \sum_{ij} \int dx_a dx_b \, F_{ij}(x_a, x_b) \int d\Phi_2 \, \frac{d\hat{\sigma}_{ij}}{d\Phi_2} , \tag{2.26}$$

with $\hat{\sigma}_{ij}$ summing all the contributions from matrix elements with the same initial state ij. Note that this subdivision into subprocesses holds at any order in perturbation theory, as QCD higher order corrections do not open new partonic channels.

2.1.4   Explicit expressions for differential cross sections

Starting with Eq. (2.23), we can find explicit expressions for cross sections double differential in p_t and y, if we are interested in studying inclusive jet cross sections, or double differential in m_{jj}^2 and y^*, if we are interested in studying dijet cross sections.

Before proceeding, note that, at leading order, final state partons and jets are equivalent concepts i.e. regardless of the jet radius R used to cluster the event, there is always a pair of jets, with momenta equal to those of the partons, as there is no possibility for extra emissions.

Inclusive jets   By using Eq. (2.1) and Eq. (2.21), and by rearranging the integration over x_a and x_b as

$$dx_a dx_b = \frac{8 p_t dp_t}{s} \cosh^2(y - Y) dY , \tag{2.27}$$

Eq. (2.23) becomes:

$$\frac{d^2\sigma^{\mathrm{LO}}}{dp_t dy} = \frac{\alpha_s^2(\mu_R)\pi}{4p_t^3} \int \frac{dY}{\cosh^4(y-Y)} \sum_{ij} x_a x_b \, F_{ij}(x_a, x_b, \mu_F) \left(\overline{\sum} |\mathcal{M}_{ij}|^2 \right) , \tag{2.28}$$

where we have introduced $\alpha_s(\mu_R) = g_s^2(\mu_R)/4\pi$. We have also multiplied and divided by $x_a x_b$, since the hadronic PDFs are usually given in terms of $x f(x)$ to kill the small-x divergent behaviour. The limits of the Y integration can be found by using Eq. (2.16) and by imposing that $0 < x_{a,b} < 1$. One obtains:

$$Y^{\mathrm{min,max}} = y \mp \frac{1}{2} \log\left(\frac{\sqrt{s}}{p_t} e^{\pm y} - 1 \right) . \tag{2.29}$$

By integrating Eq. (2.28) over the rapidity, is possible to obtain a result for the p_t spectrum at leading order:

$$\frac{d\sigma^{\mathrm{LO}}}{dp_t} = \int_{-\bar{y}}^{\bar{y}} dy \, \frac{d^2\sigma^{\mathrm{LO}}}{dp_t dy} , \tag{2.30}$$

with the absolute value of the maximum allowed rapidity given by Eq. (2.13) with $\hat{s} = s$:

$$\bar{y} = \cosh^{-1}\left(\frac{\sqrt{s}}{2p_t}\right) . \tag{2.31}$$

Dijets Here, instead, we leave the integration over x_a and x_b untouched, and use the last member of Eq. (2.21) for $d\Phi_2$. We also transform the total cross section in Eq. (2.23) in differential in m_{jj}^2, by introducing a delta function under integration. Thus, we obtain:

$$\frac{d\sigma^{\text{LO}}}{dm_{jj}^2} = \sum_{ij} \int dx_a dx_b \, f_i(x_a) f_j(x_b) \delta(x_a x_b s - m_{jj}^2) \int \frac{dy^*}{32\pi m_{jj}^2 \cosh^2(y^*)} \, g_s^4 \, \overline{\sum} |\mathcal{M}_{ij}|^2 , \tag{2.32}$$

or if we wish to be differential also in y^*:

$$\frac{d\sigma^{\text{LO}}}{dm_{jj}^2 dy^*} = \frac{\pi \alpha_s^2(\mu_R)}{2m_{jj}^2 \cosh^2(y^*)} \sum_{ij} \frac{\tau}{s} \frac{dL_{ij}}{d\tau} \left(\overline{\sum} |\mathcal{M}_{ij}|^2\right) . \tag{2.33}$$

Here $\tau = m_{jj}^2/s$ and $dL_{ij}/d\tau$ are the differential parton luminosities introduced in Eq. (1.20). In this case, the integration over partonic momentum fractions factorize, and m_{jj}^2 and y^* fix completely the Mandelstam partonic invariants, see Eq. (2.8) with $\cos\hat{\theta} = \tanh(y^*)$.

2.1.5 Analysis of the p_t spectrum

Eq. (2.28) encodes the leading behaviour of jet cross sections as a function of the jet transverse momentum. The factor p_t^3 at the denominator in Eq. (2.28), of kinematical origin, is magnified by the matrix elements, and this results in a steeply falling distribution[2]. This naive expectation is well confirmed by experimental data, as we can seen in Fig. 2.2, where we report the results of two measurements from the ATLAS and CMS collaborations (these experimental data, among others, will serve as input for the PDF fits we will present in Chapter 4).

The steepness of the p_t differential distribution has great implications for what concerns the perturbative behaviour of this observable, in that a small shift in the jet p_t value has a huge impact on the resulting contribution to the cross section. As we will discuss in Chapter 3, higher order corrections can roughly be estimated in terms of a convolution between a tree-level $2 \to 2$ scattering and subsequent emissions within or outside jet boundaries. Since out of cone radiation has the effect of altering the p_t of the emitter, this displacement is magnified by the Born scattering and results in tremendous repercussions at the level of the cross section. In fact, for later purposes, one could split Eq. (2.30) in two terms:

$$\frac{d\sigma^{\text{LO}}}{dp_t} = \frac{d\sigma_q^{\text{LO}}}{dp_t} + \frac{d\sigma_g^{\text{LO}}}{dp_t} . \tag{2.34}$$

$d\sigma_i^{\text{LO}}/dp_t$ is the LO differential cross-section for producing a parton of flavour i of transverse momentum p_t, correctly normalized in such a way that the sum over i gives the total cross-section e.g. processes like $q\bar{q} \to gg$ will contribute twice to σ_g (and similarly for process with two quarks in the final state), while processes like $qg \to qg$ will contribute both for σ_g and for

---

[2]PDFs also play a role in this context e.g the gluon PDF is suppressed at large x.
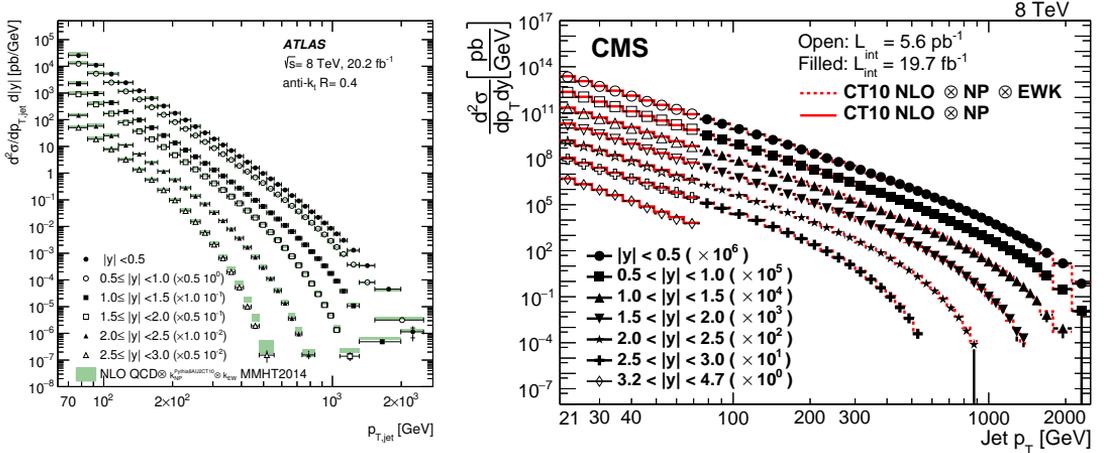
Figure 2.2: Measurements of double differential inclusive jet cross sections at the LHC (left: ATLAS collaboration, taken from ref. [86]; right: CMS collaboration, taken from ref. [87]).

σ_q, but with $\hat{t}$ and $\hat{u}$ exchanged. We will refer to $d\sigma_q$ and σ_g as the partonic spectrum for quark and gluon respectively.

We conclude the section by showing the result of the LO prediction for the p_t differential cross section, obtained through an analytic code implementing the equations written in this section (we will use and extend this code in Chapter 3). This is the blue line on the plot in Fig. 2.3, on the left. In the same plot, the green and the red lines are the partonic spectra for quark and gluon respectively. Finally, the dashed lines show the result of fitting the following power function

$$f(p_t) = \left(\frac{\text{const}}{p_t}\right)^m \tag{2.35}$$

to the tails of the distributions. We see that we need an high value of the exponent m, $m \sim 6$–7, in order to reproduce the behaviour for high value of the p_t.

2.2 Next-to-leading order (NLO) corrections

For precision phenomenology, LO results are totally unsatisfactory. In Fig. 2.3, on the right, we plot the LO and the NLO cross section, computed with NLOJET++. The central scale adopted is the individual jet p_t, and the uncertainty bands are given by the usual scale variation prescription. Two comments are in order. First, higher order corrections have the potential to change not only the size, but also the shape of the cross section, so are essential for QCD precision studies. Second, at LO the estimate of the effect of missing higher orders through scale variations is completely unreliable, as the huge uncertainty band in Fig. 2.3 testifies.

In fact, LO results represent only the first term of perturbative series in α_s:

$$d\hat{\sigma} = d\hat{\sigma}^{(0)} + d\hat{\sigma}^{(1)} + d\hat{\sigma}^{(2)} + \cdots = \sum_{k=0}^{\infty} d\hat{\sigma}^{(k)} \tag{2.36}$$

where each term is of order α_s^{2+k}, since the LO process is of order α_s^2. In the previous section
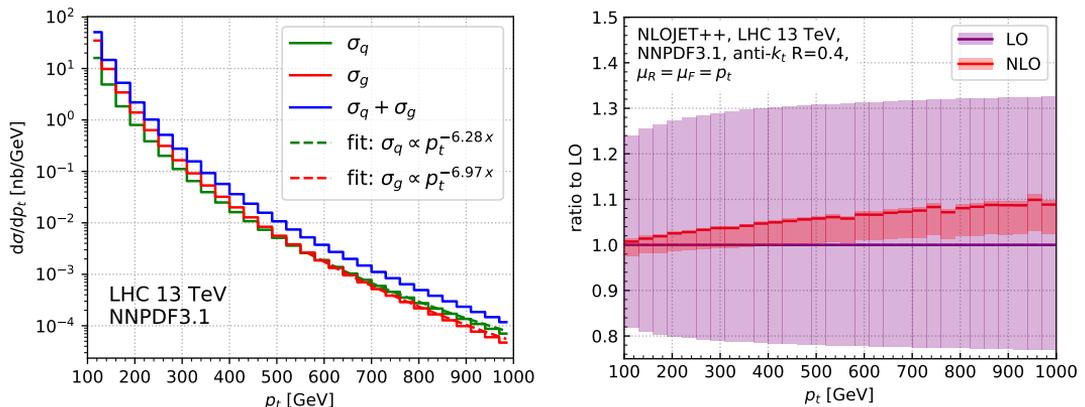
Figure 2.3: (Left) Upper panel: partonic spectra at leading-order and power law fits to the tail of the distributions. Lower panel: comparison between the result of the code implementing Eq. (2.30), a "non-averaged" version of the latter (see text) and the output of NLOJET++ at LO. (Right) Differential cross section at LO and NLO. Uncertainty bands are obtained with the 7-point scale variation rule, see Fig. 1.4.

we have explicitly calculated $d\hat{\sigma}^{(0)}$. In this section and in the forthcoming ones, we briefly discuss how leading order results have been improved in the past thirty years, by first discussing the NLO correction $d\hat{\sigma}^{(1)}$ and, in the next section, the NNLO correction $d\hat{\sigma}^{(2)}$. A systematic description of the details of the different calculations goes beyond the scope of this thesis. We limit ourselves to provide references for the interested reader.

Computation of jet processes at next-to-leading order have been made possible by the seminal work of Ellis and Sexton [88], where they provided the matrix element squared at order α_s^3, for all $2 \to 2$ (at one-loop virtual) and $2 \to 3$ (at tree-level) parton scattering subprocesses. If we have these matrix elements at our disposal, we could imagine to perform a phase space integration, similarly to what we did in Sec. 2.1 (modulo a more complex final space kinematics), and obtain for NLO contribution to the channel ij of the partonic cross section

$$d\hat{\sigma}_{ij}^{(1)} = \int d\Phi_3 \, d\hat{\sigma}_{ij}^{(1),\mathrm{R}} + \int d\Phi_2 \, d\hat{\sigma}_{ij}^{(1),\mathrm{V}} \,, \qquad (2.37)$$

where $d\hat{\sigma}_{ij}^{(1),\mathrm{R}}$ and $d\hat{\sigma}_{ij}^{(1),\mathrm{V}}$ are the real and the virtual corrections of order α_s^3. We can depict the structure of higher order corrections in a diagrams such as the one in Fig. (2.4). Both loops in virtual diagrams or addition of legs in real diagrams increase the power of the strong coupling constant. NLO corrections correspond to the first line of the inverted pyramid.

Unfortunately, we already know that, even if the sum in Eq. (2.37) if finite, each term is separately divergent. Dimensional regularization offers us one way to tackle this issue, by converting the singularities to $1/\epsilon$ poles, which then cancel between the real and virtual terms. However, the explicit $1/\epsilon$ pole in the real part does not appear until we have performed the integration, and there are no guarantees that we can actually do the integration analytically, given also the complex integration boundaries dictated by jet algorithms.

Different strategies have been devised to overcome this problem, and a simple example (taken
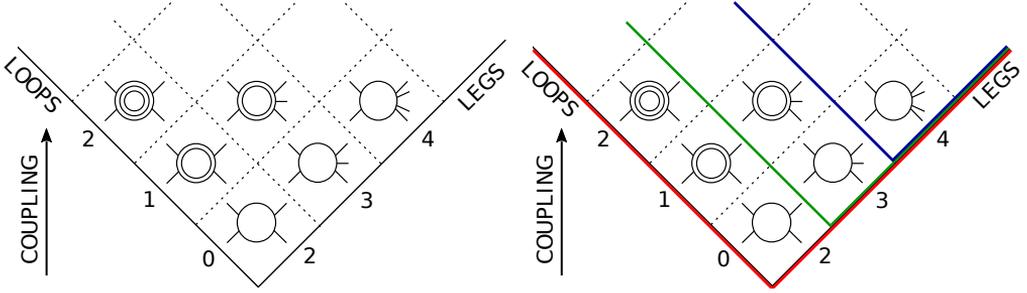
Figure 2.4: Schematic representation of higher order corrections to partonic scattering processes. The contribution in each cell is of $\mathcal{O}(\alpha_s^k)$ with $k = k_0 + \text{loops} + (\text{legs} - l_0)$ where k_0 is the power of the Born process and l_0 is the number of legs of the Born process. Higher order corrections for one process act as tree-level contribution or lower order corrections for another. For instance, on the Figure on the right, classes of matrix elements squared contributing to $2 \to 2$ (in red), $2 \to 3$ (in green) and $2 \to 4$ (in blue) scattering processes are explicitly highlighted.

from Ref. [89]) may help to understand how they work. Based on what we said above, the structure of a NLO calculation is similar to the following integral:

$$I = \lim_{\epsilon \to 0} \left\{ \int_0^1 \frac{dx}{x} x^\epsilon F(x) - \frac{1}{\epsilon} F(0) \right\} , \tag{2.38}$$

where F is a complicated function, finite in the $x \to 0$ limit, which prevents us from actually performing the integration. The first term in braces denotes the real correction to the matrix element, with x the analogous of the angle between two partons or the energy of a gluon, and therefore it would be divergent if there was no the x^ϵ term to regularize the integration. Similarly, the second term in braces stands for the virtual correction (with Born kinematics). In the limit $\epsilon \to 0$, both terms are divergent, but as their divergent behaviour is the same, the final result for I is finite.

One first strategy may be to split the integration region in two pieces, $0 < x < \delta$ and $\delta < x < 1$, with δ an arbitrary parameter. If $\delta \ll 1$, under the first integration we can approximate $F(x)$ with $F(0)$, and we can then write:

$$
\begin{aligned}
I &= \lim_{\epsilon \to 0} \left\{ \int_0^\delta \frac{dx}{x} x^\epsilon F(0) + \int_\delta^1 \frac{dx}{x} x^\epsilon F(x) - \frac{1}{\epsilon} F(0) \right\} \\
&= \lim_{\epsilon \to 0} \left\{ \frac{1}{\epsilon} F(0) + \log(\delta) F(0) + \int_\delta^1 \frac{dx}{x} x^\epsilon F(x) - \frac{1}{\epsilon} F(0) \right\} \\
&= \log(\delta) F(0) + \int_\delta^1 \frac{dx}{x} F(x) . \tag{2.39}
\end{aligned}
$$

The leftover integration can now be evaluated numerically. Of course, as long as δ is small, the final result will be independent of δ. Such phase space *slicing* method has been adopted in Ref. [90], where they first calculated analytical results at order α_s^3 for the inclusive production of a jet at large transverse momenta, in the limit of small cone opening angle. In our example, the term proportional to $\log(\delta)$ represents exactly what the Authors of Ref. [90] has found. These

results have been later numerically extended to fixed cone size [91]. The slicing method has also been employed in Ref. [92], to build a Monte Carlo program for one, two and three jet production, later used for the calculation of a dijet jet triply differential cross section [93].

A second alternative method may be the following. Instead of canceling the divergence after integration, we can alternatively add and subtract a term which mimics the divergent limit under integration, in such a way the integration can be computed numerically. Explicitly, in our example, we add and subtract $F(0)$, obtaining:

$$
\begin{aligned}
I &= \lim_{\epsilon \to 0} \left\{ \int_0^1 \frac{dx}{x} x^\epsilon (F(x) - F(0)) - \int_0^1 \frac{dx}{x} x^\epsilon F(0) - \frac{1}{\epsilon} F(0) \right\} \\
&= \int_0^1 \frac{dx}{x} (F(x) - F(0)),
\end{aligned}
\tag{2.40}
$$

and the leftover integration can be performed numerically as before. Note how the subtraction is carried out *locally* i.e. before integration, and the method has the advantage of not depending on an arbitrary sciling parameter δ. On the other hand, it relies on the fact that the subtraction term is simple enough to be integrated analytically. This *subtraction* method has been first introduced in Ref. [94], in the context of three-jet observable in e^+e^- annihilation, and later employed in Refs. [95, 96], for the calculation of the one-jet inclusive cross section at NLO. An analogous calculations have been carried out for dijets processes [97].

The early works cited above, even if pioneering in the field, were almost always considering predictions for specific differential distributions, lacking of a full generic differential treatment. In addition, calculations were based on homemade jet definitions, still not standardized, leading to difficulties when comparing predictions with the experimental data of different collaborations.

During the 1990s, different subtraction methods have been devised to tackle the calculation of NLO QCD corrections in a fully differential way and for any generic hadronic process. Two of such methods are the FKS formalism (so named because of the initials of the Authors of Ref. [98], in which the method was first introduced, in the context of the calculation of three-jet cross sections at NLO) and the Catani-Seymour dipole formalism [33].

Nowadays, one of the most used program for the calculation of order α_s^3 jet cross sections is NLOJET++, developed by Z. Nagy [99, 100], based on a slightly modified version of the Catani-Seymour subtraction method. It allows one to calculate one-, two- and three-jet cross sections at NLO. We will use this code extensively in the next chapters.

2.3   Next-to-leading order (NNLO) corrections

The calculation of the matrix elements at order α_s^4 has opened the road towards next-to-next-to-leading computations. Three pieces are needed: two-loop virtual corrections to $2 \to 2$ scattering [101–103], one-loop virtual correction to $2 \to 3$ scattering [104], double-real $2 \to 4$ scattering at tree-level [105]. The other necessary ingredient is a subtraction method at NNLO. At variance with NLO subtraction methods, we have now to deal with a more complex infrared structure, as two partons can become soft and/or collinear at the same time. We refer the reader to Ref. [106] for a detailed list of NNLO infrared subtraction methods.

One of such procedures is the antenna subtraction method, first introduced in Ref. [107] in the context of e^+e^- annihilation, and it has been applied to several observables, such as three-jet cross sections [108, 109] or event shapes [110] at NNLO. This method has later been generalized

to hadronic initial-states [111, 112].

In short, the antenna subtraction method works by writing down the analogous of Eq. (2.37):

$$d\hat{\sigma}_{ij}^{(2)} = \int d\Phi_4 \, d\hat{\sigma}_{ij}^{(2),\text{RR}} + \int d\Phi_3 \, d\hat{\sigma}_{ij}^{(2),\text{RV}} \int d\Phi_2 \, d\hat{\sigma}_{ij}^{(2),\text{VV}} \qquad (2.41)$$

and by introducing three subtraction terms:

$$\begin{aligned} d\hat{\sigma}_{ij}^{(2)} = &\int d\Phi_4 \left[d\hat{\sigma}_{ij}^{(2),\text{RR}} - d\hat{\sigma}_{ij}^{(2),\text{S}} \right] \\ &+ \int d\Phi_3 \left[d\hat{\sigma}_{ij}^{(2),\text{RV}} - d\hat{\sigma}_{ij}^{(2),\text{T}} \right] \\ &+ \int d\Phi_2 \left[d\hat{\sigma}_{ij}^{(2),\text{VV}} - d\hat{\sigma}_{ij}^{(2),\text{U}} \right] . \end{aligned} \qquad (2.42)$$

Each of these subtraction terms is intended to reproduce the singular behaviour in the regions of single or double unresolved phase space or to remove the explicit infrared poles in $1/\epsilon$, by leaving the differences under square brackets well behaved in the infrared regions, in order to yield a finite final result. Note that the functional structure of the subtraction terms is much more complex than at NLO: for instance, $d\hat{\sigma}_{ij}^{(2),\text{T}}$ contains the integration of $d\hat{\sigma}_{ij}^{(2),\text{S}}$ over a single unresolved phase space, plus an additional term for removing the remaining singularities of $d\hat{\sigma}_{ij}^{(2),\text{RV}}$; in turn, $d\hat{\sigma}_{ij}^{(2),\text{U}}$ contains the integration of $d\hat{\sigma}_{ij}^{(2),\text{S}}$ over a double unresolved phase space, plus the integration of the additional term in $d\hat{\sigma}_{ij}^{(2),\text{T}}$ over a single unresolved phase space. Thus, the subtraction terms are deeply linked to each other.

In 2017, within the antenna method, the first computation of NNLO QCD corrections to the single-jet inclusive cross section, including contributions from all partonic subprocesses, was finally reported [113]. Previous calculations with only the purely gluonic channel [114] appeared a few years before. NNLO predictions for dijet processes have been published as well [115].

In the context of these studies, the Authors of Ref. [113] and collaborators have developed a partonic Monte Carlo generators at NNLO accuracy, NNLOJET. The code is still private at the moment. The PDF fits to jet measurements we will present in Chapter 4 benefit from NNLO predictions obtained with this code, provided by the Authors of NNLOJET through K-factors.

Note that the calculation of Ref. [113] is based on on the leading-colour approximation in the case of channels involving quarks. The Authors of Ref. [116] have reported an independent calculation with exact colour at $\mathcal{O}(\alpha_s^4)$, using a different subtraction method (the sector-improved residue subtraction [117, 118]). They found negligible sub-leading colour effects in phenomenological applications.

2.4 Potential issues with the single-jet inclusive cross-section

The computation of NNLO QCD corrections to jet cross sections has opened up the possibility of doing precision phenomenology with jet observables. However, it has also settled some potential issues related to the behaviour of the perturbative series, whose full understanding has posed a theoretical challenge.

This is especially true for the single-jet inclusive cross section (we will comment later on the dijets cross section). As an observable, it is defined in a deceptively simple way: count all jets

which fall in any given kinematic bin and add them up. While this definition is remarkably simple, a minutes' reflection shows that it has a somewhat peculiar and perhaps undesirable feature. Namely, it is not unitary: each event is counted more than once, so the integral of the differential cross-section is not equal to the total cross-section. This lack of unitarity may be cause of concern: one is used to the fact that the unitarity of the total partonic cross-section is crucial in order to ensure its infrared finiteness, given that infrared singularities cancel between terms with different numbers of final–state partons. On the other hand, infrared finiteness of the N-jet cross-section is ensured by the use of a jet definition, so the question is really whether this definition leads to a good perturbative behaviour. In addition, since the contributions to the distribution come from individual jets rather than events, there could be a possible ambiguity in the choice of a proper scale. What is the degree of "hardness" of the process? The hardness of the individual jet or the hardness of the underlying Born $2 \to 2$ scattering process?
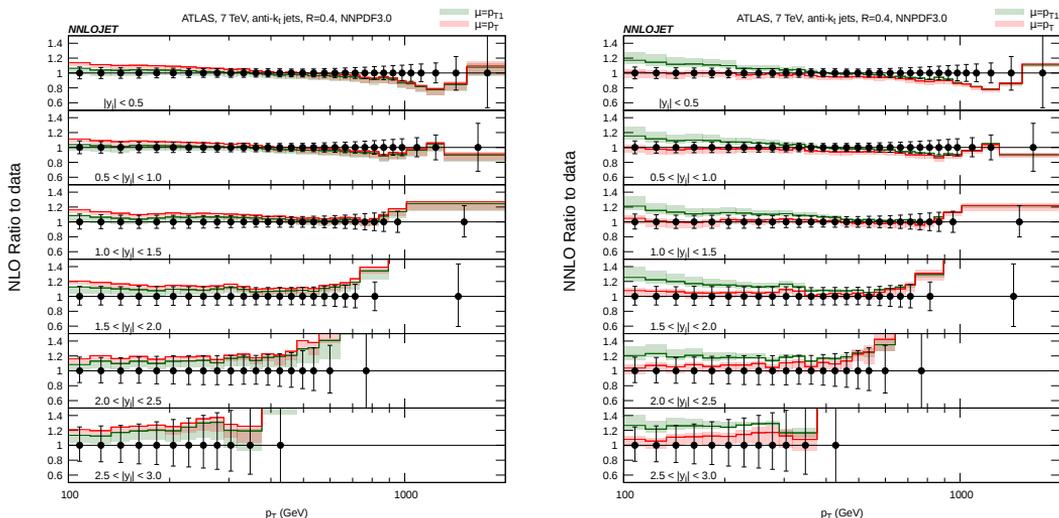

Figure 2.5: Double differential NLO and NNLO predictions for single jet inclusive production, normalised to ATLAS data [119], with reference scale choice $\mu = p_{t1}$ or $\mu = p_t$. (Taken from Ref. [120])

In Fig. 2.5, we show the NLO and NNLO predictions for single jet inclusive production, with two different commonly used scale choices: either $\mu = p_{t1}$ i.e. the transverse momentum of the hardest jet in the event, or $\mu = p_t$: the transverse momentum of each jet entering the distribution. The former is an event-based scale, in that each jet in the event is binned with the same weight; the latter is a jet-based scale, in that each jet in the event is binned with its own weight. The predictions are compared to ATLAS data [119], used as reference points. This figure is taken from Ref. [120]. While at NLO data tends to prefer the choice $\mu = p_{t1}$, and the prediction with $\mu = p_t$ is systematically above the one with $\mu = p_{t1}$ for all the rapidity slices considered, at NNLO the behaviour is exactly the opposite. The difference between the two scale choices is particularly significant in the region of low transverse momentum, where is more likely to find contributions from subleading jets, for which p_t differs significantly from p_{t1}. Moreover, the scale dependence of the result is not significantly reduced and the size of the K-factor does not significantly decrease when going from NLO to NNLO. These facts suggest a

possible perturbative instability.

In Ref. [121] the perturbative properties of this observable were extensively studied, in particular by a numerical analysis of the contributions to individual jet bins with a variety of computational setups (such as the choice of scale and of jet radius). The behaviour observed in Fig. 2.5 was traced to the infrared sensitivity of the second-jet contribution, which is aggravated by the choice $\mu = p_{t1}$. While at LO the contributions to the cross section coming from the two leading jets is the same (as p_{t1} and p_{t2} are equal, see Sec. 2.1.2), at higher orders they differ significantly (we will return in depth to this point in Chapter 3). Notably, the subleading-jet contribution is affected by drastic shifts from LO to NLO, and even from NLO to NNLO. This explains the contrasting behaviour seen between the left and the right plot in Fig. 2.5.

In this same paper, a detailed study of the scale dependence of the NNLO QCD predictions for inclusive jet production was carried out. Three different scales (and their multiples) were discussed in detail: the two scales considered above (the individual jet transverse momentum p_t and the leading jet transverse momentum p_{t1}) and the scalar sum of the transverse momenta of all partons in the event,

$$\hat{H}_T = \sum_{i \in \text{partons}} p_{ti} \,. \tag{2.43}$$

These scale choices were thus compared according to a number of criteria: perturbative convergence; scale uncertainty as error estimate; perturbative convergence of the individual jet spectra; and stability of the second jet distribution. Based on these, the event-based scale $\mu = \hat{H}_T$ and the jet-based scale $\mu = 2p_T$ were singled out as optimal choices (see Fig. 2.6).
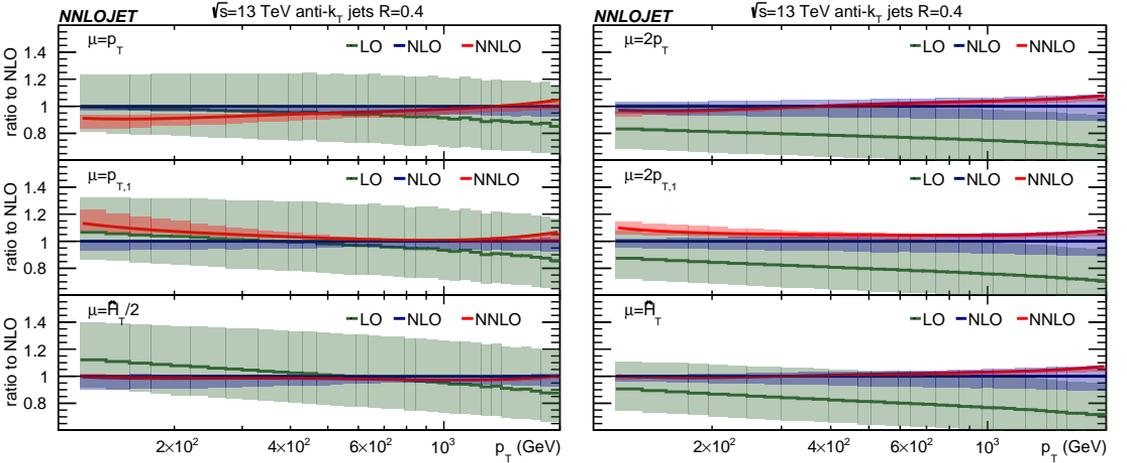
Figure 2.6: Jet p_t spectrum at LO, NLO and NNLO, normalised to the NLO prediction, for six different central scale choices. (Taken from ref. [121])

For sake of completeness, we conclude this section with a comment on the dijet cross section. Even for this observable several scale choices are possible: two popular ones are the dijet invariant mass m_{jj} and the average transverse momentum of the two leading jets, $\langle p_t \rangle = (p_{t1} + p_{t2})/2$. Theoretical predictions computed with either of these scale choices differ significantly at NLO. This difference is substantially reduced at NNLO, with $\mu = m_{jj}$ emerging as a preferred choice, based on the criteria of perturbative convergence, and residual scale dependence of the NNLO prediction [115, 122].

On the definition of single-jet inclusive cross section

Motivated by the issues about the single-jet inclusive cross section highlighted in Sec. 2.4, in this chapter we approach the problem of understanding the behaviour of this observable from a somewhat different point of view: namely, by trying to see how it behaves upon changes of its definition, specifically motivated by an attempt to correct for its non-unitarity. We study the properties of this family of new, unitary definitions both numerically, and analytically in a simple collinear approximation.

Our analysis focuses on the general properties of the observable, of which we strive to understand the main qualitative features. We thus base our discussion on NLO calculations, whose structure is easier to handle both from a numerical and an analytic point of view, though we aim at understanding their general properties at any perturbative order. We have explicitly checked their robustness in several cases at NNLO, which we have been able to obtain from a NLO code by calculating differences in which missing double-virtual contributions cancel (see Sec. 3.3).

The outline of the chapter is the following. First, in Sec. 3.1 we discuss the standard definition of the cross section and its non-unitarity, and present a family of alternative, unitary definitions. Then, in Sec. 3.2 we compare results obtained using various definitions at NLO. In Sec. 3.3 we test the validity of our results at higher orders, by carrying out the test mentioned above. Afterwards, we show how the numeric results can be understood in terms of an analytical calculation. In Sec. 3.4 we provide general (if somewhat formal) arguments, whereas in Sec. 3.5 we perform more explicit calculations, using a soft-collinear approximation to derive, and understand, qualitative features of the NLO results. Finally, we provide a short summary of our results in Sec. 3.6.

3.1 Unitary weighted definitions

The single-jet inclusive cross section is defined in terms of the differential cross section for producing N jets (after cuts) with transverse momenta p_{ti}, as

$$\frac{d\sigma}{dp_t} = \sum_N \frac{d\sigma_{N\,\text{jets}}}{dp_t} \tag{3.1}$$

$$\frac{d\sigma_{N\,\text{jets}}}{dp_t} = \int dp_{t1} \ldots dp_{ti} \ldots dp_{tN} \, \frac{d\sigma_{N\,\text{jets}}}{dp_{t1} \ldots dp_{ti} \ldots dp_{tN}} F_N\left[p_{t1}, \ldots, p_{tN}; p_t\right], \tag{3.2}$$

where F_N, for a standard definition, is given by

$$F_N^{\text{std}}\left[p_{t1}, \ldots, p_{tN}; p_t\right] = \sum_{i=1}^{N} \delta(p_{ti} - p_t), \tag{3.3}$$

and it fills the bin with transverse momentum p_t by picking all contributions from the fully differential N-jets cross section. The sum in Eq. (3.1) runs over the number of jets in each event that pass some kinematic cut. The sum over the total number of jets starts with $N = 1$ (the $N = 0$ case gives of course no contribution) and goes up to two at leading order (LO), three at NLO, and generally $p + 2$ at N^{p}LO.

It is clear that the inclusive-jet cross section defined in this way is not unitary, in that its integral over p_t does not give the total number of scattering events per unit flux per unit time within a given fiducial region. Indeed, with this definition, when filling a histogram in p_t, an event with N jets is binned N times.

We generalize the definition of the single-jet inclusive cross section by introducing jet weights that render the cross section unitary. Namely, we modify the definition Eq. (3.2) by introducing weights in the definition of the function F_N, Eq. (3.3):

$$F_N\left[p_{t1}, \ldots, p_{tN}; p_t\right] = \sum_{i=1}^{N} \delta(p_{ti} - p_t) \, w^{(N)}(p_t; p_{t1}, \ldots, p_{tN}) \tag{3.4}$$

The choice $w^{(N)} = 1$ represents the standard non unitary definition Eq. (3.3). The choice $w^{(N)} = 1/N$ restores unitarity, but has undesirable discontinuities whenever the kinematics of the final state changes in such a way that the number of jets jumps from N to $N + 1$. In this chapter, we consider a set of weights defined as

$$w^{(N)}(p_t; p_{t1}, \ldots, p_{tN}) = \begin{cases} 1 & \text{(standard)} \\ \dfrac{p_t^r}{\sum_{j=1}^{N} p_{tj}^r} & \text{(weighted)} \end{cases} \tag{3.5}$$

where p_{tj} is the transverse momentum of the j-th jet. All weighted choices lead to a unitary definition. We consider specifically three families of definitions of these weights, according to which jets are included when constructing the weights.

- *A: jets above p_t^{cut}*
  Only jets with $p_t \geq p_t^{\text{cut}}$ are included in the definitions of F_N Eq. (3.4). In particular, this implies that the sum in the denominator of Eq. (3.5) includes only jets for which

$p_{tj} \geq p_t^{\mathrm{cut}}$. When $r = 0$ this reduces to the simplest unitary choice with all weights equal to $1/N$.

- *B: all jets*
  F_N includes all the jets but the numerator in the weight definition, Eq. (3.5), only includes jets above p_t^{cut}. In particular, the denominator in Eq. (3.5) sums over all jets. This definition is infrared safe only for $r > 0$. While this definition may seem unphysical, in practice it corresponds to having a p_t^{cut} that is small compared to the p_t value of the first bin one is interested in.

- *C: two leading jets*
  Only the first two leading jets in p_t are included in the definition of both F_N and the weights, so $N = 2$ in both Eqs. (3.4) and (3.5). In this case we consider the two leading jets independently on whether their p_t is larger or smaller than a possible p_t^{cut}.

These definitions are "unitary" in the sense that the weights add up to one. This implies that, with the first definition, integrating over p_t gives the total cross section to have at least one jet above p_t^{cut}. For the second definition (with $p_t^{\mathrm{cut}} \to 0$ or an explicit underflow bin) and for the third definition, one instead gets the total pp cross section. To keep the discussion simple, we do not impose any rapidity cut in the studies carried on in this thesis. Nevertheless each of the previous definitions could be extended to the case in which a rapidity cut is introduced. Note that in the case of the third definition, a rapidity cut could change what the leading jets are. To avoid potential issues, in particular for $r < 0$ which is more sensitive to small p_t, one might have in practice to impose an additional dijet selection cut (similar to what is already done when studying e.g. the dijet invariant mass).

To highlight the various features we are interested in studying, it is useful to consider different ways of organizing the perturbative calculation of the single-jet inclusive cross section at N^{p}LO accuracy. This can, in fact, be written as a sum of contributions, each of order α_s^{2+k}, $k = 0, \ldots, p$, assuming that the leading-order (LO) process is of order α_s^2:

$$\frac{d\sigma^{\mathrm{N}^p\mathrm{LO}}}{dp_t} = \sum_{k=0}^{p} \frac{d\sigma^{(k)}}{dp_t}. \tag{3.6}$$

It is useful to think about the order α_s^{k+2} contribution in two different ways. The first is as a sum of contributions with a different number of jets, as we have done in Eq. (3.2). In such a case, the k-th order contribution to the cross section is built out of terms containing at most $k + 2$ jets i.e. two at LO ($k = 0$), three at NLO ($k = 1$) and so forth:

$$\frac{d\sigma^{(k)}}{dp_t} = \sum_{N=1}^{k+2} \frac{d\sigma_{N\,\mathrm{jets}}^{(k)}}{dp_t}. \tag{3.7}$$

Eq. (3.7) is the same as Eq. (3.1), but for the k-th order contribution only. However, in order to understand the perturbative behaviour of the cross section it also useful to break it up into the contribution from the jet with the largest p_t (leading, or first jet), the jet with the second

largest p_t (subleading, or second jet), and so on:

$$\frac{d\sigma^{(k)}}{dp_t} = \sum_{n=1}^{k+2} \frac{d\sigma^{(k)}_{n\text{-th jet}}}{dp_t}. \tag{3.8}$$

In Eq. (3.7), $d\sigma^{(k)}_{N\,\text{jets}}/dp_t$ is the contribution to the cross section coming from configurations with N jets, while in Eq. (3.8) $d\sigma^{(k)}_{n\text{-th jet}}/dp_t$ is the contribution coming from the n-th leading jet. The range of the sum is the same in both cases and it is equal to the maximum number of jets that can be produced at a given perturbative order k.

3.2 Comparing definitions of the cross section

In order to study the effects of the various unitary definitions, Eq. (3.5), we start by simply comparing results obtained in each case. We focus on two observables: the total NLO K factor, and the individual n-th-leading jet NLO K factor as a function of p_t,

$$K = \sum_{n=1}^{3} K_n, \quad \text{with} \quad K_n = \frac{d\sigma^{\text{NLO}}_{n\text{-th jet}}}{d\sigma^{\text{LO}}}. \tag{3.9}$$

This way, we can see how imposing unitarity affects the p_t distribution of the single-jet inclusive cross section. In Section 3.4 we then turn to analytic arguments, both in general and in a collinear approximation.

All results presented in this section are obtained using the following setup. Computations up to NLO are performed using NLOJET++ (v4.1.3) for pp collisions, with center of mass energy $\sqrt{s} = 13$ TeV. Parton distribution functions are taken from the NNPDF3.1 [58] set at NNLO, with $\alpha_s(M_Z) = 0.118.$, and interfaced using the LHAPDF library (v6.1.6) [123]. Jets are clustered using the anti-k_t algorithm [68], as implemented in FASTJET (v3.3.2) [70], with $R = 0.4$, unless otherwise specified.

The dependence on the choice of central factorization and renormalization scale is studied by considering three options: (i) the *average dijet scale*,

$$p_t^{(\text{avg})} = \frac{p_{t1}^{(R=1)} + p_{t2}^{(R=1)}}{2}, \tag{3.10}$$

where $p_{t1,2}^{(R=1)}$ are the transverse momenta of the two leading jets clustered with a radius $R = 1$, (ii) the *partonic scalar k_t halved*,

$$\frac{\hat{H}_T}{2} = \frac{1}{2} \sum_{i=1}^{n\text{-partons}} k_{ti}, \tag{3.11}$$

suggested as an optimal[1] scale choice in Ref. [121], and (iii) the *leading jet p_t*, $p_t^{(\text{max})}$, defined as

$$p_t^{(\text{max})} = p_{t1}^{(R=1)}. \tag{3.12}$$

---

[1] Actually, as we report in Sec. 2.4, the scale suggested in Ref. [121] is $\hat{H}_T$. The reason why here we are considering $\hat{H}_T/2$ is to guarantee the same LO result with our three scale options.
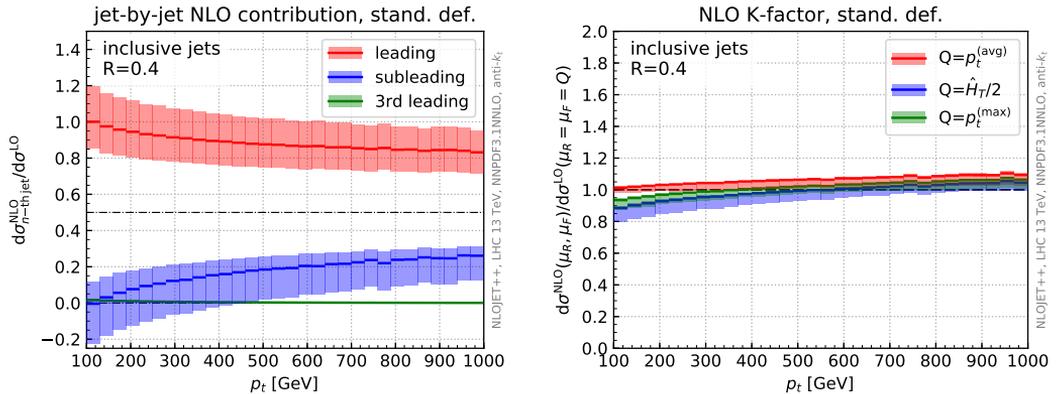
Figure 3.1: Left: Contributions from the leading, subleading, and third-leading jet to the NLO inclusive cross section, with central scale choice $\mu_R = \mu_F = p_t^{(\text{avg})}$ Eq. (3.10). Right: Inclusive NLO K factors, with three different central scale choices (see text).

Note the unusual choice to evaluate the p_{ti} values entering the expression of the scales in Eqs. (3.10)-(3.12) using jets with radius $R = 1$. This was suggested in ref. [124], in order to disentangle the scale of the hard scattering from the jet reconstruction process, and then obtain a choice independent of R.

3.2.1 Standard (non-unitary) definition

We start by discussing results for the standard definition, shown in Fig. 3.1. Three main features are apparent.

1. While the total NLO K factor is quite close to one (see the right plot in Fig. 3.1), the individual K_n for the leading and subleading jet deviate from their leading order value, $1/2$, by sizable amounts (see the left plot in Fig. 3.1). However, they almost exactly compensate when added up into the total cross section, yielding a total NLO K factor close to 1, as well as a scale uncertainty much smaller than those of the individual K_n. This almost exact compensation is largely accidental as it depends on the value of the jet radius. This can be seen in Fig. 3.2, where we plot the K factor for the total cross section as a function of R: the leading and the second leading jet K factors only compensate (up to a residual $\sim 10\%$ effect) in the region $R \sim 0.3$–0.6. This effect has also been noticed in Refs. [124, 125].

    The behaviour of the individual jet K factors can be explained in a simple fashion. At NLO, the K factor of the leading jet K_1 is substantially larger than one, most likely a consequence of recoil effects from initial state radiation that lift the born level p_t degeneracy. This has the effect of increasing the p_t of a jet that then becomes the leading one, and it is amplified by the fact that the LO cross section is steeply falling in p_t, as already noted in Sec. 2.1.5. Furthermore, at NLO, K_1 does not depend on R, as explicitly visible in Fig. 3.2 and as we show analytically in Sec. 3.4 below. However K_2 decreases at small R since out-of-cone final state radiation depends on the jet radius and has the effect of lowering the p_t of the
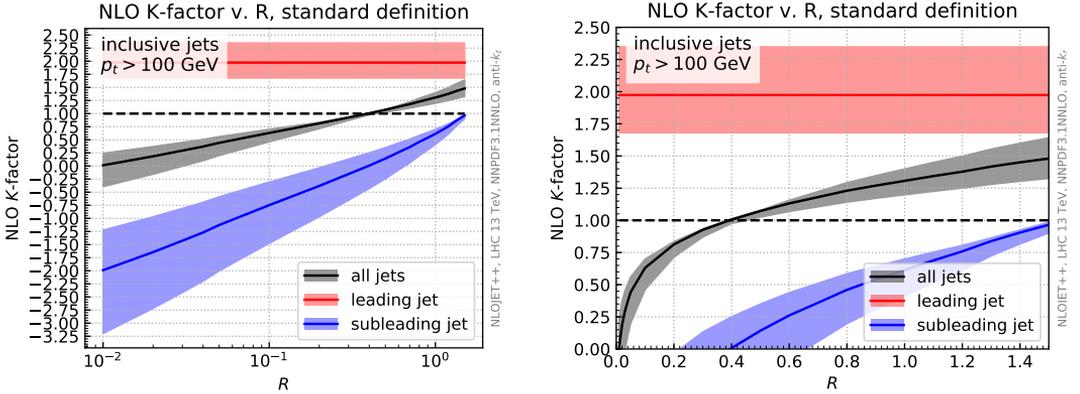
Figure 3.2: The NLO K factor for the single-inclusive total jet cross section as a function of the radius R of the jet (black). The contributions from the leading jet (red) and the subleading jet (blue) are also shown. Results are plotted both with a logarithmic (left) and linear (right) scale.

emitter. This effect is again drastically enhanced by the steeply-falling nature of the LO differential cross section in p_t.

It can be seen from the logarithmic scale that the dependence of the cross section on $\ln R$ becomes linear only for $R \lesssim 0.2$: hence, the logarithmic contribution dominates the cross section only in the very small R region, and indeed resummation was shown to be necessary in this region in Refs. [124, 126, 127]. For larger R the $\ln R$ term is still sizable, but the bulk of the $\ln R$ effects is captured by the exact NLO result, and for $R \gtrsim 0.4$ there is a modest benefit in resumming them, as also shown in previous cited References, where this resummation was performed explicitly.

2. While the leading and second jet account for most of the cross section, the contribution of the third jet to the total K factor is much smaller (giving a correction of less than 2% of the LO cross section) and almost completely negligible. The dominance of the first two jets as p_t grows is important in determining the qualitative features of the standard definition and also plays a role for the various other definitions that we consider below. It persists at NNLO, as shown in Ref. [121], and it is in fact to be expected to persist to all orders, as a consequence of the dominance of soft radiation which, combined with the transverse-momentum conservation, favors configurations in which two hard jets are back-to-back while all the others are softer. This is based on the observation that the multi-differential cross section, by dimensional analysis and transverse momentum conservation, must scale like

$$\frac{d\sigma}{dp_{t1}\dots dp_{tn}} \propto \frac{1}{p_{t1}\dots p_{tn}}\delta(p_{t1}+\dots+p_{tn})\,, \qquad (3.13)$$

which in turn implies that the cross section is the largest when there is a back-to-back configuration with all the other jets being very soft.

3. By inspecting the uncertainty bands shown in Fig. 3.1, one can see that scale variation bands for $R = 0.4$ for different central scale choices do not overlap in the small p_t region.

As we have mentioned in Sec. 2.4, an in-depth discussion of this problem and how this changes when including even higher order QCD corrections is given in Ref. [121]. It is however clear that this is a consequence of the accidental compensation of the two leading jets discussed above, which then propagates onto the scale variation. It follows that theoretical uncertainties obtained by performing standard scale variation for fixed $R \sim 0.4$ are unrealistically small. A more reliable estimate can be obtained performing uncorrelated scale variation (see Section 3.2.3 below).

All this shows that the putative perturbative instability of the standard definition is in fact a byproduct of an entirely accidental cancellation which happens only at NLO in a given R range. Because this cancellation is not protected by a symmetry, one should not expect it to persist with a different definition or at higher perturbative orders.

3.2.2 Weighted (unitary) definitions

We now turn to the study of the weighted (unitary) definitions of the single inclusive-jet cross section introduced in Sec. 3.1. We start our discussion with case (A), in which a p_t^{cut} is adopted, and we show that in fact this unitary definition appears to display a somewhat problematic behaviour, whose origin is discussed analytically in Sec. 3.4. We then turn to cases (B) and (C) which provide a natural way to alleviate this problematic behaviour.

A. Jets above p_t^{cut}. In Fig. 3.3 we show again the individual jet contributions and K factor, now using weighted definitions of type (A), with a positive ($r = 2$) and a negative ($r = -4$) value for the exponent in the weights. Note that the K_n, and hence the total K factor, are normalized to the LO weighted jet cross section which is exactly half of the LO jet cross section obtained with the standard definition. Indeed, at LO we have $w_1 = w_2 = 1/2$, by kinematic constraint, for the weighted definition, independently of r.

We first discuss the behaviour for p_t far above p_t^{cut}. Broadly speaking, positive weights enhance the difference between leading and second leading jets, with features that resemble those of the standard definition for the individual K_n factors. This is also true, in particular, for the total K factor for p_t sufficiently larger than p_t^{cut} (top row of Fig. 3.3). Negative values of r, on the other hand, have the effect of balancing the difference between leading and subleading jets. This results in more similar individual K_n factors, at the price of an overall larger total K factor (bottom row of Fig. 3.3). At very large p_t this effect becomes very large, which can be easily understood as follows: whenever we have three jets passing the p_t cut with $p_{t1,2} \gg p_{t3}$ we have

$$w_{1,2}^{(3)}(r < 0) = \frac{p_{t1,2}^r}{p_{t1}^r + p_{t2}^r + p_{t3}^r} \sim \left(\frac{p_{t3}}{p_{t1,2}} \right)^{|r|} \ll 1 \,, \tag{3.14}$$

$$w_3^{(3)}(r < 0) = \frac{p_{t3}^r}{p_{t1}^r + p_{t2}^r + p_{t3}^r} \sim 1 \,. \tag{3.15}$$

The contributions of the two leading jets to the inclusive cross section, which are strongly dominating the NLO cross section for the standard definition (or for the weighted definition with $r \geq 0$), are now power suppressed by the weights. Furthermore, corresponding virtual corrections have two jets in the final state with $w_{1,2}^{(2)}(r < 0) = 1/2$. At large p_t real and virtual
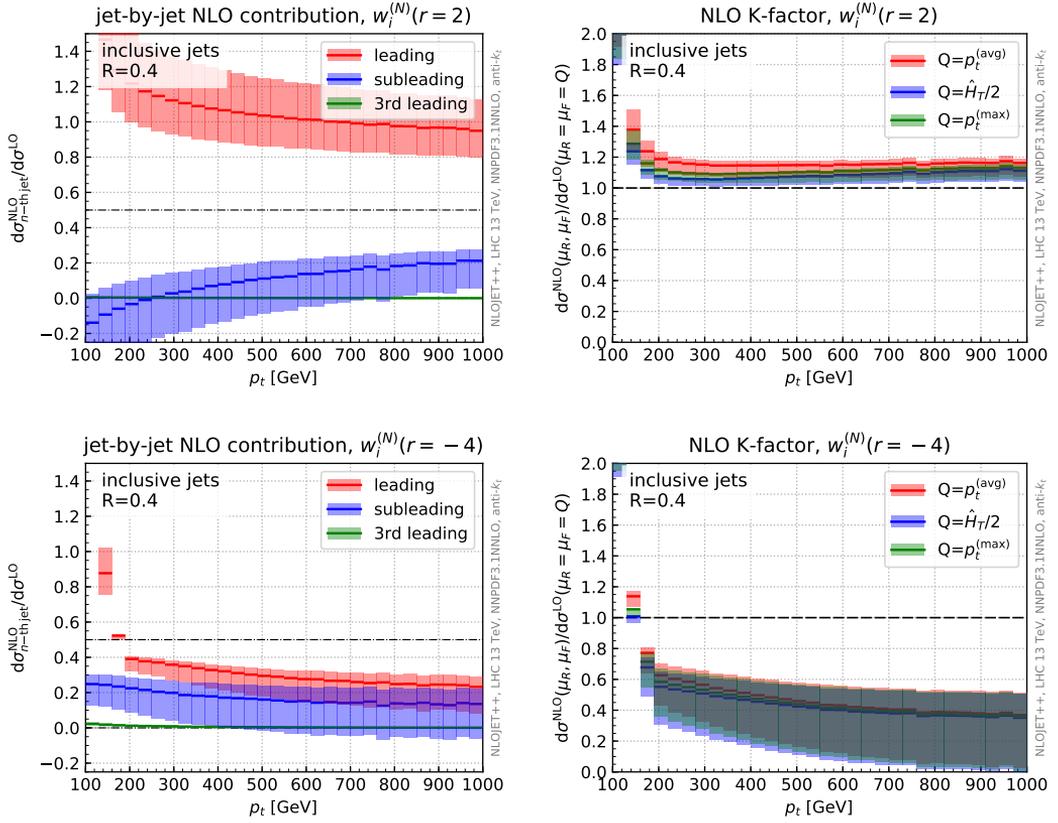
Figure 3.3: Same as Fig. 3.1 but using the weighted definitions of type A (see text: jets above p_t^{cut}) for $r = 2$ (top) and $r = -4$ (bottom).

corrections with $p_t^{\mathrm{cut}} \ll p_{t3} \ll p_{t1,2} \sim p_t$ therefore yield, after integration over p_{t3}, a negative contribution enhanced by $\log(p_t/p_t^{\mathrm{cut}})$, corresponding to the large corrections seen in Fig. 3.3.

Now turning to the region where $p_t \to p_t^{\mathrm{cut}}$, we see from Fig. 3.3 that this weighted definition (for both positive and negative r) develops a singular behaviour. The origin of this behaviour is explained analytically in Section 3.4. For the time being, we note that these singularities, both for $p_t \gg p_t^{\mathrm{cut}}$ and for $p_t \to p_t^{\mathrm{cut}}$, are of logarithmic origin and could in principle be dealt with resummation.

In summary, the weighted definitions of type (A) (with p_t^{cut}) have the undesirable feature of developing problematically unstable behaviours for p_t close to the p_t cut as well as at large p_t for $r < 0$. In the other p_t regions their perturbative behaviour now shows large K factors also at NLO since the accidental cancellation of the standard definition is spoiled; while this is perhaps more natural, it does not suggest an improvement in perturbative behaviour over the standard definition.

B. All jets. A natural way of curing the logarithmic divergence observed when $p_t \to p_t^{\mathrm{cut}}$ using weights of type (A) is to include all jets down to a p_t much smaller than the first bin of
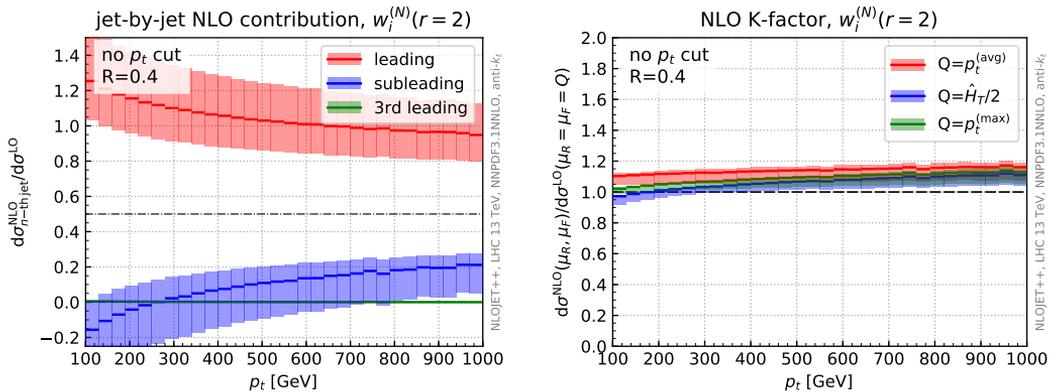
Figure 3.4: Same as Fig. 3.1, but using a weighted definition of type B (all jets) for $r = 2$.

the distribution. Based on Fig. 3.3, taking a p_t^{cut} two or three times smaller than the first bin of the distribution would already get rid of most of the sensitivity to p_t^{cut}, e.g. without any need for an additional resummation. One can view the weighted definition of type (B) as simply taking the limit $p_t^{\text{cut}} \to 0$ and one should not expect our conclusions to change as long as p_t^{cut} remains much smaller than the first bin of the distribution, say $p_t^{\text{cut}} \sim 20 - 30$ GeV. This possibility is only sensible for positive weights, for which the low p_t part of the spectrum is suppressed. For negative weights this choice is infrared unsafe.

Results are shown in Fig. 3.4 for $r = 2$. As expected, the singular behaviour of the K factor for p_t close to p_t^{cut} is now absent, and features similar to those of the standard definition are now recovered. Specifically, non-overlapping scale variation bands are observed in the low p_t region, though to a smaller extent than in the standard case. As a last comment, we have checked that this definition does not suffer from large non-perturbative corrections, such as those coming from underlying events, despite involving low-p_t jets. In a practical experimental context, one would still need to make sure that this remains true with realistic pileup conditions.

C. Two leading jets. An alternative choice, motivated by the observation that the contribution of the third jet to the inclusive jet cross section is much smaller than that of the first two jets (see Fig. 3.1) is to switch to definitions of type (C), in which only the two leading jets are included in the weights, whether or not they pass a given p_t^{cut}. Clearly this should also remove the problem of the behaviour for $p_t \sim p_t^{\text{cut}}$ of definitions of type (A). This approach is similar in spirit to what is done when looking at the dijet cross section. Results in this case are presented in Fig. 3.5 for the individual K factors K_n and the total K factor. The situation for positive r is again similar to what we observe for the standard definition: in particular there seems to be a large compensation between the leading and subleading jets, leading to a rather flat K factor, though larger than in the standard case.

As explained above, negative values of r have the effect of normalizing the individual K_n factors for the leading and subleading jets, reducing the effect of the compensation seen in the standard case. Furthermore, the uncertainty bands obtained for the three different scale choices now overlap. Nevertheless, the inclusive K factor is relatively larger than for the standard definition and shows a somewhat strong p_t dependence.
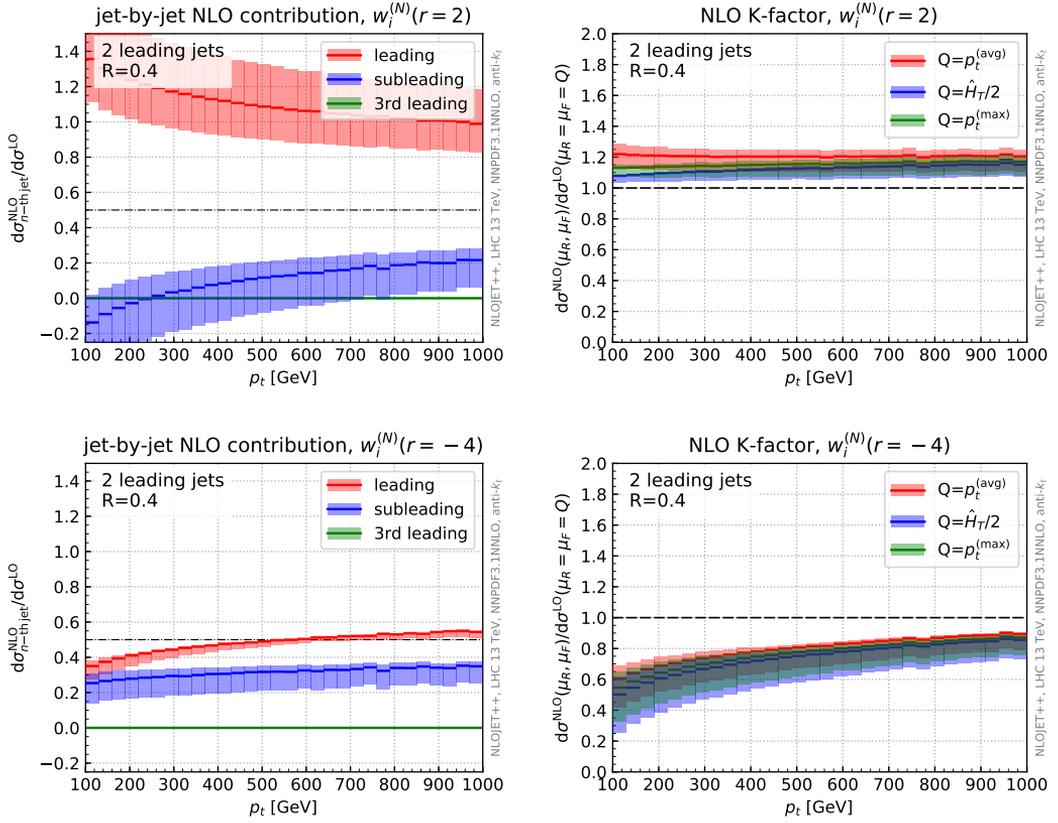
Figure 3.5: Same as Fig. 3.1, but using the weighted definitions of type C (two leading jets) for $r = 2$ (top) and $r = -4$ (bottom).

Comparing these results to the other weighted definitions, we see that the logarithmic divergence for p_t close to p_t^{cut} which is observed in Fig. 3.3 when using the jets above p_t^{cut} has now disappeared for both positive and negative r. This, as discussed above, is expected: the weights do not depend on whether one or two of the two leading jets passes the p_t^{cut}, so the definition becomes independent of the cut. Furthermore, the issue with large K factors at large p_t for negative r when including jets above p_t^{cut} has also disappeared. This is simply because the third jet no longer contributes to the weights and therefore the large contribution seen in Eq. (3.14) is absent.

In summary, weighted definitions of type (C) behave similarly to the standard definition for positive r. The perturbative behaviour for negative r changes, with some desirable features (the individual K factors K_1 and K_2 are similar, and the scale uncertainty bands for different scale choices overlap), and some undesirable ones (the overall K factor is larger).

3.2.3   Uncorrelated scale variations

An alternative way of estimating missing higher order uncertainties, through an uncorrelated scale variation procedure, has been suggested in Ref. [124], in the context of small-R resumma-

tion. This prescription arises from the observation that we can rearrange the NLO cross section at radius R as follows:

$$
\begin{aligned}
\sigma^{\text{NLO}}(R) &= \sigma^{(0)} + \sigma^{(1)}(R) \\
&= \sigma^{(0)} + \sigma^{(1)}(R_0) - \sigma^{(1)}(R_0) + \sigma^{(1)}(R) \\
&= \left[\sigma^{(0)} + \sigma^{(1)}(R_0)\right] \left[1 + \frac{\sigma^{(1)}(R) - \sigma^{(1)}(R_0)}{\sigma^{(0)}}\right] + \mathcal{O}(\alpha_s^4),
\end{aligned} \tag{3.16}
$$

where we have introduced $\sigma^{(1)}(R_0)$, the NLO correction evaluated with a (larger) value of jet radius $R_0 \sim 1$. Thus we see that

$$
\sigma^{\text{NLO,mult}}(R) = \sigma^{\text{NLO}}(R_0) \times \left(1 + \Delta^{(1)}(R, R_0)\right) \tag{3.17}
$$

differs from $\sigma^{\text{NLO}}(R)$ only by higher order terms, with

$$
\Delta^{(1)}(R, R_0) = \frac{\sigma^{(1)}(R) - \sigma^{(1)}(R_0)}{\sigma^{(0)}} . \tag{3.18}
$$

It is possible to see Eq. (3.17) as the fixed order version of a multiplicative matching between the exact NLO result and the leading-logarithmic resummation of the $\alpha_s^n \ln^n R^2$ terms, performed in Ref. [124]. However, even at fixed order Eq. (3.17) has a nice physical interpretation: the cross section for production of jets with radius R is obtained as a product between the cross section for the production of partons (which can be though of as jets with a larger radius $R_0 \sim 1$) and their fragmentation into jets, with $\Delta^{(1)}(R, R_0)$ accounting for the different radius value.

The uncertainty bands around the NLO prediction are thus unnaturally small because of an unphysical cancellations in scale dependence between these two processes. A more reliable estimate could be obtained starting with eq. (3.17), considering separately the uncertainties of the two factors — adopting the standard 7-point rule or another prescription — and eventually summing them in quadrature.

We can check the effects of decorrelated scale variation both on the standard and on the weighted definitions. This study is performed in Fig. 3.6. We now note overlapping scale uncertainties across the whole p_t spectrum, notably for the standard definition and for the weighted definition of type B. This is analogous to what happens in the context of jet vetoing, where decorrelated scale variation also leads to more realistic uncertainty estimates in the presence of cancellations [128].

As a final remark, note that it is possible to write down an equation similar to Eq. (3.17) at NNLO. By using the fixed order NNLO results, the option of uncorrelated scale variations has been recently studied in great details in [125]. In particular, the Authors of Ref. [125] propose alternative choices about the actual implementation of the factorization in eq. (3.17), such as

$$
\sigma^{\text{(N)NLO,mult}}(R) = \sigma^{\text{(N)NLO}}(R_0) \times \frac{\sigma^{\text{(N)NLO}}(R)}{\sigma^{\text{(N)NLO}}(R_0)} \tag{3.19}
$$

i.e. a non-expanded version of Eq. (3.17). This prescription has the potential advantage of conserving the central value of the original (N)NLO prediction.
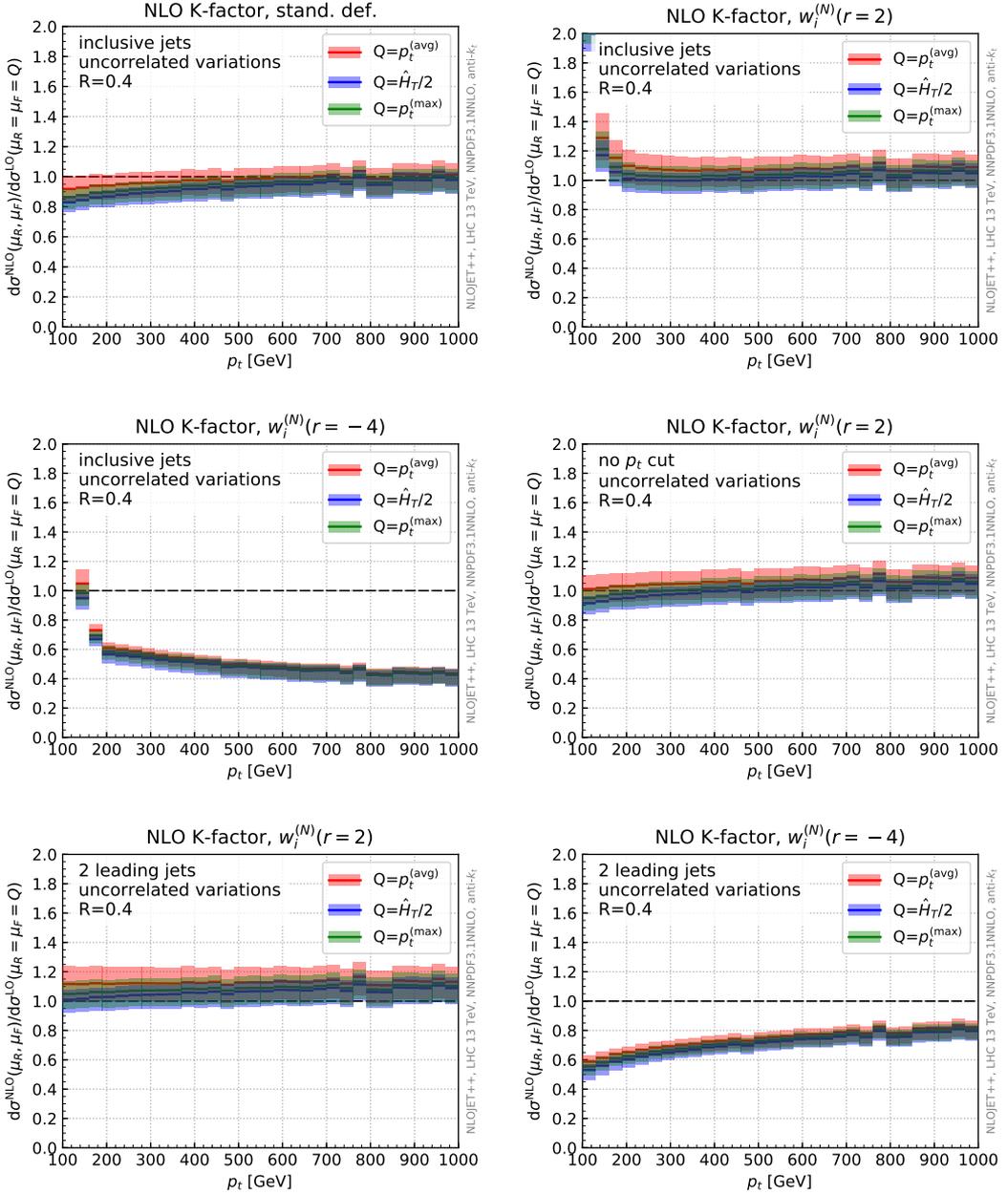
Figure 3.6: Same as Fig. 3.1, but with uncertainty bands determined by uncorrelated scale variation procedure.

3.3   Hints on NNLO results

While the discussion presented in the previous sections is mostly at NLO, in this section we would like to comment on the validity of our results at higher orders. This is to some extent

possible, as we are able to evaluate some specific difference between cross sections at NNLO by using public NLO codes. For instance, the following relation holds:

$$\frac{d\sigma_{\text{1-st jet}}^{\text{NNLO, 2j}}}{dp_t} - \frac{d\sigma_{\text{2-nd jet}}^{\text{NNLO, 2j}}}{dp_t} = \frac{d\sigma_{\text{1-st jet}}^{\text{NLO, 3j}}}{dp_t} - \frac{d\sigma_{\text{2-nd jet}}^{\text{NLO, 3j}}}{dp_t}\,. \tag{3.20}$$

This is due to the fact that virtual corrections to 2-jet production process cancel in the difference, just because $p_{t1} = p_{t2}$, yielding the same result as the difference between NLO corrections to the 3-jet production process. This cancellation is easy to understand given Fig. 2.4 on the right. If we take difference between two red triangles (corrections to $2 \to 2$ scattering), the contributions coming from the "loop row" on the left cancel each other, and we are left with exactly the difference between two green triangles (corrections to $2 \to 3$ scattering).

In practice, we use NLOJET++ to generate 3-jet events at NLO. For each event, we bin directly the difference between the weights associated to the leading and the subleading jet. Note that we are not imposing any cut on the presence of a third jet in the event: this is necessary to ensure the validity of Eq. (3.20).

In this way, we can explicitly check if some of our results persist through NNLO. In Fig. 3.7 we compare such a difference at NLO and at NNLO. For instance, the red curves in Fig. 3.7 as simply obtained as the difference between the red and the blue curve in the left plot of Figs. 3.1, 3.3, 3.5 respectively. We see that at NNLO there is still a cancellation between the two contributions, although it is less pronounced than at NLO. This attenuation seems to be independent of the definition, and therefore to be due to a pure dynamic NNLO effect. In the end, we conclude that the features observed at NLO are still there at NNLO, though slightly reduced.

3.4 Analytic arguments at NLO

We now show how several features of the results presented in the previous sections can be understood on the basis of simple analytic arguments. Specifically, we show that the behaviour in the vicinity of p_t^{cut} is strongly tied to the unitarity, or lack thereof, of the various definitions. Here we provide some general, exploiting the fact that at NLO the jet functions used for partitioning the phase-space have a compact and manageable form.

In order to understand the behaviour of various definitions we need an explicit expression for the contribution to the N-jet cross section of Eq. (3.7) and to the n-th jet cross section of Eq. (3.8). These can be constructed in terms of parton-level cross sections by introducing explicit jet functions that cluster final-state partons into jets, in the latter case further supplemented by a function that selects the n-th leading jet, and bins the result into a fixed p_t bin. In order to cancel infrared singularities, the k-th order contribution must be constructed by adding up contributions coming from final states with a number of final-state partons that goes from two (with k virtual loops), up to $k + 2$ (with k real emissions on top of the Born level). For instance the NLO $k = 1$ term receives contributions both from a two-parton final state with one loop, and from a real emission three-parton state, and so on.

Explicitly, we can write the N-exclusive jets contribution, Eq. (3.7), as a sum of terms where
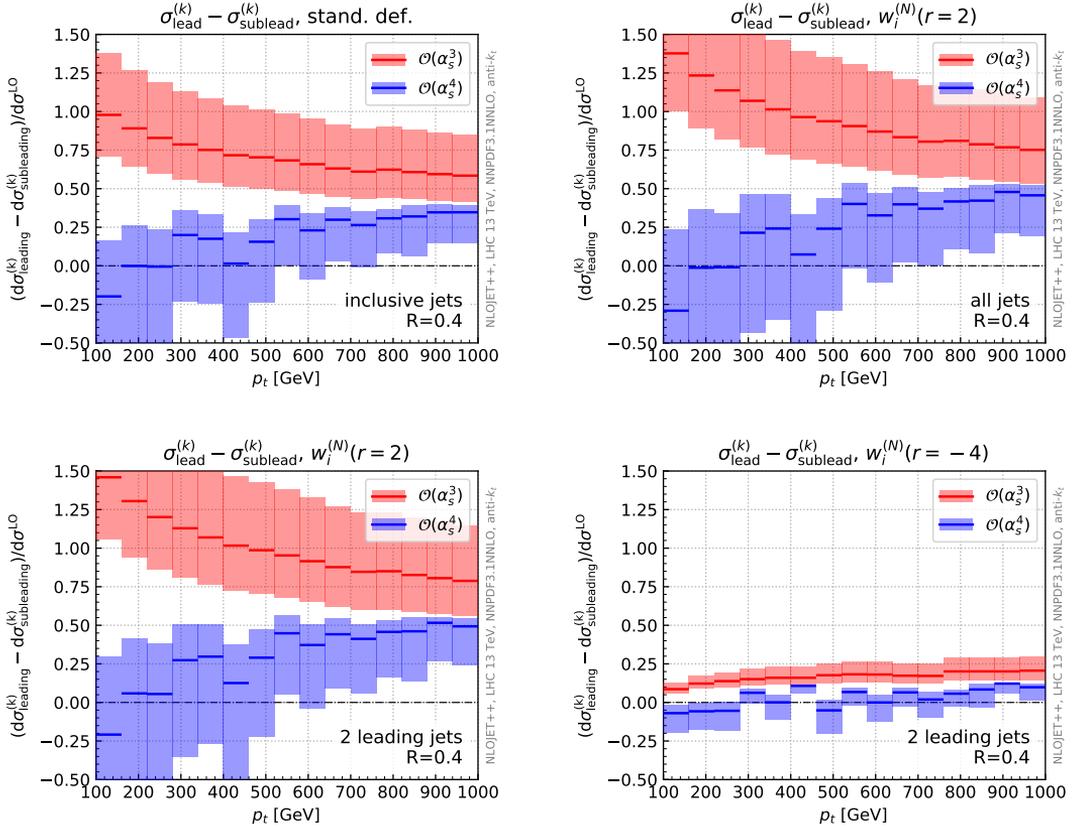
Figure 3.7: Summary of tests carried out at NNLO ($\mathcal{O}(\alpha_s^4)$) by taking the difference between the leading and the second leading jets. Respectively: top left is for the standard definition, top right for definition (B) with $r = 2$, bottom left for definition (C) with $r = 2$, and bottom right for definition (C) with $r = -4$.

the N jets are produced from an m parton final-state, $\mathrm{d}\Phi_m$,

$$\frac{\mathrm{d}\sigma_{N\,\mathrm{jets}}^{(k)}}{\mathrm{d}p_t} = \sum_{m=2}^{k+2} \int \mathrm{d}\Phi_m \frac{\mathrm{d}\hat{\sigma}_m^{(k)}}{\mathrm{d}\Phi_m} G_{m\to N\,\mathrm{jets}}(\Phi_m, p_t) \,. \tag{3.21}$$

where $G_{m\to N\,\mathrm{jets}}$ is the jet function which cluster m partons into N jets. $G_{m\to N\,\mathrm{jets}}$ contains the function F_N, Eq. (3.4), which in turn includes the possible weights. The jet function thus depends on the jet momentum p_t, and on the partonic phase space variables $\mathrm{d}\Phi_m$.

We can give an explicit expression of $G_{m\to N}$ at NLO ($k = 1$). For this, let us denote by k_{ti} the parton transverse momenta, with $k_{t1} \geq k_{t2} \geq k_{t3}$. Using the anti-$k_t$ [68] jet clustering with $R < \pi/2$, one has

$$G_{2\to 1} = G_{2\to 3} = 0 \tag{3.22}$$

$$G_{2\to 2} = \Theta(p_t > p_t^{\text{cut}}) \left\{ 2\, w^{(2)}(p_t; p_t, p_t)\, \delta(p_t - k_{t1}) \right\} \tag{3.23}$$

$$G_{3\to 1} = \Theta(\Delta R_{23} > R)\, \Theta(k_{t1} > p_t^{\text{cut}} > k_{t2} > k_{t3}) \left\{ w^{(1)}(p_t; p_t)\, \delta(p_t - k_{t1}) \right\} \tag{3.24}$$

$$G_{3\to 2} = \Theta(\Delta R_{23} > R)\, \Theta(k_{t1} > k_{t2} > p_t^{\text{cut}} > k_{t3}) \left\{ \sum_{i=1}^{2} w^{(2)}(p_t; k_{t1}, k_{t2})\, \delta(p_t - k_{ti}) \right\}$$

$$+ \Theta(\Delta R_{23} < R)\, \Theta(p_t > p_t^{\text{cut}}) \left\{ 2\, w^{(2)}(p_t; p_t, p_t)\, \delta(p_t - k_{t1}) \right\} \tag{3.25}$$

$$G_{3\to 3} = \Theta(\Delta R_{23} > R)\, \Theta(k_{t1} > k_{t2} > k_{t3} > p_t^{\text{cut}}) \left\{ \sum_{i=1}^{3} w^{(3)}(p_t; k_{t1}, k_{t2}, k_{t3})\delta(p_t - k_{ti}) \right\}, \tag{3.26}$$

where we have defined, as is customary, $\Delta R_{ij} = \sqrt{(\Delta\phi_{ij})^2 + (\Delta y_{ij})^2}$, as the distance between parton i and parton j in the rapidity-azimuth plane, with y and ϕ the rapidity and the azimuthal angle respectively. Note also that, due to momentum conservation, it is sufficient to consider the recombination of the two softest partons. The second line of Eq. (3.25) corresponds to the case where the two softest partons cluster, yielding two back-to-back jets of momentum k_{t1}.

Using Eqs. (3.22)-(3.26), the issue of unitarity vs. cancellation of the dependence on p_t is easily understood. On the one hand, it is clear that the standard definition is not unitary and only the weighted definitions are unitary because

$$\int \mathrm{d}\, p_t\, G_{3\to 1} + G_{3\to 2} + G_{3\to 3}\big|_{\text{wgt}} = \Theta(k_{t1} > p_t^{\text{cut}}). \tag{3.27}$$

This result, valid for any r, means that integrating the single-jet cross section over p_t yields the total cross section for producing (at least) one jet above p_t^{cut} (with definitions of type (A) in the sense of Sec. 3.1) or the total cross section (for definitions of type (B) or of type (C)). Hence these choices are unitary, and thus the standard choice cannot be.

On the other hand, it is clear that the inclusive cross section is independent of p_t^{cut} when using the standard definition. Indeed, in this case one has

$$G_{3\to 1} + G_{3\to 2} + G_{3\to 3}\big|_{\text{std}} = \Theta(p_t > p_t^{\text{cut}}) \Bigg\{ \Theta(\Delta R_{23} > R) \left[\sum_i^3 \delta(p_t - k_{ti}) \right]$$

$$+ \Theta(\Delta R_{23} < R) 2\, \delta(p_t - k_{t1}) \Bigg\}, \tag{3.28}$$

where now the subscript "std" denotes that in the definition of F_N, Eq. (3.4), the standard case in Eq. (3.5) has been selected. The result Eq. (3.28) is manifestly independent of p_t^{cut} since all the dependence on p_t^{cut} is factored in an overall Θ function which is always satisfied as long as one has at least one jet in the event. In practice, the dependence on p_t^{cut} disappears since, when integrating over the partonic transverse momenta, the N-jet contribution has p_t^{cut} as a lower bound of integration while the $N-1$-jet contribution has p_t^{cut} as an upper bound. When summing both contributions, the p_t^{cut} dependence cancels.

When one instead uses a unitary definition which explicitly introduces a p_t^{cut} dependence, such as definition (A), this cancellation is spoiled: whether a jet passes a cut or not changes the weights of all the other jets, thereby introducing a cutoff dependence of the observable. The lack of cancellation then propagates into the individual n-th jet cross sections, thus explaining

the singular behaviour observed in Fig. 3.3 when $p_t \sim p_t^{\text{cut}}$. Of course this cutoff dependence is not present for the two other weighted definitions, (B) and (C), even if the weight associated to a jet still depends on the other jets in the event, which is needed to eventually ensure the unitarity of the cross section.

We can similarly understand the R dependence or lack thereof of the leading jet contribution, which as discussed in Sec. 3.2.1 controls the behaviour of the NLO K factor, by introducing explicit expressions for individual jet functions. We now need to consider the n-th leading jet contribution, Eq. (3.8)

$$\frac{\mathrm{d}\sigma_{n\text{-th jet}}^{(k)}}{\mathrm{d}p_t} = \sum_{m=2}^{k+2} \int \mathrm{d}\Phi_m \frac{\mathrm{d}\hat{\sigma}_m^{(k)}}{\mathrm{d}\Phi_m} S_{m\to n\text{-th jet}}(\Phi_m, p_t), \tag{3.29}$$

where the functions $S_{m\to n\text{-th jet}}$ are defined summing the contributions coming from the n-th jet in the functions G given above. By direct calculation, we find

$$S_{2\to p_{t1}} = S_{2\to p_{t2}} = \frac{1}{2} G_{2\to 2} \tag{3.30}$$

$$\begin{aligned} S_{3\to p_{t1}} = \Theta(p_t > p_t^{\text{cut}})\, \delta(p_t - k_{t1}) \\ \times \Bigg\{ \Theta(\Delta R_{23} > R) \Big[\Theta(p_t^{\text{cut}} > k_{t2} > k_{t3}) w^{(1)}(p_t; k_{t1}) \\ + \Theta(k_{t2} > p_t^{\text{cut}} > k_{t3})\, w^{(2)}(p_t; k_{t1}, k_{t2}) \\ + \Theta(k_{t2} > k_{t3} > p_t^{\text{cut}}) w^{(3)}(p_t; k_{t1}, k_{t2}, k_{t3}) \Big] \\ + \Theta(\Delta R_{23} < R)\, w^{(2)}(p_t; p_t, p_t) \Bigg\} \end{aligned} \tag{3.31}$$

$$\begin{aligned} S_{3\to p_{t2}} = \Theta(k_{t1} > p_t > p_t^{\text{cut}}) \\ \times \Bigg\{ \Theta(\Delta R_{23} > R)\, \delta(p_t - k_{t2}) \Big[\Theta(p_t^{\text{cut}} > k_{t3})\, w^{(2)}(p_t; k_{t1}, k_{t2}) \\ + \Theta(k_{t3} > p_t^{\text{cut}})\, w^{(3)}(p_t; k_{t1}, k_{t2}, k_{t3}) \Big] \\ + \Theta(\Delta R_{23} < R)\, \delta(p_t - k_{t1})\, w^{(2)}(p_t; p_t, p_t) \Bigg\} \end{aligned} \tag{3.32}$$

$$S_{3\to p_{t3}} = \Theta(k_{t1} > k_{t2} > p_t > p_t^{\text{cut}})\, \delta(p_t - k_{t3})\, \Theta(\Delta R_{23} > R)\, w^{(3)}(p_t; k_{t1}, k_{t2}, k_{t3}). \tag{3.33}$$

If now one sets all weights $w = 1$, Eq. (3.31) takes the form

$$S_{3\to p_{t1}}\big|_{\text{std}} = \Theta(p_t > p_t^{\text{cut}})\, \delta(p_t - k_{t1}) = S_{2\to p_{t1}}\big|_{\text{std}}, \tag{3.34}$$

where the subscript "std" again denotes that in the definition of F_N, Eq. (3.4), the standard case in Eq. (3.5) has been selected. This means that all the Θ functions simplify, leading to an overall factor providing a condition that is always satisfied if at least one jet in the event is above p_t^{cut}. At NLO, the leading jet contribution is therefore always given by the transverse momentum of the hardest parton (this is valid for both the real contribution with three partons in the final state and the virtual corrections with two partons), independently of the jet radius R. Note that, one can similarly see that for any weighted definition, at NLO, corrections to the

leading jet are R-dependent for the same reason that the weighted definitions depend on p_t^{cut}: the value of the weights depend on how many partons have $\Delta R_{ij} > R$. Furthermore, the NLO corrections for the subleading and third-leading jet also depend on R. This is trivial for the latter which shows an explicit R dependence in (3.33). For the subleading jet, this is due to the fact that the p_t of the jet changes (between k_{t1} and k_{t2}) depending on how ΔR_{23} compares to R.

3.5 Soft-collinear approximation to NLO cross section

The arguments outlined in the previous section may seem somewhat formal. To gain further analytic insight, it is useful to take a soft-collinear approximation in which case Eqs. (3.21),(3.29) simplify considerably. Indeed, if one considers a collinear splitting at a small angle ϑ, the NLO contribution from a real emission can be written in simple form by parametrising the final-state momenta as

$$p_1^\mu = \tilde{p}_a^\mu + \mathcal{O}(k_\perp^2), \quad p_2^\mu = (1-z)\tilde{p}_b^\mu + k_\perp^\mu + \mathcal{O}(k_\perp^2), \quad p_3^\mu = z\,\tilde{p}_b^\mu - k_\perp^\mu + \mathcal{O}(k_\perp^2), \tag{3.35}$$

where $\tilde{p}_a^\mu$ and $\tilde{p}_b^\mu$ are the Born final-state hard directions, z is the longitudinal momentum fraction of the splitting, and the transverse momentum $k_\perp$ satisfies $k_\perp \cdot \tilde{p}_a = k_\perp \cdot \tilde{p}_b = 0$; $k_\perp$ can then be parametrized by the angle ϑ between p_2 and p_3 and an azimuthal angle φ. Including only terms that produce a logarithmic enhancement in the limit $\vartheta \to 0$, the real emission contribution takes the form

$$d\Phi_3 \frac{d\hat{\sigma}_3^{(1)}}{d\Phi_3} = \sum_{i=q,g} \left[\frac{d\sigma_i^{\text{LO}}}{dp_t}(\tilde{p}_t) \right] \left[\frac{\alpha_s C_i}{\pi} P_i(z) \right] d\tilde{p}_t \, dz \frac{d\vartheta^2}{\vartheta^2} \frac{d\varphi}{2\pi}. \tag{3.36}$$

Note that within this approximation recoil effects on p_1 become negligible. They could be addressed using a similar formalism but going beyond the small-angle approximation that we adopt here.

In Eq. (3.36) $d\sigma_i^{\text{LO}}/dp_t$, with $i = q, g$, is the LO partonic spectrum for producing a quark or a gluon of transverse momentum $\tilde{p}_t$, already introduced in Eq. (2.34). $P_i(z)$ corresponds to the Altarelli-Parisi splitting function in Eqs. (1.91)-(1.92), with z the momentum fraction of the collinear splitting. Note that we have explicitly factored out a colour factor $2\,C_i$ ($C_i = C_F$ for quarks and $C_i = C_A$ for gluons). At this accuracy, the NLO one-loop virtual correction has exactly the same form as Eq. (3.36) integrated over the full phase-space of the extra real emission, but with the opposite sign. In what follows, we further assume that the extra emission is soft so we can approximate $P_i(z) \approx 1/z$, as in Eq. (1.93). This soft approximation is made for the sake of simplicity and can easily be lifted to include the full splitting function.

The soft-collinear approximation is sufficient to obtain results in fair agreement with the full calculation, and specifically reproduce three important aspects discussed in Sec. 3.2. First, we can see explicitly how the cancellation of the p_t^{cut} dependence which happens in the standard case is spoiled for the weighted definition (A) and restored with definitions (B) and (C). Second, we are able to identify the R dependence of the second leading jet with out-of-cone radiation. Third, we can further study the impact of weighted definitions at large p_t. Conversely, working in a soft-collinear approximation, we are neglecting all recoil effects. This means in particular that the calculation below will not reproduce the large K_1 factor for the leading jet. The text below outlines the structure of the calculation and our main results, deferring additional details to Appendix 3.A.

The fact that the real and virtual contributions have the opposite sign implies that the N-jet contribution Eq. (3.21) and the n-th jet contribution Eq. (3.29) take respectively the simple form

$$\frac{d\sigma_{N\,\text{jets}}^{(1)}}{dp_t} \approx \sum_{i=q,g} \frac{C_i}{\pi} \int d\tilde{p}_t \, dz \frac{d\vartheta^2}{\vartheta^2} \left[\frac{d\sigma_i^{\text{LO}}}{dp_t}(\tilde{p}_t) \right] \alpha_s \, P_i(z) \, \{ G_{3\to N\,\text{jets}} - G_{2\to N\,\text{jets}} \} \tag{3.37}$$

$$\equiv \sum_{i=q,g} \left[\frac{d\sigma_i^{\text{LO}}}{dp_t}(p_t) \right] \frac{C_i}{\pi} \ln \left(\frac{R_{\max}^2}{R^2} \right) I_N \tag{3.38}$$

and

$$\frac{d\sigma_{n\,\text{-th jet}}^{(1)}}{dp_t} \approx \sum_{i=q,g} \frac{C_i}{\pi} \int d\tilde{p}_t \, dz \frac{d\vartheta^2}{\vartheta^2} \left[\frac{d\sigma_i^{\text{LO}}}{dp_t}(\tilde{p}_t) \right] \alpha_s \, P_i(z) \, \{ S_{3\to n\,\text{-th jet}} - S_{2\to n\,\text{-th jet}} \} \tag{3.39}$$

$$\equiv \sum_{i=q,g} \left[\frac{d\sigma_i^{\text{LO}}}{dp_t}(p_t) \right] \frac{C_i}{\pi} \ln \left(\frac{R_{\max}^2}{R^2} \right) J_n \, , \tag{3.40}$$

where in both cases $R_{\max}$ is the upper limit of the ϑ integration. The functions I_N and J_n can be cast in a simple closed analytic form by writing the LO cross section as a power law

$$\frac{d\sigma_i^{\text{LO}}}{dp_t}(\tilde{p}_t) \sim \tilde{p}_t^{-m_i} \, , \tag{3.41}$$

where m_i is, in general, different for the quark and gluon case (see end of Sec. 2.1.5). In Appendix 3.A explicit analytic expressions are given for the standard definition, with the general definitions easily amenable to numerical treatment.

We can now use Eqs. (3.38),(3.40) to address the issues mentioned above. We start by investigating the behaviour in the $p_t \to p_t^{\text{cut}}$ limit and focus on the leading jet. J_1 receives real contributions from $S_{3\to p_{t1}}$, Eq. (3.31), and virtual corrections from $S_{2\to p_{t1}}$, Eq. (3.30). The latter contribution cancels against the real one in the region $\Delta R_{23} \equiv \vartheta < R$. Up to power corrections in z, we can set $k_{t2} = (1-z)k_{t1}$ and $k_{t3} = zk_{t1}$. For $p_t \to p_t^{\text{cut}}$ we can then assume $k_{t3} < p_t^{\text{cut}}$ and we are left with two terms:

$$J_1 \overset{p_t \to p_t^{\text{cut}}}{\sim} \int_{1-p_t^{\text{cut}}/p_t}^{p_t^{\text{cut}}/p_t} dz \, P(z) \, w^{(1)}(p_t; p_t) - \int_{1-p_t^{\text{cut}}/p_t}^{p_t^{\text{cut}}/p_t} dz \, P(z) \, w^{(2)}(p_t; p_t, p_t). \tag{3.42}$$

The first term corresponds to $k_{t2} < p_t^{\text{cut}}$ while the second term includes the real emissions with $k_{t2} > p_t^{\text{cut}}$ as well as the remaining virtual corrections. After integration over z, we thus find

$$J_1 = \log \left(\frac{p_t^{\text{cut}}}{p_t - p_t^{\text{cut}}} \right) - \omega \log \left(\frac{p_t^{\text{cut}}}{p_t - p_t^{\text{cut}}} \right) = \begin{cases} 0 & \text{[standard]} \\ -\frac{1}{2} \log \left(\frac{p_t - p_t^{\text{cut}}}{p_t^{\text{cut}}} \right) & \text{[weighted (A)]} \end{cases} , \tag{3.43}$$

where $\omega = 1$ for the standard definition and $\omega = \frac{1}{2}$ for the weighted definition (A), independently of the exponent r which enters the definition of the weights, Eq. (3.4). In the same limit it turns out that J_2 and J_3 are non-singular. This explains our findings from Sec. 3.2: the unitary definition suffers from a logarithmic divergence close to p_t^{cut} while the standard definition is independent of the value of p_t^{cut}. Furthermore, this behaviour (see Fig. 3.3), only affects the

leading jet, whose properties are encoded in J_1. Of course it also follows from Eq. (3.43) that when $p_t^{\mathrm{cut}} \ll p_t$, corresponding to using definitions of the weights of type (B), the singular behaviour disappears. A similar conclusion can be reached for the definition of type (C).

Next, we can also use Eq. (3.40) to predict the small-R behaviour of the second and third leading jet contributions. In both cases one would get a logarithmic enhancement at small R. Note that at first sight Eq. (3.40) seems to imply that the leading jet contribution also has a logarithmic R dependence in the standard case, in contradiction to the behaviour observed in Fig. 3.2, and to our previous general conclusion based on Eq. (3.34). However, one should realize that, in the small-R limit where Eq. (3.40) holds, Eq. (3.43) implies that J_1 is zero, and thus obviously R-independent in the standard case. In all weighted cases J_1 is non-vanishing, and thus the leading jet contribution becomes R-dependent in agreement with our previous analytic and numerical arguments, with a logarithmic dependence on R in the small-R limit.

Finally, we can study the limit of the functions J_n when $p_t \gg p_t^{\mathrm{cut}}$, in the weighted case with r negative and $|r| \sim |m_i|$. In this case, we find that the contributions from the leading and the subleading jet are comparable (see Eqs. (3.67)-(3.68)), partially solving the problem of the large compensation seen in the standard definition or for positive values of r, as observed in Sec. 3.2.2, Fig. 3.3.

Results obtained for the leading, subleading and third-leading jet contributions using the approximation Eqs. (3.38),(3.40) are shown in Fig. 3.8 for a representative set of cases, to be compared to Figs. 3.1,3.3-3.5. All plots have been produced implementing Eq. (3.39), with Eq. (3.41) and $m = 5$. Note that this parametrization of the LO p_t spectrum already includes initial state PDFs. We have checked that using the exact LO partonic cross section and the full Altarelli-Parisi splitting functions yields similar results. We choose $R_{\mathrm{max}} = 1$ and use $R = 0.4$ to allow for a comparison with the full results presented in Section 3.2. Tests of consistency between the the full NLO result in a proper collinear limit and our code implementing Eq. (3.36) are carried out in Appendix 3.B. Finally, we set $\alpha_s(p_{t1})$. As anticipated, it is clear that the main qualitative features of the exact results are reproduced by the soft-collinear approximation.

3.6 Summary of results

In this chapter we have addressed the potential issue of the non-unitarity of the single-jet inclusive cross section, by introducing a series of alternative weighted definitions of this observable which are unitary in the sense that upon integration they lead to the total cross section. The main features of the various definitions we have considered are summarised in Table 3.1.

Our conclusion is that a naive weighted approach [type (A) of Sec. 3.1] in which one simply introduces a weighting of all jets above a certain p_t^{cut} is flawed, in the sense that it develops logarithmic singularities associated with the transverse momentum cut on jets, p_t^{cut}. More sophisticated definitions avoid this problem by setting p_t^{cut} to zero [type (B)] or by considering only the two leading jets [type (C)]. Both these definitions could be more challenging to implement in a practical (experimental) environment.

Additionally, even leaving aside practical considerations, there does not seems to be any real advantage in adopting these definitions in term of perturbative stability. In particular, all weighted definitions with positive r show features at best similar to the standard definition. Furthermore, the apparent perturbative instability of the conventional definition appears in fact to be the manifestation of an unnatural smallness of the NLO K factors which only happens for a limited range of jet radius $R \sim 0.4$. It is a consequence of an accidental cancellation
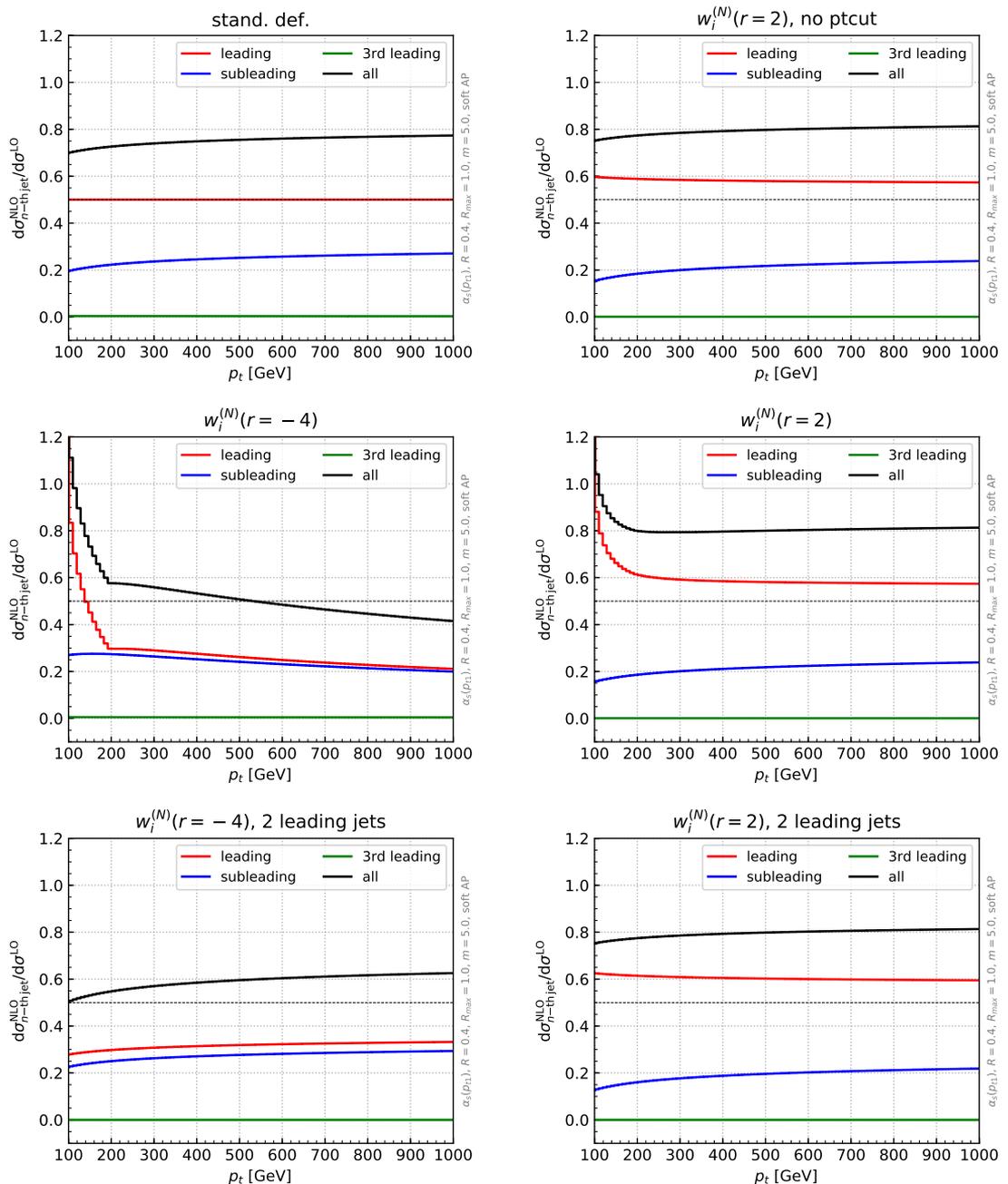
Figure 3.8: Contributions from the leading, subleading, and third-leading jets to the NLO inclusive K factors in the soft-collinear approximation. The standard definition (top left) is compared to weighted definition of type (B) (no p_t^{cut}) with $r = 2$ (top right), weighted definitions of type (A) (with p_t^{cut}) with $r = -4$ (middle left) and $r = 2$ (middle right) and of type (C) (two jets) also with with $r = -4$ (bottom left) and $r = 2$ (bottom right).

Definition	standard	weighted		
		(A) $> p_t^{\mathrm{cut}}$	(B) all jets	(C) two lead.
Reference plot	Fig. 3.1	Fig. 3.3	Fig. 3.4	Fig. 3.5
unitarity	no	yes	yes	yes
no large logs close to p_t^{cut}	✓	✗	✓	✓
no large logs at large p_t	✓	✓ for $r > 0$ ✗ for $r < 0$	✓	✓
overlapping scale variation bands	✗ ✓ uncer. as [124]	✓	✓	✓
no large cancellations between K_1 and K_2	✗	✗	✗	✗ for $r > 0$ ✓ for $r < 0$

Table 3.1: Summary of the main properties of the various single-inclusive jet definitions studied in this chapter.

which makes standard scale variation unreliable as a means of estimating missing higher order corrections. This apparent issue for example disappears with more conservative estimates of the perturbative uncertainties. One possible case of interest is the definition of type (C), focusing on the two leading jets, with $r < 0$. Compared to the standard definition, it has the potential advantage of reducing the large difference between the K factor of the leading and subleading jets, at the cost of having a larger overall NLO K factor.

Our final conclusion is both negative, and positive. On the negative side, we conclude that unitary definitions of the jet inclusive cross section are at best as good as the standard definition, while being rather more contrived. On the positive side, we conclude that the standard definition shows no critical sign of pathological features or problems, other than its unitarity, which however is *per se* not causing any perturbative problem.

To return to the question we asked ourselves in the introduction of the thesis, on what is the best observable between single-jet or dijet processes, the findings of this chapter give clues for a possible answer. Among the unitary definitions, the weighted definitions based on including only the two leading jets appear to be particularly well-behaved. After all, these definitions are a way to study the dijet system as a function of the p_t and rapidity of the individual jets, rather than adopting event-wide variables. This is in agreement with previous studies [121] in which dijet observables are also found to have better perturbative stability, see end of Sec. 2.4. The PDF fits in the next chapter will add an extra piece of information to this debate.

3.A Explicit analytic expressions

The N-jet contribution and the n-th jet contribution to the differential cross section at NLO in the soft-collinear approximation are given by Eq. (3.37) and Eq. (3.39) respectively. Using an explicit expression for the splitting functions P_i, Eqs. (1.91)-(1.92), and for the G or the S functions in the collinear limit we can perform the phase-space integration explicitly.

By adopting the parametrization of the final-state given in Eq. (3.35), the jet functions G and S can be rewritten in the collinear and small R limit, i.e. $\Delta R_{23} = \vartheta \ll 1$. For the weighted

definition with jets above p_t^{cut} we have:

$$G_{2\to1} = G_{2\to3} = 0 \tag{3.44}$$

$$G_{2\to2} = \Theta(\tilde{p}_t > p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,\tilde{p}_t)[\delta(p_t - \tilde{p}_t) + \delta(p_t - \tilde{p}_t)] \tag{3.45}$$

$$G_{3\to1} = \Theta(\vartheta^2 > R^2)\Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t < p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t < p_t^{\mathrm{cut}})w^{(1)}(p_t|\tilde{p}_t)[\delta(p_t - \tilde{p}_t)] \tag{3.46}$$

$$\begin{aligned}
G_{3\to2} = & \Theta(\vartheta^2 < R^2)\Theta(\tilde{p}_t > p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,\tilde{p}_t)[\delta(p_t - \tilde{p}_t) + \delta(p_t - \tilde{p}_t)] \\
& + \Theta(\vartheta^2 > R^2)\Theta(\tilde{p}_t > p_t^{\mathrm{cut}}) \\
& \qquad \Big\{ \Theta(z\tilde{p}_t < p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t > p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,(1-z)\tilde{p}_t) \\
& \qquad \quad \times [\delta(p_t - \tilde{p}_t) + \delta(p_t - (1-z)\tilde{p}_t)] \\
& \qquad \quad + \Theta(z\tilde{p}_t > p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t < p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,z\tilde{p}_t)[\delta(p_t - \tilde{p}_t) + \delta(p_t - z\tilde{p}_t)] \Big\}
\end{aligned} \tag{3.47}$$

$$\begin{aligned}
G_{3\to3} = & \Theta(\vartheta^2 > R^2)\Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t > p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t > p_t^{\mathrm{cut}}) \\
& \times w^{(3)}(p_t|\tilde{p}_t,z\tilde{p}_t,(1-z)\tilde{p}_t)[\delta(p_t - \tilde{p}_t) + \delta(p_t - z\tilde{p}_t) + \delta(p_t - (1-z)\tilde{p}_t)].
\end{aligned} \tag{3.48}$$

and

$$S_{2\to p_{t1}} = S_{2\to p_{t2}} = \Theta(\tilde{p}_t > p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,\tilde{p}_t)\delta(p_t - \tilde{p}_t) \tag{3.49}$$

$$\begin{aligned}
S_{3\to p_{t1}} = & \Theta(\vartheta^2 < R^2)\Theta(\tilde{p}_t > p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,\tilde{p}_t)\delta(p_t - \tilde{p}_t) + \Theta(\vartheta^2 > R^2) \\
& \times \Big[\Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t < p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t < p_t^{\mathrm{cut}})w^{(1)}(p_t|\tilde{p}_t)\delta(p_t - \tilde{p}_t) \\
& \quad + \Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t > p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t < p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,z\tilde{p}_t)\delta(p_t - \tilde{p}_t) \\
& \quad + \Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t < p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t > p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,(1-z)\tilde{p}_t)\delta(p_t - \tilde{p}_t) \\
& \quad + \Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t > p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t > p_t^{\mathrm{cut}}) \\
& \qquad \times w^{(3)}(p_t|\tilde{p}_t,z\tilde{p}_t,(1-z)\tilde{p}_t)\delta(p_t - \tilde{p}_t) \Big]
\end{aligned} \tag{3.50}$$

$$\begin{aligned}
S_{3\to p_{t2}} = & \Theta(\vartheta^2 < R^2)\Theta(\tilde{p}_t > p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,\tilde{p}_t)\delta(p_t - \tilde{p}_t) + \Theta(\vartheta^2 > R^2) \\
& \times \Big\{ \Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t > p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t < p_t^{\mathrm{cut}})w^{(2)}(p_t|\tilde{p}_t,z\tilde{p}_t)\delta(p_t - z\tilde{p}_t) \\
& \quad + \Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t < p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t > p_t^{\mathrm{cut}}) \\
& \qquad \times w^{(2)}(p_t|\tilde{p}_t,(1-z)\tilde{p}_t)\delta(p_t - (1-z)\tilde{p}_t) \\
& \quad + \Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t > p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t > p_t^{\mathrm{cut}})w^{(3)}(p_t|\tilde{p}_t,z\tilde{p}_t,(1-z)\tilde{p}_t) \\
& \qquad \times [\Theta(z > 1/2)\delta(p_t - z\tilde{p}_t) + \Theta(z < 1/2)\delta(p_t - (1-z)\tilde{p}_t)] \Big\}
\end{aligned} \tag{3.51}$$

$$\begin{aligned}
S_{3\to p_{t3}} = & \Theta(\vartheta^2 > R^2)\Theta(\tilde{p}_t > p_t^{\mathrm{cut}}; z\tilde{p}_t > p_t^{\mathrm{cut}}; (1-z)\tilde{p}_t > p_t^{\mathrm{cut}})w^{(3)}(p_t|\tilde{p}_t,z\tilde{p}_t,(1-z)\tilde{p}_t) \\
& \times [\Theta(z < 1/2)\delta(p_t - z\tilde{p}_t) + \Theta(z > 1/2)\delta(p_t - (1-z)\tilde{p}_t)]
\end{aligned} \tag{3.52}$$

The standard definition can trivially be recovered by setting the weights to 1, while the case of the weighted definition including all jets can be obtained by taking the limit $p_t^{\mathrm{cut}} \to 0$. Similarly, the weighted definition with 2 leading jets is instead obtained by firstly taking the limit $p_t^{\mathrm{cut}} \to 0$

and by then keeping the terms proportional to $\delta(p_t - \tilde{p}_t)$ as well as the terms proportional to either $\delta(p_t - z\tilde{p}_t)$ if $z > 1/2$, or $\delta(p_t - (1-z)\tilde{p}_t)$ if $z < 1/2$, modifying the weights accordingly.

The $\tilde{p}_t$ integration in Eqs. (3.37)-(3.39) can be simplified using the delta functions $\delta(p_t - \tilde{p}_t)$, $\delta(p_t - z\tilde{p}_t)$ and $\delta(p_t - (1-z)\tilde{p}_t)$. The ϑ integration leads to a logarithmic dependence on the jet radius R. The only nontrivial integral is over z, thereby leading to a final result of the form of Eqs. (3.38),(3.40). Explicitly, I_N and J_n there present are given by:

$$I_1 = \Theta(p_t < 2p_t^{\text{cut}}) \int_{1-p_t^{\text{cut}}/p_t}^{p_t^{\text{cut}}/p_t} dz\, P(z)\, [1]\, \tilde{\sigma}(p_t) \tag{3.53}$$

$$I_2 = \Theta(p_t < 2p_t^{\text{cut}}) \left[\int_{p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{1+z^r} \right] \tilde{\sigma}(p_t) \right. \tag{3.54}$$
$$+ \int_{0}^{1-p_t^{\text{cut}}/p_t} dz\, P(z) \left(\left[\frac{1}{1+(1-z)^r} \right] \tilde{\sigma}(p_t) - \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right)$$
$$\left. - \int_{1-p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right]$$
$$+ \Theta(p_t > 2p_t^{\text{cut}}) \left[\int_{1-p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{1+z^r} \right] \tilde{\sigma}(p_t) \right.$$
$$+ \int_{0}^{p_t^{\text{cut}}/p_t} dz\, P(z) \left(\left[\frac{1}{1+(1-z)^r} \right] \tilde{\sigma}(p_t) - \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right)$$
$$\left. - \int_{p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right]$$
$$+ \int_{p_t/(p_t+p_t^{\text{cut}})}^{1} dz\, P(z) \left[\frac{z^r}{1+z^r} \right] \frac{1}{z} \tilde{\sigma}\left(\frac{p_t}{z} \right)$$
$$+ \int_{0}^{p_t^{\text{cut}}/(p_t+p_t^{\text{cut}})} dz\, P(z) \left(\left[\frac{(1-z)^r}{1+(1-z)^r} \right] \frac{1}{1-z} \tilde{\sigma}\left(\frac{p_t}{1-z} \right) - \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right)$$
$$- \int_{p_t^{\text{cut}}/(p_t+p_t^{\text{cut}})}^{1} dz\, P(z) \left[\frac{1}{2} \right] \tilde{\sigma}(p_t)$$

$$I_3 = \Theta(p_t > 2p_t^{\text{cut}}) \int_{p_t^{\text{cut}}/p_t}^{1-p_t^{\text{cut}}/p_t} dz\, P(z) \left[\frac{1}{1+z^r+(1-z)^r} \right] \tilde{\sigma}(p_t) \tag{3.55}$$
$$+ \int_{0}^{p_t/(p_t+p_t^{\text{cut}})} dz\, P(z) \left[\frac{z^r}{1+z^r+(1-z)^r} \right] \frac{1}{z} \tilde{\sigma}\left(\frac{p_t}{z} \right)$$
$$+ \int_{p_t^{\text{cut}}/(p_t+p_t^{\text{cut}})}^{1} dz\, P(z) \left[\frac{(1-z)^r}{1+z^r+(1-z)^r} \right] \frac{1}{1-z} \tilde{\sigma}\left(\frac{p_t}{1-z} \right)$$

$$J_1 = \Theta(p_t < 2p_t^{\text{cut}}) \left[\int_{1-p_t^{\text{cut}}/p_t}^{p_t^{\text{cut}}/p_t} dz\, P(z)\, [1]\, \tilde{\sigma}(p_t) - \int_{1-p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right. \tag{3.56}$$
$$\left. + \int_{0}^{1-p_t^{\text{cut}}/p_t} dz\, P(z) \left(\left[\frac{1}{1+(1-z)^r} \right] \tilde{\sigma}(p_t) - \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right) \right.$$

$$+ \int_{p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{1+z^r} \right] \tilde{\sigma}(p_t) \Bigg]$$

$$+ \Theta(p_t > 2p_t^{\text{cut}}) \Bigg[\int_{p_t^{\text{cut}}/p_t}^{1-p_t^{\text{cut}}/p_t} dz\, P(z) \left[\frac{1}{1+z^r+(1-z)^r} \right] \tilde{\sigma}(p_t)$$

$$- \int_{p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{2} \right] \tilde{\sigma}(p_t)$$

$$+ \int_{0}^{p_t^{\text{cut}}/p_t} dz\, P(z) \left(\left[\frac{1}{1+(1-z)^r} \right] \tilde{\sigma}(p_t) - \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right)$$

$$+ \int_{1-p_t^{\text{cut}}/p_t}^{1} dz\, P(z) \left[\frac{1}{1+z^r} \right] \tilde{\sigma}(p_t) \Bigg]$$

$$J_2 = \int_{p_t/(p_t+p_t^{\text{cut}})}^{1} dz\, P(z) \left[\frac{z^r}{1+z^r} \right] \frac{1}{z} \tilde{\sigma}\left(\frac{p_t}{z} \right) - \int_{p_t^{\text{cut}}/(p_t+p_t^{\text{cut}})}^{1} dz\, P(z) \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \tag{3.57}$$

$$+ \int_{0}^{p_t^{\text{cut}}/(p_t+p_t^{\text{cut}})} dz\, P(z) \left(\left[\frac{(1-z)^r}{1+(1-z)^r} \right] \frac{1}{1-z} \tilde{\sigma}\left(\frac{p_t}{1-z} \right) - \left[\frac{1}{2} \right] \tilde{\sigma}(p_t) \right)$$

$$+ \int_{1/2}^{p_t/(p_t+p_t^{\text{cut}})} dz\, P(z) \left[\frac{z^r}{1+z^r+(1-z)^r} \right] \frac{1}{z} \tilde{\sigma}\left(\frac{p_t}{z} \right)$$

$$+ \int_{p_t^{\text{cut}}/(p_t+p_t^{\text{cut}})}^{1/2} dz\, P(z) \left[\frac{(1-z)^r}{1+z^r+(1-z)^r} \right] \frac{1}{1-z} \tilde{\sigma}\left(\frac{p_t}{1-z} \right)$$

$$J_3 = \int_{0}^{1/2} dz\, P(z) \left[\frac{z^r}{1+z^r+(1-z)^r} \right] \frac{1}{z} \tilde{\sigma}\left(\frac{p_t}{z} \right) \tag{3.58}$$

$$+ \int_{1/2}^{1} dz\, P(z) \left[\frac{(1-z)^r}{1+z^r+(1-z)^r} \right] \frac{1}{1-z} \tilde{\sigma}\left(\frac{p_t}{1-z} \right)$$

where the terms in squared brackets correspond to the weights, here given for a definition of type (A), and we have set the running coupling scale to $p_t^{\max} \equiv p_{t1} = \tilde{p}_t$ and introduced

$$\tilde{\sigma}(x) \equiv \frac{d\sigma^{\text{LO}}}{dp_t}(x)\, \alpha_s(x). \tag{3.59}$$

In the fixed coupling approximation or if we take $\alpha_s(p_t)$, the coupling can be factorized out of the integration and directly moved to Eq. (3.38) or Eq. (3.40). Note that the above expressions do not assume $z \ll 1$. Keeping the full z dependence of the splitting functions would therefore account for hard-collinear splittings.

In the general weighted case, these integrals can only be computed numerically. Results for the standard (unweighted) definition are found by simply removing all terms in square brackets. In this case, by using Eq. (3.41) for the Born cross section and the soft approximation of the splitting functions these integrals can be computed exactly in the fixed coupling approximation and their expressions are

$$I_1^{(\text{std})} = \Theta(p_t < 2p_t^{\text{cut}}) \ln \left(\frac{p_t^{\text{cut}}}{p_t - p_t^{\text{cut}}} \right) \tag{3.60}$$

$$I_2^{\text{(std)}} = \Theta(p_t > 2p_t^{\text{cut}}) \ln\left(\frac{p_t^{\text{cut}}}{p_t - p_t^{\text{cut}}}\right) - \Theta(p_t < 2p_t^{\text{cut}}) \ln\left(\frac{p_t^{\text{cut}}}{p_t - p_t^{\text{cut}}}\right) \tag{3.61}$$

$$+ \ln\left(\frac{p_t^{\text{cut}}}{p_t + p_t^{\text{cut}}}\right) + \frac{1}{m-1}\left(1 - \left(\frac{p_t + p_t^{\text{cut}}}{p_t}\right)^{1-m}\right)$$

$$- (m-1)\left(\frac{p_t^{\text{cut}}}{p_t + p_t^{\text{cut}}}\right) {}_3F_2\left(1, 1, 2-m; 2, 2; \frac{p_t^{\text{cut}}}{p_t + p_t^{\text{cut}}}\right)$$

$$I_3^{\text{(std)}} = -\Theta(p_t > 2p_t^{\text{cut}}) \ln\left(\frac{p_t^{\text{cut}}}{p_t - p_t^{\text{cut}}}\right) - \ln\left(\frac{p_t^{\text{cut}}}{p_t + p_t^{\text{cut}}}\right) + \frac{1}{m-1}\left(\frac{p_t + p_t^{\text{cut}}}{p_t}\right)^{1-m} \tag{3.62}$$

$$- H_{m-1} + (m-1)\left(\frac{p_t^{\text{cut}}}{p_t + p_t^{\text{cut}}}\right) {}_3F_2\left(1, 1, 2-m; 2, 2; \frac{p_t^{\text{cut}}}{p_t + p_t^{\text{cut}}}\right)$$

and

$$J_1^{\text{(std)}} = 0 \tag{3.63}$$

$$J_2^{\text{(std)}} = -\frac{1}{2}(m-1) \, {}_3F_2\left(1, 1, 2-m; 2, 2; \frac{1}{2}\right) - \frac{2^{1-m}-1}{m-1} - \log 2 \tag{3.64}$$

$$J_3^{\text{(std)}} = \frac{1}{2}(m-1) \, {}_3F_2\left(1, 1, 2-m; 2, 2; \frac{1}{2}\right) - H_{m-1} + \frac{2^{1-m}}{m-1} + \log 2 \tag{3.65}$$

where H_n are harmonic numbers, ${}_pF_q$ is a generalized hypergeometric function, and m is the power of the LO cross section in Eq. (3.41), which can in principle differ for quarks and gluons.

Adding up all contributions we get:

$$\frac{d\sigma^{(1)}}{dp_t} = \sum_{i=q,g}\left[\frac{d\sigma_i^{\text{LO}}}{dp_t}(p_t)\right]\frac{\alpha_s C_i}{\pi}\ln\left(\frac{R_{\max}^2}{R^2}\right)\left[\frac{1}{m_i - 1} - H_{m_i - 1}\right]. \tag{3.66}$$

For $m_q = m_g$, the K factor is flat, since both the p_t and the p_t^{cut} dependence have canceled completely in the square bracket in the last line. The only remaining dependence on p_t would therefore come either from differences between the quark and gluon contributions ($m_q \neq m_g$) or from the running of α_s which was neglected in the above result.

We conclude by studying the large p_t limit of J_n in the weighted case. When $p_t \to \infty$, from Eqs. (3.56)-(3.57) we get

$$J_1^{\text{(wgt)}} \overset{p_t\to\infty}{\sim} -\int_0^1 dz\, P(z)\left[\frac{1}{2}\right]\tilde{\sigma}(p_t) + \int_0^1 dz\, P(z)\left[\frac{1}{1 + z^r + (1-z)^r}\right]\tilde{\sigma}(p_t) \tag{3.67}$$

$$J_2^{\text{(wgt)}} \overset{p_t\to\infty}{\sim} -\int_0^1 dz\, P(z)\left[\frac{1}{2}\right]\tilde{\sigma}(p_t) + \int_{1/2}^1 dz\, P(z)\left[\frac{z^r}{1 + z^r + (1-z)^r}\right]\frac{1}{z}\tilde{\sigma}\left(\frac{p_t}{z}\right)$$

$$+ \int_0^{1/2} dz\, P(z)\left[\frac{(1-z)^r}{1 + z^r + (1-z)^r}\right]\frac{1}{1-z}\tilde{\sigma}\left(\frac{p_t}{1-z}\right) \tag{3.68}$$

while J_3 in Eq. (3.58) does not depend on p_t and it is always negligible. Assuming that the LO cross section behaves accordingly to the power law Eq. (3.41), and choosing a negative exponent $r \sim -m$ for the weights, it appears that J_1 and J_2 become the same in the $p_t \to \infty$ limit. Hence, the effect of the weight is to balance the leading and the second leading jet contributions.

3.B   Test of collinear approximation

In this appendix we present some simple tests of consistency between the full NLO result, obtained as output of the NLOJET++ code, and our code implementing the collinear approximation. In particular, we focus on the 3-jet cross section at LO, $\sigma_{3\,\text{jets}}^{\text{LO}}$, which does not involve virtual corrections, whose finite part is missing in our code (we are canceling the infrared divergences imposing unitarity). In practice, for each NLOJET++ event, we cluster the final-state partons with the anti-k_t algorithm and $R = 0.2$, and we impose the presence of 3 jets in the event. In this way, we have equivalence between jets and partons, and we can check whether, in some angular limit, our approximation agree with the full result.

For instance, we can constrain real emissions by imposing that the angular distance between the subleading and the third leading jet, ΔR_{23}, defined after Eq. (3.26), is less than some angular cut R_{max} i.e. $R^2 < \Delta R_{23}^2 < R_{\text{max}}^2$. In our collinear approximation, this is this the condition we have to fulfill to obtain a 3-jet event, according to Eq. (3.48). The result of this test is shown in Fig. 3.9 on the left, by comparing the p_t spectrum as a function of R_{max} for several p_t values. We clearly see that, as R_{max} decreases, the outputs of our code and NLOJET++ tend to agree.

Furthermore, we can verify this angular dependence at the differential level i.e. we can check whether

$$\frac{d\sigma_{3\,\text{jets}}^{\text{LO}}}{\Delta R_{23}^2} \propto \frac{1}{\Delta R_{23}^2} \text{ in the limit } \Delta R_{23}^2 \to 0 \,. \tag{3.69}$$

Note that in our code, we have implemented exactly this scaling behaviour, as dictated by the collinear approximation (see Eq. (3.36) with $\vartheta^2 = \Delta R_{23}^2$). This test is performed in Fig. 3.9 on the right: again, when the angular distance decreases, the NLO full result tends towards the collinear approximation characterized by the $1/\Delta R_{23}^2$ scaling.
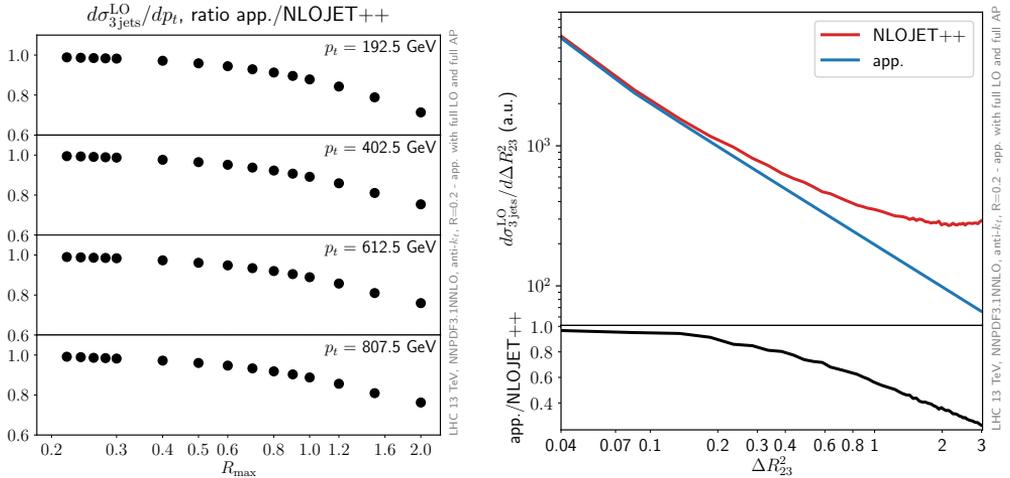

Figure 3.9: Comparisons between the output of NLOJET++ and results in the collinear approximation. Left: 3-jet p_t-differential cross section as a function of R_{max} for several p_t values. Right: 3-jet ΔR_{23}^2-differential cross section.

Impact of jet measurements on parton distributions

Inclusive jets have been used for the determination of the parton distributions of the proton for over thirty years [129]. However, as we have discussed in the previous chapters, there are a number of questions related to the definition of the observable which is most promising and appropriate for precision QCD studies, such as the determination of the PDFs and of the strong coupling constant α_s. In this chapter, we address these issues from a phenomenological point of view, specifically within the context of a global PDF determination. In particular, we study the effect of adding jet cross-sections to a global dataset, with various choices of the observable (single-jet inclusive, or dijet) and of the scale. In each case, we assess the fit quality and the impact of the jet data on the PDFs, at various perturbative orders. We can test phenomenologically the conclusions of the previous chapters, by checking which observable and which scale choice leads to better perturbative stability, better PDF compatibility with other data and better fit quality, and more stringent constraints on the PDFs.

We will consider the complete inclusive jet [86, 87, 130, 131] and dijet [131–133] dataset from ATLAS and CMS at $\sqrt{s}$ = 7 and 8 TeV. Whereas most recent global determinations of the proton PDFs [56–58, 134] include some of these jet datasets (for instance, NNPDF3.1 included the ATLAS and CMS single-inclusive data with $\sqrt{s}$ = 2.76 and 7 TeV), and other studies have assessed the impact of some jet measurements on smaller datasets [133, 135], this is the first time that the full LHC-Run I jet dataset is being considered, and specifically the first time dijets are included in a modern global PDF determination.

The outline of this chapter is as follows. In Sec. 4.1 we discuss the experimental data for single-inclusive jet and dijet production that will be used. In Sec. 4.2 we present the theory that will be used, in particular with respect to the role of NNLO QCD corrections and to the choice of scales, and we briefly discuss how theoretical calculations are implemented by means of fast interpolation grids and K-factors. In Sec. 4.3 we shortly introduce the reader to the NNPDF methodology, based on a Monte Carlo approach to experimental data and on the adoption of neural networks as universal interpolants. The settings of the global PDF analyses of this study

Experiment	Measurement	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [fb^{-1}]	R	Distribution	n_{dat}	Reference
ATLAS	Inclusive jets	7	4.5	0.6	$d^2\sigma/dp_t d|y|$	140	[130]
CMS	Inclusive jets	7	4.5	0.7	$d^2\sigma/dp_t d|y|$	133	[131]
ATLAS	Inclusive jets	8	20.2	0.6	$d^2\sigma/dp_t d|y|$	171	[86]
CMS	Inclusive jets	8	19.7	0.7	$d^2\sigma/dp_t d|y|$	185	[87]
ATLAS	Dijets	7	4.5	0.6	$d^2\sigma/dm_{jj} d|y^*|$	90	[132]
CMS	Dijets	7	4.5	0.7	$d^2\sigma/dm_{jj} d|y_{\mathrm{max}}|$	54	[131]
CMS	Dijets	8	19.7	0.7	$d^3\sigma/d\langle p_t\rangle\, dY dy^*$	122	[133]

Table 4.1: The LHC inclusive jet and dijet cross-section measurements that will be used in this study. For each dataset we indicate the experiment, the measurement, the center of mass energy $\sqrt{s}$, the luminosity $\mathcal{L}$, the jet radius R, the measured distribution, the number of data points n_{dat} and the reference.

are then spelled out in Sec. 4.4, where we also present the corresponding results, and discuss their implications. Finally, in Sec. 4.5 we draw general comparative conclusions on the behaviour of different observables at different perturbative orders.

4.1 Experimental data

We now discuss the inclusive single-jet and dijet data. We summarize the available jet production data from the LHC, and we provide details on the treatment and kinematic coverage of the single-jet and dijet datasets that we will use in this study.

4.1.1 Jet production at the LHC

The ATLAS and CMS collaborations have performed a number of measurements of the single-inclusive and dijet cross-sections at different center of mass energies, ranging from $\sqrt{s} = 2.76$ TeV to 13 TeV. We will focus on the 7 and 8 TeV data, for which single-inclusive and dijet data corresponding to the same underlying dataset and integrated luminosity can be compared.

The $\sqrt{s} = 7$ and 8 TeV data are summarized in Table 4.1, where for each dataset we indicate the experiment, the measured quantity, the center of mass energy $\sqrt{s}$, integrated the luminosity $\mathcal{L}$, the number of data points n_{dat}, and the published reference.

We recall to the reader the relevant kinematic variables, some of them already introduced in chapter 2. For inclusive jets, p_t and y are the jet transverse momentum and rapidity. For dijets, m_{jj} is the dijet invariant mass, $y^* = |y_1 - y_2|/2$ and $|y_{\mathrm{max}}| = \max(|y_1|, |y_2|)$ are respectively the absolute rapidity difference and maximum absolute rapidity of the two leading jets of the event. Finally, for dijet triple-differential distributions, $\langle p_t \rangle = (p_{t1} + p_{t2})/2$ is the average transverse momentum of the two leading jets, and $Y = |y_1 + y_2|/2$ is the boost of the dijet system.

In addition to those listed in Table 4.1, ATLAS and CMS have also performed measurements at $\sqrt{s} = 13$ TeV, though so far with smaller integrated luminosities than for their Run I counterparts: at Run II, the single-inclusive jet measurements from ATLAS [136] and CMS [137] have $\mathcal{L} = 3.2$ fb^{-1} and $\mathcal{L} = 71$ pb^{-1} respectively, while the dijet measurements from ATLAS [136]

and CMS [138] have $\mathcal{L} = 3.2$ fb^{-1} and $\mathcal{L} = 2.3$ fb^{-1}. For this reason, we do not include these datasets. Very recently, CMS has presented a single-inclusive jet measurement at $\sqrt{s} = 13$ TeV, based on a luminosity of $\mathcal{L} = 35.9$ fb^{-1} [139].

In addition, ATLAS and CMS have also presented several measurements of multijet (≥ 3 jets) production. For example, ATLAS has provided measurements of the three jet cross-sections at 7 TeV [140], differential in three-jet mass and the sum of the absolute rapidity separations between the three leading jets; and of four-jet cross-sections at 8 TeV [141], differential in the p_t of the four leading jets in the event. CMS has also a measurement of the 3-jet production cross section at 7 TeV [142] differential in the invariant mass of the three jets m_{jjj}. Because theoretical predictions are currently only available up to NLO for these observables, they will not be considered here, though they are important for other applications such as the validation of Monte Carlo event generators and searches for physics beyond the Standard Model.

4.1.2 Jet data in this analysis

The single-inclusive jet data from ATLAS and CMS used in this work are the double-differential (y, p_t) distributions listed in Table 4.1. The ATLAS 7 TeV data cover the range 100 GeV $\leq p_t \leq 1.992$ TeV and $0 \leq |y| \leq 3$, while the ATLAS 8 TeV data cover the same rapidity range, but an extended range of transverse momenta, namely 70 GeV $\leq p_t \leq 2.5$ TeV. In our default fit we include only the central rapidity bin ($y_{\text{jet}} \leq 0.5$) of the ATLAS 7 TeV, due to the aforementioned issues in the covariance matrix of this data. This is not expected to affect results, as in Ref. [58] it was shown that PDFs fit to the central rapidity bin provide an equally good fit to all other rapidity bins, and in Ref. [143] it was checked explicitly that PDFs determined including each rapidity bin from this data in turn are indistinguishable.

The CMS 7 TeV data cover the range 100 GeV $\leq p_t \leq 2.0$ TeV and $0 \leq |y| \leq 2.5$, and the CMS 8 TeV data the extended range 74 GeV $\leq p_t \leq 2.5$ TeV and $0 \leq |y| \leq 3.0$. We note that in the case of the CMS 8 TeV single-inclusive jets, measurements for $p_t < 74$ GeV are also available, but these are excluded from the fit because non-perturbative and resummation corrections, not accounted for by fixed-order computations, are large at small p_t. We therefore retain only 185 points out of a total of 239.

For the dijet cross-sections we consider three Run I measurements from ATLAS and CMS, specifically the ATLAS and CMS 7 TeV [131,132] double-differential distributions and the CMS 8 TeV triple-differential distributions [133]. Note that currently ATLAS dijet measurements are only available at 7 and 13 TeV, and not at 8 TeV. The ATLAS data are double-differential in m_{jj} and $|y^*|$. The corresponding ranges are 260 GeV $\leq m_{jj} \leq 4.27$ TeV and $0 \leq y^* \leq 3.0$. The CMS 7 TeV data [131] are instead double-differential in m_{jj} and $|y_{\text{max}}|$. The ranges are 200 GeV $\leq m_{jj} \leq 5$ TeV and $0 \leq |y|_{\text{max}} \leq 2.5$. The CMS 8 TeV [133] data are triple differential in $\langle p_t \rangle$, Y, and $|y^*|$. The ranges are 133 GeV $\leq \langle p_t \rangle \leq 1.78$ TeV and $0 \leq Y, y^* \leq 3$.

For all these measurements, we will use the complete set of systematic uncertainties and correlations available from HEPDATA.

4.2 Theoretical calculations and implementation

In this section we present the available theoretical predictions for jet processes required for state-of-the-art precision phenomenology. We mainly focus on QCD correction, in keeping with Chapter 2, but we also briefly discuss electroweak (EW) corrections. We first show how NLO

QCD calculations of hadronic observables can be significantly speed up by means of fast interpolation grids. We then discuss the NNLO QCD corrections and the NLO EW corrections, provided in the form of K-factors. Finally, we show how to combine these ingredients through a multiplicative prescription.

Note that fast interpolation grids for jet processes accurate to NNLO are not yet publicly available. The NNLOJET+APPLFAST fast interpolation tables with NNLO QCD corrections are so far only available for jet production in deep-inelastic scattering [144].

4.2.1 NLO QCD fast interpolation grids

The partonic generator NLOJET++ we have used extensively in Chapter 3 is a powerful tool, but too slow for PDF fitting purposes. A standard PDF fit include thousands of hadronic data points (the fits of the present study have 4333 and 4079 data points, if we include inclusive jets or dijets measurements, respectively), and predictions has to be computed thousands of times during the minimization process, with the time required to obtain sufficiently accurate results of the order of a few hours per data point.

However, since during the fit we are only interested in varying PDFs, partonic matrix elements can be precomputed in such a way that the numerical convolution with generic input PDFs can be efficiently approximated by means of interpolation techniques. Packages such as APPLGRID [145] and FASTNLO [146], once interfaced with NLOJET++ or any other partonic Monte Carlo generator, fill PDF- and α_s-independent look-up tables of cross sections weights. Such a procedure is required to speed up the computation of hadronic observables when the fit is performed.

We can understand the basic principle behind fast interpolation grids in the following way. A total hadronic cross section (or the prediction for the specific bin of a differential distribution) can be written the same way as Eq. (2.26):

$$d\sigma_{pp} = \sum_s \sum_p \int dx_1 dx_2 \, d\hat{\sigma}^{(p)(s)} \left[\alpha_s^p(Q^2) F^{(s)}(x_1, x_2, Q^2) \right] , \qquad (4.1)$$

where we are summing over the perturbative orders p up to the accuracy required, and over the partonic subprocesses s contributing to the process. Q^2 is the typical hard scale of the process, and we have set $\mu_R^2 = \mu_F^2 = Q^2$. As usual, $d\hat{\sigma}^{(p)(s)}$ is the contribution of order α_s^p to the partonic cross section for the subprocess s.

In APPLGRID or FASTNLO, the square bracket under integration in Eq. (4.1) is represented in terms of interpolating functions, spanning the Q^2, x_1 and x_2 values:

$$\alpha_s^p(Q^2) F^{(s)}(x_1, x_2, Q^2) = \sum_{\alpha, \beta, \tau} \alpha_s^p(Q_\tau^2) F^{(s)}(x_\alpha, x_\beta, Q_\tau^2) \mathcal{I}_\tau(Q^2) \mathcal{I}_\alpha(x_1) \mathcal{I}_\beta(x_2) , \qquad (4.2)$$

so that the convolution in Eq. (4.1) is reduced to a simple product:

$$d\sigma_{pp} = \sum_s \sum_p \sum_{\alpha, \beta, \tau} \alpha_s^p(Q_\tau^2) \, F^{(s)}(x_\alpha, x_\beta, Q_\tau^2) \, W_{\alpha\beta, \tau}^{(p)(s)} \qquad (4.3)$$

where

$$W_{\alpha\beta, \tau}^{(p)(s)} = \int dx_1 dx_2 \, d\hat{\sigma}^{(p)(s)} \mathcal{I}_\tau(Q^2) \mathcal{I}_\alpha(x_1) \mathcal{I}_\beta(x_2) \qquad (4.4)$$

is the cross section weight matrix, the result of the convolution of the partonic cross section with the interpolation polynomials. If we pre-compute $W_{\alpha\beta,\tau}^{(p)(s)}$ in a look-up table, then by using Eq. (4.3) the user can *a posteriori* include PDFs, modify α_s or vary the renormalization and factorization scales by arbitrary factor ξ_R and ξ_F.

Within the NNPDF framework, there is another task which may be pre-computed. As we will discuss in Sec. 4.3, during the fit, PDFs are parametrized at a fixed scale Q_0^2, while in Eq. (4.3) PDFs are required at a generic scales Q_τ^2, so an evolution from the scale Q_0^2 to the scale Q_τ^2 is needed. By directly including the DGLAP evolution kernel — responsible for this evolution — in Eq. (4.3), after some straightforward algebra one arrives at:

$$d\sigma_{pp} = \sum_{k,l} \sum_{\gamma,\delta} \widetilde{W}_{kl,\gamma\delta} f_k(x_\gamma, Q_0^2) f_l(x_\delta, Q_0^2) . \tag{4.5}$$

Compared to Eq. (4.3), $\widetilde{W}_{kl,\gamma\delta}$ encodes more information that can be calculated before fitting, thus minimizing the number of operations required during the fit and therefore decreasing the computing time. The APFELGRID [147] package takes care of the combination between fast interpolation grids in the APPLGRID format and PDF evolution kernels provided by the APFEL package [52]. The output of APFELGRID is nothing but an implementation of $\widetilde{W}_{kl,\gamma\delta}$, in a format compliant with the NNPDF framework.

In the current study, for each dataset, we have produced fast interpolation grids, accurate to NLO in QCD, by using NLOJET++ interfaced to FASTNLO. The computation is performed with the scale choices discussed in Sec. 4.2.2. These fast interpolation grids are then combined with PDF evolution kernels using APFELGRID. However, since APFELGRID is based on grids provided in the APPLGRID format, we implemented an APPLGRID interface to the weight tables in the FASTNLO format. We provide details about this script in Appendix 4.A.

4.2.2 NNLO QCD K-factors

As we have seen in the previous chapters, single-inclusive and dijet observables display a somewhat different perturbative behaviour. Single-jet cross section is in general rather sensitive to the choice of central scale, even at NNLO, as pointed out in Sec. 2.4: the commonly used scale choices $\mu = p_t$ or $\mu = p_{t1}$ lead to predictions which even at NNLO may differ by an amount which is comparable to, or larger than, their scale dependence. Still in Sec. 2.4, we presented the findings of Ref. [121], whereby the event-based scale $\mu = \hat{H}_T$, defined in Eq. (2.43), was singled out as optimal choices. Following this prescription, here we will adopt $\mu = \hat{H}_T$ as central scale choice; results obtained with this scale choice will be compared in Sec. 4.4 to those found using $\mu = p_t$, which was the baseline choice adopted in previous NNPDF determinations, specifically NNPDF3.1. For what concerns dijets, as commented at the end of Sec. 2.4, the NNLO computation has essentially settled the issue of scale choice, with the dijet invariant mass m_{jj} emerged as the preferred choice [122]. This is the scale choice which we will adopt in the sequel.

NNLO QCD corrections computed with NNLOJET [148] will be included by supplementing theoretical predictions accurate to NLO QCD with K-factors defined as

$$K_{\mathrm{NNLO}}^{\mathrm{QCD}} \equiv \frac{\sum_{ij} \widetilde{\sigma}_{ij}^{\mathrm{NNLO}} \otimes \mathcal{L}_{ij}^{\mathrm{NNLO}}}{\sum_{ij} \widetilde{\sigma}_{ij}^{\mathrm{NLO}} \otimes \mathcal{L}_{ij}^{\mathrm{NNLO}}} , \tag{4.6}$$

where the sum runs over partonic subchannels, $\widetilde{\sigma}_{ij}$ are partonic cross sections, and $\mathcal{L}_{ij}$ the
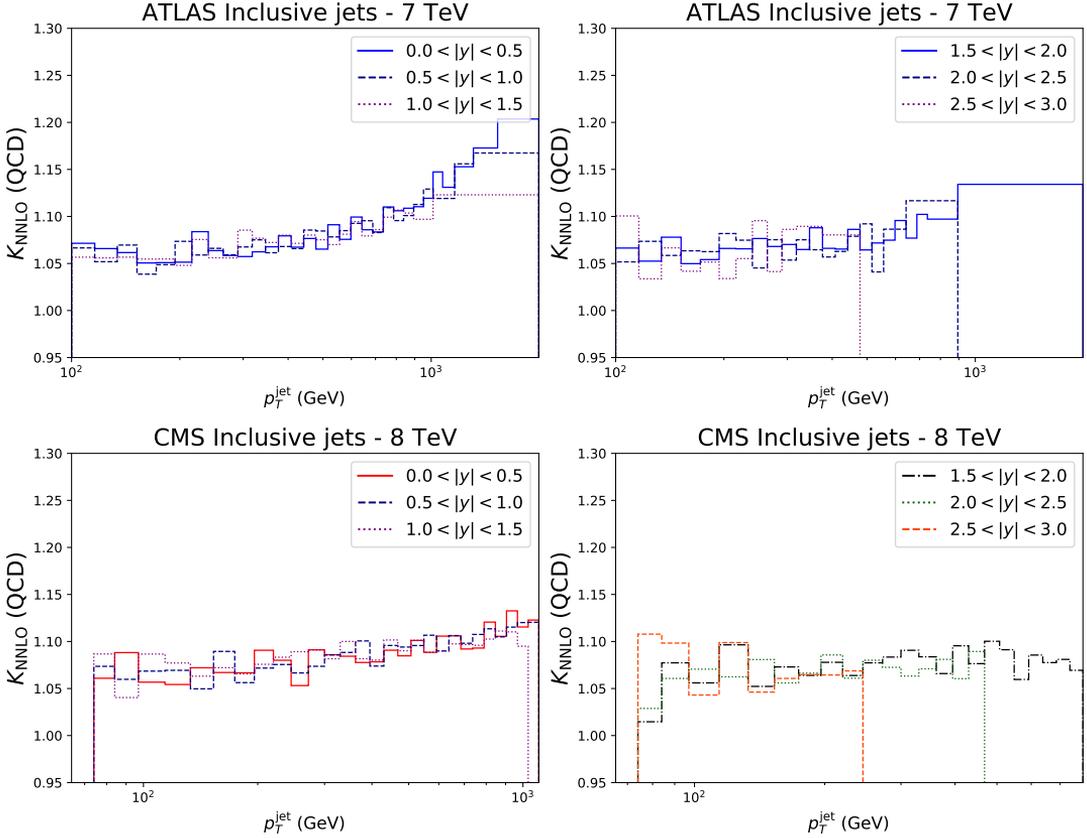
Figure 4.1: The NNLO QCD K-factors, Eq. (4.6), for the ATLAS 7 TeV (top) and CMS 8 TeV (bottom) single-inclusive jet cross sections evaluated using NNPDF3.1 PDF at the scale and $\mu = \hat{H}_T$. Results are shown as function of the jet p_t in different jet rapidity bins, with the central (forward) bins shown in the left (right) plot.

corresponding parton luminosities, computed both in the numerator and the denominator using NNPDF3.1 NNLO as a fixed input PDF set.

In Fig. 4.1 we show the NNLO QCD K-factors, Eq. (4.6), corresponding to the ATLAS 7 TeV and CMS 8 TeV inclusive jet cross sections evaluated with the NNPDF3.1 NNLO PDF set and $\mu = \hat{H}_T$ as central scale. Results are shown as a function of the jet p_t in different jet rapidity bins, with the central (forward) bins in the left (right) plot. At central rapidities, the NNLO K-factor increases monotonically with p_t from about 5% to about 20–25%. This growth with p_t becomes less marked as the jet rapidity increases: in fact at 8 TeV for $|y| \geq 1.5$ the K-factor depends only mildly on the jet p_t. The K-factors display moderate point-to-point fluctuations, especially in the forward rapidity bins.

In Fig. 4.2 we display the NNLO QCD K-factors, Eq. (4.6), computed with this scale choice and the NNPDF3.1 NNLO PDF set, for the ATLAS 7 TeV and CMS 8 TeV dijet cross sections. For ATLAS, the K-factors at small rapidity separations are somewhat below unity for low invariant masses, then grow monotonically with m_{jj} up to about $K \sim 1.15$ at the highest $m_{jj} \sim 4$ TeV. For larger rapidity separations, $1.5 \leq |y^*| \leq 3.0$, the K-factors are less sensitive to
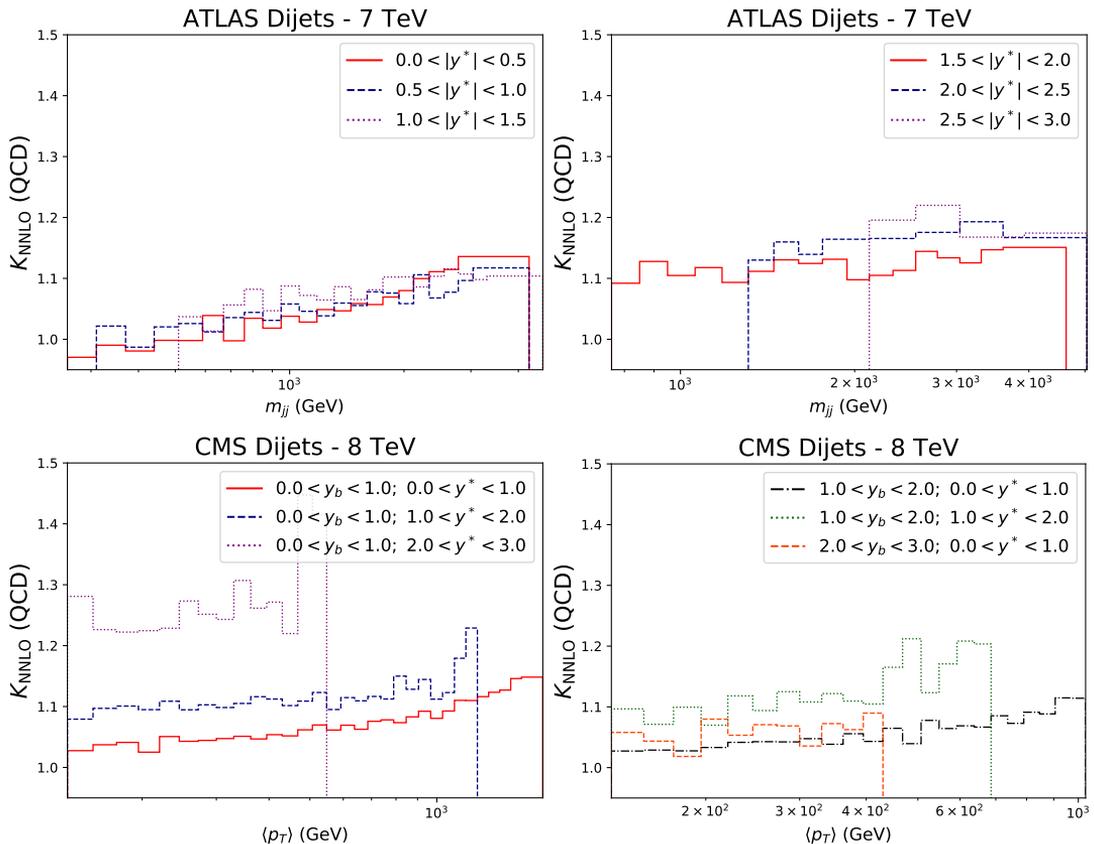
Figure 4.2: The NNLO QCD K-factors, Eq. (4.6), corresponding to the ATLAS 7 TeV (top) and CMS 8 TeV (bottom) dijet cross sections evaluated with NNPDF3.1 PDF at the scale $\mu = m_{jj}$. Results are shown as function of the jet p_t in different jet rapidity bins, with the most central (forward) bins in the left (right) plot.

m_{jj}, and their value corresponds to corrections between 10% and 20%. For CMS, as mentioned, the measurement is presented as a triple differential distribution in $\langle p_t \rangle$, y^*, and Y. As seen in Fig. 4.2, the qualitative behaviour of the K-factors is similar in all rapidity bins, and shows a monotonic growth with $\langle p_t \rangle$. However, the value depends strongly on the rapidity difference, with the K factor larger at larger y^*. For example, in the $0 \leq Y, y^* \leq 1$ bin the K-factor ranges from a few percent at low $\langle p_t \rangle$ to up to 15%, while in the $0 \leq Y \leq 1$ and $2 \leq y^* \leq 3$ bin it goes up to 25%. Also these K-factors display sizable point-to-point fluctuations.

These point-to-point fluctuations of QCD K-factors reveal an underlying numerical uncertainty. For illustration purposes, this uncertainty is displayed in Fig. 4.3 for the central rapidity bins of the ATLAS 7 TeV single jet and of the CMS 8 TeV dijet distributions. We have estimated this uncertainty through the procedure for the suppression of outliers as described in Ref. [149]. When performing PDF fits, this numerical uncertainty is added in quadrature to the experimental uncertainty, fully uncorrelated datapoint by datapoint.
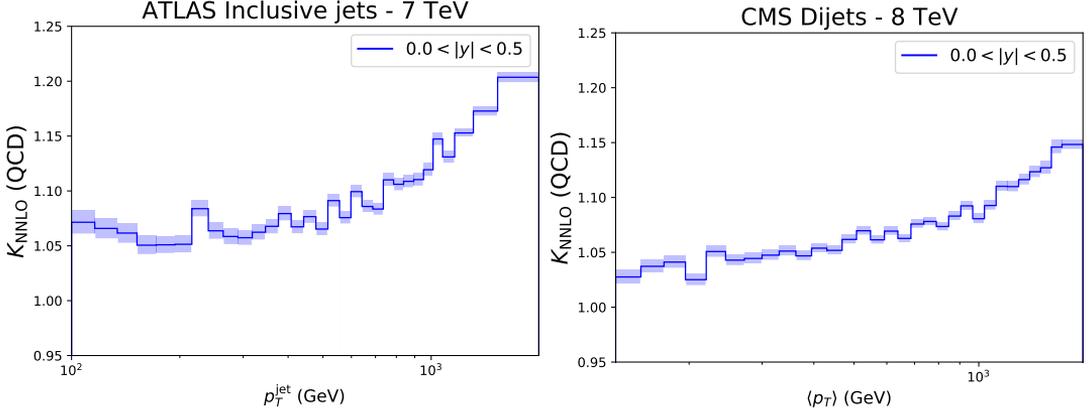
Figure 4.3: The NNLO QCD K-factors for the central rapidity bins of the ATLAS 7 TeV single-inclusive jets (left) and CMS 8 TeV dijets (right), with the Monte Carlo numerical uncertainties shown as filled bands around the central result.

4.2.3 Electroweak corrections

EW corrections for jet processes have been computed in Ref. [150]. This calculation include $O(\alpha\alpha_s)$ and $O(\alpha^2)$ tree level contributions (where α and α_s are the electromagnetic and strong couplings, respectively), and the weak radiative corrections of $O(\alpha\alpha_s^2)$. In particular, they include the virtual exchange of weak bosons that give rise to the dominant EW Sudakov logarithms, suitably combined with the respective hard QCD emissions to cancel infrared singularities. They will be collectively referred to as NLO EW corrections in the remainder of this chapter.

As done for the NNLO QCD corrections, we include EW corrections through a K-factor defined as

$$K^{\mathrm{EW}} \equiv \frac{\sum_{ij} \widetilde{\sigma}_{ij}^{\mathrm{LO\ QCD+EWK}} \otimes \mathcal{L}_{ij}^{\mathrm{NNLO}}}{\sum_{ij} \widetilde{\sigma}_{ij}^{\mathrm{LO\ QCD}} \otimes \mathcal{L}_{ij}^{\mathrm{NNLO}}} \,, \qquad (4.7)$$

where the partonic cross sections in the numerator are obtained combining the contributions computed in Ref. [150] with the LO QCD computation. The K-factor defined in Eq. (4.7) has been computed using a proprietary code [150]. Electroweak K-factors have been evaluated using consistently the NNPDF3.1 NNLO PDF set, and the same scale choice as that of the corresponding NNLO QCD predictions. Note that because of cancellations between (negative) Sudakov logarithms and (positive) subleading Born contributions, the K-factors are quite sensitive to the underlying parton decomposition, and it is consequently important to make a consistent choice of PDFs in the computation of QCD and EW K-factors.

The K-factors thus computed are shown in Fig. 4.4 for the ATLAS 7 TeV and CMS 8 TeV single-inclusive jet cross sections and for the ATLAS and CMS 7 TeV dijet cross sections. Results are shown as a function of p_t for single-inclusive jets and as a function of m_{jj} for dijets, in bins of rapidity y (single-inclusive), absolute rapidity difference y^* (ATLAS dijets) or maximum absolute rapidity y_{max} (CMS dijets). In all cases the qualitative behaviour is similar: the K-factor is close to unity for small values of p_t or m_{jj}; it is flat (in fact slightly decreasing) for large values of the rapidity variable; and it grows with respectively p_t or m_{jj} at central rapidity, the growth being stronger at smaller rapidity. The largest EW correction can reach 20% or more for transverse momenta or invariant masses in the TeV range and the smallest rapidity.
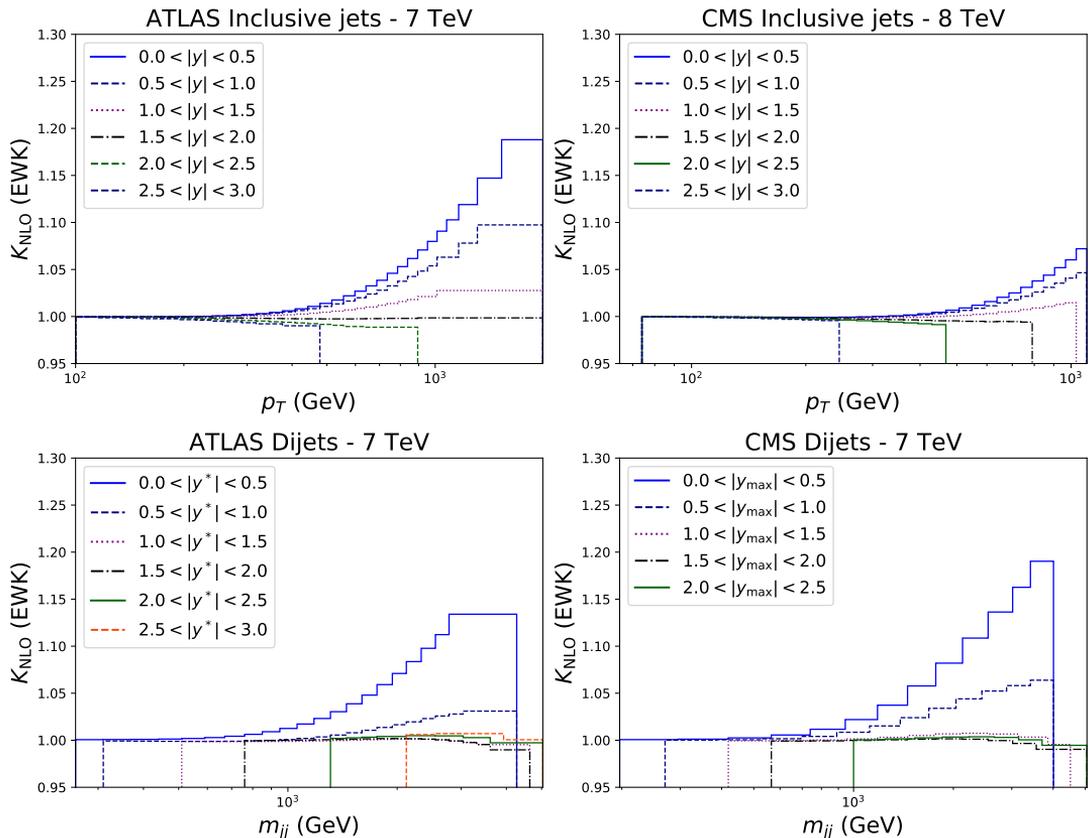
Figure 4.4: The EW K-factors, Eq. (4.7), for the ATLAS and CMS single-inclusive (top) and dijet (bottom) measurements. For single-inclusive jets the K-factors are shown as a function of jet p_t in six different rapidity bins. For dijets they are shown as a function of the dijet invariant mass m_{jj} for different y^* bins for ATLAS (left) or y_{max} bins for CMS (right).

4.2.4 Combining grids with K-factors

In the previous subsections, we have presented all the components required for state-of-the-art precision predictions: NLO QCD grids, NNLO QCD K-factors and NLO EW K-factors. In order to combine these ingredients, we adopt the following multiplicative prescription:

$$\left. \frac{d^2\sigma}{dp_t dy} \right|_{\mathrm{NNLO_{QCD}+EWK}} = \left. \frac{d^2\sigma}{dp_t dy} \right|_{\mathrm{NLO_{QCD}}} \times K_{\mathrm{NNLO}}^{\mathrm{QCD}}(p_t, y, \sqrt{s}) \times K^{\mathrm{EW}}(p_t, y, \sqrt{s}). \qquad (4.8)$$

The first term on the right-hand side of the equation is the output of the NLO computation, while the second and third terms are the bin-by-bin QCD and EW K-factors defined in Eqs. (4.6) and (4.7), respectively. If the EW K-factor is not included, Eq. (4.8) exactly reproduces the NNLO results obtained with NNLOJET.

We conclude this section by noting that the theoretical computations of single- and dijet observables are subject to non-perturbative corrections and to missing higher order uncertainties (MHOU). The former arise from the underlying event and multiple parton interactions, and are

estimated by the experimental collaborations by comparing predictions obtained from different Monte Carlo parton shower generators. In the case of all of the CMS measurements, they are provided in terms of point-by-point rescaling factors, which we apply to the data together with an additional, fully correlated, systematic uncertainty, which we estimate as the difference between the rescaled and unrescaled value of each datapoint. The estimate of MHOUs requires some care, especially for single-inclusive jets. In Chapter 3 we observed that there are accidental cancellations which occur for values of the jet radius $R \sim 0.5$ which are close to the values adopted by ATLAS and CMS, where the NLO scale dependence evaluated in a standard way is artificially small, and thus is not a good estimator of the MHOU. A more reliable estimate of the MHOU requires performing uncorrelated scale variation, as documented in Sec. 3.2.3. The inclusion of MHOU in PDF fits, though in principle possible using the formalism of Refs. [151, 152], goes beyond the scope of this analysis, and we will not consider it further.

4.3 Principles of NNPDF methodology

In this section we briefly introduce the main components of a NNPDF fit. The interested reader can find more information in Refs. [153–155]. In a nutshell, the NNPDF methodology is based on a Monte Carlo treatment of experimental data, with PDFs parametrized by means of neural networks (NN). NN are introduced with the specific purpose of being as agnostic as possible about the assumed parametrization of PDFs, thus reducing the associated bias. Indeed, NN feature a great number of parameters, and are able to reproduce the behaviour of a very wide class of functions (more details below). As for the Monte Carlo method, the final aim is to provide a faithful representation of uncertainties on parton distributions, or in other words, to devise an uncertainty with statistical interpretation in the PDF space of functions, which is *a priori* infinite-dimensional. A parametrization as the one introduced in Eq. (1.88), with a handful of fit parameters $\vec{a}$, reduces this infinite-dimensional space to a finite space of parameters, and one can then apply standard error propagation techniques on $\vec{a}$. However, in case of NN, where the number of parameters is deliberately redundant, a Monte Carlo approach offers a more natural solution, and in addition does not rely on linear error propagation.

4.3.1 Monte Carlo replicas

Specifically, the Monte Carlo approach consists in the generation of artificial data replicas, starting from the original experimental data points. Experimental collaborations usually provide data in the form of a set measured central value O_i, with $i = 1, \ldots, N_{\text{dat}}$, supplemented by the experimental covariance matrix cov_{ij} between each pair of data points i and j:

$$\text{cov}_{ij} = \delta_{ij} \left(\delta O_i^{\text{stat}} \right)^2 + \sum_{\alpha=1}^{N_A} \delta O_{i,\alpha}^{\text{sysA}} \delta O_{j,\alpha}^{\text{sysA}} + O_i O_j \sum_{\beta=1}^{N_M} \delta O_{i,\beta}^{\text{sysM}} \delta O_{j,\beta}^{\text{sysM}}, \tag{4.9}$$

There are several kind of uncertainties appearing in the previous equation:

- δO_i^{stat} is sum in quadrature of the statistical uncertainty and of the uncorrelated systematic uncertainties;

- $\delta O_{i,\alpha}^{\text{sysA}}$ are the additive correlated uncertainties;

- $\delta O_{i,\beta}^{\text{sysM}}$ are the multiplicative (correlated) uncertainties.

The Monte Carlo replicas are generated accordingly to multi-Gauss distribution centered on the vector of central values O_i and with variances based on experimental covariance matrix cov_{ij}. The total number of Monte Carlo replicas, if large enough, reproduce the statistical properties of the original data set at the desired accuracy.

4.3.2 PDF parametrization

PDFs are parametrized at a reference scale Q_0^2 (usually 1–2 GeV). The result of the fit will not depend on this choice, as DGLAP equations evolves this input parametrization to the energy of each experimental data point. The parametrization adopted for the PDF with flavour i is[1]:

$$f_i(x, Q_0^2) = A_i\, x^{-\alpha_i}\, (1-x)^{\beta_i}\, \text{NN}_i(x)\,. \tag{4.10}$$

The core of Eq. (4.10) lies in the NN_i term, which encodes the neural network, or more precisely a feedforward multi-layer perceptron. The reader not acquainted with NNs can find a pedagogical introduction in Ref. [156]. Here, suffice it to say that a sufficiently large neural network (i.e. with enough parameters) is able to reproduce the behaviour of a very wide class of functions without the need to adjust the form of the parametrization. A NN is then a functional form with a much greater flexibility than a polynomial form, reducing the bias associated to the specific parametrization chosen. For what concerns the other terms present in Eq. (4.10), the factor $x^{-\alpha_i}\, (1-x)^{\beta_i}$ is often present in traditional PDF parametrizations, see Eq. (1.88), but its purpose here is only to speed up the minimization process. The exponents α_i and β_i are randomly chosen within a suitable range and assigned to each replica. Finally, A_i is an overall normalization constant, calculated at each iteration of the fit, which enforces QCD sum rules.

4.3.3 Fitting procedure

The minimization procedure is run in parallel for all of the artificial replicas of experimental data. Given a replica r, we first initialize the PDFs, by randomly choosing the neural network weights. Then we compute the theoretical prediction for the i-th data point, $T_i^{(r)}$, by performing the convolution between the interpolation grid and the PDFs. At this point, we can compare the result obtained for $T_i^{(r)}$ with the replica central value $O_i^{(r)}$ for the i-th data point, evaluating the loss function defined as

$$\chi^{2,(r)} = \frac{1}{N_{\text{dat}}} \sum_{i,j}^{N_{\text{dat}}} \left(O_i^{(r)} - T_i^{(r)} \right) \left. \text{cov}_{ij}^{-1} \right|_{t_0} \left(O_j^{(r)} - T_j^{(r)} \right)\,. \tag{4.11}$$

Here, cov_{ij} is the experimental[2] covariance matrix, as defined in Eq. (4.9), and the subscript t_0 denotes that it has been evaluated by following the t_0 prescription [157] (this is required in case of multiplicative uncertainties, to avoid systematic biases). Within a given replica, we adopt a genetic algorithm as optimizer. It is based on the generation of "mutants", where a mutant is a new set of neural networks with weights altered by mutations. Starting with the initial neural networks, $n = 80$ mutants are generated. The best fit mutant i.e. the mutant with the smallest

---

[1]Actually, the fit basis for the PDFs is different from the physical one i.e. f in Eq. (4.10) is linear combination of the quark, antiquark and gluon PDFs. These two basis are related by a rotation in the flavour space.

[2]Note that, in principle, this covariance matrix may include as well a theory covariance matrix added in quadrature, due to the uncertainties associated to missing higher order in perturbation theory [151, 152].

$\chi^{2,(r)}$ value, is selected, while the other mutants are discarded. The best fit mutant acts as basis for the 80 mutants of the next generation, and the algorithm is iterated for a fixed number (30k) of times. In order to avoid overfitting, the set of experimental data is randomly split into a training and into a validation set. The minimization is performed only on the training set, but the loss function is evaluated on the validation set. At the end, with a simple *look-back* method, the best configuration of weights yielding the lowest value of $\chi^{2,(r)}$ is selected as the final result of the training for the replica r.

Finally, once we have fit results for the whole set of Monte Carlo replicas, we perform some quality tests. We first discard replicas yielding a too unlikely value of χ^2. We also discard replicas non passing some positivity constraints. As it is well known, beyond LO, PDFs loose a probabilistic interpretation and do not need to be positive definite; however, through convolution with partonic cross sections, PDFs enter in the calculation of physical cross section, which of course should be positive. The "generalized" positivity of the PDFs is imposed by means of pseudo-observables, which not are measurable in practice, but are required to be positive for reasons of principle. Such pseudo-observables are present as Lagrange multipliers during the fit i.e. the χ^2 is large in case of negative values for their predictions.

The ensemble of PDFs passing the post-fit tests is the final result of the parton determination. At this point, we can compute the mean value and the standard deviation of this ensemble, and provide the PDFs with the corresponding one-sigma uncertainties. Unlike in a Hessian approach — where we estimate the covariance matrix of the fitting parameters as the (Hessian) matrix of second derivatives around the minimum of Eq. (4.11) — the Monte Carlo method does not rely on linear error propagation. However, we can check at the PDF level whether the one-sigma uncertainty bands correspond to the 68% confidence level intervals, as is the case for a Gaussian distribution i.e. whether, given $N = 100$ replicas, 68 replicas lie in the one-sigma uncertainty band around the central value. This test has been performed [71] and the result confirm this expectation, apart from the regions near the kinematical limits, where there are significant deviation from gaussianity, due to the positivity constraints imposed on the PDFs.

Note that the fitting code and the part of the methodology has been recently revised [158], in order to exploit the new technologies and algorithms provided by the machine learning community. The NNPDF regression framework has been rewritten in Python and make use of modern ML libraries, such as Keras [159] and Tensorflow [160]. The genetic algorithm has been replaced by gradient descent methods, which lead to more stable replicas and, combined with the optimized code, reduce the computational time by a considerable amount (~ 20 faster on average). In addition, a systematic scan of NN hyperparameters (e.g. number of layers, activation functions, learning rate etc.) is now possible, with the aim of further reducing possible systematic biases related to the choice of the NN model itself.

4.4   Results

We now present our main results. They consist of a set of global PDF determinations, in which the NNPDF3.1 global dataset is supplemented by the single-inclusive jet and inclusive dijet data presented in Sec. 4.1.2: by comparing fit results, we study the impact of varying the jet observable, the data, and the theory settings. Specifically, we have performed fits including either single-inclusive or dijet data, in each case either at 7 TeV or both at 7 and at 8 TeV, and with theory at pure NLO QCD, pure NNLO QCD, or NNLO QCD supplemented by EW

	NNLO$_{\text{QCD}}$+EW	NNLO$_{\text{QCD}}$	NLO$_{\text{QCD}}$
baseline (see text)	—	bn	b
ATLAS & CMS jets 7-8 TeV	janw	—	—
ATLAS & CMS jets 7 TeV	j7nw	j7n	j7
ATLAS & CMS jets 7 TeV ($\mu = p_t^{\text{jet}}$)	—	j7n-pt	j7-pt
ATLAS & CMS jets 8 TeV	j8nw	j8n	j8
ATLAS & CMS dijets 7-8 TeV	danw	—	—
ATLAS & CMS dijets 7 TeV	d7nw	d7n	d7
CMS dijets 8 TeV	d8nw	d8n	d8

Table 4.2: The PDF determinations discussed in this study and their IDs. Each row corresponds to a different input jet dataset or fit settings (listed in the first column), and each column corresponds to a different theory accuracy (listed in the first row). The ID encodes the process used (j for single inclusive jets and d for dijets); the data used (a for all, 7 or 8 for the 7 TeV or 8 TeV datasets); the perturbative accuracy (n for QCD NNLO, w if EW corrections included); the choice of scale (pt when $\mu = p_t^{\text{jet}}$).

corrections as discussed in Sec. 4.2.3. For the single-inclusive 7 TeV data we have also performed fits with alternative choices of central scale. ,

We will first present in Sec. 4.4.2 all PDF sets based on single-inclusive data, including variations of scale choice, then in Sec. 4.4.3 PDF sets based on inclusive dijet data.

4.4.1 PDF sets

The inclusion of jet data in a global NNPDF3.1-like PDF determinations essentially impacts only the gluon PDF, as was shown in Ref. [58], while leaving other PDFs essentially unchanged. The correlation between each data point and the gluon PDF may be quantified by computing the correlation coefficient [153, 161], defined as:

$$\rho(\mathcal{F}, g) = \frac{N_{\text{rep}}}{N_{\text{rep}} - 1} \frac{\langle \mathcal{F} g \rangle_{\text{rep}} - \langle \mathcal{F} \rangle_{\text{rep}} \langle g \rangle_{\text{rep}}}{\sigma_{\mathcal{F}} \sigma_g} , \qquad (4.12)$$

where g is the gluon PDF, $\mathcal{F}$ is the prediction for each data point (dependent on g), and the average is taken over the set of N_{rep} Monte Carlo replicas. The denominator is the product of the standard deviations of the two ensembles. By computing such a coefficient, using the default baseline NNLO PDF set (before inclusion of the jet data, #bn, see Table 4.2 below), correlations are seen to be large or very large (up to almost one) for all $x \gtrsim 10^{-2}$.

The full list of PDF determinations that we will discuss is given in Table 4.2, together with an ID that will be used to identify them. Each row corresponds to a different choice of dataset or methodological settings, while columns correspond to the theory adopted: QCD at NLO or NNLO, without or with EW corrections included. By "NLO" or "NNLO" we mean that we incorporate jets predictions at NLO or NNLO in a global QCD fit in which everything is consistently done at NLO or NNLO respectively. Instead, by "EW" we mean that EW corrections are possibly included for jet predictions only.

The jet data of Table 4.1 are added to a baseline dataset, which essentially coincides with the NNPDF3.1 dataset. This dataset includes: fixed-target neutral-current (NC) DIS structure function data; charged-current (CC) DIS structure function data; fixed-target Drell-Yan data; collider Drell-Yan data; inclusive gauge boson and top-pair production data. In total this baseline dataset contains $n_{\mathrm{dat}} = 3813$ datapoints, see Ref. [58] for more details. The number of datapoints corresponding to the jet data included in the various fits of Table 4.2 is given in Table 4.3 and in Table 4.4 below.

In all of these fits, experimental systematic uncertainties are fully correlated across bins of different kinematic variables, while statistical uncertainties coming from the unfolding are correlated only across bins of transverse momentum (for jets) or invariant mass (for dijets), but not across rapidity bins. For jet or dijet data, non-perturbative corrections are included by default, as are Monte Carlo uncertainties due to finite numerical precision of NNLO QCD K-factor computations (see Sec. 4.2 for details). The factorization and renormalization scales are by default taken to be $\mu = \hat{H}_T$ for single-inclusive jets, and $\mu = m_{jj}$ for dijets (see the discussion in Sec. 4.2). An alternative choice of scale for single-inclusive jets will be considered in Sec. 4.4.2 below.

All the fits listed in Table 4.2 otherwise closely follow the NNPDF3.1 analysis [58]. Specifically, the same settings and codes are used for the computation of physical observables in the baseline dataset, and the same choice of kinematic cuts, of values of physical parameters, and of fitting methodology are adopted. All PDF sets include N_{rep}=100 Monte Carlo replicas. The REPORTENGINE software [162] is used in the sequel to analyze each fit and compute various fit metrics. Specifically, we will consider:

- the χ^2 of the theory prediction for each dataset or combinations of datasets, defined according to Eqs. (4.11),(4.9);

- the distance d between two fits f and h, defined as the absolute difference of the PDF central values in units of the standard deviation of the mean:

$$d = \frac{|\langle f \rangle - \langle h \rangle|}{\sqrt{\sigma_f^2 + \sigma_h^2}} \,. \tag{4.13}$$

For instance, for a sample of 100 replicas, one expect the distance to fluctuate around $d \sim 1$ if the PDFs are statistically identical (replicas extracted from the same underlying distribution). On the other hand, $d \sim \sqrt{N_{\mathrm{rep}}} = 10$ corresponds to PDFs that differ by one standard deviation.

The values of the χ^2 per datapoint for all fits with default settings at NLO and NNLO with or without EW corrections and single-inclusive jet or dijet data are collected in Tab. 4.3 and in Tab. 4.4 respectively; χ^2 values are shown for all data in the global dataset, grouped by process kind (DIS NC, DIS CC, Drell-Yan, $Z\ p_t$, top pair) and for all jet data, both those which are and those which are not included in each fit. The values of χ^2 per datapoint for all jet data (included or not included) for all fits performed with alternative choices of central scale are collected in Tab. 4.5. In these tables, χ^2 values corresponding to data not included in each fit are enclosed in square brackets.

Dataset	n_{dat}	b	bn	janw	j7	j7n	j7nw	j8	j8n	j8nw
DIS NC	2103	1.17	1.17	1.18	1.17	1.18	1.17	1.17	1.17	1.18
DIS CC	989	1.06	1.10	1.11	1.06	1.11	1.10	1.08	1.11	1.11
Drell-Yan	577	1.35	1.33	1.30	1.35	1.31	1.31	1.34	1.31	1.31
$Z\ p_T$	120	1.84	1.01	1.02	1.85	1.02	1.02	1.89	1.03	1.03
Top pair	24	1.10	1.05	1.25	1.09	1.06	1.02	2.00	1.61	1.24
ATLAS $\sigma_{t\bar{t}}$	3	2.02	0.90	0.70	1.68	0.74	0.72	1.70	0.79	0.78
ATLAS $t\bar{t}$ rap	9	1.12	1.22	2.01	1.25	1.38	1.31	2.93	2.78	1.96
CMS $\sigma_{t\bar{t}}$	3	0.53	0.22	0.21	0.42	0.24	0.31	0.34	0.17	0.19
CMS $t\bar{t}$ rap	9	0.98	1.17	0.98	0.96	1.09	1.04	1.65	1.12	0.99
Jets (all)	520	[1.48]	[2.60]	1.88	[1.86]	[2.45]	[2.53]	[1.20]	[1.75]	[1.89]
Jets (fitted)		—	—	1.88	0.79	1.15	1.12	1.40	2.05	2.20
ATLAS 7 TeV	31	[1.26]	[1.87]	1.59	1.12	1.73	1.15	[1.07]	[1.69]	[1.62]
ATLAS 8 TeV	171	[2.60]	[5.01]	3.22	[3.55]	[4.76]	[4.58]	2.03	3.18	3.25
CMS 7 TeV	133	[0.60]	[1.06]	1.09	0.71	1.01	1.11	[0.72]	[0.94]	[1.14]
CMS 8 TeV	185	[1.10]	[1.59]	1.25	[1.24]	[1.47]	[1.80]	0.81	1.01	1.23
Dijets (all)	266	[3.49]	[3.07]	[2.10]	[4.16]	[2.96]	[2.56]	[3.34]	[2.21]	[2.22]
Dijets (fitted)		—	—	—	—	—	—	—	—	—
ATLAS 7 TeV	90	[1.49]	[2.47]	[1.95]	[1.77]	[2.46]	[1.97]	[1.43]	[2.28]	[2.01]
CMS 7 TeV	54	[2.06]	[2.40]	[2.08]	[2.43]	[2.50]	[2.12]	[1.65]	[2.00]	[2.15]
CMS 8 TeV	122	[5.60]	[3.81]	[2.21]	[6.70]	[3.53]	[3.20]	[5.48]	[2.26]	[2.39]
Total		1.20	1.18	1.28	1.17	1.17	1.17	1.39	1.27	1.27

Table 4.3: The χ^2 per datapoint for all fits of Table 4.2 including single-inclusive jet data, with default settings. Results are shown for all datasets, aggregated by process type. For jets, results are shown both for the data included in each fit, and also for those not included, enclosed in square brackets. Combined results are also shown for all single-inclusive jet and for all dijet data, both for the full set, and for those included in each fit. The number of datapoints in each dataset is also shown.

4.4.2 Single-inclusive jets

We first present PDF sets obtained by including single-inclusive jet data. We discuss in turn the impact and consistency of individual datasets; perturbative QCD stability and the impact of EW corrections; the choice of central scale.

Impact and consistency of datasets

We provide a general comparative assessment of the impact of single-inclusive jet data on PDFs by comparing fits performed with the default theory settings of Sec. 4.2 and the highest theory accuracy, i.e. NNLO QCD theory used throughout in the fit, and EW corrections included in the jet predictions. According to the data included, these correspond to the fits #bn, #janw, #j7nw, and #j8nw of Table 4.2.

First, we compare fit #janw, that contains all of the single-inclusive jet data, to the baseline

Dataset	n_{dat}	b	bn	danw	d7	d7n	d7nw	d8	d8n	d8nw
DIS NC	2103	1.17	1.17	1.18	1.17	1.17	1.17	1.21	1.18	1.18
DIS CC	989	1.06	1.10	1.12	1.07	1.09	1.09	1.11	1.11	1.12
Drell-Yan	577	1.35	1.33	1.29	1.36	1.33	1.32	1.32	1.28	1.28
$Z\ p_T$	120	1.84	1.01	1.07	1.85	1.03	1.03	2.06	1.07	1.08
Top pair	24	1.10	1.05	1.14	1.16	1.06	1.04	1.57	1.34	1.26
ATLAS $\sigma_{t\bar{t}}$	3	2.02	0.90	0.66	1.79	0.74	0.73	0.80	0.68	0.69
ATLAS $t\bar{t}$ rap	9	1.12	1.22	1.57	1.26	1.34	1.32	2.41	2.02	1.82
CMS $\sigma_{t\bar{t}}$	3	0.53	0.22	0.53	0.48	0.29	0.28	0.01	0.74	0.67
CMS $t\bar{t}$ rap	9	0.98	1.17	1.04	1.07	1.09	1.07	1.42	1.04	1.04
Jets (all)	520	[1.48]	[2.60]	[2.06]	[1.62]	[2.75]	[2.70]	[1.42]	[1.94]	[2.14]
Jets (fitted)	—	—	—	—	—	—	—	—	—	—
ATLAS 7 TeV	31	[1.26]	[1.87]	[1.63]	[1.26]	[1.86]	[1.74]	[1.00]	[1.70]	[1.61]
ATLAS 8 TeV	171	[2.60]	[5.01]	[3.36]	[2.62]	[4.80]	[4.65]	[2.18]	[3.30]	[3.55]
CMS 7 TeV	133	[0.60]	[1.06]	[1.06]	[0.71]	[1.13]	[1.14]	[0.77]	[0.97]	[1.07]
CMS 8 TeV	185	[1.10]	[1.59]	[1.64]	[1.42]	[2.16]	[2.17]	[1.27]	[1.41]	[1.68]
Dijets (all)	266	[3.49]	[3.07]	1.65	[3.03]	[2.21]	[2.16]	[2.38]	[1.74]	[1.71]
Dijets (fitted)		—	—	1.65	1.33	1.79	1.72	3.69	1.59	1.68
ATLAS 7 TeV	90	[1.49]	[2.47]	1.76	1.20	1.94	1.78	[1.04]	[1.96]	[1.78]
CMS 7 TeV	54	[2.06]	[2.40]	1.60	1.54	1.55	1.63	[1.67]	[1.70]	[1.66]
CMS 8 TeV	122	[5.60]	[3.81]	1.58	[5.03]	[2.70]	[2.67]	3.69	1.59	1.68
Total		1.20	1.18	1.22	1.33	1.20	1.19	1.33	1.20	1.20

Table 4.4: Same as Table 4.3, but now for dijets. The baseline is repeated for ease of reference.

#bn, which does not include any jet data. Note that, as discussed in Sec. 4.1.2, in our default global dataset only the central rapidity bin of the ATLAS 7 TeV data is included. In Fig. 4.5 we display the distance between the PDF central values for the two fits, and the gluon PDF in both fits, normalized to the baseline, both at $Q = 100$ GeV. From Table 4.3, we note that individual jet datasets are well described (with χ^2 per datapoint of order one), except the 8 TeV ATLAS data ($\chi^2 = 3.22$), to be investigated in greater detail below. In comparison to the baseline fit, the inclusion of the single-inclusive jet data leads to a slight deterioration in the description of the ATLAS top pair rapidity distributions, whose χ^2 per datapoint increases from 1.22 to 2.01. On the other hand, it leads to an improvement in the description of the dijet data, especially the 8 TeV CMS data, which are not included in any of these fits. This suggests that the inclusion of single-inclusive and dijet data have a similar impact on PDFs, as we shall also see in Sec. 4.4.3 and discuss in greater detail in Sec. 4.5 below.

As mentioned above, and as it is clear from the distance plot in Fig. 4.5, single-inclusive jet data only have an impact on the gluon. The regions which are most affected are $x \simeq 0.05$, $0.1 \lesssim x \lesssim 0.2$, and $0.3 \lesssim x \lesssim 0.5$ in these regions the gluon PDF changes by up to slightly more than half of one sigma. In comparison to the baseline, the central gluon PDF is suppressed by about 2% in the small x region and enhanced by about 4% in the large x regions, though it always remains within the uncertainty band of the baseline.
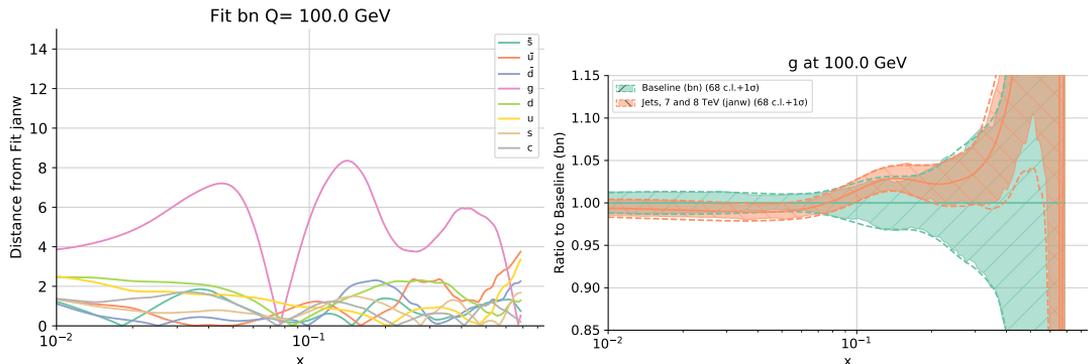
Figure 4.5: Comparison between the baseline fit with no jet data (#bn) and the fit with all single-inclusive jet data included (#janw); both with default settings and the most accurate theory (NNLO QCD, including EW corrections for jets). The distance (see text) between all PDFs (left) and the ratio of the gluon PDF to the baseline (right) are shown at the scale $Q = 100$ GeV. The shaded band is the 68% confidence interval, while the outer dashed line is the one sigma interval.
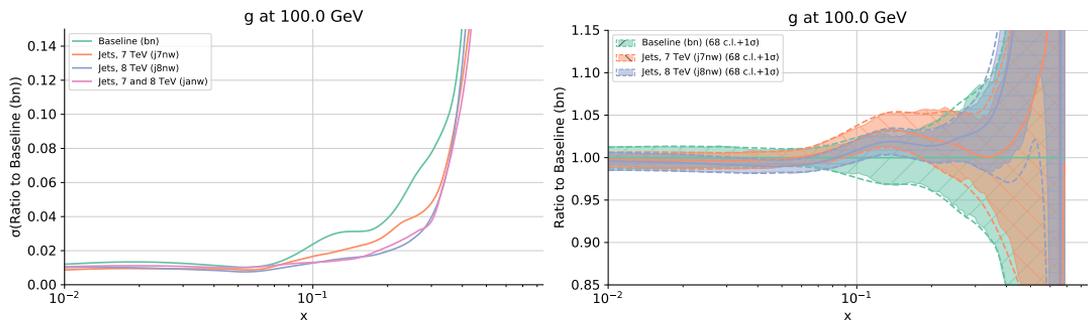

Figure 4.6: Comparison between the baseline fit with no jet data (#bn), and the fits with only 7 TeV (#j7nw) or only 8 TeV (#j8nw) jet data included. The relative uncertainty on the gluon PDF (left) and the ratio of the gluon PDF to the baseline (right) are shown at $Q = 100$ GeV. All results are shown as ratios to the baseline.

We next assess the relative impact of different jet datasets, by adding to the comparison of the baseline (#bn) and the fit with all single-inclusive jet data (#janw) also fits in which only 7 TeV (#j7nw) or 8 TeV (#j8nw) jet data are included, all with the same settings (NNLO QCD+EW). The comparison is shown in Fig. 4.6, where we compare the gluon and its relative uncertainty. Here and henceforth, when comparing relative uncertainties, the uncertainties shown are computed as a ratio to a common baseline, i.e. the plot displays all uncertainties as a percentage of the same reference fit. From Table 4.3, we note that the unsatisfactory description of the ATLAS 8 TeV data persists even when the 7 TeV data are not included in the fit, and the deterioration in fit quality for the ATLAS top data in the global fit is also similar. On the other hand, the fit in which only 7 TeV are included shows excellent fit quality both for the jet data and the global dataset. This suggests tension between the ATLAS 8 TeV data and the rest of the global dataset.

The relative pull of the jet datasets at 7 TeV and 8 TeV can be inferred from Fig. 4.6. They
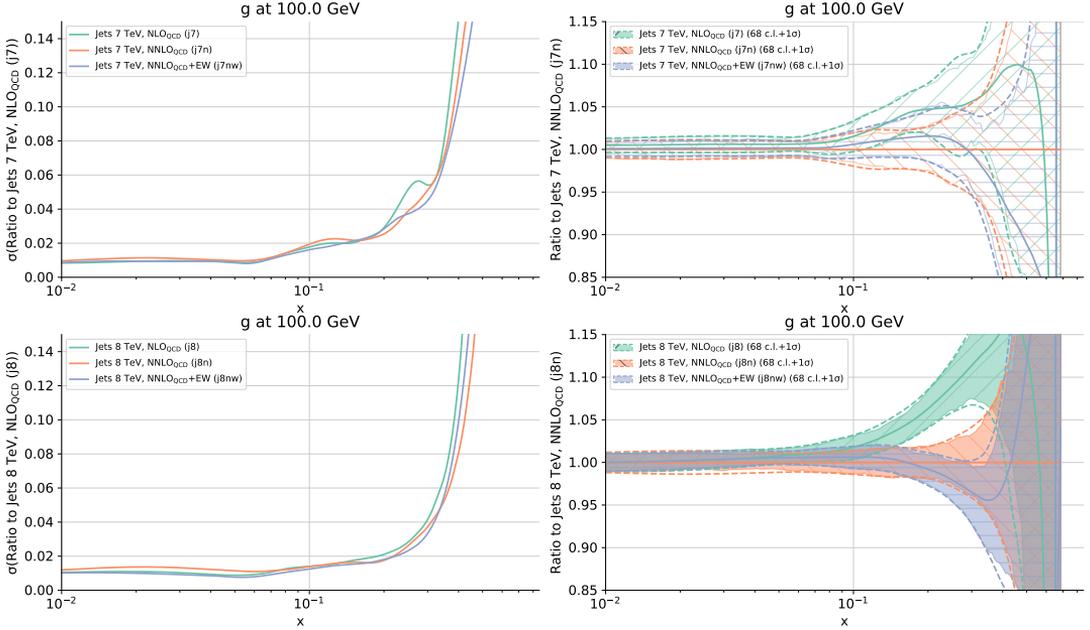
Figure 4.7: Same as Fig. 4.6, but now comparing fits with NLO, NNLO and NNLO+EW theory for 7 TeV (top: fits #j7, #j7n, #j7nw respectively) and 8 TeV (bottom: fits #j8, #j8n, #j8nw). All results are now shown as ratios to the NNLO fit.

both lead to a comparable suppression of the gluon PDF of about 1% in the region $0.3 \lesssim x \lesssim 0.5$, while they respectively enhance it by 4% and 2% in the region $0.1 \lesssim x \lesssim 0.2$. However, the decrease in gluon uncertainty is rather more marked upon inclusion of the 8 TeV data, and in fact, results obtained including all jet data, or only 8 TeV are almost identical. Specifically, in comparison to the baseline, inclusion of the 8 TeV data results in a reduction of the relative gluon uncertainty at $x \simeq 0.2$ from 4% to 1.5%, to be compared to the reduction 4% to 3% when the 7 TeV data are included. A similar behaviour was observed in the recent CT18 global PDF determination [57], which includes the ATLAS and CMS jet datasets at 7 TeV and the CMS jet dataset at 8 TeV.

Impact of higher-order QCD and EW corrections

Having assessed the impact of various single-inclusive jet data on PDFs with optimal theory settings, we now turn to the assessment of the perturbative stability of results. To this purpose, we compare fits at NLO, NNLO and with EW corrections (included for jet data only), separately for the 7 TeV and 8 TeV data, i.e. respectively, fits #j7, #j7n, #j7nw; and #j8, #j8n, #j8nw. The gluon PDFs at $Q = 100$ GeV and their uncertainty for these fits are compared in Fig. 4.7.

It is clear from the figure that, both for 7 TeV and 8 TeV data, at NLO the gluon undergoes a significant distortion in the region $0.1 \lesssim x \lesssim 0.5$ in comparison to the NNLO results shown in Fig. 4.6. Specifically, at the peak, $x \sim 0.3$ the NLO gluon turns out to be by 30-40% larger than the baseline. This effect is driven by the jet data: we have verified that in the baseline (without jet data) the NLO gluon does show some distortion in comparison to the NNLO baseline, but by a much smaller amount, with the largest enhancement of order 5%. This is thus evidence

Dataset	n_{dat}	j7	j7-pt	j7n	j7n-pt
ATLAS jets 7 TeV	31	1.12	1.13	1.13	1.13
ATLAS jets 8 TeV	171	[3.55]	[3.93]	[4.76]	[4.99]
CMS jets 7 TeV	133	0.71	0.91	0.95	0.94
CMS jets 8 TeV	185	[1.24]	[1.16]	[1.47]	[1.81]
ATLAS dijets 7 TeV	90	[1.77]	[1.98]	[2.46]	[2.55]
CMS dijets 7 TeV	54	[2.43]	[2.52]	[2.50]	[2.57]
CMS dijets 8 TeV	122	[6.70]	[7.48]	[3.53]	[3.89]

Table 4.5: Same as Table 4.3 for fits performed with alternative choices of central scale. Now only χ^2 values for jet data are shown. Results for the fits with default settings #j7 and #j7n, already shown in Table 4.3 are included for ease of reference.

for large missing NNLO corrections to the single-inclusive jet cross section in the NLO fit. The effect is more pronounced for the 8 TeV data, which can be understood as a consequence of their greater precision.

The effect of EW corrections is rather more moderate, with the shift of the central value always within the NNLO uncertainty band. Also, EW corrections seem to have an opposite effect when added to the fit to the 7 TeV or the 8 TeV data, leading to a slight enhancement of the gluon in the former case and a significant suppression in the latter for $x \gtrsim 0.2$. For both datasets, the uncertainty on the gluon for $x \gtrsim 0.1$, where the jet data have an impact, is reduced by a non-negligible amount by the inclusion of NNLO corrections. On the other hand the impact of the EW corrections is less clear. All this suggests that NNLO corrections have a significant impact by affecting the best-fit large-x gluon shape and improving its precision, while the impact of EW corrections is minor, and not clear-cut.

The effect of the inclusion of the NNLO and EW corrections on fit quality is less clear. Indeed, from Table 4.3, we observe that generally the fit quality to jet data deteriorates somewhat upon inclusion of NNLO corrections, and a little more upon inclusion of EW corrections. On the other hand, the global fit quality, as measured by the total χ^2, is unchanged for the 7 TeV data, and it improves significantly, from 1.39 to 1.27, for the more precise 8 TeV data, with the improvement mostly driven by the top and Z p_t data which are most sensitive to the gluon. However, as already noted in Sec. 4.4.2, the χ^2 of the top data deteriorates when adding the jet data to the baseline, and the fit quality to the ATLAS 8 TeV data remains unchanged. This suggests that, for the more precise 8 TeV data, the NNLO corrections reduce a tension between top and jets (especially ATLAS).

In summary, we conclude that, NNLO corrections have a sizable impact on single-inclusive jets, and in particular their inclusion leads to a reduction of the uncertainty on the large-x gluon PDF and an improved consistency of the jet data with the rest of the global dataset, demonstrated by a reduction of the shift of the gluon central value upon inclusion of jets, and as an improvement of the global χ^2 (for the more precise 8 TeV jet data), when going from NLO to NNLO. Electroweak corrections do not appear to lead to improvements either in terms of fit quality or PDF uncertainty.
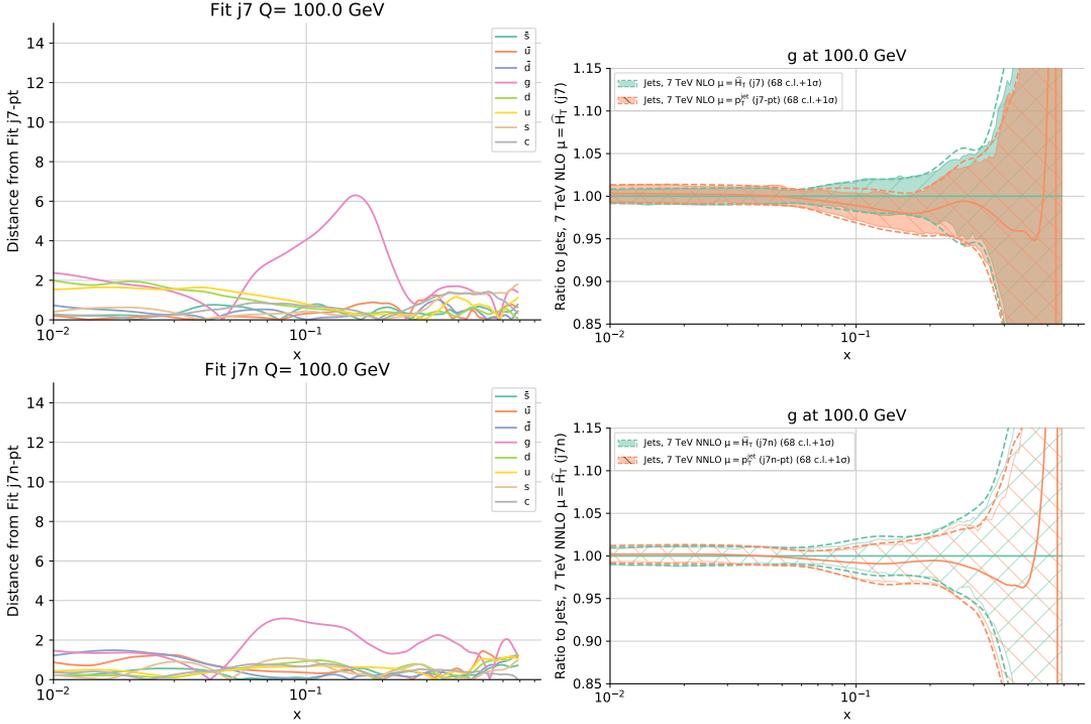
Figure 4.8: Same as Fig. 4.5, but now comparing fits to the 7 TeV data with the choices of central renormalization and factorization scale scale $\mu = \hat{H}_T$ (as shown in Fig. 4.7, top) and $\mu = p_t^{\text{jet}}$ at NLO (fits #j7 and #j7-pt) and NNLO (fits #j7n and #j7n-pt). The gluon is shown as ratio to the fits with $\mu = \hat{H}_T$.

Impact of the choice of scale

We now turn to an assessment of the impact of the choice of central scale: specifically, we compare results obtained by fitting with our default scale choice $\mu = \hat{H}_T$ and with the scale choice $\mu = p_t^{\text{jet}}$. The comparison is performed for fits to the 7 TeV data (fits #j7 and #j7-pt at NLO, and #j7n and #j7n-pt at NNLO). In Fig. 4.8 we show the distance between PDF central values of the two pairs of fits, at NLO and NNLO, and compare the corresponding gluon PDFs.

Inspection of Table 4.5 shows that at NLO the scale choice $\mu = \hat{H}_T$ leads to a better description of the jet data, both included and not included in the fits, with respect to $\mu = p_t^{\text{jet}}$. However, the effect of the scale choice on the PDFs is very mild (see Fig. 4.8), with a localized modification of the gluon below the half sigma level for $x \simeq 0.2$ and no effect on the other PDFs. On the other hand, at NNLO the two scale choices lead to almost indistinguishable results, both in terms of fit quality and PDF shape, with the scale choice $\mu = \hat{H}_T$ leading to a slightly better description of data not included in the fit, and a difference in gluon central values barely above statistical indistinguishability.

We conclude that the scale choice $\mu = \hat{H}_T$ is perturbatively more stable, in that it leads to a better NLO fit, but that at NNLO the choice of central scale is not an issue.
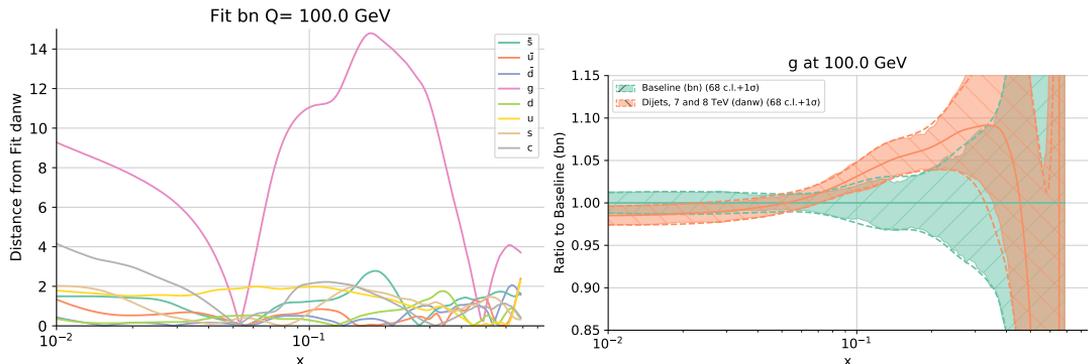
Figure 4.9: Same as Fig. 4.5, but now for dijets.

4.4.3 PDF fits with dijet data

We now turn to PDF fits in which dijet, rather that single-inclusive jet data are included. Also in this case, we first discuss the impact and compatibility of this data, and then the perturbative stability of results.

Impact and consistency of datasets

We assess the impact of dijet data on PDFs by comparing fits with optimal settings, i.e. with QCD corrections up to NNLO and for jet data also EW corrections included, and either the full dataset (#danw), or the 7 TeV (#d7nw) or 8 TeV (#d8nw) data included in turn.

We start by comparing to the baseline #bn, with no jet data, fit #danw in which all dijet data are included; PDFs are compared in Fig. 4.9. From Table 4.4, we see that individual dijet datasets are overall fairly well described (the χ^2 per datapoint is around 1.5 for each of them). Inclusion of the dijet datasets in the baseline leads to an improved description of single-inclusive jet data, just like (see Sec. 4.4.2) inclusion of single-inclusive jet data leads to an improved description of dijets. This confirms consistency of the single-inclusive and dijet data. Unlike in the case of single-inclusive jet data, no tension is observed between dijet data and the rest of the global dataset (specifically top rapidity distributions), whose χ^2 is left almost unchanged.

As in the case of single-inclusive jets, only the gluon PDF is affected by the inclusion of dijet data, with the strongest impact observed in the regions $x \simeq 0.01$ and $0.06 \lesssim x \lesssim 0.4$ (see Fig. 4.9). In the former region the gluon is suppressed by about 2%, corresponding to a shift in central value by about one sigma; in the latter it is enhanced by up to 10% around $x \sim 0.3$, corresponding to a shift by about one and a half sigma, hence outside the error band of the baseline. These shifts are qualitatively similar to those observed upon inclusion of the single-inclusive jet data, but somewhat more pronounced and in a somewhat wider kinematic region.

We then turn to the assessment of the relative impact of different datasets, by comparing to the baseline (#bn) the fits in which only 7 TeV (#d7nw) or only 8 TeV (#d8nw) dijet data are included, see Fig. 4.10. From Table 4.4, we see that the fit quality is equally good for 7 TeV or 8 TeV data, however the fit to the 8 TeV dijet data is closer to the fit in which all dijet data are included, in that it leads to a similar description of all of the jet and dijet data, including those that are not included in either fit. Such a description is better in both fits than in the fit
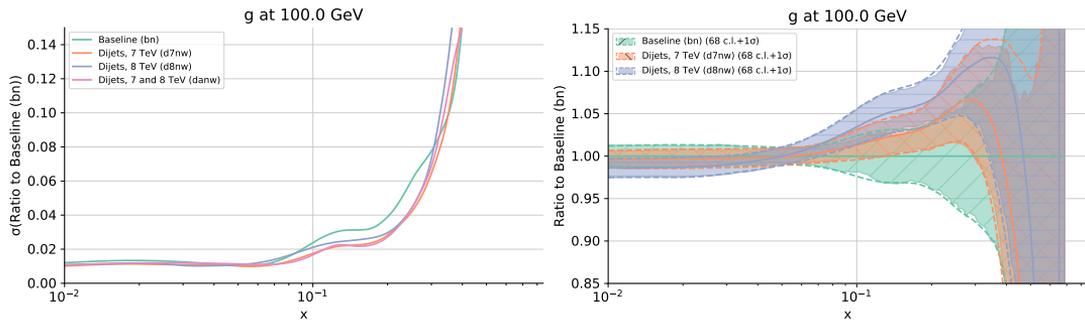
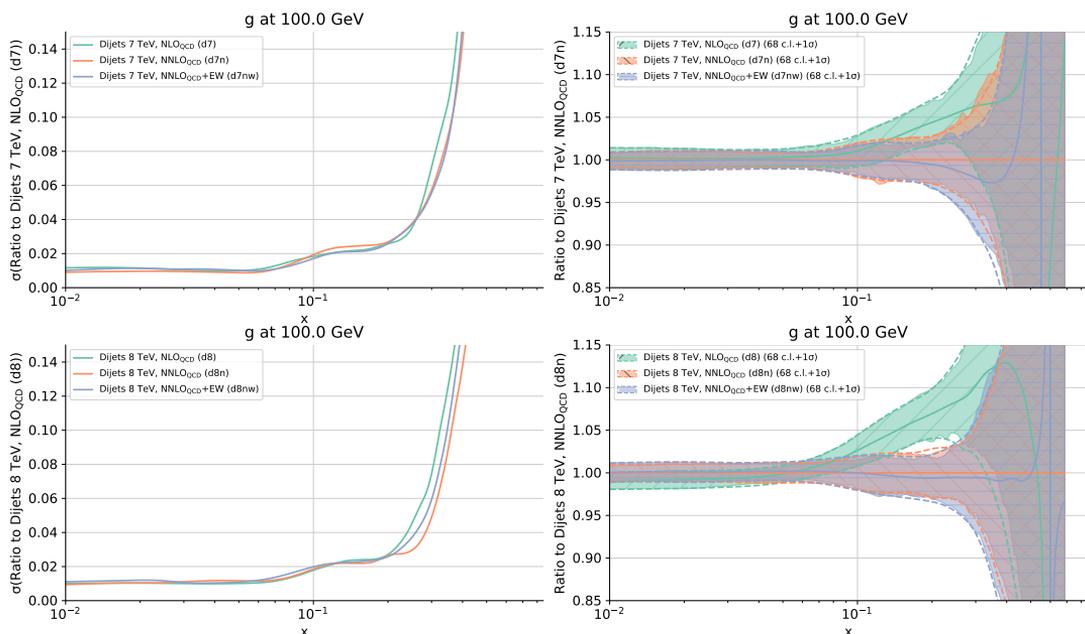**Figure 4.10:** Same as Fig. 4.6, but now for dijets.

Figure 4.11: Same as Fig. 4.7, but now for dijets.

to the 7 TeV dijet datasets only, and is accompanied by a similar change in the description of the ATLAS top pair differential rapidity distributions. This suggests that among the dijet data, the 8 TeV data provide the dominant contribution.

The relative impact of the 7 TeV and 8 TeV data on the gluon central values and uncertainty can be directly inferred from Fig. 4.10. The impact of the two datasets on the gluon central value is qualitatively the same, and thus also the same as that of the full dijet dataset, but with the 8 TeV data having a stronger impact, almost equivalent to the impact of the full dataset. The reduction in uncertainty in comparison to the baseline due to either dataset is roughly the same, by about 3-4% to 3% at $x \simeq 0.2$. Consequently, the gluon PDF determined when including all of the dijet data is very close to that found when including only the 8 TeV data, thus confirming that the 8 TeV data have a dominant impact in the dijet dataset.

Impact of higher order QCD and EW corrections

As for single-inclusive jets, we assess the perturbative stability of fits with dijet data by comparing fits at NLO, NNLO and with EW corrections, separately for the 7 TeV and 8 TeV data, i.e. respectively, fits #d7, #d7n, #d7nw; and #d8, #d8n, #d8nw. The gluon PDFs for these fits are compared in Fig. 4.11.

The figure shows that the perturbative behaviour of the gluon upon inclusion of the dijet data is very similar to what observed when including single-inclusive jets. Namely, at NLO the gluon is distorted in the region $0.1 \lesssim x \lesssim 0.5$ in comparison to the NNLO results already shown in Fig. 4.10, with the effect more pronounced for the dominant and more precise 8 TeV data, again providing evidence for large missing NNLO corrections. The effect of the EW corrections is even less marked than in the case of single-inclusive jets: in fact, their inclusion leaves the gluon PDF almost unchanged. For both datasets, inclusion of the NNLO corrections leads to a reduction in uncertainty, more marked for 8 TeV data, while inclusion of the EW corrections has no clear effect; in fact, for the 8 TeV data it leads to a slight increase of the uncertainty. As in the case of single-inclusive jets, we conclude that NNLO corrections have a strong impact by modifying the gluon shape and reducing its uncertainty, while EW corrections have essentially no impact.

Unlike in the case of single-inclusive jets, where the inclusion of NNLO corrections did not have a clear impact on fit quality, for dijets at NNLO there is a clear improvement in χ^2 values (see Table 4.4). Specifically, when all dijet data are included at NLO, the χ^2 of the global fit deteriorates significantly in comparison to the baseline, with the largest effect seen in data which are most sensitive to the gluon, such as the Z p_t distribution and the top rapidity distribution. This deterioration goes away upon inclusion of NNLO corrections. Namely, when NNLO corrections are included, the quality of the global fit including dijets improves considerably, corresponding now to a fit quality which is essentially the same for the fits with or without the dijet data. Accordingly, the fit quality to the dijet data is significantly better at NNLO than at NLO. The effect is driven by the more precise 8 TeV data. Indeed, the same pattern is observed when only 8 TeV data are included, while with 7 TeV data only fit quality to the dijet data at NLO and NNLO is essentially the same, and so is the fit quality with or without dijet data.

This means that inclusion of NNLO corrections is crucial in order to ensure compatibility of the dijet data with the rest of the global dataset. Interestingly, when fitting dijet data no clear effect is seen when going from NLO to NNLO in the fit quality of single-inclusive jet data (not fitted). Inclusion of EW corrections has no significant effect on fit quality.

We conclude that for dijets NNLO corrections have a significant impact on both fit quality, the central value of the gluon PDF and its uncertainty, with a clear pattern of improvement when going from NLO to NNLO.

4.5 Single-inclusive jets vs. dijets: a comparative assessment

Having assessed the impact on PDFs of jet and dijet datasets separately, we now assess them comparatively, in terms of perturbative stability, fit quality, and impact on PDFs. Specifically, we compare directly PDFs obtained in fits to all single-inclusive (#janw) and dijet (#danw) datasets with the most accurate NNLO+EW theory and default settings in Figs. 4.12-4.13, where the baseline fit (with no jet data) and, in the latter case, the CT18 PDF fit [57] are
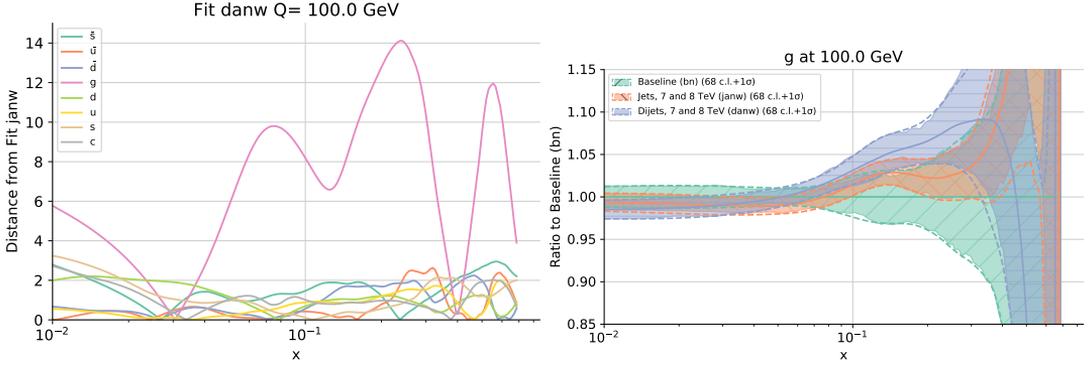
Figure 4.12: Same as Fig. 4.5, but now comparing the fits with all single-inclusive jet data (#janw), and that with all dijet data (#danw) and highest theory accuracy (NNLO QCD+ EW) and default settings. In the gluon comparison (right) results are displayed as a ratio to the baseline with no jet data included (also shown for reference).
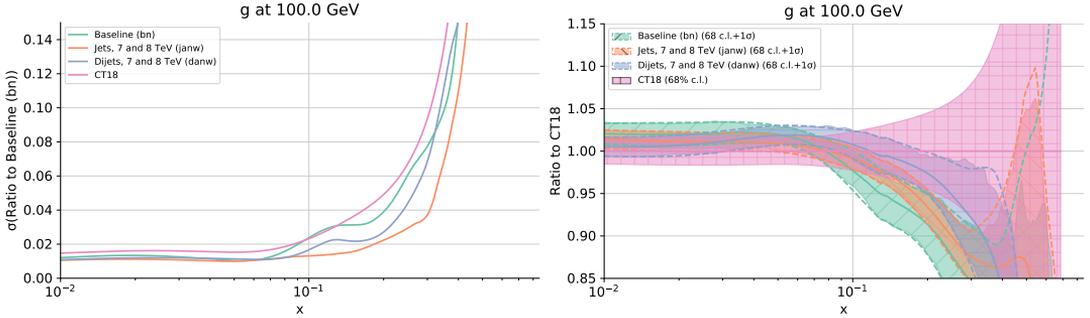

Figure 4.13: Same as Fig. 4.6, but now comparing the baseline (#bn) to the fits with all single-inclusive jet (#janw) and dijet data (#danw) of Fig. 4.12. All results are shown as a ratio to the CT18 fit (also shown for reference).

also shown for reference. Also, in Fig. 4.14 we compare to a representative set of datapoints from each of the single-inclusive jet and dijet datasets predictions obtained using PDFs from the baseline fit, the fit with single-inclusive jets, and the fit with dijets.

Based on the χ^2 values from Tables 4.3-4.4 and the PDF comparisons in Figs. 4.12-4.14, our conclusions are the following.

1. Concerning the relative impact on PDFs of single-inclusive jets and dijets:

   (a) The effect on PDFs of the inclusion of jet and dijet data in the NNPDF3.1 global dataset is qualitatively the same. Namely, they only affect the gluon, by leading to an enhancement of its central value in the region $0.1 \lesssim x \lesssim 0.4$, accompanied by a suppression in the region $0.01 \lesssim x \lesssim 0.1$. The suppression is by about 1%, while the enhancement at the peak, localized at $x \simeq 0.3$ is by about 2.5% for single-inclusive jets, but stronger, by about 7.5% for dijets. An enhanced gluon is also present in the CT18, which, as mentioned, includes the 8 TeV CMS single-inclusive jet data, and whose gluon PDF is consistent with our result within its rather larger uncertainty.

(b) The inclusion of either single-inclusive or dijets leads to a reduction in the gluon uncertainty, with a somewhat stronger reduction observed for single-inclusive jets. It should be noted in this respect that for the most accurate 8 TeV dijet dataset, which as shown in Secs. 4.4.2-4.4.3 is mostly responsible for the reduction in gluon uncertainty, only CMS data are currently available. The constraining power of the dijet dataset is consequently at present more limited than that of the single-inclusive jet dataset.

(c) The inclusion of single-inclusive jet or dijet data does not lead to a deterioration in the description of the rest of the data in comparison to the baseline fit: almost all χ^2 values for other datasets are unchanged. This shows that the single-inclusive and dijet data are not only consistent with each other, but also with the rest of the global dataset, and their impact on the gluon central value, accompanied by a reduction in uncertainty, corresponds to a genuine addition of new information in the fit. Indeed, a comparative assessment of the impact of jet, Z p_t and top production data on the gluon distribution in Ref. [143] showed good consistency, specifically wish the top data also leading to an enhancement of the gluon in the $x \gtrsim 0.1$ region. An exception is the ATLAS top rapidity distributions, which seem to be in tension with the ATLAS 8 TeV single-inclusive jet data, as discussed in Sec. 4.4.2. The quality of the fit to this data also deteriorates, though by smaller amount, when dijet data are fitted; note however that in this case the quality of the fit to CMS top rapidity data improves.

2. Concerning relative fit quality:

(a) The quality of the fit to single-inclusive jet data and dijet data when each of them is fitted is comparable, though somewhat better for dijets ($\chi^2 = 1.65$ vs. $\chi^2 = 1.88$). The quality of the fit to dijets when single inclusive jets are fitted and conversely are almost identical ($\chi^2 = 2.10$ for dijets when fitting single inclusive jets vs. $\chi^2 = 2.06$ for single inclusive jets when fitting dijets), and only marginally worse than the quality of the fit to each dataset when it is fitted. This confirms the full consistency of the two datasets, with a marginal preference for dijets.

(b) The fit including dijet data is also somewhat more internally consistent than the fit including single-inclusive jet data. Indeed, the χ^2 per datapoint of the global fit is closer to one (1.22 vs 1.28), and also, the χ^2 for individual datasets is generally better. In particular, this happens for top production data, also sensitive to the large-x gluon. It is unclear whether this is due to a greater theoretical accuracy of the NNLO dijet observable, or to better quality of the dijet data (specifically a better control of correlated systematics). However, the issue is phenomenologically immaterial, given that the shape and size of the data to theory ratio are qualitatively comparable for all of the jet and dijet data (including for the rapidity bins not displayed in Fig. 4.14), regardless of which dataset is actually fitted.

3. Concerning relative perturbative stability:

(a) When fitting the dijet data, fit quality to the fitted data improves significantly from NLO to NNLO ($\chi^2 = 2.44$ at NLO vs. 1.65 at NNLO), but the fit quality to the single-inclusive jet data actually deteriorates from NLO to NNLO (from $\chi^2 = 1.54$ to 2.06). When fitting the single-inclusive jet data, the fit quality to the fitted data

does not improve and actually deteriorates from NLO to NNLO (from $\chi^2 = 1.25$ to $\chi^2 = 1.88$) but, perhaps surprisingly, the fit quality to the dijet data, not fitted, does improve (from $\chi^2 = 3.29$ at NLO to the NNLO $\chi^2 = 2.10$). Whereas this shows a good theoretical consistency of the dijet data, it is unclear whether the lack of improvement of the single-inclusive jet data is due to a less stable perturbative behaviour of the jet observable, or to issues with data.

(b) As already noted in Sec. 4.4.3, the fit quality to all other data included in the global datasets deteriorates at NLO when including jet data, with a greater deterioration seen in the case of dijets, and more moderate for single-inclusive jets: the total χ^2 per datapoint for the global fit goes from $\chi^2 = 1.20$ of the baseline to 1.28 in the former case and 1.33 in the latter. At NNLO, when dijets are fitted the global fit quality significantly improves and becomes almost the same as that of the baseline ($\chi^2 = 1.22$, in comparison to $\chi^2 = 1.18$ of the baseline) while for the fit to single-inclusive jets it does not improve. The greater deterioration of fit quality at NLO for dijets can be understood as a consequence of the fact, observed in point 1.a above, that dijets have a greater pull on the gluon: hence missing NNLO corrections lead to a stronger loss of accuracy. The lack of improvement in the description of single-inclusive jets shows again that this observable seems to be somewhat less well-behaved, either for theoretical or experimental reasons.

We generally conclude that single-inclusive jets and dijets are mutually consistent and at NNLO consistent with the global dataset and have a similar impact on the gluon. The dijet observable has a better behaved perturbative behaviour and a stronger pull on the gluon PDF and it appears to be marginally preferable, though it leads to a less pronounced decrease of the gluon uncertainty, possibly because ATLAS dijet measurements are not yet available at 8 TeV, while single-inclusive jet measurements are available both from ATLAS and CMS.

4.A APPLGRID interface to FASTNLO

In this appendix, we want to briefly highlight the main differences between grids in the APPLGRID format (version 1.4.7) and in the FASTNLO format (series 2.x with fixed additive contributions[3]). Taking into account these differences, it has been possible to provide APPLGRID with the possibility to read FASTNLO tables.

- In the case of multi-differential distribution e.g. $d^2\sigma/dp_t dy$, in APPLGRID any rapidity bin defines a different observable. Instead, in FASTNLO, the grid is only one and each point has more kinematical coordinates e.g. (p_t, y). This implies that a single FASTNLO table will return a vector of APPLGRID grids.

- In FASTNLO, predictions for different values of $\mu_R = \xi_R Q^2$ and $\mu_F = \xi_F Q^2$ are explicitly saved into the grid. In APPLGRID only the prediction for the central scale i.e. $\mu_R = \mu_F = Q^2$ is needed, as it has the ability to *a posteriori* vary both scales.

---

[3]Starting from version 2, FASTNLO introduces also alternative *flexible-scale* tables [163], which allow one to *a posteriori* choose as renormalization and factorization scale any function of user-defined observables s_1 and s_2, calculated at filling time.
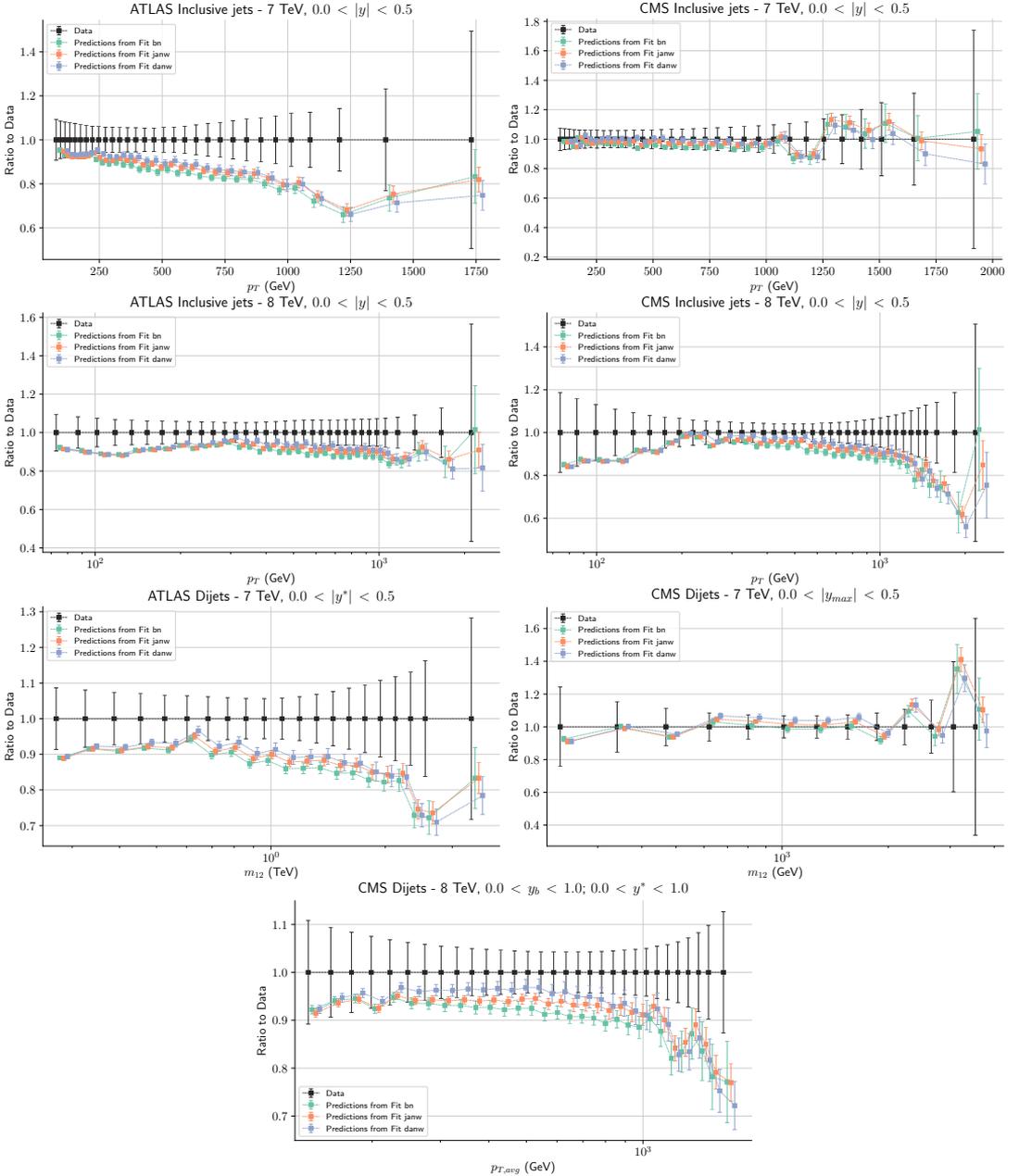
Figure 4.14: The theory to data ratio for a representative set of points from each dataset introduced in Sec. 4.1. Specifically, we show the central rapidity bins for all the ATLAS and CMS jet and dijet datasets at 7 and 8 TeV. Theoretical predictions are computed from fits #bn, #janw and #danw with corresponding theoretical accuracy.
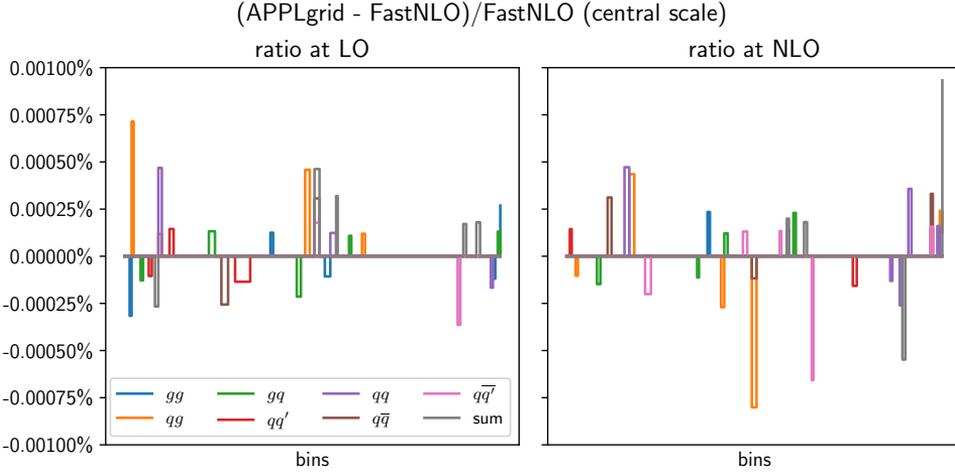
Figure 4.15: Benchmark test of our APPLGRID interface to FASTNLO tables. Results are shown both at LO and at NLO, for each subprocess. See text for details.

- In APPLGRID, each grid has two x-coordinates, x_α and x_β, as in Eq. (4.4). In FASTNLO tables, there is only one x-coordinate. This is possible thanks to the fact that if for any grid we adopt the same range of x_α and x_β values, we have a symmetric 2-dim matrix, and then we can build a 1-dim vector with just the elements in upper right triangle. The single x-coordinate of FASTNLO spans such a vector.

- In case of jet production in pp collisions, both APPLGRID and FASTNLO adopt the same set of seven subprocesses, defined in Eq. (2.24). However, there is a reshuffle of the indices used in APPLGRID and FASTNLO for denoting the same subprocess: $s^{\mathrm{AG}} = \pi(s^{\mathrm{FN}})$.

- Both in APPLGRID and in FASTNLO, the grids are not saved in terms of x and Q^2, but as a function of more natural variables y and τ, defined by the relations:

$$y(x) = \ln \frac{1}{x} + a(1-x), \quad \tau(Q^2) = \ln \ln \frac{Q^2}{\Lambda^2}. \tag{4.14}$$

- In FASTNLO, the weight $W_{\alpha\beta,\tau}^{(p)(s)(b)}$ for the bin b in a differential distribution is divided by the bin size (in case of multi-differential distributions, the bin size is the width of the most internal bin).

- There is a factor $x_\alpha x_\beta$ of difference between the FASTNLO and the APPLGRID weights, as the FASTNLO ones are designed to be convoluted with $x_1 f(x_1)$ and $x_2 f(x_2)$.

In Fig. (4.15) we show a comparison between the output of APPLGRID interfaced to a FASTNLO table, and the original output of FASTNLO with the same input table. We consider predictions for the the dijet process measured in Ref. [133] with central scale choice ($\xi_R = 1$, $\xi_F = 1$). We observe a relative agreement for the two predictions below 10^{-5}, both at LO and at NLO, for the whole set of data points and for different partonic channels.

<div align="right">CHAPTER **5**</div>

Towards machine learning analytics for jet substructure

The rapid development of machine-learning (ML) techniques is having a profound impact on particle physics. For instance, as we have seen in Sec. 4.3, the NNPDF collaboration heavily exploits ML techniques in its fitting procedure e.g. neural networks as universal interpolant and a genetic algorithm as optimizer. In addition, with the recent revision of the whole methodology, more efficient ML software libraries have been adopted and NN hyperparameters have been optimized.

The same is true for jet physics: ML is deeply affecting jet substructure studies. We have briefly introduced the main concepts behind jet substructure in Sec. 1.9, and we have seen how jet substructure has emerged as an important tool in the context of analyses involving hadronic final states at the LHC. In the context of jet physics, ML algorithms are typically trained on a control sample, which could be either Monte Carlo pseudo-data or high-purity dataset, and then applied to an unknown sample to classify its properties. Because of limitations on the algorithms' efficiency and on computers' power, ML algorithms were applied to relatively low-dimensional projections of the full radiation pattern that one wished to classify. Even so, such projections usually correspond to physically-motivated observables, such as, e.g. the jet mass, jet shapes, and therefore limitation in performance were mitigated with physics understanding. The field of jet substructure is currently undergoing a ML revolution, moving away from low-dimensional projections and exploiting deep neural network to perform classification. Early progress was made on supervised classification of known particles [164–168] to the point where powerful network architectures are now available [169–172] and the focus lies on improving the stability [173–177] of these approaches.

Despite the success of ML, physicists — and theorists in particular — often feel uneasy about relying on black-box techniques. The situation is not too dissimilar to the one at the birth of jet substructure. At the time there was a plethora of new tagging algorithms and a deeper understanding was reached only when theoretical studies based on QCD were performed [75, 76, 178–188]. Similarly, we would like to accomplish a first-principle understanding of ML, made

possible by the fact that we do have a QCD expert-knowledge of the underlying phenomena behind the classification problem. In particular, we would like to find an answer to the following question, which naturally arise when employing NN techniques for classification. Does the NN, after proper training, lead to the same classifier as the one dictated by the likelihood ratio, which is the optimal single-variable discriminant? In other words, is the NN tuned with the weights corresponding to the global minimum of the cost function equivalent to a cut on the likelihood ratio? In this chapter we are going to find an answer to this question, by looking for a situation where most studies can be performed analytically, and afterwards by investing how our findings in the simplest case compare to a more general setup.

To this purpose, in Section 5.1, we introduce a variant of the well-known N-subjettiness variable (see Sec. 1.9) based on the primary Lund plane declustering tree [189]. This *primary* N-subjettiness $\mathfrak{T}_N$ is more amenable to an all-order QCD analysis, which we perform at leading logarithmic (LL) accuracy, while maintaining, if not improving, the discrimination power. This definition is such that, if we measure the N-subjettiness variables $\{\mathfrak{T}_1...\mathfrak{T}_n\}$, then, at LL, a cut on the likelihood ratio simply corresponds to a simple cut on $\mathfrak{T}_n$. Furthermore, with this definition and within the stated accuracy, we are able to find analytic expressions for the cumulative distributions and for the so-called area under the curve (AUC), which measures the discriminating power of the observable, for any value of n. We then move in Section 5.2 to use the aforementioned primary N-subjettiness variables as input to a NN. Because we expect the optimal discriminant to be a cut on $\mathfrak{T}_n$, we can start by considering the simplest possible NN, namely a single neuron, or *perceptron* with n inputs and one output. This opens up the interesting possibility of performing analytic calculations that describe the behaviour of the perceptron at LL accuracy in QCD. In Sections 5.3 and 5.4, we compare our theoretical findings with actual networks with architecture of increasing complexity using pseudodata, which we generate either according to a leading-logarithmic distribution in QCD or using a general-purpose Monte Carlo parton shower.

5.1   Primary N-subjettiness

A general approach to quantitatively assess the power of quark/gluon discrimination by using the N-subjettiness variables τ_N (introduced in Sec. 1.9) was put forward in Ref. [78]. The main idea underlying this approach is as follows. If we measure the first n variables $\{\tau_1 \ldots \tau_n\}$, we are able to resolve n emissions in the jet. In particular, the Authors of Ref. [78] calculated explicitly the N-subjettiness distributions

$$p_i(\tau_1,\ldots,\tau_n) = \frac{1}{\sigma_i}\frac{d\sigma_i}{d\tau_1\cdots d\tau_n}, \quad i=q,g, \tag{5.1}$$

at LL accuracy, i.e. in the limit in which all emissions are strongly ordered, for the cases $n = 1, 2, 3$. Given that the likelihood ratio — which according to the Neyman-Pearson lemma [190] provides us with the best single-variable discriminant — is simply given by:

$$\mathcal{L}(\tau_1,\ldots,\tau_n) = \frac{p_{\mathcal{B}}}{p_{\mathcal{S}}} = \frac{p_g(\tau_1,\ldots,\tau_n)}{p_q(\tau_1,\ldots,\tau_n)}, \tag{5.2}$$

the Authors of Ref. [78] has been able to determine the best single-variable discriminant for n emissions resolved in a jet.

However, even in the limit where the emissions are strongly ordered, p_i results in a rather complicated structure. This is because the emissions that set the values of the observables τ_i, which are always gluons at LL accuracy, can either be *primary* emissions, i.e. they originate from the original hard parton, which could be a quark or a gluon, or they can originate from subsequent gluon splittings. If we consider, for instance, the case of a jet initiated by a hard quark, one ends up with n contributions with colour factors $C_F^{n-i} C_A^i$, $i = 0, \ldots n-1$. One would also end up with a Sudakov form factor with both C_F and C_A contributions with a structure depending on the complete tower of n emissions. It is clear that this intricate structure does not facilitate analytical calculations.

Therefore, we find convenient to introduce a variant of N-subjettiness that is sensitive, at LL accuracy, only to primary emissions, such that the distributions p_i are determined by strongly-ordered gluon emissions off the initial hard parton. Specifically, we define the new observable, *primary N-subjettiness*, as follows. Starting from a jet of radius R_0, one first builds the list of primary Lund declusterings [189]:

1. Recluster the jet constituents with the Cambridge/Aachen algorithm [66, 67].

2. Iteratively undo the last step of the clustering $j \rightarrow j_1 + j_2$, with $p_{t1} > p_{t2}$. At step i $(i = 1, \ldots, m)$, define

$$\tilde{p}_{ti} = p_{t2} \qquad \text{and} \qquad \Delta_i = \sqrt{\Delta y_{12}^2 + \Delta \phi_{12}^2}. \tag{5.3}$$

   Repeat the procedure with $j = j_1$, i.e. following the harder branch of the de-clustering.

3. When the de-clustering terminates, i.e. when j is no longer coming from a $j_1 + j_2$ clustering, define $\tilde{p}_{t0}$ as the transverse momentum of j.

From the set of transverse momenta, we can define the momentum fractions

$$z_i = \frac{\tilde{p}_{ti}}{\sum_{i=0}^{m} \tilde{p}_{ti}} \qquad i = 1, \ldots, m, \tag{5.4}$$

where we note that the final hard momentum $\tilde{p}_{t0}$ is included in the normalisation. This produces a set of values (z_i, Δ_i) that we order such that $z_1 \Delta_1^\beta \geq z_2 \Delta_2^\beta \geq \cdots \geq z_m \Delta_m^\beta$. The primary N-subjettiness is then defined as [1].

$$\mathcal{T}_N = \sum_{i=N}^{m} z_i \left(\frac{\Delta_i}{R_0} \right)^\beta. \tag{5.5}$$

Note that $\mathcal{T}_1 \geq \cdots \geq \mathcal{T}_n$, like with the standard N-subjettiness τ_N. The primary N-subjettiness definition in Eq. (5.5) is very similar to the standard N-subjettiness definition with the main difference that it is computed based on primary Lund declusterings. The definition of the momentum fractions z_i is such that $\sum_{i=0}^{m} z_i = 1$.

By construction the LL expression for the n-dimensional differential distribution is obtained by considering independent emissions off the originating hard parton (either a quark or a gluon). Of this infinite tower of emission, n of them set the values of $\mathcal{T}_1, \ldots, \mathcal{T}_n$, while all the other

---

[1]Note that an alternative definition, equivalent at leading-logarithmic accuracy, but different beyond, would be to define $\mathcal{T}_N^{(\text{max})} = z_N (\Delta_N/R_0)^\beta$ i.e. the maximum value instead of the sum.

ones remain unresolved. Furthermore, we have to consider virtual corrections, which naturally exponentiate in the soft and collinear limit. Both unresolved and virtual corrections are divergent order by order in perturbation theory and we regularise such singularities with and infra-red regulator ϵ. We have

$$p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n) = \int_0^1 \frac{d\rho_1}{\rho_1} C_i \mathcal{R}'(\rho_1) \cdots \int_0^{\rho_{n-1}} \frac{d\rho_n}{\rho_n} C_i \mathcal{R}'(\rho_n) \prod_{j=1}^n \delta \left(\mathcal{T}_j - \sum_{k \geq j} \rho_k \right)$$

$$\cdot \lim_{\epsilon \to 0} \sum_{p=0}^{\infty} \frac{1}{p!} \left(\int_\epsilon^{\rho_n} \frac{d\rho}{\rho} C_i \mathcal{R}'(\rho) \right)^p \exp \left[-\int_\epsilon^1 \frac{d\rho}{\rho} C_i \mathcal{R}'(\rho) \right], \qquad (5.6)$$

where $i = q, g$ (with $C_q = C_F$ and $C_g = C_A$), and $\rho_k = z_k \theta_k^\beta$, where as before θ_k is the angular distance between parton k and the jet axis, in units of the jet radius R_0. Since we are working in the strongly order limit $\rho_k \gg \rho_{k+1}$, we can approximate the delta function in Eq. (5.6) as

$$\delta \left(\mathcal{T}_j - \sum_{k \geq j} \rho_k \right) = \delta \left(\mathcal{T}_j - \rho_j \right). \qquad (5.7)$$

Furthermore, we have introduced the LL radiator function $\mathcal{R}$, defined in Eq. (1.99), and its derivative

$$\mathcal{R}'(\rho) = \frac{d\mathcal{R}(\rho)}{d \log(1/\rho)}, \qquad (5.8)$$

where with $\log x$ we always denote the natural logarithm of x. Going back to Eq. (5.6), we note that the unresolved emission contribution exponentiates too and combines with the virtual corrections to produce a finite exponent, giving rise to the so-called Sudakov form factor. Thus, we arrive at a resummed expression for the probability distribution $p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n)$, which is valid to LL accuracy:

$$p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n) = \left(\prod_{j=1}^n \frac{\mathcal{R}'(\mathcal{T}_j)}{\mathcal{T}_j} \right) C_i^n \exp \left[-C_i \mathcal{R}(\mathcal{T}_n) \right]. \qquad (5.9)$$

An important observation is the following. From Eq. (5.9) we note that the structure of the probability distributions at LL in QCD for primary definition of N-subjettiness is the *same* for quark and gluon jets, with a parametric difference encoded in the colour factor C_F or C_A (this is not case with the standard definition of N-subjettiness). Consequently, the likelihood ratio Eq. (5.2) at LL becomes

$$\mathcal{L}^{\mathrm{LL}} = \left(\frac{C_A}{C_F} \right)^n \exp \left[-2 \frac{C_A - C_F}{\beta} \int_{\mathcal{T}_n}^1 \frac{d\rho}{\rho} \int_\rho^1 \frac{dz}{z} \frac{\alpha_s \left(z^{\frac{\beta-1}{\beta}} \rho^{\frac{1}{\beta}} p_t R_0 \right)}{\pi} \right], \qquad (5.10)$$

which is monotonic in $\mathcal{T}_n$. Therefore, at LL, a cut on the likelihood ratio is equivalent to a cut on $\mathcal{T}_n$. The remarkable simplicity of this result is the strongest motivation for introducing *primary* N-subjettiness. This observable is then the ideal laboratory to study in an analytical fashion how a neural network that takes the *primary* N-subjettiness variables as inputs performs. Note

that because of the simplicity of the classifier, i.e. a cut on a single variable, we expect that even the simplest network, i.e. the one formed by a single neuron, should lead to meaningful results.

Analytic studies of a perceptron will be the topic of Section 5.2, but, before moving to that, let us derive a couple of results that allow us to establish *primary N-subjettiness* as an appealing observable on its own, rather than just a shortcut to more tractable analytic results. It is interesting to obtain an analytic expression for the cumulative distribution with a cut $\mathcal{T}$ on $\mathcal{T}_n$. As we have just seen, this is equivalent to a cut on the likelihood. We have

$$
\begin{aligned}
\Sigma_i(\mathcal{T}_n < \mathcal{T}) &= \int_0^1 d\mathcal{T}_1 \int_0^{\mathcal{T}_1} d\mathcal{T}_2 \cdots \int_0^{\mathcal{T}_{n-2}} d\mathcal{T}_{n-1} \int_0^{\mathcal{T}_{n-1}} d\mathcal{T}_n\, p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n)\, \Theta(\mathcal{T}_n < \mathcal{T}) \\
&= \int_0^1 d\mathcal{T}_1 \int_0^{\mathcal{T}_1} d\mathcal{T}_2 \cdots \int_0^{\mathcal{T}_{n-2}} d\mathcal{T}_{n-1} \int_0^{\min[\mathcal{T}_{n-1}, \mathcal{T}]} d\mathcal{T}_n\, p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n)\,.
\end{aligned} \tag{5.11}
$$

The latter expression splits naturally in two terms: if $\mathcal{T}_{n-1} < \mathcal{T}$, we simply find $\Sigma_i(\mathcal{T}_{n-1} < \mathcal{T})$; if $\mathcal{T}_{n-1} > \mathcal{T}$, the exponential factors out and we obtain

$$
e^{-C_i \mathcal{R}(\mathcal{T})} \int_{\mathcal{T}}^1 d\mathcal{T}_1 \int_{\mathcal{T}}^{\mathcal{T}_1} d\mathcal{T}_2 \cdots \int_{\mathcal{T}}^{\mathcal{T}_{n-2}} d\mathcal{T}_{n-1}\, p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n) = e^{-C_i \mathcal{R}(\mathcal{T})} \frac{C_i^{n-1} \mathcal{R}^{n-1}(\mathcal{T})}{(n-1)!}\,. \tag{5.12}
$$

By induction we arrive at

$$
\Sigma_i(\mathcal{T}_n < \mathcal{T}) = e^{-C_i \mathcal{R}(\mathcal{T})} \sum_{k=1}^n \frac{C_i^{k-1} \mathcal{R}^{k-1}(\mathcal{T})}{(k-1)!} = \frac{\Gamma(n, C_i \mathcal{R}(\mathcal{T}))}{\Gamma(n)}, \tag{5.13}
$$

where $\Gamma(n, x)$ is the incomplete Gamma function.

It is also possible to find an analytic expression for the AUC. Exploiting the ROC curve definition in Eq. (1.105), we can write the AUC as an integral of the quark and gluon distributions:

$$
\begin{aligned}
\mathrm{AUC} = \int_0^1 d\mathcal{T}_{1q} \int_0^{\mathcal{T}_{1q}} d\mathcal{T}_{2q} \cdots \int_0^{\mathcal{T}_{n-1,q}} d\mathcal{T}_{nq} \int_0^1 d\mathcal{T}_{1g} \int_0^{\mathcal{T}_{1g}} d\mathcal{T}_{2g} \cdots \int_0^{\mathcal{T}_{n-1,g}} d\mathcal{T}_{ng} \\
\cdot\, p_q(\mathcal{T}_{1q}, \mathcal{T}_{2q}, \cdots, \mathcal{T}_{nq})\, p_g(\mathcal{T}_{1g}, \mathcal{T}_{2g}, \cdots, \mathcal{T}_{ng})\, \Theta(\mathcal{T}_{nq} > \mathcal{T}_{ng})\,.
\end{aligned} \tag{5.14}
$$

At LL accuracy, all the above integrals can be performed analytically. The necessary steps are not entirely trivial, so we collect some details in Appendix 5.A for the interested reader. The LL expression for the area under the ROC curve is

$$
\mathrm{AUC} = 1 - \left(\frac{C_F C_A}{(C_F + C_A)^2} \right)^n \frac{\Gamma(2n)}{\Gamma(n)\Gamma(1+n)}\, {}_2F_1\left(1, 2n, 1+n; \frac{C_A}{C_F + C_A} \right). \tag{5.15}
$$

We can compare the values given by this expression with the ones computed in [78] for the standard definition of N-subjettiness.

$$
\begin{array}{lcll}
n = 1 & \rightarrow & \mathrm{AUC}_{\mathrm{standard}} = 0.308, & \mathrm{AUC}_{\mathrm{primary}} = 0.308 \\
n = 2 & \rightarrow & \mathrm{AUC}_{\mathrm{standard}} = 0.256, & \mathrm{AUC}_{\mathrm{primary}} = 0.226 \\
n = 3 & \rightarrow & \mathrm{AUC}_{\mathrm{standard}} = 0.231, & \mathrm{AUC}_{\mathrm{primary}} = 0.173
\end{array}
$$

We conclude that, at least at LL, a cut on the *primary* N-subjettiness $\mathcal{T}_n$ provides better quark/gluon discrimination power than a cut on the standard N-subjettiness. The comparison of the two different definitions when evaluated on Monte-Carlo generated pseudodata will be discussed in Sec. 5.4.

So far we have obtained analytic expressions for primary N-subjettiness distributions that are valid at LL and therefore include running-coupling effects. Henceforth, for sake of simplicity, we are going to consider the fixed coupling limit of Eq. (1.99). In this limit, the probability distributions for quarks and gluons Eq. (5.9) is

$$p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n) = \left(\frac{\alpha_S}{\pi\beta}\right)^n (2C_i)^n \prod_{j=1}^{n} \left(\frac{\log\left(1/\mathcal{T}_j\right)}{\mathcal{T}_j}\right) \exp\left[-\frac{\alpha_S}{\pi\beta} C_i \log^2 \mathcal{T}_n\right], \qquad (5.16)$$

and the likelihood ratio Eq. (5.10) consequently becomes:

$$\mathcal{L}^{\text{LL-f.c.}} = \left(\frac{C_A}{C_F}\right)^n \exp\left[-\frac{C_A - C_F}{\beta} \frac{\alpha_s}{\pi} \log^2 \mathcal{T}_n\right]. \qquad (5.17)$$

In order to further simplify our notation, we can also imagine to reabsorb the factor $\alpha_s/(\pi\beta)$ in the variables $\mathcal{T}_i$, by an appropriate redefinition of the N-subjettiness, or, equivalently, we can imagine to define the colour factors C_F and C_A in units of $\alpha_s/(\pi\beta)$. Thus, we rewrite Eq. (5.16) as

$$p_i(\mathcal{T}_1, \ldots, \mathcal{T}_n) = \left(2\widetilde{C}_i\right)^n \frac{\log\left(1/\mathcal{T}_1\right)}{\mathcal{T}_1} \cdots \frac{\log\left(1/\mathcal{T}_n\right)}{\mathcal{T}_n} \exp\left[-\widetilde{C}_i \log^2 \mathcal{T}_n\right], \qquad (5.18)$$

with $\widetilde{C}_i = \frac{\alpha_S C_i}{\pi\beta}$. Finally, for later purposes, it is convenient to express the probability distribution Eq. (5.18) in terms of logarithmic variables $l_i = \log(1/\mathcal{T}_i)$:

$$p_i(l_1, \ldots, l_n) = \frac{1}{\sigma_i} \frac{d\sigma_i}{dl_1 \cdots dl_n} = (2\widetilde{C}_i)^n \, l_1 \cdots l_n \, \exp\left[-\widetilde{C}_i \, l_n^2\right], \qquad (5.19)$$

or double logarithmic variables $L_i = \log^2(1/\mathcal{T}_i)$:

$$p_i(L_1, \ldots, L_n) = \frac{1}{\sigma_i} \frac{d\sigma_i}{dL_1 \cdots dL_n} = (\widetilde{C}_i)^n \, \exp\left[-\widetilde{C}_i \, L_n\right], \qquad (5.20)$$

where in both cases we have taken into account the appropriate Jacobian factor.

5.2   Perceptron analytics

In this section we will investigate the simplest neural network, namely the *perceptron*, where the input layer is directly linked to the output through a single neuron, as depicted in Fig. 5.1, on the left. Analytically, this means that the network output is a function of a weighted linear combination of the inputs:

$$f(\vec{x} \cdot \vec{a} + b) \qquad (5.21)$$

with $\vec{x}$ vector of the input variables, while $\vec{a}$ is the vector of the *weights* of the neuron and b is a *bias*. The function f, called *activation function*, allows us to introduce an element of non-linearity in the network. We will focus on the *sigmoid*, which is widely adopted in the ML
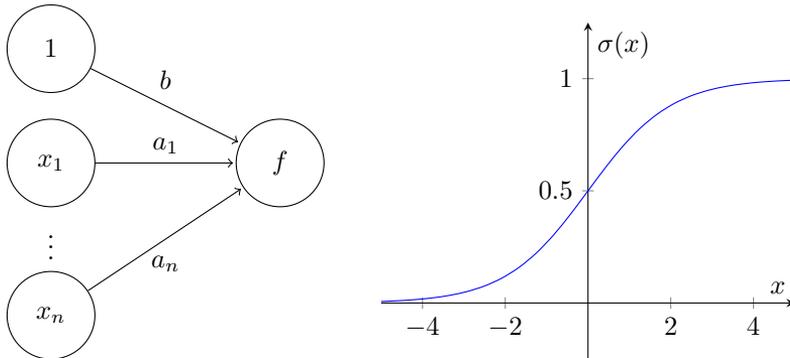
Figure 5.1: Left: The perceptron consists of n input units and one output $y = f(\vec{a} \cdot \vec{x} + b)$. Right: The sigmoid activation function, Eq. (5.22).

community. The sigmoid is defined as

$$f(x) \equiv \sigma(x) = \frac{1}{1 + e^{-x}}\,, \tag{5.22}$$

and it is shown in Fig. 5.1, on the right. Clearly, one can also choose alternative functional form, with similar behaviour. However, note that, in a classification context, the output of the network is required to be a single value bounded between 0 and 1 in order to acquire a probabilistic interpretation. In the perceptron case, this implies that this property should also apply to the image of the activation function. For instance, in this context, one could replace the sigmoid with a rectified version (*hard-sigmoid*):

$$f(x) = \max(0, \min(x, 1))\,. \tag{5.23}$$

We note the standard *rectifier linear unit (ReLU)*, defined as $f(x) = \max(0, x)$, does not meet this requirement.

The neural network learns the best choice of the weights by minimising the *cost* function, which quantifies the difference between the expected value of the network $\hat{y}$ and the predicted value y (the latter in the perceptron case is simply equal to Eq. (5.21)). In this study we are focusing on the issue of quark versus gluon discrimination, which is an example of a binary classification problem, where we have $\hat{y} = 0, 1$. In this context, the *cross-entropy* loss is one of the most common functional form employed for the cost function

$$C(y, \hat{y}) = -(1 - \hat{y}) \log(1 - y) - \hat{y} \log(y)\,. \tag{5.24}$$

In our study we will focus on a cost function defined with the cross-entropy loss. However, many of the results we obtain also applies to other loss functions, such as, for instance, the *quadratic* loss:

$$C(y, \hat{y}) = (y - \hat{y})^2\,. \tag{5.25}$$

In order to train the NN, one usually starts with a so-called training sample, i.e. a collection of input vectors $\{\vec{x}_i\}$, each labelled as a quark jet or as a gluon jet. If we have a training sample of $2N$ input vectors, equally divided between signal and background labels, we can write the

cost function as:

$$\widetilde{C}(\vec{a}, b) = \frac{1}{2N} \sum_{i=1}^{N} \left[C\left(f(\vec{x}_i^{(q)} \cdot \vec{a} + b), 0 \right) + C\left(f(\vec{x}_i^{(g)} \cdot \vec{a} + b), 1 \right) \right]. \qquad (5.26)$$

The input variables $\vec{x}^{(i)}$ can be thought as being generated according to a probability distribution $p_i(\vec{x})$, with p_q being the probability distribution of the inputs $\vec{x}$ for quark jets and p_g for gluon jets. If the training sample is rather large, as it usually is, we can re-write the above equation in the continuous limit:

$$\widetilde{C}(\vec{a}, b) = \frac{1}{2} \int d\vec{x} \left[p_q(\vec{x}) \, C(f(\vec{x} \cdot \vec{a} + b), 0) + p_g(\vec{x}) \, C(f(\vec{x} \cdot \vec{a} + b), 1) \right]. \qquad (5.27)$$

In a general classification problem, the probability distributions of the inputs are unknown. However, in the context of QCD studies, we can exploit expert-knowledge: if we choose the input variables $\vec{x}$ as IRC safe observables, we can apply the powerful machinery of perturbative quantum field theory to determine these distributions at a well-defined and, in principle, systematically improvable accuracy.

In what follows, we are going to use the *primary N-subjettiness* variables $\{\mathcal{T}_i\}$ as input variables for a perceptron. Thanks to the results obtained in Section 5.1, we will be able to evaluate Eq. (5.27) using probability distributions for the inputs calculated at LL accuracy. We will study the global minimum point of Eq. (5.27), by adopting the sigmoid, Eq. (5.22), as activation function f, and the cross-entropy loss, Eq. (5.24), as cost function (analogous results can be found with the quadratic loss, Eq. (5.25)). In particular, we want to establish whether the set of weights $\vec{a}$ and bias b that minimises Eq. (5.27) does correspond to a cut on $\mathcal{T}_n$, which, in turns, is equivalent at LL to a cut on the likelihood ratio, as we have already deduced from Eq. (5.10). We are going to discover that the ability of the perceptron of finding the likelihood ratio crucially depends on the functional form of the input variables. In our studies, we shall consider three cases, all based on the *primary N-subjettiness*, namely $\mathcal{T}_i$, $\log \mathcal{T}_i$ and $\log^2 \mathcal{T}_i$.

5.2.1 Minimisation of the cost function

In order to find the extrema of the cost function Eq. (5.27), we consider the partial derivatives with respect to a generic weight a_i or b. With a simple application of the chain rule, we find a set of $n + 1$ simultaneous equations

$$\frac{\partial \widetilde{C}}{\partial a_i} = \frac{1}{2} \int d\vec{x} \, x_i f'(\vec{x} \cdot \vec{a} + b)$$
$$\left[p_q(\vec{x}) \, C'(f(\vec{x} \cdot \vec{a} + b), 0) + p_g(\vec{x}) \, C'(f(\vec{x} \cdot \vec{a} + b), 1) \right] = 0,$$
$$\frac{\partial \widetilde{C}}{\partial b} = \frac{1}{2} \int d\vec{x} \, f'(\vec{x} \cdot \vec{a} + b)$$
$$\left[p_q(\vec{x}) \, C'(f(\vec{x} \cdot \vec{a} + b), 0) + p_g(\vec{x}) \, C'(f(\vec{x} \cdot \vec{a} + b), 1) \right] = 0, \qquad (5.28)$$

where $i = 1, \ldots, n$ and the prime indicates the derivative of a function with respect to its (first) argument.

In general, in order to solve the above system of simultaneous equations, we have to explicitly

compute the n-dimensional integral in Eq. (5.28). However, in some cases, it might be also possible to find directly a solution at the integrand level. To this purpose, we observe that if the equality

$$\frac{p_q(\vec{x})}{p_g(\vec{x})} = -\frac{C'(f(\vec{x} \cdot \vec{a} + b), 1)}{C'(f(\vec{x} \cdot \vec{a} + b), 0)} \tag{5.29}$$

is satisfied for any value of $\vec{x}$, then the system of equations is fulfilled. In addition, because of the properties of the cross-entropy loss, Eq. (5.24),

$$C'(y, 1) = -\frac{1}{y}, \quad C'(y, 0) = -C'(1 - y, 1), \tag{5.30}$$

the dependence of Eq. (5.29) on the loss function drops out, i.e.

$$-\frac{C'(f(\vec{x} \cdot \vec{a} + b), 1)}{C'(f(\vec{x} \cdot \vec{a} + b), 0)} = \frac{1 - f(\vec{x} \cdot \vec{a} + b)}{f(\vec{x} \cdot \vec{a} + b)}. \tag{5.31}$$

We note that the above equation also holds in the case of the quadratic loss.

For later purposes, it may be useful to specialise Eq. (5.28) to our probabilities p_i – given respectively in Eq. (5.18), Eq. (5.19) and Eq. (5.20), to the sigmoid activation function and to the cross-entropy loss. The probability densities p_i share the common structure

$$p_i(x_1, \ldots, x_n) = r'(x_1) \cdots r'(x_n) \left(\widetilde{C}_i^n \exp\left[-\widetilde{C}_i \, r(x_n) \right] \right), \tag{5.32}$$

where the function $r(x)$ is essentially the radiator introduced in Eq. (1.99), but with its argument dictated by the functional form of the perceptron input variables. Given the following identities for the sigmoid function:

$$1 - \sigma(x) = \sigma(-x) \quad \sigma'(x) = \sigma(x) \, \sigma(-x), \tag{5.33}$$

and the properties of the cross-entropy loss in Eq. (5.30), the system in Eq. (5.28) may be alternatively rewritten as:

$$\frac{\partial \widetilde{C}}{\partial a_i} = \frac{1}{2} \int d\vec{x} \, x_i \, r'(x_1) \cdots r'(x_n)$$
$$\left[\widetilde{C}_F^n \, e^{-\widetilde{C}_F \, r(x_n)} \sigma(\vec{x} \cdot \vec{a} + b) - \widetilde{C}_A^n \, e^{-\widetilde{C}_A \, r(x_n)} \sigma(-\vec{x} \cdot \vec{a} - b) \right] = 0,$$
$$\frac{\partial \widetilde{C}}{\partial b} = \frac{1}{2} \int d\vec{x} \, r'(x_1) \cdots r'(x_n)$$
$$\left[\widetilde{C}_F^n \, e^{-\widetilde{C}_F \, r(x_n)} \sigma(\vec{x} \cdot \vec{a} + b) - \widetilde{C}_A^n \, e^{-\widetilde{C}_A \, r(x_n)} \sigma(-\vec{x} \cdot \vec{a} - b) \right] = 0. \tag{5.34}$$

Log-square inputs. Let us start with considering $L_i = \log^2 \mathfrak{T}_i$ as inputs to the perceptron. In this case the probability distributions for quarks and gluons in the fixed-coupling limit are given by Eq. (5.20). This is a very lucky scenario where we can determine the minimum at the integrand level. Indeed, by simply equating the l.h.s. of Eq. (5.29) and the r.h.s. of Eq. (5.31),

we obtain

$$\left(\frac{\widetilde{C}_F}{\widetilde{C}_A}\right)^n \exp\left[-(\widetilde{C}_F - \widetilde{C}_A)L_n\right] = \exp\left[-\vec{a}\cdot\vec{L} - b\right], \tag{5.35}$$

leading to the following solution

$$a_1 = \cdots = a_{n-1} = 0, \quad a_n = \widetilde{C}_F - \widetilde{C}_A, \quad b = n\log\left(\frac{\widetilde{C}_A}{\widetilde{C}_F}\right). \tag{5.36}$$

Hence, for the log-square inputs, the minimum of the cost function does agree with the optimal cut on $\mathcal{T}_n$ dictated by the likelihood. Moreover, as $\widetilde{C}_F < \widetilde{C}_A$, the weight a_n is negative. This has to be expected, since the sigmoid function is monotonic and we have mapped the gluon (quark) sample to output 1 (0), see Eq. (5.26), whereas the gluon sample has larger $\mathcal{T}_i$ and thus smaller L_i: the negative sign of a_n restores the proper ordering between inputs and output of the perceptron.

If we restrict ourselves to the case $n = 2$, it is also possible to explicitly perform the integrals that appear in Eq. (5.27) and arrive at an analytic expression for the cost function. We report this calculation in Appendix 5.B.

Log inputs.   We now turn our attention to logarithmic inputs $l = -\log\mathcal{T}_i$. In this case we are not able to determine the position of the minimum at the integrand level and we are forced to address the actual integrations in Eqs. (5.28) In particular, we would like to check whether the likelihood condition $a_1 = \cdots = a_{n-1} = 0$ is still present in the solution of the system of simultaneous equations. To this purpose, we use Eq. (5.34) with the probability distribution given by Eq. (5.19), then we set $a_1 = \cdots = a_{n-1} = 0$, thus explicitly obtaining

$$\frac{\partial\widetilde{C}}{\partial a_i} = 2^{n-1}\int_0^\infty dl_n\, l_n\, I_i(l_n)\left[\frac{\widetilde{C}_F^n\exp(-\widetilde{C}_F l_n^2)}{1+\exp(-a_n l_n - b)} - \frac{\widetilde{C}_A^n\exp(-\widetilde{C}_A l_n^2)}{1+\exp(a_n l_n + b)}\right] = 0,$$

$$\frac{\partial\widetilde{C}}{\partial b} = 2^{n-1}\int_0^\infty dl_n\, l_n\, I_0(l_n)\left[\frac{\widetilde{C}_F^n\exp(-\widetilde{C}_F l_n^2)}{1+\exp(-a_n l_n - b)} - \frac{\widetilde{C}_A^n\exp(-\widetilde{C}_A l_n^2)}{1+\exp(a_n l_n + b)}\right] = 0, \tag{5.37}$$

where $I_i(l_n)$ is the result of the integration over $l_1, \ldots, l_{n-1}$

$$I_i(l_n) = \int_0^{l_n} dl_{n-1}\, l_{n-1}\cdots\int_0^{l_2} dl_1\, l_1\, v_i, \tag{5.38}$$

where $v_0 = 1$, $v_i = l_i$. These integrals can be straightforwardly evaluated and, up to an irrelevant constant, we have

$$I_0(l_n) \propto l_n^{2n-2}, \quad I_i(l_n) \propto l_n^{2n-1} \quad \text{for} \quad i = 1, \ldots, n \tag{5.39}$$

Replacing this result in Eq. (5.37), we see that all of the derivatives with respect to a_i, $i = 1, \ldots, n$ give rise to the same equation, and thus the system of $n+1$ simultaneous equation reduces to a system of just two independent equations, for any n. These two equations correspond to two lines in the (a_n, b) plane. If these two lines never cross, then the system has no solution, and the minimum of the cost function is not at $a_1 = \ldots a_{n-1} = 0$ and thus the perceptron is

not able to correctly reproduce the likelihood. If instead these lines meet at some $\bar{a}_n, \bar{b}$, then the minimum of the cost function does correspond to the likelihood ratio. Despite numerous attempts, we have not been able to perform the final integration over l_n in Eq. (5.37). However, we can perform the integration numerically and plot the result in the (a_n, b) plane. This is done in Fig. 5.2, on the left, where, without loss of generality, we have concentrated on the case $n = 2$. It is clear from the plot that the two curves do meet in a point and hence the perceptron is able to find the correct minimum, i.e. the one dictated by the likelihood with $a_1 = 0$.
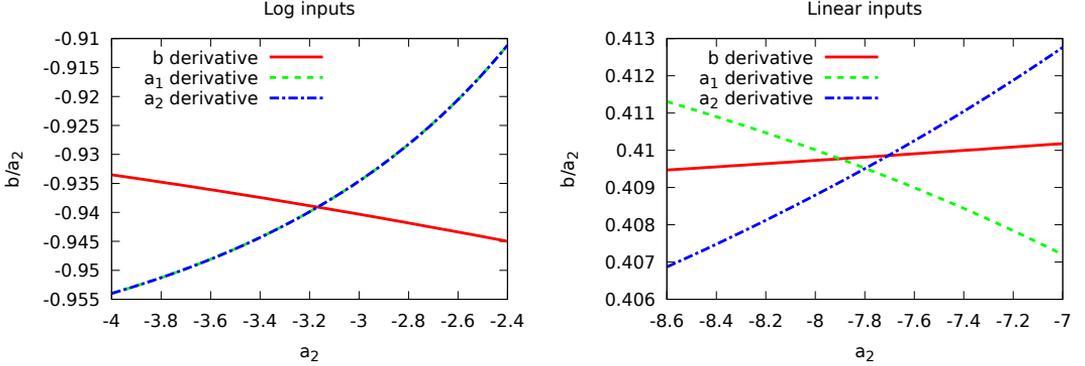

Figure 5.2: Solutions of Eq. (5.34), in the case $n = 2$ with $a_1 = 0$, plotted as points in the $(a_2, b/a_2)$ plane, for log (left) and linear (right) inputs.

The explicit n-dependence of the coefficients a_n and b_n can be found numerically by solving the system of equations. Furthermore, it is possible to obtain analytically the scaling of a_n with respect the colour factors, as detailed in Appendix 5.C. We find that, once we have factored out $\widetilde{C}_F^{-1/2}$, the resulting coefficient only depends on the ratio of colour factors:

$$\frac{a_n}{\sqrt{\widetilde{C}_F}} = F\left(\frac{\widetilde{C}_A}{\widetilde{C}_F}\right), \tag{5.40}$$

where F stands for a unspecified function that we have not determined.

Linear inputs. We now move to consider linear inputs of the perceptron, i.e. the variables $\mathcal{T}_i$ directly. We follow the same logic as in the case of the logarithmic inputs, namely we want to check whether at the position of the minimum of the cost function we have $a_1 = \ldots a_{n-1} = 0$. Following the same steps as before, we have

$$\frac{\partial \widetilde{C}}{\partial a_i} = 2^{n-1} \int_0^1 \frac{d\mathcal{T}_n}{\mathcal{T}_n} \log \frac{1}{\mathcal{T}_n} I_i(\mathcal{T}_n) \left[\frac{\widetilde{C}_F^n \exp(-\widetilde{C}_F \log^2 \mathcal{T}_n)}{1 + \exp(-a_n \mathcal{T}_n - b)} - \frac{\widetilde{C}_A^n \exp(-\widetilde{C}_A \log^2 \mathcal{T}_n)}{1 + \exp(a_n \mathcal{T}_n + b)} \right] = 0,$$

$$\frac{\partial \widetilde{C}}{\partial b} = 2^{n-1} \int_0^1 \frac{d\mathcal{T}_n}{\mathcal{T}_n} \log \frac{1}{\mathcal{T}_n} I_0(\mathcal{T}_n) \left[\frac{\widetilde{C}_F^n \exp(-\widetilde{C}_F \log^2 \mathcal{T}_n)}{1 + \exp(-a_n \mathcal{T}_n - b)} - \frac{\widetilde{C}_A^n \exp(-\widetilde{C}_A \log^2 \mathcal{T}_n)}{1 + \exp(a_n \mathcal{T}_n + b)} \right] = 0,$$

$$\tag{5.41}$$

with

$$I_i(\mathfrak{T}_n) = \int_{\mathfrak{T}_n}^1 \frac{d\mathfrak{T}_{n-1}}{\mathfrak{T}_{n-1}} \log \frac{1}{\mathfrak{T}_{n-1}} \cdots \int_{\mathfrak{T}_2}^1 \frac{d\mathfrak{T}_1}{\mathfrak{T}_1} \log \frac{1}{\mathfrak{T}_1} v_i \, , \tag{5.42}$$

where now $v_0 = 1 \, , v_i = \mathfrak{T}_i$. There is a crucial difference between the integrals in Eq. (5.42) and the corresponding ones in the case of the logarithmic inputs Eq. (5.38): the cases with $i = n$ and with $i < n$ do not lead to the same functional form of the results. For sake of simplicity, let us specialise to $n = 2$. We have

$$I_i(\mathfrak{T}_2) = \int_{\mathfrak{T}_2}^1 \frac{d\mathfrak{T}_1}{\mathfrak{T}_1} \log \frac{1}{\mathfrak{T}_1} v_i = \begin{cases} \frac{1}{2} \log^2 \mathfrak{T}_2 \, , & i = 0, \\ 1 - \mathfrak{T}_2 + \mathfrak{T}_2 \log \mathfrak{T}_2 \, , & i = 1, \\ \frac{1}{2} \mathfrak{T}_2 \log^2 \mathfrak{T}_2 \, , & i = 2. \end{cases} \tag{5.43}$$

Thus, all the three equations appearing in (5.41), i.e. $i = 0, 1, 2$ provide independent conditions. Then, the system has solutions if the corresponding three curves in the (a_2, b) plane meet in one point. By numerically performing the integrations, we can check whether this happens or not. This is done in Fig. 5.2, on the right. It is clear from the plot that the three curves do not meet in a point and hence the perceptron is unable to find the correct minimum, i.e. the one dictated by the likelihood.

Let us summarise the findings of this section. We have analytically studied the behaviour of a perceptron in the context of a binary classification problem. The perceptron features a sigmoid activation function and a cross-entropy cost function. We have explicitly considered three variants of the primary N-subjettiness inputs: squared logarithms, logarithms and linear inputs. In the first two cases the minimum of the cost function does correspond to the configuration dictated by the likelihood, i.e. $a_1 = \ldots a_{n-1} = 0$, while this does not happen with linear inputs. This is most likely due to the fact that the simple perceptron with linear inputs struggles to correctly *learn* the probability distributions of signal and background, that are intrinsically logarithmic, although the configuration is $a_1 = \ldots a_{n-1} = 0$ is within the reach of the network. A more complex network would be needed in this case.

5.3 Perceptron numerics

In this section we validate our analytic findings with an actual implementation of a perceptron. Because our first-principle analysis has been developed at LL accuracy, in order to numerically test the perceptron performance we generate a sample of pseudo-data according to the QCD LL distribution for quark and gluon jets. The perceptron is trained using the same home-brewed implementation based on Ref. [156], with a training over 15 epochs. We consider the three different input variants also used in the analytic study, namely square logarithms, logarithms and the linear version of the N-subjettiness inputs. We use 1M events for the training in the first two cases, and 16M in the for linear inputs, unless otherwise stated. Furthermore, we perform our study as a function of number of N-subjettiness variables. Specifically, when we quote a given value of n, we imply that all $\mathfrak{T}_i$ with $i \leq n$ have been employed. The results of this study are collected in Fig. 5.3. Each plot shows the value of the network weights a_i and b after training, i.e. at the minimum of the cost function that has been found by the network through the steepest-descent algorithm. The plot on the left is for log-square inputs, the one in the middle for log inputs and the one on the right for linear inputs. The values of the weights
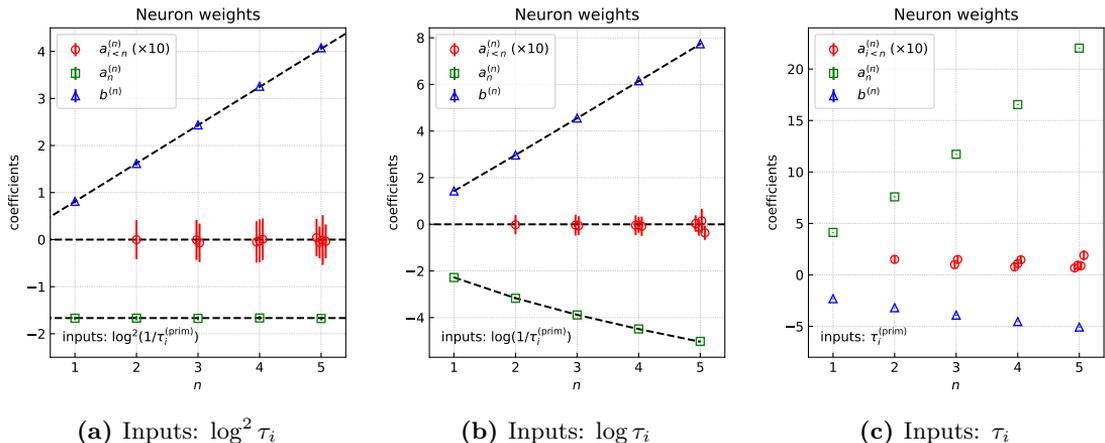
(a) Inputs: $\log^2 \tau_i$ **(b)** Inputs: $\log \tau_i$ **(c)** Inputs: τ_i

Figure 5.3: Perceptron parameters after training. When available, expected analytic results are shown as dashed lines.

determined by the network are shown as circles, squared and triangles, with the respective numerical uncertainties. In the case of log-square and log inputs, we also show the theoretically expected behaviour, derived in the previous section. We find perfect agreement. In particular, the minimum of the cost function which is found by the network exhibits $a_1 \cdots a_{n-1} = 0$, as dictated by the likelihood. This does not happen in the case of linear inputs, although the discrepancy is tiny.

It is interesting to investigate whether the theoretical issues for the linear inputs, which we have found analytically and confirmed with the study that we have just discussed, have some visible effect on the network performance. In order to do so, we first perform a study of the perceptron convergence as a function of the training sample size N_{train}. As before, we repeat this study for each of the input variants previously introduced, namely, square logarithms, logarithms and linear inputs. We also consider two different values of n: for the case $n = 3$ we build our inputs from $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$, while for $n = 5$, we also include $\mathcal{T}_4$ and $\mathcal{T}_5$. In Fig. 5.4 we plot the cost function as a function of the training sample size. Fig. 5.4 is obtained by training the perceptron with a progressively increasing sample of pseudo-data generated on the fly according to the QCD LL distribution. At fixed values of N_{train}, the cost function $\widetilde{C}$ of the trained network is then evaluated on a test sample of fixed dimension. This procedure is iterated a number of times, and at the end, for each N_{train}, we take the central value of the envelope (the solid line in Fig. 5.4), and the standard deviation of the results as a measure of the uncertainty (the error band in Fig. 5.4). The plots clearly show that the convergence with linear inputs is rather slower than in the cases of log-squares and logs, exposing the fact that the single-neuron network struggles to learn intrinsically logarithmic distributions with linear inputs.

Furthermore, with the same set up, we can study the actual performance of the network in terms of ROC curves. In order to expose possible, albeit small, deviations from in ideal case, instead of ROC curves, the plots Fig. 5.4 show:

$$\Delta\epsilon_g = \frac{\text{ROC}_{NN}}{\text{ROC}_{LL}} - 1, \tag{5.44}$$

121

(a) Inputs: $n = 3$

(b) Inputs: $n = 5$
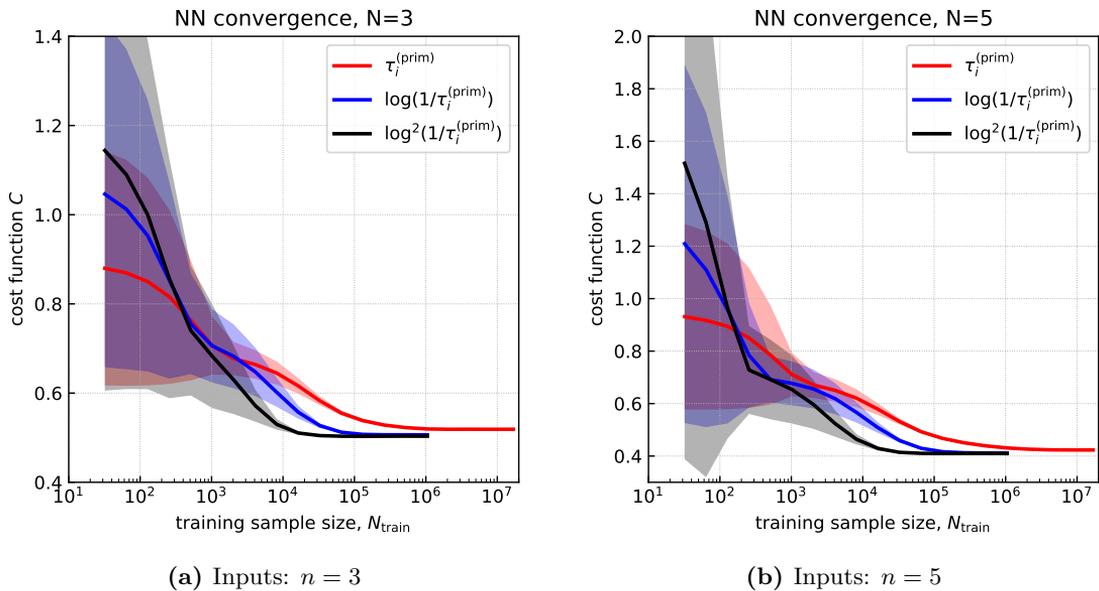
Figure 5.4: Convergence of the network as a function of the training sample size.
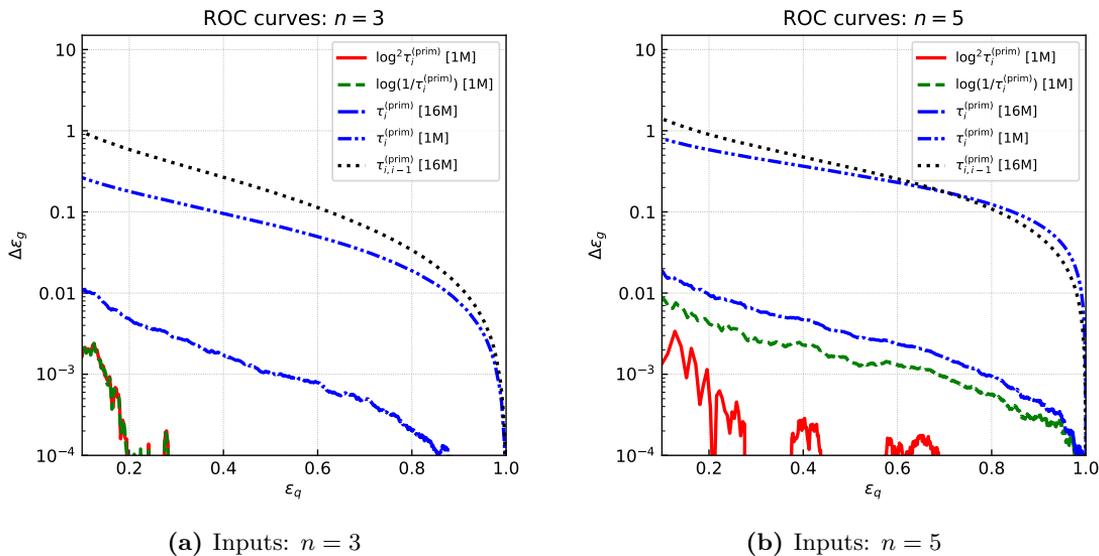

(a) Inputs: $n = 3$

(b) Inputs: $n = 5$

Figure 5.5: ROC curves obtained from a trained perceptron with different inputs.

where ROC_{NN} is the network ROC curve, while ROC_{LL} is the ROC curve that corresponds to a cut on $\mathcal{T}_n$. We perform this study for two different sizes of the training sample: 1M and 16M. We know from Fig.5.4 that former is enough to achieve convergence in the case of log-square and log inputs but not for the linear inputs. Thus, we expect to see larger differences in this case. This is indeed confirmed by the plots in Fig. 5.5. The dash-dot-dotted curve in blue indicates that the difference with respect the optimal case is of order 10% for $\epsilon_q < 0.6$ in the case of $n = 3$ and exceeds 30% in the same region, for $n = 5$. On the other hand, if the training sample size is increased to 16M, the perceptron with linear inputs always performs worse than the other choices, although the difference after training is rather minimal and never exceeds 1%. For comparison, we also show the results for N-subjettiness ratios as inputs. In this case, for $n = 3$, we include $\mathcal{T}_1$, $\mathcal{T}_{2,1} = \frac{\mathcal{T}_2}{\mathcal{T}_1}$ and $\mathcal{T}_{3,2} = \frac{\mathcal{T}_3}{\mathcal{T}_2}$. The perceptron performance in this case is much worse because a single-neutron network is not able to reconstruct the likelihood, i.e. a cut on $\mathcal{T}_3$, if it is passed N-subjettiness ratios as inputs.

5.4 Monte Carlo studies

In the previous section, we have validated the results of Sec. 5.2, using the same setting of the analytical calculations: namely, we considered a single-neuron network and we fed it with primary N-subjettiness variables (or functions thereof) evaluated on samples of pseudo-data generated according to QCD LL distribution. This setup is somehow oversimplified, and one may wonder how our results compare in term of performance and convergence to a fully-fledged neural network trained on Monte Carlo pseudo-data. In addition, we have adopted the primary N-subjettiness definition, whose nice analytical properties naturally justifies its use in our theoretical studies. However, this alternative definition so far has been compared to the standard definition only in terms of AUC at LL accuracy. Even though the purpose of this work is not a detailed comparison of the two definitions, we need to make sure that primary N-subjettiness performs sensibly as a quark/gluon tagger.

The purpose of this section is to address some of these concerns. We report some preliminary results obtained by extending the setup of Sec. 5.3 to a more realistic scenario, both in terms of the generation of pseudo-data and of the network architecture employed in the analysis.

First, we generate pseudo-data with Pythia 8.230 [191]. We simulate dijet events $qq \to qq$ and $gg \to gg$ and we cluster jets with the anti-k_t algorithm [68], as implemented in FastJet [70], with jet radius $R_0 = 0.4$. We keep jets with $p_t > 2$ TeV and $|y| < 4$. We then compute τ_N and $\mathcal{T}_N$ (i.e. both N-subjettiness definitions) up to the desired $N = n$ for each jet. We set the N-subjettiness parameter β to 1. For the standard N-subjettiness, we use the reference axes obtained from the exclusive k_t axis with winner-takes-all [192] recombination and a one-pass minimisation [193]. In addition to linear, log-square and log functional forms and N-subjettiness ratios already adopted in the previous section, we also consider a running-coupling-like input. This is obtained by evaluating the radiator $\mathcal{R}(\tau_i)$ of Eq. (1.99) with the one-loop approximation for the running coupling.[2]

The samples thus obtained are used to train either a single perceptron (as in the previous section) or a full neural network. The latter is a home-brewed implementation based on Ref. [156] with a fully connected network with 3 layers of 64 (sigmoid) neurons each and standard gradient descend. We plan to compare our results with a fully fledged NN as implemented in public codes.

---

[2]We set $\alpha_s(p_t R_0) = 0.09$, and we regulate the divergence at the Landau pole by freezing the coupling at a scale k_t such that $2\alpha_s(p_t R)\beta_0 \log(p_t R_0/k_t) = 0.99$.
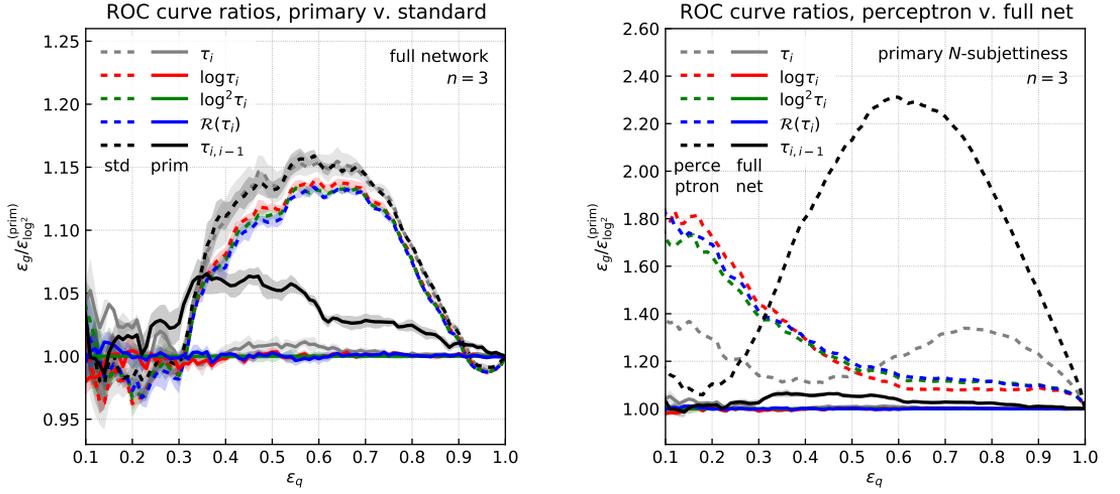
(a) Performance comparison between the standard (full) N-subjettiness (dashed lines) and the primary N-subjettiness (solid lines). All results are obtained using a full neural network.

(b) Performance comparison between a single perceptron (dashed lines) and a full neural network (solid lines). All results are obtained using primary N-subjettiness.

Figure 5.6: ROC curves obtained after training on a Pythia8 sample, using τ_1, τ_2 and τ_3. To better highlight differences the vertical axes show the gluon rate normalised by what is obtained using $\log^2 \mathcal{T}_i$ inputs for primary N-subjettiness trained on a full neural network. The different colours correspond to different inputs to the neural network.

We start with a comparison of the primary and the standard definition of N-subjettiness in terms of network performance. In Fig. 5.6a we plot the ROC curves for the primary (solid lines) and full (dashed line) definitions of N-subjettiness, and for different choices of the input functional form (different colours). The gluon efficiencies are normalised by the central value obtained with primary N-subjettiness with $\log^2(\mathcal{T}_i)$ inputs trained on a full NN. The central values and the error bands are calculated by taking the average and standard deviation of five different runs.

We see that the performance of NN trained with the full definition is worse by 10-15% compared to the primary definition for mid-values of the quark efficiency, $0.4 \lesssim \varepsilon_q \lesssim 0.8$ and comparable elsewhere except for a small region at large quark efficiency, $\varepsilon_q \gtrsim 0.9$. Even though the benefit of using primary N-subjettiness is not as large as one could have expected based on our leading-logarithmic calculations, we still observe a performance improvement. We note also that at small quark efficiency non-perturbative effects would have a sizeable effect, invalidating our arguments purely based on a perturbative calculation. Furthermore, large ε_q correspond to the regime where the cut on N-subjettiness is no longer small and our arguments based on the resummation of large logarithms no longer apply, and one should instead consider the full structure of hard matrix elements. However, as Fig. 5.6a testifies, it seems that for the most of the range in quark efficiency, a better discrimination is reached if we adopt the primary-N-subjettiness definition.

It is also interesting to compare the results obtained with different inputs in Fig. 5.6a. Although one should expect that full NN should ultimately converge to the log-likelihood inde-

pendently on the choice of inputs, our practical setup still shows a mild dependence on the choice of inputs. The key point is that the favoured inputs agree with our analytic expectations from Sec. 5.2 with logarithmic inputs ($\log \tau_i$, $\log^2 \tau_i$ and $\mathcal{R}(\tau_i)$) showing an equally good performance, the linear inputs only marginally worse and the ratio inputs showing a (few percent) worse discriminating power. This shows that an understanding of the underlying physical behaviour of the problem one addresses is helpful for deep-learning. We note however that even though the convergence of the neural network is more delicate for the case of ratio inputs (and, to a lesser extent, for linear inputs), a performance similar to the optimal one reported in Fig. 5.6a for logarithmic inputs could be achieved with a careful optimisation of the hyperparameters.

We now move to Fig. 5.6b, where we compare the ROC curves obtained with a full neural network to those of a simple perceptron. In this plot we select the primary N-subjettiness definition and we display results for the usual input functional forms. The solid lines in Fig. 5.6b coincide with the ones in Fig. 5.6a. First, we observe that a perceptron trained with Monte Carlo pseudo-data performs worse compared to the full NN for all the considered input types. This is not surprising as the arguments in Sec. 5.2 are based on a simplified, leading-logarithmic, approach and subleading correction can be expected to come with additional complexity that a single perceptron would fail to capture. It is actually remarkable that for $\varepsilon_q \gtrsim 0.6$ the performance of a single perceptron trained with logarithmic inputs gives performances which are only 10% worse that the full network. Comparing the different input types, we see that the perceptron performances in case of N-subjettiness ratios as inputs are remarkably worse compared to the performances we get by using the full NN, and this in agreement with the observation made near the end of Sec. 5.3: the more complex architecture of the NN is able to learn the correct weights for the ratio inputs, while the perceptron is not. The behaviour of the other functional forms is less clear, and dependent on the value of quark efficiency. For values of quark efficiency $0.5 \leq \varepsilon_q \leq 0.9$, which are phenomenologically relevant, we qualitatively recover the hierarchy among the different input functional forms already observed in Fig. 5.6a and in Sec. 5.3. Namely, the log-like inputs perform better than the linear inputs, which in turn perform better than the N-subjettiness ratios. However, for smaller values of quark efficiencies, this hierarchy seems to be inverted.[3]

5.A Details of the analytic calculations

In this appendix we report details of the analytic calculation of the area under the ROC curve (AUC). We start from the generic definition of the AUC and we exploit the definition of the ROC curve given in Eq. (1.105):

$$\text{AUC} = \int_0^1 dx \, \text{ROC}(x) = \int_0^1 dx \, \Sigma_\mathcal{B} \left(\Sigma_\mathcal{S}^{-1}(x) \right), \tag{5.45}$$

where, for a given observable v, the cumulative distribution for signal or background as a function of a cut v_{cut} reads:

$$\Sigma_i(v_{\text{cut}}) = \int d\mathbf{t} \, p_i(\mathbf{t}) \, \Theta(v(\mathbf{t}) < v_{\text{cut}}), \quad i = \mathcal{S}, \mathcal{B}. \tag{5.46}$$

---

[3]One possible reason for this is that at smaller quark efficiencies one is more sensitive to non-perturbative effects and that these are better characterised in terms of linear inputs, but a more detailed study would be needed to better understand this behaviour.

We know perform a change of variable from the efficiency x to the cut on the observable v_{cut}, which results in the following Jacobian factor:

$$\frac{dx}{dv_{\text{cut}}} = \frac{dx}{d\Sigma_{\mathcal{S}}^{-1}(x)} = \frac{d\Sigma_{\mathcal{S}}(v_{\text{cut}})}{dv_{\text{cut}}}. \tag{5.47}$$

Thus, we have

$$\text{AUC} = \int_0^1 dv_{\text{cut}} \, \Sigma_{\mathcal{B}}(v_{\text{cut}}) \, \frac{d\Sigma_{\mathcal{S}}(v_{\text{cut}})}{dv_{\text{cut}}} \tag{5.48}$$

$$= \int_0^1 dv_{\text{cut}} \int d\mathbf{t}_{\mathcal{B}} \, p_{\mathcal{B}}(\mathbf{t}_{\mathcal{B}}) \, \Theta(v(\mathbf{t}_{\mathcal{B}}) < v_{\text{cut}}) \int d\mathbf{t}_{\mathcal{S}} \, p_{\mathcal{S}}(\mathbf{t}_{\mathcal{S}}) \, \delta(v(\mathbf{t}_{\mathcal{S}}) - v_{\text{cut}}) \tag{5.49}$$

$$= \int d\mathbf{t}_{\mathcal{B}} \int d\mathbf{t}_{\mathcal{S}} \, p_{\mathcal{B}}(\mathbf{t}_{\mathcal{B}}) p_{\mathcal{S}}(\mathbf{t}_{\mathcal{S}}) \, \Theta(v(\mathbf{t}_{\mathcal{B}}) < v(\mathbf{t}_{\mathcal{S}})), \tag{5.50}$$

which is the expression presented in Eq. (5.14).

We now specialise to our case and we consider the probability distributions p_q and p_g for primary N-subjettiness at LL accuracy, which are given in Eq. (5.9). We first integrate over quark variables and exploiting the result found in Eq. (5.13), we obtain

$$\int_0^1 d\mathcal{T}_{1q} \cdots \int_0^{\mathcal{T}_{n-1,q}} d\mathcal{T}_{nq} \, p_q(\mathcal{T}_{1q}, \mathcal{T}_{2q}, \cdots, \mathcal{T}_{nq}) \, \Theta(\mathcal{T}_{nq} > \mathcal{T}_{ng}) = 1 - \frac{\Gamma(n, C_F \mathcal{R}(\mathcal{T}_{ng}))}{\Gamma(n)}. \tag{5.51}$$

The integral over $\mathcal{T}_{ng}$ can be performed using the following integration-by-parts identity:

$$\int_0^{\mathcal{T}} d\mathcal{T}' \, C_A \frac{\mathcal{R}'(\mathcal{T}')}{\mathcal{T}'} e^{-C_A \mathcal{R}(\mathcal{T}')} \, \Gamma(n, C_F \mathcal{R}(\mathcal{T}'))$$

$$= e^{-C_A \mathcal{R}(\mathcal{T})} \Gamma(n, C_F \mathcal{R}(\mathcal{T})) - \frac{C_F^n}{(C_F + C_A)^n} \Gamma(n, (C_F + C_A)\mathcal{R}(\mathcal{T})). \tag{5.52}$$

We see that iteratively the following kind of integrals appear:

$$\int_0^{\mathcal{T}} d\mathcal{T}' \, C_A \frac{\mathcal{R}'(\mathcal{T}')}{\mathcal{T}} \, \mathcal{R}(\mathcal{T}')^m \, \Gamma(n, (C_F + C_A)\mathcal{R}(\mathcal{T}'))$$

$$= \frac{C_A}{1+m} \left(\frac{\Gamma(n+1+m, (C_F + C_A)\mathcal{R}(\mathcal{T}))}{(C_F + C_A)^{1+m}} - \mathcal{R}(\mathcal{T})^{1+m} \Gamma(n, (C_F + C_A)\mathcal{R}(\mathcal{T})) \right). \tag{5.53}$$

In the end we obtain:

$$\int_0^1 d\mathcal{T}_{1g} \cdots \int_0^{\mathcal{T}_{n-1,g}} d\mathcal{T}_{ng} \, p_g(\mathcal{T}_{1g}, \mathcal{T}_{2g}, \cdots, \mathcal{T}_{ng}) \, \Gamma(n, C_F \mathcal{R}(\mathcal{T}_{ng}))$$

$$= \Gamma(n) - \frac{C_F^n}{(C_F + C_A)^{2n-1}} \sum_{k=0}^{n-1} \frac{\Gamma(n+k)}{\Gamma(k+1)} C_A^k (C_F + C_A)^{n-1-k}. \tag{5.54}$$

The sum can be performed explicitly, leading to the result presented in Eq. (5.15).

5.B   Analytic results for the cost function

In Section 5.2 we have looked for a minimum of the cost function, Eq. (5.27), by computing from the very beginning the derivatives with respect to the NN weights. This procedure allowed us to determine analytically the position of the minimum for the case of log-square inputs. An alternative approach would be to determine an analytic expression for the cost function, as a function of the NN weights and bias, which we would have to, in turn, minimise.

As we shall see in this appendix, this approach is not as successful as the one presented in the main text. The integrals that appear in the determination of the cost function Eq. (5.27) are rather challenging and we have been able to solve them only in a couple of simple cases. Namely, we limit ourselves to the case $n = 2$ with log-square inputs, by using the sigmoid as activation function, and we report results for the cross-entropy loss, Eq. (5.24), and the quadratic loss, Eq. (5.25). Even if the NN setup is minimal, the derivation of explicit expressions for the cost function may be instructive as they represent a first-principle determination of a NN behaviour and they could be valuable in the context of comparisons between experimental measurements that make use of machine-learning algorithms and theoretical predictions. Furthermore, it gives us the opportunity to analytically study how the neural network approaches the minimum during learning.

5.B.1   Cross-entropy loss

We start by considering the cost function Eq. (5.27) with the cross-entropy loss. Explicitly:

$$
\widetilde{C}^{(\mathrm{XE})}(a_1, a_2, b) = \frac{1}{2} \int_0^\infty dL_2 \int_0^{L_2} dL_1 \Bigg[\widetilde{C}_F^2 \, e^{-\widetilde{C}_F L_2} \left(-\log\left(\frac{1}{1 + e^{a_1 L_1 + a_2 L_2 + b}} \right) \right)
$$
$$
+ \widetilde{C}_A^2 \, e^{-\widetilde{C}_A L_2} \left(-\log\left(\frac{1}{1 + e^{-a_1 L_1 - a_2 L_2 - b}} \right) \right) \Bigg]. \quad (5.55)
$$

Note that the replacement $\widetilde{C}_F \leftrightarrow \widetilde{C}_A$ is equivalent to $(a_1, a_2, b) \leftrightarrow (-a_1, -a_2, -b)$, due to the symmetries of the functions involved. We first observe that the integral over L_1 gives rise to a dilogarithmic function. In order to evaluate the integral over L_2 we make use of the following result:

$$
\mathcal{I}^{(\mathrm{XE})}(C_R, c, d) = C_R^2 \int_0^\infty dx \, e^{-C_R x} \operatorname{Li}_2\left(-e^{cx+d} \right) \quad (5.56)
$$
$$
= C_R \operatorname{Li}_2\left(-e^d \right) + \begin{cases} -c \log\left(1 + e^d\right) - \frac{c^2}{C_R}\left(1 - {}_2F_1\left(1, -\frac{C_R}{c}, 1 - \frac{C_R}{c}; -e^d\right)\right) & (c < 0) \\ 0 & (c = 0) \\ -c \log\left(1 + e^d\right) - \frac{c^2}{C_R} {}_2F_1\left(1, \frac{C_R}{c}, 1 + \frac{C_R}{c}; -e^{-d}\right) & (c > 0) \end{cases}
$$

The analytic expression that we obtain for the cost function is

$$
\widetilde{C}^{(\mathrm{XE})}(a_1, a_2, b) = \frac{1}{2a_1} \Bigg[-(\mathcal{I}^{(\mathrm{XE})}(\widetilde{C}_F, a_1 + a_2, b) - \mathcal{I}^{(\mathrm{XE})}(\widetilde{C}_F, a_2, b))
$$
$$
+ (\mathcal{I}^{(\mathrm{XE})}(\widetilde{C}_A, -(a1 + a_2), -b) - \mathcal{I}^{(\mathrm{XE})}(\widetilde{C}_A, -a_2, -b)) \Bigg]. \quad (5.57)
$$

127

From Section 5.2, we already now that the minimum is located at $a_1 = 0$. However, given the structure of the explicit result after integration, Eq. (5.57), it is highly nontrivial to recover the position of the minimum analytically, due to the presence of a_1 both at denominator and in the arguments of hypergeometric function.

5.B.2   Quadratic loss

An analogous calculation can be performed in the case of the quadratic loss Eq. (5.25). We have to calculate

$$\widetilde{C}^{(\chi^2)}(a_1, a_2, b) = \frac{1}{2} \int_0^\infty dL_2 \int_0^{L_2} dL_1 \left[\frac{\widetilde{C}_F^2 \, e^{-\widetilde{C}_F L_2}}{(1 + e^{-a_1 L_1 - a_2 L_2 - b})^2} + \frac{\widetilde{C}_A^2 \, e^{-\widetilde{C}_A L_2}}{(1 + e^{a_1 L_1 + a_2 L_2 + b})^2} \right]. \quad (5.58)$$

As in the case of the cross-entropy loss, the integral over L_1 is straightforward. We then make use of the following identity to perform the remaining integral over L_2:

$$\mathcal{I}^{(\chi^2)}(C_R, c, d) = C_R^2 \int_0^\infty dx \, e^{-C_R x} \left[\log\left(1 + e^{cx+d}\right) + \frac{1}{1 + e^{cx+d}} \right]$$

$$= C_R \log\left(1 + e^d\right) + \begin{cases} c + (C_R - c) \, {}_2F_1\left(1, -\frac{C_R}{c}, 1 - \frac{C_R}{c}; -e^d\right) & (c < 0) \\ \dfrac{C_R}{1 + e^d} & (c = 0) \\ C_R + (c - C_R) \, {}_2F_1\left(1, \frac{C_R}{c}, 1 + \frac{C_R}{c}; -e^{-d}\right) & (c > 0) \end{cases} \quad (5.59)$$

We find

$$\widetilde{C}^{(\chi^2)}(a_1, a_2, b) = \frac{1}{2a_1} \left[\left(\mathcal{I}^{(\chi^2)}(\widetilde{C}_F, a_1 + a_2, b) - \mathcal{I}^{(\chi^2)}(\widetilde{C}_F, a_2, b) \right) \right.$$

$$\left. - \left(\mathcal{I}^{(\chi^2)}(\widetilde{C}_A, -(a1 + a_2), -b) - \mathcal{I}^{(\chi^2)}(\widetilde{C}_A, -a_2, -b) \right) \right]. \quad (5.60)$$

As for Eq. (5.57), the position of the minimum is hard to find analytically. To check whether we recover the result obtained in Eq. (5.36), we can numerically look for the minimum of Eqs. (5.57) and (5.60). For illustration purposes, we fix $\widetilde{C}_A$ and $\widetilde{C}_F$ to C_A and C_F respectively. In Fig. 5.7 we plot each cost function around the found minimum point. We see that the condition $a_1 = 0$ and the (negative) value of $a_2 = C_F - C_A \simeq -1.67$ are indeed confirmed.

5.C   Scaling of a_n in the log inputs case.

In this appendix we would like to study the scaling of a_n with respect to $\widetilde{C}_F$ and $\widetilde{C}_A$. In order to simplify the notation, in this appendix we will rename $x \equiv l_n$ and $a \equiv a_n$.

We first expand in series around $x = 0$ the sigmoid functions which appear in Eq. (5.37):

$$\frac{1}{1 + e^{-a\,x - b}} = \frac{e^b}{1 + e^b} + \sum_{k=1}^\infty a^k \, f^{(k)}(e^b) \, x^k \,, \quad (5.61)$$

$$-\frac{1}{1 + e^{a\,x + b}} = -\frac{1}{1 + e^b} + \sum_{k=1}^\infty a^k \, f^{(k)}(e^b) \, x^k \,, \quad (5.62)$$

(a) Cross-entropy loss
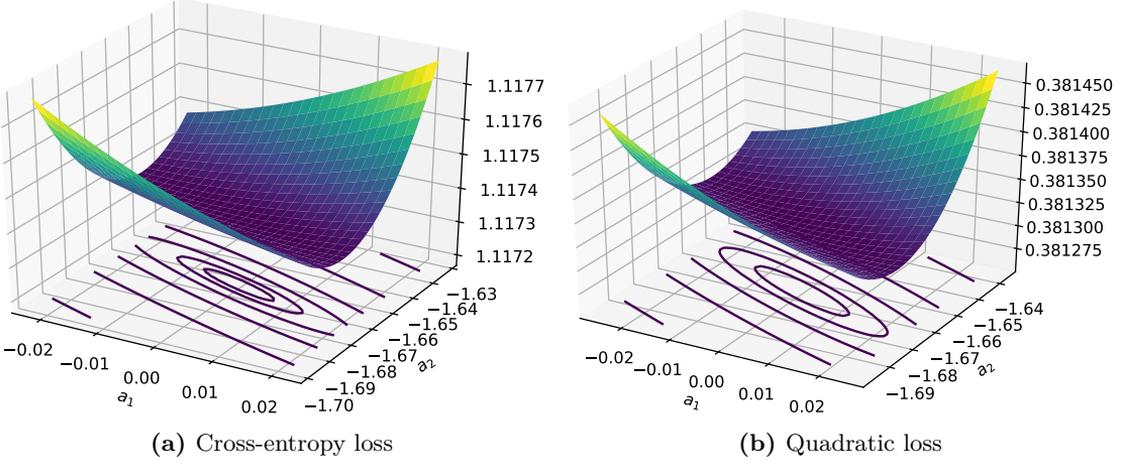
(b) Quadratic loss

Figure 5.7: Cost function as a function of a_1 and a_2 around the minimum point. b has been accordingly fixed to the value in Eq. (5.36).

where the explicit form of $f^{(k)}(e^b)$ is not of relevance to our purposes. Note that the two series with index k appearing on the r.h.s. of Eqs. (5.61)-(5.62) are the same. By substituting Eqs. (5.61)-(5.62) under integration, Eq. (5.37) becomes:

$$\int_0^\infty dx\, x^m \left[\widetilde{C}_F^n\, e^{-\widetilde{C}_F x^2} \left(\frac{e^b}{1+e^b} + \sum_{k=1}^\infty a^k\, f^{(k)}(e^b)\, x^k \right) \right.$$
$$\left. + \widetilde{C}_A^n\, e^{-\widetilde{C}_A x^2} \left(-\frac{1}{1+e^b} + \sum_{k=1}^\infty a^k\, f^{(k)}(e^b)\, x^k \right) \right] = 0 \tag{5.63}$$

with $m = 2n$ or $m = 2n-1$. Given the following integration identity:

$$\int_0^\infty dx\, x^m\, e^{-c\,x^2} = \frac{1}{2\sqrt{c^{m+1}}} \Gamma\left(\frac{m+1}{2} \right), \quad c > 0 \tag{5.64}$$

by also dividing by $\sqrt{\widetilde{C}_F^{2n-m-1}}$ we obtain:

$$\frac{\Gamma\left(\frac{m+1}{2}\right)}{2(1+e^b)} \left(e^b - \left(\frac{\widetilde{C}_A}{\widetilde{C}_F} \right)^{(2n-m-1)/2} \right)$$
$$+ \sum_{k=1}^\infty \left(\frac{a}{\sqrt{\widetilde{C}_F}} \right)^k f^{(k)}(e^b) \frac{\Gamma\left(\frac{m+1+k}{2}\right)}{2} \left(1 + \left(\frac{\widetilde{C}_A}{\widetilde{C}_F} \right)^{(2n-m-1-k)/2} \right) = 0 . \tag{5.65}$$

This equation suggests the relation quoted in the main text:

$$\frac{a_n}{\sqrt{\widetilde{C}_F}} = F\left(\frac{\widetilde{C}_A}{\widetilde{C}_F} \right) . \tag{5.66}$$

CHAPTER 6

Electron PDFs

As we have discussed at length in the previous chapters, theoretical predictions for processes at hadron colliders always imply the introduction of long-distance objects, like the hadronic PDFs, whose presence is required by the non-perturbative nature of QCD. Conversely, the typical cross section relevant to e^+e^- collisions is in principle entirely computable as a perturbative series in the QED coupling constant α. In practice, however, this is hardly useful. The coefficients of such a series are very large and prevent the series from being well behaved. The problem stems from the fact that the incoming $e^\pm$ particles tend to copiously radiate photons at small angles w.r.t. the beamline. In perturbation theory, any zero-angle emission would induce a divergent cross section, were it not for the screening effect provided by the mass of the emitter and/or the emitted particle. This has already been observed in Sec. 1.4, where we also showed that, when integrating over all possible emissions, the cross section will contain logarithms of the ratio m^2/E^2, where E is a scale of the order of the hardness of the process, and m is the screening mass i.e. that of the electron. It is these logarithms that, by growing large when $m^2/E^2 \ll 1$, give the dominant contributions to the perturbative coefficients, thus compensating the suppression due to α. In other words, all terms of the series might be of the same order numerically, which leads to a complete loss of predictive power.

Fortunately, such $\log(m^2/E^2)$ terms are universal, and because of this they can be taken into account to all orders in α by a process-independent resummation procedure. In the so-called *structure-function* approach, the physical cross section is written by means of a factorisation formula that recalls the standard QCD factorization formula at hadron colliders, Eq. (1.64). Order by order, one collects all of the logarithmic terms in the Parton Distribution Functions (PDFs) of the electron (or of the positron). Then, by means of the DGLAP evolution equations, Eq. (1.70), one is able to resums such terms to all orders.

QED electron PDFs, at variance with hadronic PDFs, are entirely calculable with perturbative techniques. Presently they are known in close analytical forms [194–196] which are leading-logarithmic (LL) accurate, and that include all-order in α contributions in the region $z_\pm \simeq 1$ (which is responsible for the bulk of the cross section), matched with up to $\mathcal{O}(\alpha^3)$ terms for any values of $z_\pm$; both of these forms exploit leading-order (LO) initial conditions.

The goal of this chapter is to improve on the results of Refs. [194–196] by extending them to the next-to-leading logarithm accuracy (NLL) starting from the next-to-leading order (NLO) initial conditions computed in Ref. [42]. In keeping with what was done in the literature, we shall present predictions both for all-order PDFs in the $z_\pm \simeq 1$ region, and for up to $\mathcal{O}(\alpha^3)$ NLL terms valid for any $z_\pm$. By working at the NLL+NLO accuracy, the mixing between the electron/positron and the photon PDFs is taken into proper account, as are running-α effects. Our results are obtained with both analytical and numerical methods, which are compared and used to validate each other.

The chapter is organised as follows. In Sec. 6.1 we report the initial conditions for the evolution of the electron PDFs, while Sec. 6.2 briefly describes the evolution-operator formalism. We then move to the actual solution of the evolution equations. We obtain the NLL-accurate PDFs of the electron in closed analytical forms in two different ways. First, in Sec. 6.3, by solving the evolution equations order by order in perturbation theory. Second, in Sec. 6.4 and in Sec. 6.5 for the non-singlet and the singlet-photon sector respectively, by using the properties of the evolution operator to obtain the asymptotic behaviour in the $z \to 1$ region to all orders in α. These two analytical results can then be combined in order to obtain predictions which are numerically well-behaved in the whole of the z range. This is done in Sec. 6.7 by means of an additive matching. Afterwards, in Sec. 6.8, we describe the code employed to obtain numerical results. Finally, our analytical and numerical predictions are extensively compared in Sec. 6.9.

6.1 Initial conditions for the electron PDFs

In the structure-function approach, we write a generic cross section at e^+e^- colliders as follows:

$$
\begin{aligned}
d\bar{\sigma}_{e^+e^-}(p_{e^+}, p_{e^-}, m^2) &= \sum_{ij=e^\pm,\gamma} \int dz_+ dz_-\, \Gamma_{i/e^+}(z_+, \mu^2, m^2)\, \Gamma_{j/e^-}(z_-, \mu^2, m^2) \\
&\quad \times d\hat{\sigma}_{ij}(z_+ p_{e^+}, z_- p_{e^-}, \mu^2).
\end{aligned}
\tag{6.1}
$$

Let us describe the various terms present in Eq. (6.1).

- $d\bar{\sigma}_{e^+e^-}$ is a the *particle-level* cross section, defined so as to retain only terms that do not vanish in the $m/E \to 0$ limit. In other words, it is obtained by computing the cross section $d\sigma_{e^+e^-}$ with massive electrons, by Taylor-expanding the result, and by keeping only the terms either proportional to a logarithm (possibly to some power) of m^2/s, or are independent of m.

- $d\hat{\sigma}_{ij}$ is a *parton-level* cross section, understood to be computed with massless electrons. As in QCD, a suitable zero-mass subtraction scheme must be introduced (e.g. $\overline{\text{MS}}$) and a factorization scale μ appears. $d\hat{\sigma}_{ij}$ does not contain any logarithmic term, and is expected to be well-behaved order by order in perturbation theory.

- $\Gamma_{i/e^\pm}$ are the Parton Distribution Functions (PDFs) of the electron or the positron, a name that originates from the analogy of Eq. (6.1) with its QCD counterpart. By working in QED the cases of the electron and of the positron PDFs are identical. Thus, in this chapter we shall only consider the PDFs of the electron, which we simply denote as Γ_i, with parton index i equal to $i = e^-, e^+, \gamma$.

Note that neither $\Gamma_{i/e^\pm}$ nor $d\hat{\sigma}_{ij}$ are physical quantities; their definitions always involve some degree of arbitrariness, which is parametrised by the mass scale μ, that is only constrained by the requirement $\mu \sim E$, and by the chosen factorisation scheme. Fuller details on the usage of the factorisation formula (6.1) in calculations relevant to e^+e^- colliders and on its physical meaning can be found in Ref. [42].

A necessary ingredient for the collinear evolution through DGLAP equations are the initial conditions for such an evolution. By "initial condition" we mean the expression for Γ_i at the initial scale μ_0. The electron PDFs can be expanded perturbatively, and we denote such an expansion as:

$$\Gamma_i(z, \mu_0^2) = \Gamma_i^{[0]}(z, \mu_0^2) + \frac{\alpha}{2\pi}\, \Gamma_i^{[1]}(z, \mu_0^2) + \mathcal{O}(\alpha^2)\,. \tag{6.2}$$

If the initial scale μ_0 is chosen to be of the same order of the electron mass m, the higher order terms in Eq. (6.2) can be safely neglected, and the fixed order expansion of Γ_i at the scale μ_0 is assumed to be the initial condition for the evolution. Needless to say, the accuracy required for initial condition depends on the accuracy of the resummation we are interested in. For instance, if we aim at a LL resummation, we only need the zeroth-order term, whose expression is given by physically obvious condition:

$$\Gamma_i^{[0]}(z, \mu_0^2) = \delta_{ie^-}\delta(1-z)\,, \tag{6.3}$$

i.e. at leading order no collinear splittings are allowed. Instead, the prerequisite for an evolution accurate at NLL level is the explicit expression for the NLO coefficient, $\Gamma_i^{[1]}$.

Such a coefficient has been calculated in Ref. [42], within the FKS subtraction method [98, 197], in two different ways:

- the factorization formula Eq. (6.1) is seen as a definition for Γ_i at a given μ. Thus, one finds both $d\bar{\sigma}_{e^+e^-}$ and $d\hat{\sigma}_{ij}$, and then solves for Γ_i. This requires an explicit computation of the particle-level and parton-level cross sections for a specific process; however, the final result does not depend on the arbitrary process chosen in the calculation.

- by exploiting the universal factorization properties in the collinear limit, one is able to calculate the difference between the massive and the massless cross sections — which is the only thing that matters for the determination of Γ_i — for z strictly less than 1. The distributional part of Γ_i is eventually determined by the momentum-conservation condition:

$$\int_0^1 dz\, z \Big[\Gamma_{e^-}(z) + \Gamma_\gamma(z) + \Gamma_{e^+}(z)\Big] = 1\,, \tag{6.4}$$

or by the charge-conservation condition:

$$\int_0^1 dz \Big[\Gamma_{e^-}(z) - \Gamma_{e^+}(z)\Big] = 1\,. \tag{6.5}$$

133

With both strategies we get the same $\mathcal{O}(\alpha)$ initial conditions, which read as follows:

$$\Gamma_{e^-}^{[1]}(z, \mu_0^2) = \left[\frac{1 + z^2}{1 - z}\left(\log\frac{\mu_0^2}{m^2} - 2\log(1 - z) - 1\right)\right]_+ + K_{ee}(z)\,, \tag{6.6}$$

$$\Gamma_{\gamma}^{[1]}(z, \mu_0^2) = \frac{1 + (1 - z)^2}{z}\left(\log\frac{\mu_0^2}{m^2} - 2\log z - 1\right) + K_{\gamma e}(z)\,, \tag{6.7}$$

$$\Gamma_{e^+}^{[1]}(z, \mu_0^2) = 0\,, \tag{6.8}$$

where $\mu_0 \simeq m$, and m is the electron mass. The rightmost terms on the r.h.s. of Eqs. (6.6) and (6.7) are associated with, and fully determined by, the scheme used to subtract the initial-state collinear singularities. In this chapter, we work in the $\overline{\text{MS}}$ scheme, which implies:

$$K_{ee}(z) = K_{\gamma e}(z) = 0 \quad \Longleftrightarrow \quad \overline{\text{MS}}\,. \tag{6.9}$$

Eq. (6.6), apart from colour factors, coincides with the initial condition for the perturbative heavy quark[1] fragmentation function in QCD, first calculated in Ref. [198]. This is hardly surprising. The way the structure-function approach in QED addresses the problem of potentially large logarithms of collinear origin is indeed QCD-inspired, as the factorization formula Eq. (6.1) conveys. Both in the case of electronic PDFs and in the case of heavy quark fragmentation functions, one wants to resum the collinearly-enhanced perturbative contributions, in the former case due to initial state photon emissions, in the latter case due to final state gluon emissions.

6.2 Evolution operator formalism

As far as the evolution in μ is concerned, DGLAP equations for singlet, Eq. (1.77), and non-singlet, Eq. (1.75) are identical. We shall thus start dealing with the former one, which has a more involved flavour structure; the results will then be applied to the non-singlet case as well, by simply considering a one-dimensional flavour space. We re-write Eq. (1.77) by means of a simpler notation, where all of the irrelevant indices are dropped:

$$\frac{\partial\Gamma(z, \mu^2)}{\partial\log\mu^2} = \frac{\alpha(\mu)}{2\pi}\big[\mathbb{P} \otimes \Gamma\big](z, \mu^2)\,, \tag{6.10}$$

and Γ is a column vector. As already pointed out in Sec. 1.6, in Mellin space the convolution becomes a standard matrix multiplication:

$$\frac{\partial\Gamma_N(\mu^2)}{\partial\log\mu^2} = \frac{\alpha(\mu)}{2\pi}\,\mathbb{P}_N(\mu)\,\Gamma_N(\mu^2) = \sum_{k=0}^{\infty}\left(\frac{\alpha(\mu)}{2\pi}\right)^{k+1}\mathbb{P}_N^{[k]}\,\Gamma_N(\mu^2)\,. \tag{6.11}$$

By denoting by $\Gamma_{0,N} \equiv \Gamma_N(\mu_0^2)$ the PDF initial conditions at the reference scale μ_0, and by introducing an evolution operator $\mathbb{E}_N(\mu^2, \mu_0^2)$ such that:

$$\Gamma_N(\mu^2) = \mathbb{E}_N(\mu^2, \mu_0^2)\,\Gamma_{0,N}\,, \qquad \mathbb{E}_N(\mu_0^2, \mu_0^2) = I\,, \tag{6.12}$$

---

[1]Heavy quarks are quarks such as charm or bottom whose mass m is large enough ($m \gg \Lambda_{\text{QCD}}$) to be considered in the perturbative domain.

Eq. (6.11) becomes:

$$\frac{\partial \mathbb{E}_N(\mu^2, \mu_0^2)}{\partial \log \mu^2} \Gamma_{0,N} = \sum_{k=0}^{\infty} \left(\frac{\alpha(\mu)}{2\pi} \right)^{k+1} \mathbb{P}_N^{[k]} \, \mathbb{E}_N(\mu^2, \mu_0^2) \, \Gamma_{0,N} \,. \tag{6.13}$$

Since Eq. (6.13) must be true regardless of the specific choice for $\Gamma_{0,N}$, it is equivalent to:

$$
\begin{aligned}
\frac{\partial \mathbb{E}_N(\mu^2, \mu_0^2)}{\partial \log \mu^2} &= \sum_{k=0}^{\infty} \left(\frac{\alpha(\mu)}{2\pi} \right)^{k+1} \mathbb{P}_N^{[k]} \, \mathbb{E}_N(\mu^2, \mu_0^2) \\
&= \frac{\alpha(\mu)}{2\pi} \left[\mathbb{P}_N^{[0]} + \frac{\alpha(\mu)}{2\pi} \mathbb{P}_N^{[1]} \right] \mathbb{E}_N(\mu^2, \mu_0^2) + \mathcal{O}(\alpha^2) \,.
\end{aligned} \tag{6.14}
$$

Following Ref. [199], it is appropriate to introduce the variable[2]:

$$t = \frac{1}{2\pi b_0} \log \frac{\alpha(\mu)}{\alpha(\mu_0)} \,. \tag{6.15}$$

Equation (6.15) implies that:

$$\frac{\partial}{\partial \log \mu^2} = \frac{1}{2\pi b_0} \frac{\beta(\alpha(\mu))}{\alpha(\mu)} \frac{\partial}{\partial t} \,, \tag{6.16}$$

and thus:

$$\frac{\partial \alpha(\mu)}{\partial t} = 2\pi b_0 \alpha(\mu) \quad \Longrightarrow \quad \alpha(\mu) = \alpha(\mu_0) e^{2\pi b_0 t} \,. \tag{6.17}$$

With Eq. (6.16), Eq. (6.14) becomes[3]:

$$
\begin{aligned}
\frac{\partial \mathbb{E}_N(t)}{\partial t} &= \frac{b_0 \alpha^2(\mu)}{\beta(\alpha(\mu))} \sum_{k=0}^{\infty} \left(\frac{\alpha(\mu)}{2\pi} \right)^k \mathbb{P}_N^{[k]} \, \mathbb{E}_N(t) \\
&= \left[\mathbb{P}_N^{[0]} + \frac{\alpha(\mu)}{2\pi} \left(\mathbb{P}_N^{[1]} - \frac{2\pi b_1}{b_0} \mathbb{P}_N^{[0]} \right) \right] \mathbb{E}_N(t) + \mathcal{O}(\alpha^2) \,.
\end{aligned} \tag{6.18}
$$

Note that, from Eq. (6.12), $\mathbb{E}_N(t = 0) = I$.

If the flavour space is one-dimensional (as for the non-singlet evolution), Eq. (6.18) can be solved analytically. Notation-wise, we deal with this case by performing the formal replacements:

$$\mathbb{E}_N \longrightarrow E_N \,, \qquad \mathbb{P}_N^{[k]} \longrightarrow P_N^{[k]} \,. \tag{6.19}$$

By exploiting Eq. (6.17), one readily obtains:

$$\log E_N = P_N^{[0]} \, t + \frac{1}{4\pi^2 b_0} \left(\alpha(\mu) - \alpha(\mu_0) \right) \left(P_N^{[1]} - \frac{2\pi b_1}{b_0} P_N^{[0]} \right) + \mathcal{O}(\alpha^3) \,. \tag{6.20}$$

By construction, the $\mathcal{O}(\alpha^3)$ terms neglected in Eq. (6.20) stem from the truncation of the series that gives the evolution kernels in Eq. (1.71); conversely, the relationship between $\alpha(\mu)$ and $\alpha(\mu_0)$

---

[2] This differs by a minus sign w.r.t. that of QCD, since it is convenient to still have $t > 0$ for $\mu > \mu_0$.

[3] As the argument of the evolution operator, we shall use t interchangeably with the pair (μ, μ_0).

is treated exactly thanks to the usage of the variable t. If one wants to expose explicitly the large logarithms that originate from having $\mu \gg \mu_0$, one can use the following series expansions:

$$\alpha(\mu_0) = \alpha(\mu) - \alpha^2(\mu)b_0 L + \alpha^3(\mu)\left(b_0^2 L^2 - b_1 L\right) + \mathcal{O}(\alpha^4)\,, \tag{6.21}$$

$$t = \frac{\alpha(\mu)}{2\pi}L - \frac{\alpha^2(\mu)}{4\pi}\left(b_0 L^2 - \frac{2b_1}{b_0}L\right) + \mathcal{O}(\alpha^3)\,, \tag{6.22}$$

having defined:

$$L = \log\frac{\mu^2}{\mu_0^2}\,. \tag{6.23}$$

By employing these results, Eq. (6.20) becomes:

$$\log E_N = \frac{\alpha(\mu)}{2\pi}P_N^{[0]}L + \left(\frac{\alpha(\mu)}{2\pi}\right)^2\left(P_N^{[1]}L - \pi b_0 P_N^{[0]}L^2\right) + \mathcal{O}(\alpha^3)\,. \tag{6.24}$$

This result is useful because, at variance with that of Eq. (6.20), it allows one to consider the case of a non-running α, which can simply be obtained from Eq. (6.24) in the limit $b_0 \to 0$. As a consistency check, it is immediate to verify that, by taking such a limit, one arrives at a form for $\log E_N$ which could have been directly obtained from Eq. (6.14), by working in a one-dimensional flavour space and by freezing $\alpha(\mu)$ there.

6.3   Analytical recursive solutions

Following Ref. [196], perturbative solutions for the evolution equations can conveniently be obtained by re-writing Eq. (6.10) in an integral form:

$$\frac{\partial \mathcal{F}(z,\mu^2)}{\partial \log \mu^2} = \frac{\alpha(\mu)}{2\pi}\left[\mathbb{P}\,\overline{\otimes}\,\mathcal{F}\right](z,\mu^2)\,, \tag{6.25}$$

with[4]:

$$\mathcal{F}(z,\mu^2) = \int_0^1 dy\,\Theta(y-z)\,\Gamma(y,\mu^2) \quad\Longrightarrow\quad \Gamma(z,\mu^2) = -\frac{\partial}{\partial z}\mathcal{F}(z,\mu^2)\,, \tag{6.26}$$

and the modified convolution operator defined as follows:

$$g\,\overline{\otimes}\,_z h = \int_0^1 dx\,\Theta(x-z)\,g(x)\,h(z/x) = \bar{g}\otimes_z h\,, \qquad \bar{g}(x) = xg(x)\,, \tag{6.27}$$

which is a valid definition regardless of whether $g(x)$ is a distribution or an ordinary function. Note that $\mathcal{F}$ is a column vector, and that Eq. (6.25) has a matrix structure, in the flavour space. As was the case for Eq. (6.10), this implies that all of the results to be obtained in the following can be applied to the limiting situation of a one-dimensional flavour space as well.

The procedure of Ref. [196] is LL-accurate. In order to generalise it to the NLL accuracy we are interested in in this work, it is best to first consider the case of non-running α. With this

---

[4]The use of a Θ function in Eq. (6.26) guarantees its validity also when Γ is a distribution, and thus allows one to take into account its possible endpoint contributions. Conversely, while $\mathcal{F}$ should also be treated as a distribution, we shall regard it as an ordinary function, because in the large-z region we shall in any case employ the asymptotic solutions whose results, given in Sec. 6.4, are more accurate there.

assumption, the solution of Eq. (6.25) can formally be written as follows:

$$\mathcal{F}(z,\mu^2) = \mathcal{F}(z,\mu_0^2) + \frac{\alpha}{2\pi} \int_{\log \mu_0^2}^{\log \mu^2} d\log {\mu'}^2 \, [\mathbb{P} \,\overline{\otimes}\, \mathcal{F}] \, (z,{\mu'}^2) \, . \tag{6.28}$$

From this equation, $\mathcal{F}$ can be obtained by representing it by means of a power series:

$$\mathcal{F}(z,\mu^2) = \sum_{k=0}^{\infty} \frac{\eta_0^k}{2^k k!} \left(\mathcal{I}_k^{\mathrm{LL}}(z) + \frac{\alpha}{2\pi} \mathcal{I}_k^{\mathrm{NLL}}(z) \right) , \tag{6.29}$$

where:

$$\eta_0 = \frac{\alpha}{\pi} L \, , \qquad L = \log \frac{\mu^2}{\mu_0^2} \, , \tag{6.30}$$

and with $\mathcal{I}_k^{\mathrm{LL}}$ and $\mathcal{I}_k^{\mathrm{NLL}}$ two sets of unknown functions. More precisely, $\mathcal{I}_k^{\mathrm{LL}}$ and $\mathcal{I}_k^{\mathrm{NLL}}$ are two-dimensional column vectors in the singlet-photon flavour space, whose elements are functions of z, and c-number functions in the non-singlet flavour space. By replacing Eq. (6.29) into Eq. (6.28), the two sides of the latter equation become two series in η_0: one then equates the coefficients relevant to the same power of η_0 on the two sides, thereby obtaining equations that can be solved for $\mathcal{I}_k^{\mathrm{LL}}$ and $\mathcal{I}_k^{\mathrm{NLL}}$ (recursively in k). The r.h.s. of Eq. (6.29) is simply an expansion in terms of αL, and thus η_0 is a convenient expansion parameter, irrespective of the logarithmic accuracy one is working at. Indeed, Eq. (6.29) can be extended by adding further contributions to its r.h.s., that are suppressed by higher powers of α. Conversely, by keeping only the $\mathcal{I}_k^{\mathrm{LL}}$ contributions, one recovers what was done in Ref. [196]. The recursive solutions for $\mathcal{I}_k^{\mathrm{LL}}$ and $\mathcal{I}_k^{\mathrm{NLL}}$ stemming from Eq. (6.29) read as follows:

$$\mathcal{I}_k^{\mathrm{LL}} = \mathbb{P}^{[0]} \,\overline{\otimes}\, \mathcal{I}_{k-1}^{\mathrm{LL}} \, , \tag{6.31}$$

$$\mathcal{I}_k^{\mathrm{NLL}} = \mathbb{P}^{[0]} \,\overline{\otimes}\, \mathcal{I}_{k-1}^{\mathrm{NLL}} + \mathbb{P}^{[1]} \,\overline{\otimes}\, \mathcal{I}_{k-1}^{\mathrm{LL}} \, , \tag{6.32}$$

with:

$$\mathcal{I}_0^{\mathrm{LL}} = \mathcal{F}^{[0]}(z,\mu_0^2) \, , \qquad \mathcal{I}_0^{\mathrm{NLL}} = \mathcal{F}^{[1]}(z,\mu_0^2) \, . \tag{6.33}$$

The quantities in Eq. (6.33) must be obtained by direct computation by using the definition in Eq. (6.26), with the perturbative expansion of Eq. (6.2) and the initial conditions of Eqs. (6.3)–(6.8). By doing so, we obtain:

$$\mathcal{I}_{\mathrm{S},0}^{\mathrm{LL}} = \mathcal{I}_{\mathrm{NS},0}^{\mathrm{LL}} = 1 \, , \tag{6.34}$$

$$\mathcal{I}_{\gamma,0}^{\mathrm{LL}} = 0 \, , \tag{6.35}$$

$$\mathcal{I}_{\mathrm{S},0}^{\mathrm{NLL}} = \mathcal{I}_{\mathrm{NS},0}^{\mathrm{NLL}} = 2z + (1 - 2z - z^2) \log(1-z) - 2\log^2(1-z)$$
$$+ \left[z + z^2/2 + 2\log(1-z) \right] \log \frac{\mu_0^2}{m^2} \, , \tag{6.36}$$

$$\mathcal{I}_{\gamma,0}^{\mathrm{NLL}} = -2(1-z) + (2 - 4z + z^2) \log z + 2\log^2 z$$
$$- \left[\frac{1}{2} (3 - 4z + z^2) + 2\log z \right] \log \frac{\mu_0^2}{m^2} \, . \tag{6.37}$$

The key to the simplicity of the solutions in Eqs. (6.31) and (6.32) is the fact that the dependence on μ on the r.h.s. of Eq. (6.29) is entirely parametrised by L, which in turn allows

one to compute the integral on the r.h.s. of Eq. (6.28) in a trivial manner:

$$\int_{\log \mu_0^2}^{\log \mu^2} d\log {\mu'}^2 \; \eta_0^k \big|_{\mu \to \mu'} = \left(\frac{\alpha}{\pi}\right)^{-1} \frac{\eta_0^{k+1}}{k+1} \,. \tag{6.38}$$

Unfortunately, things are not so simple when α is running. In this case, as was already done in Sec. 6.2, it is convenient to use the variable t introduced in Eq. (6.15). Owing to Eq. (6.16), the analogue of Eq. (6.28) reads as follows:

$$\mathcal{F}(z,t) = \mathcal{F}(z,0) + \int_0^t du \, \frac{b_0 \alpha^2(u)}{\beta(\alpha(u))} \left[\mathbb{P} \overline{\otimes} \mathcal{F}\right](z,u) \,. \tag{6.39}$$

As a consequence of this, we shall use the representation:

$$\mathcal{F}(z,t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left(\mathcal{J}_k^{\mathrm{LL}}(z) + \frac{\alpha(t)}{2\pi} \mathcal{J}_k^{\mathrm{NLL}}(z)\right), \tag{6.40}$$

rather than that of Eq. (6.29). Thus:

$$\frac{b_0 \alpha^2(t)}{\beta(\alpha(t))} \mathbb{P} \overline{\otimes} \mathcal{F} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left\{ \mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_k^{\mathrm{LL}} + \frac{\alpha(t)}{2\pi} \left[\mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_k^{\mathrm{NLL}} + \mathbb{P}^{[1]} \overline{\otimes} \mathcal{J}_k^{\mathrm{LL}} \right. \right.$$
$$\left. \left. - \frac{2\pi b_1}{b_0} \mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_k^{\mathrm{LL}} \right] + \mathcal{O}(\alpha^2) \right\} \,. \tag{6.41}$$

The r.h.s. of Eq. (6.39) therefore features two independent classes of integrals, namely:

$$a_k = \int_0^t du \, u^k = \frac{t^{k+1}}{k+1} \,, \tag{6.42}$$

$$b_k = \int_0^t du \, u^k \alpha(u) \,. \tag{6.43}$$

In order to evaluate Eq. (6.43), we make repeated use of Eq. (6.17). Then:

$$b_k = \alpha(0) \int_0^t du \, u^k \, e^{2\pi b_0 u} = \alpha(t) e^{-2\pi b_0 t} \sum_{j=0}^{\infty} \frac{(2\pi b_0)^j}{j!} \int_0^t du \, u^k \, u^j \,. \tag{6.44}$$

By direct computation:

$$e^{-2\pi b_0 t} \sum_{j=0}^{\infty} \frac{(2\pi b_0)^j}{(k+1+j)j!} \, t^{k+1+j} = \frac{t^{k+1}}{k+1} \sum_{p=0}^{\infty} d_{k,p} \, t^p \,, \tag{6.45}$$

with:

$$d_{k,p} = (-)^p (2\pi b_0)^p \frac{\Gamma(k+2)}{\Gamma(k+2+p)} \,. \tag{6.46}$$

We have thus:

$$\mathcal{F}(z,t) - \mathcal{F}(z,0) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(g_k \frac{t^{k+1}}{k+1} + \frac{\alpha(t)}{2\pi} h_k \sum_{p=0}^{\infty} \frac{t^{k+1+p}}{k+1} d_{k,p} \right), \qquad (6.47)$$

where:

$$g_k = \mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_k^{\text{LL}}, \qquad (6.48)$$

$$h_k = \mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_k^{\text{NLL}} + \mathbb{P}^{[1]} \overline{\otimes} \mathcal{J}_k^{\text{LL}} - \frac{2\pi b_1}{b_0} \mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_k^{\text{LL}}. \qquad (6.49)$$

The r.h.s. of Eq. (6.47) can be simplified by means of algebraic manipulations of the summation indices:

$$\sum_{k=0}^{\infty} \frac{1}{k!} g_k \frac{t^{k+1}}{k+1} = \sum_{k=1}^{\infty} \frac{t^k}{k!} g_{k-1}, \qquad (6.50)$$

and:

$$\sum_{k=0}^{\infty} \frac{1}{k!} h_k \sum_{p=0}^{\infty} \frac{t^{k+1+p}}{k+1} d_{k,p} = \sum_{k=1}^{\infty} \frac{t^k}{k!} \sum_{p=0}^{k-1} (-)^p (2\pi b_0)^p h_{k-1-p}, \qquad (6.51)$$

since from Eq. (6.46):

$$\frac{d_{k-1-p,p}}{(k-p)!} = \frac{(-)^p (2\pi b_0)^p}{k!}. \qquad (6.52)$$

The initial conditions must then be written as follows:

$$\begin{aligned} \mathcal{F}(z,0) &= \mathcal{F}^{[0]}(z,\mu_0^2) + \frac{\alpha(t)e^{-2\pi b_0 t}}{2\pi} \mathcal{F}^{[1]}(z,\mu_0^2) \\ &= \mathcal{F}^{[0]}(z,\mu_0^2) + \frac{\alpha(t)}{2\pi} \mathcal{F}^{[1]}(z,\mu_0^2) \sum_{k=0}^{\infty} \frac{(-)^k (2\pi b_0)^k}{k!} t^k. \end{aligned} \qquad (6.53)$$

By replacing the results of Eqs. (6.50), (6.51), and (6.53) into Eq. (6.47), and by using the representation of Eq. (6.40) for $\mathcal{F}(z,t)$, we obtain the sought recursion relations:

$$\mathcal{J}_k^{\text{LL}} = \mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_{k-1}^{\text{LL}}, \qquad (6.54)$$

$$\begin{aligned} \mathcal{J}_k^{\text{NLL}} = &\ (-)^k (2\pi b_0)^k \mathcal{F}^{[1]}(\mu_0^2) \\ &+ \sum_{p=0}^{k-1} (-)^p (2\pi b_0)^p \left(\mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_{k-1-p}^{\text{NLL}} + \mathbb{P}^{[1]} \overline{\otimes} \mathcal{J}_{k-1-p}^{\text{LL}} \right. \\ &\left. - \frac{2\pi b_1}{b_0} \mathbb{P}^{[0]} \overline{\otimes} \mathcal{J}_{k-1-p}^{\text{LL}} \right), \end{aligned} \qquad (6.55)$$

with:

$$\mathcal{J}_0^{\text{LL}} = \mathcal{F}^{[0]}(z,\mu_0^2), \qquad \mathcal{J}_0^{\text{NLL}} = \mathcal{F}^{[1]}(z,\mu_0^2). \qquad (6.56)$$

These results generalise those obtained in the case of non-running α, which can be obtained from them. Indeed, in the limit of fixed α, which at the NLL can be achieved by letting $b_0 \to 0$ and $b_1 \to 0$ (with $b_1/b_0 \to 0$), we have $t \to \eta_0/2$, thereby Eq. (6.40) coincides with Eq. (6.29), if

139

one identifies $\mathcal{J}^{\text{LL}}$ with $\mathcal{I}^{\text{LL}}$ and $\mathcal{J}^{\text{NLL}}$ with $\mathcal{I}^{\text{NLL}}$. This is justified, since Eqs. (6.31) and (6.54) are identical, and the recursive relation of Eq. (6.55) coincides with that of Eq. (6.32) when α is not running.

After solving Eqs. (6.31) and (6.32) for $\mathcal{I}_k^{\text{LL}}$ and $\mathcal{I}_k^{\text{NLL}}$, with the definition in Eq. (6.26) one arrives at the following representation of the PDF in the case of fixed α:

$$\Gamma(z, \mu^2) = \sum_{k=0}^{\infty} \frac{\eta_0^k}{2^k k!} \left(I_k^{\text{LL}}(z) + \frac{\alpha}{2\pi} I_k^{\text{NLL}}(z) \right), \tag{6.57}$$

where

$$I_k^{\text{LL}}(z) = -\frac{d}{dz} \mathcal{I}_k^{\text{LL}}(z), \qquad I_k^{\text{NLL}}(z) = -\frac{d}{dz} \mathcal{I}_k^{\text{NLL}}(z). \tag{6.58}$$

Analogously, in the case of running α:

$$\Gamma(z, \mu^2) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left(J_k^{\text{LL}}(z) + \frac{\alpha(t)}{2\pi} J_k^{\text{NLL}}(z) \right), \tag{6.59}$$

with

$$J_k^{\text{LL}}(z) = -\frac{d}{dz} \mathcal{J}_k^{\text{LL}}(z), \qquad J_k^{\text{NLL}}(z) = -\frac{d}{dz} \mathcal{J}_k^{\text{NLL}}(z). \tag{6.60}$$

We point out that with, for example, $\alpha(\mu) = 1/128$ and $\alpha(\mu_0) = 1/137$ we have $t \simeq 0.1/n_F$. Furthermore, since:

$$2\pi b_0 = \frac{2}{3} n_F, \qquad \frac{2\pi b_1}{b_0} = \frac{3}{2}, \tag{6.61}$$

the numerical coefficients in front of the convolution products and of the initial conditions in Eq. (6.55) are of order one. Therefore, the series of Eq. (6.40) is expected to be poorly convergent only for $z \to 1$ and $z \to 0$, owing to the possible presence of $\log^p(1-z)$ and $\log^p z$ terms in the $\mathcal{J}^{\text{LL}}$ and $\mathcal{J}^{\text{NLL}}$ functions.

The recursive solutions are thus obtained by means of the following procedure. One first computes the $\mathcal{J}_k^{\text{LL}}$ and $\mathcal{J}_k^{\text{NLL}}$ functions, by employing Eqs. (6.54) and (6.55). These equations must be applied recursively, by working one's own way up in k from the known $k = 0$ results (given in Eqs. (6.34)–(6.37)). The expressions for the relevant Altarelli-Parisi kernels are reported in Appendix 6.C. Finally, the J_k^{LL} and J_k^{NLL} functions are obtained by derivation, according to Eq. (6.60).

We have computed the J_k^{LL} and J_k^{NLL} basis functions that appear in Eq. (6.59) for $0 \le k \le 3$ and $0 \le k \le 2$, respectively, i.e. up to $\mathcal{O}(\alpha^3)$. We write the actual recursive solution as follows:

$$\Gamma(z, \mu^2) = \sum_{k=0}^{k_{\text{max}}^{\text{LL}}} \frac{t^k}{k!} J_k^{\text{LL}}(z) + \frac{\alpha(t)}{2\pi} \sum_{k=0}^{k_{\text{max}}^{\text{NLL}}} \frac{t^k}{k!} J_k^{\text{NLL}}(z), \tag{6.62}$$

with

$$k_{\text{max}}^{\text{LL}} = 3, \qquad k_{\text{max}}^{\text{NLL}} = 2. \tag{6.63}$$

We also remind the reader that from Eq. (6.62) one can obtain the solution in the case of

non-running α, by replacing J_k^{LL} with I_k^{LL} and J_k^{NLL} with I_k^{NLL}, where:

$$
\begin{aligned}
I_k^{\mathrm{LL}}(z) &= J_k^{\mathrm{LL}}(z)\,, & (6.64)\\
I_k^{\mathrm{NLL}}(z) &= J_k^{\mathrm{NLL}}(z)\big[b_0 \to 0\,, b_1 \to 0\,, b_1/b_0 \to 0\big]\,. & (6.65)
\end{aligned}
$$

It is convenient, also in view of the matching with the large-z solution in Sec. 6.7, to present the results for the basis functions by writing them as follows:

$$
\begin{aligned}
J_k^{\mathrm{LL}}(z) &= \bar{J}_k^{\mathrm{LL}}(z) + \hat{J}_k^{\mathrm{LL}}(z)\,, & (6.66)\\
J_k^{\mathrm{NLL}}(z) &= \bar{J}_k^{\mathrm{NLL}}(z) + \hat{J}_k^{\mathrm{NLL}}(z)\,. & (6.67)
\end{aligned}
$$

By definition, $\hat{J}_k^{\mathrm{LL}}$ and $\hat{J}_k^{\mathrm{NLL}}$ collect all of the terms of J_k^{LL} and J_k^{NLL}, respectively, that vanish at $z = 1$:

$$
\lim_{z\to 1}\hat{J}_k^{\mathrm{LL}}(z) = \lim_{z\to 1}\hat{J}_k^{\mathrm{NLL}}(z) = 0\,. \qquad (6.68)
$$

It then follows that $\bar{J}_k^{\mathrm{LL}}$ and $\bar{J}_k^{\mathrm{NLL}}$ include all contributions that are either divergent (which then feature all the $\log^p(1-z)$ terms) or equal to a non-null constant at $z = 1$. Because of this, it is useful to introduce the following auxiliary functions:

$$
\begin{aligned}
\ell_i(z) &= \frac{\log^i(1-z)}{1-z}\,, & i \ge 0\,, & (6.69)\\
q_i(z) &= \log^i(1-z)\,, & i \ge 0\,, & (6.70)
\end{aligned}
$$

and write:

$$
\begin{aligned}
\bar{J}_k^{\mathrm{LL}}(z) &= \sum_{i=0}^{i_{\max}^{\mathrm{LL}}(k)}\Big[b_{k,i}^{\mathrm{LL}}\,\ell_i(z) + c_{k,i}^{\mathrm{LL}}\,q_i(z)\Big]\,, & k \ge 1\,, & (6.71)\\
\bar{J}_k^{\mathrm{NLL}}(z) &= \sum_{i=0}^{i_{\max}^{\mathrm{NLL}}(k)}\Big[b_{k,i}^{\mathrm{NLL}}\,\ell_i(z) + c_{k,i}^{\mathrm{NLL}}\,q_i(z)\Big]\,, & k \ge 0\,. & (6.72)
\end{aligned}
$$

with:

$$
\begin{aligned}
i_{\max}^{\mathrm{LL}}(k) &= k - 1\,, & (6.73)\\
i_{\max}^{\mathrm{NLL}}(k) &= k + 1\,. & (6.74)
\end{aligned}
$$

In addition to this, one must take into account that, at $\mathcal{O}(\alpha^0)$:

$$
J_0^{\mathrm{LL}}(z) = \bar{J}_0^{\mathrm{LL}}(z) = \hat{J}_0^{\mathrm{LL}}(z) = 0\,. \qquad (6.75)
$$

The contribution to $\Gamma(z)$ that does not vanish at $z \to 1$ is then written as follows:

$$
\overline{\Gamma}(z,\mu^2) = \sum_{k=0}^{k_{\max}^{\mathrm{LL}}}\frac{t^k}{k!}\,\bar{J}_k^{\mathrm{LL}}(z) + \frac{\alpha(t)}{2\pi}\sum_{k=0}^{k_{\max}^{\mathrm{NLL}}}\frac{t^k}{k!}\,\bar{J}_k^{\mathrm{NLL}}(z)\,. \qquad (6.76)
$$

The expressions of the $b_{k,i}^{\mathrm{(N)LL}}$ and $c_{k,i}^{\mathrm{(N)LL}}$ coefficients for the non-singlet, singlet, and photon PDFs

will be presented in Appendix 6.B. The expressions of the functions $\hat{J}^{(N)LL}(z)$ are lengthy and not relevant to the matching. Some of them receive contributions that we have not computed analytically, as detailed in Appendix 6.A. For these reasons, they are only reported in an ancillary file that accompanies the submission of Ref. [2] to the arXiv.

In order to document the effect of increasing the number of terms included in the recursive solutions, we plot in Fig. 6.1 the ratio of the result of Eq. (6.62) over the numerical predictions[5] minus one; Eq. (6.62) is computed by setting:

$$k_{max}^{NLL} = k_{max}^{LL} - 1 , \qquad k_{max}^{LL} = 1, 2, 3 . \tag{6.77}$$

The ratios are displayed as green dot-dashed lines ($k_{max}^{LL} = 1$), blue dashed lines ($k_{max}^{LL} = 2$), and red solid lines ($k_{max}^{LL} = 3$). In order for the results to fit into the layout of the figures, the green and blue curves are multiplied by a constant factor equal to 10^{-2} and 10^{-1}, respectively.
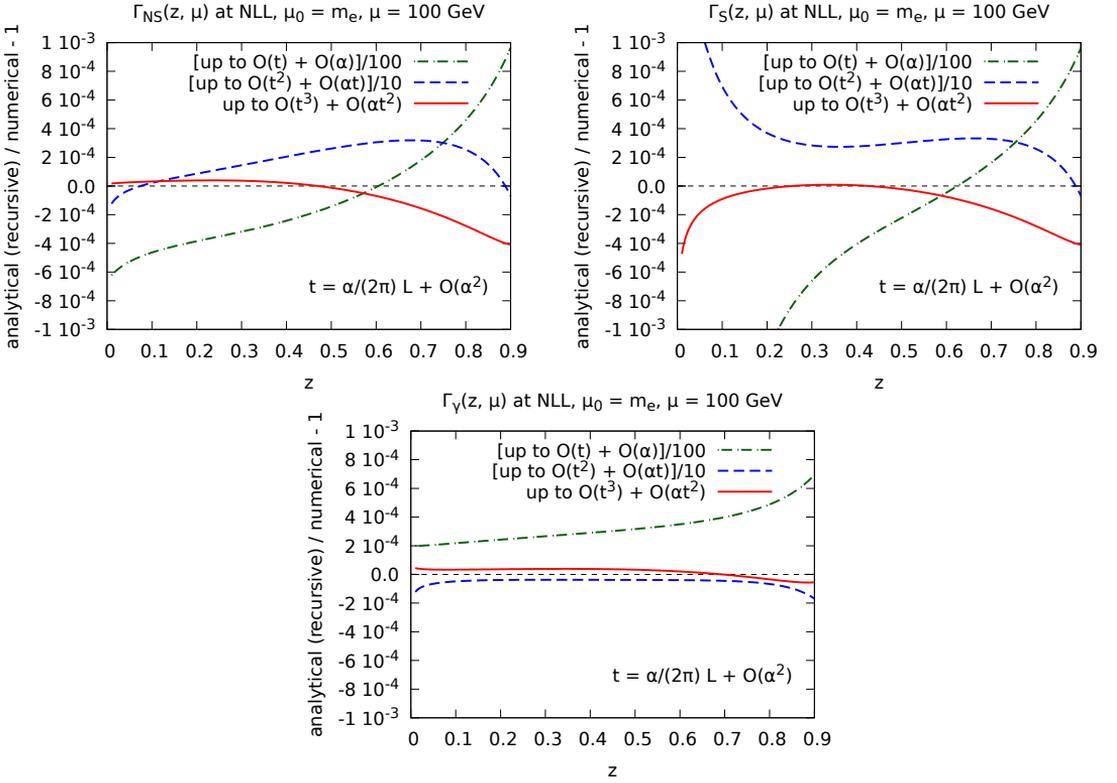

Figure 6.1: Agreement between recursive solutions of various accuracies, and the numerical predictions, for the non-singlet (top left panel), singlet (top right panel), and photon (bottom panel), for $\mu = 100$ GeV. See the text for details.

We see that our most accurate recursive predictions ($k_{max}^{LL} = 3$) agree with the numerical results at the level of a few 10^{-4} at the worst. Note that since here we are dealing only with the recursive solutions we have limited ourselves to plotting the PDFs in the range $z \in (0, 0.9)$

---

[5]The numerical solutions will be discussed in Sec. 6.8.

– at the upper end of the range, the absence of the contribution from the asymptotic solution starts to be felt. The information stemming from Fig. 6.1 is that, if we had only computed either the first term or the first two terms in the sums of Eq. (6.62), the $\mathcal{O}(10^{-4})$ agreement remarked above would actually have been roughly equal to, but generally worse than, 10^{-2} and 10^{-3}, respectively. The figure also shows that, for any given accuracy of the recursive solution, the agreement with the numerical prediction marginally worsens towards $z \to 0$ in the case of the singlet, owing to the presence of $\log z$ terms which are not resummed.

6.4 Analytical large-z solution for non-singlet PDF

The electron PDF is equal to $\delta(1 - z)$ at the LO (see Eq. (6.3)); while the LL evolution of such an initial condition does smooth its behaviour, resulting in a tail that extends down to $z = 0$ [194–196], the PDF remains very peaked towards $z = 1$, where it has an integrable singularity. This implies that the perturbative expansion of the LL-accurate solution features $\log(1 - z)$ terms at each order: if one truncates such a perturbative series, one exposes a non-integrable divergence at $z = 1$, regardless of the order at which the truncation occurs. The same is true when NLO initial conditions and NLL-accurate evolution are considered.

In order to address this issue, the $\log(1-z)$ terms must be resummed. This can conveniently be done by exploiting the evolution-operator formalism presented in Sec. 6.2, whose usage is simplified by the observation that the large-z region corresponds to the large-N region in Mellin space:

$$z \to 1 \quad \longleftrightarrow \quad N \to \infty \,. \tag{6.78}$$

Thus, when dealing with Mellin transforms and their inverse, we shall often implicitly assume Eq. (6.78). In this section, we study the $z \to 1$ asymptotic behaviour of the non-singlet component; we shall return to and comment on the singlet-photon case in Sec. 6.5.

6.4.1 LL solution

Given that the LL-accurate result has been available for a long while [44], this case is presented here only to show how the evolution-operator formalism helps find the asymptotic solution in a straightforward manner. At the LL we are entitled to neglect the running[6] of α. Thus, the appropriate form for the evolution operator is obtained by keeping only the $\mathcal{O}(\alpha)$ term of Eq. (6.24), with $\alpha(\mu) \to \alpha$ there, supplemented by the LO initial condition:

$$\Gamma_{0,N}^{[0]} = 1 \,. \tag{6.79}$$

From Eqs. (6.12) and (6.79) we obtain:

$$\Gamma(z, \mu^2) = M^{-1}\big[\exp\big(\log E_N\big)\big] \,. \tag{6.80}$$

A direct calculation in the large-N region leads to:

$$P_N^{[0]} \xrightarrow{N \to \infty} -2\log \bar{N} + 2\lambda_0 \,, \tag{6.81}$$

---

[6]Whenever the coupling constant is not running, we simply denote its fixed value by α, i.e. we remove its argument μ from the notation.

where all terms suppressed by at least one inverse power of N have been neglected, and we have defined:

$$\bar{N} = N \, e^{\gamma_{\mathrm{E}}} \, , \qquad \lambda_0 = \frac{3}{4} \, . \tag{6.82}$$

We point out that $\bar{N}$ is a quantity that routinely appears in the computation of Mellin transforms, and which helps retain some universal subleading terms. Therefore:

$$\log E_N = \frac{\alpha}{2\pi} \, P_N^{[0]} \, L \stackrel{N \to \infty}{\longrightarrow} -\eta_0 \left(\log \bar{N} - \lambda_0 \right) \, , \tag{6.83}$$

with η_0 defined in Eq. (6.30). Equation (6.83), when substituted into Eq. (6.80), implies:

$$M \left[\Gamma(z, \mu^2) \right] = N^{-\eta_0} e^{-\gamma_{\mathrm{E}} \eta_0} e^{\lambda_0 \eta_0} \, . \tag{6.84}$$

The inverse Mellin transform can now be evaluated by using the following result, valid for any $\kappa > 0$:

$$M \left[(1-z)^{-1+\kappa} \right] = \frac{\Gamma(\kappa)\Gamma(N)}{\Gamma(\kappa + N)} \stackrel{N \to \infty}{\longrightarrow} \Gamma(\kappa) N^{-\kappa} \, . \tag{6.85}$$

The comparison of Eq. (6.85) with Eq. (6.84) allows one to arrive at the final result [44]:

$$\Gamma(z, \mu^2) = \frac{e^{-\gamma_{\mathrm{E}} \eta_0} e^{\lambda_0 \eta_0}}{\Gamma(1 + \eta_0)} \, \eta_0 (1-z)^{-1+\eta_0} \, . \tag{6.86}$$

This is identical to what is nowadays a rather standard form, except for an exponentiated term of pure-soft origin (stemming from the use of $\beta_{exp} = \beta$, rather than of $\beta_{exp} = \eta$, as defined e.g. in Eq. (67) of Ref. [200]). Such a term clearly cannot be obtained by means of the collinear resummation carried out here.

6.4.2 $\overline{\mathrm{MS}}$ NLL solution

At the NLL, the PDF initial conditions must be set as given in Eqs. (6.3) and (6.6), with $K_{ee} = 0$ in the latter equation (see Eq. (6.9)). By exploiting the property of the Mellin transform of Eq. (1.87), we have:

$$
\begin{aligned}
\Gamma(z, \mu^2) &= \left(\delta(1-x) + \frac{\alpha(\mu_0^2)}{2\pi} \left[\frac{1+x^2}{1-x} \left(\log \frac{\mu_0^2}{m^2} - 2\log(1-x) - 1 \right) \right]_+ \right) \\
&\otimes_z M^{-1} \left[\exp \left(\log E_N \right) \right],
\end{aligned} \tag{6.87}
$$

with $\log E_N$ given in Eq. (6.20) (where running-α effects are also included). With Eq. (6.81) and its NLO analogue:

$$P_N^{[1]} \stackrel{N \to \infty}{\longrightarrow} \frac{20}{9} n_F \log \bar{N} + \lambda_1 \, , \tag{6.88}$$

where:

$$\lambda_1 = \frac{3}{8} - \frac{\pi^2}{2} + 6\zeta_3 - \frac{n_F}{18} (3 + 4\pi^2) \, , \tag{6.89}$$

we can cast the logarithm of the evolution operator in the same form as in Eq. (6.83), namely:

$$\log E_N \stackrel{N \to \infty}{\longrightarrow} -\xi_1 \log \bar{N} + \hat{\xi}_1 \, , \tag{6.90}$$

having defined:

$$\xi_1 = 2t - \frac{\alpha(\mu)}{4\pi^2 b_0}\left(1 - e^{-2\pi b_0 t}\right)\left(\frac{20}{9}n_F + \frac{4\pi b_1}{b_0}\right) \tag{6.91}$$

$$= 2\left[1 - \frac{\alpha(\mu)}{\pi}\left(\frac{5}{9}n_F + \frac{\pi b_1}{b_0}\right)\right]t$$

$$+ \frac{\alpha(\mu)}{\pi}\left(\frac{10}{9}\pi b_0 n_F + 2b_1\pi^2\right)t^2 + \mathcal{O}(t^3), \tag{6.92}$$

$$\hat{\xi}_1 = \frac{3}{2}t + \frac{\alpha(\mu)}{4\pi^2 b_0}\left(1 - e^{-2\pi b_0 t}\right)\left(\lambda_1 - \frac{3\pi b_1}{b_0}\right) \tag{6.93}$$

$$= \frac{3}{2}\left[1 + \frac{\alpha(\mu)}{\pi}\left(\frac{\lambda_1}{3} - \frac{\pi b_1}{b_0}\right)\right]t$$

$$- \frac{\alpha(\mu)}{\pi}\left(\frac{\pi b_0}{2}\lambda_1 - \frac{3}{2}\pi^2 b_1\right)t^2 + \mathcal{O}(t^3). \tag{6.94}$$

Equation (6.90) implies that we can follow the same steps that have led us to Eq. (6.86), and therefore:

$$M^{-1}\big[\exp\big(\log E_N\big)\big] = \frac{e^{-\gamma_E \xi_1}e^{\hat{\xi}_1}}{\Gamma(1 + \xi_1)}\,\xi_1(1 - y)^{-1+\xi_1}. \tag{6.95}$$

We must now replace this result into Eq. (6.87). In this way, two independent convolution integrals emerge:

$$I_+(z) = \frac{1}{2}\left[\frac{1 + x^2}{1 - x}\right]_+ \otimes_z (1 - y)^{-1+\kappa}, \tag{6.96}$$

$$I_L(z) = \frac{1}{2}\left[\frac{1 + x^2}{1 - x}\log(1 - x)\right]_+ \otimes_z (1 - y)^{-1+\kappa}. \tag{6.97}$$

A tedious but otherwise relatively straightforward procedure leads to the following results:

$$I_+(z) = (1 - z)^{-1+\kappa}\left[A(\kappa) + \log(1 - z) + \frac{3}{4}\right], \tag{6.98}$$

$$I_L(z) = (1 - z)^{-1+\kappa}\left[B(\kappa) + A(\kappa)\log(1 - z) + \frac{1}{2}\log^2(1 - z) - \frac{7}{8}\right], \tag{6.99}$$

where, inside the square brackets, we have neglected terms that vanish at $z \to 1$. We have introduced the two functions:

$$A(\kappa) = \sum_{k=1}^{\infty}\frac{1}{k\,k!}\frac{\Gamma(1 - \kappa + k)}{\Gamma(1 - \kappa)} = -\gamma_E - \psi_0(\kappa), \tag{6.100}$$

$$B(\kappa) = -\sum_{k=1}^{\infty}\frac{1}{k^2\,k!}\frac{\Gamma(1 - \kappa + k)}{\Gamma(1 - \kappa)}$$

$$= \frac{1}{2}\gamma_E^2 + \frac{\pi^2}{12} + \gamma_E\,\psi_0(\kappa) + \frac{1}{2}\psi_0(\kappa)^2 - \frac{1}{2}\psi_1(\kappa), \tag{6.101}$$

where:

$$\psi_j(z) = \frac{d^{j+1} \log \Gamma(z)}{dz^{j+1}} \, . \tag{6.102}$$

By putting everything back together, we arrive at the final result:

$$\Gamma(z, \mu^2) = \frac{e^{-\gamma_E \xi_1} e^{\hat{\xi}_1}}{\Gamma(1 + \xi_1)} \xi_1 (1 - z)^{-1 + \xi_1} \tag{6.103}$$

$$\times \left\{ 1 + \frac{\alpha(\mu_0)}{\pi} \left[\left(\log \frac{\mu_0^2}{m^2} - 1 \right) \left(A(\xi_1) + \frac{3}{4} \right) - 2B(\xi_1) + \frac{7}{4} \right.\right.$$

$$\left.\left. + \left(\log \frac{\mu_0^2}{m^2} - 1 - 2A(\xi_1) \right) \log(1 - z) - \log^2(1 - z) \right] \right\},$$

which is therefore the NLL-accurate counterpart of Eq. (6.86).

A couple of observations about Eq. (6.103) are in order. Firstly, owing to Eqs. (6.91) and (6.22), we have $\xi_1 \simeq \eta_0$. With μ_0 and μ of the order of the electron mass and of a few hundred GeV's, respectively, one obtains $\eta_0 \sim 0.05$. Therefore, both the LL and the NLL solutions are still very peaked towards $z = 1$, where they diverge with an integrable singularity. Furthermore, the $z \to 1$ behaviour of Eq. (6.103) is worse than that of Eq. (6.86) because of the presence of the explicit $\log^p(1 - z)$ terms in the former equation. Secondly, the small numerical value of ξ_1 just mentioned implies that the following expansions:

$$A(\kappa) = \frac{1}{\kappa} + \mathcal{O}(\kappa) \, , \tag{6.104}$$

$$B(\kappa) = -\frac{\pi^2}{6} + 2\zeta_3 \kappa + \mathcal{O}(\kappa^2) \, , \tag{6.105}$$

are rather accurate approximations of the complete results of Eqs. (6.100) and (6.101). Equation (6.104), in particular, implies that numerically the $\log(1 - z)$ term is much larger than the (formally dominant) $\log^2(1 - z)$ one, even for z values that are *extremely* close to one. This fact might be significant when performing the integral of the convolution between electron PDFs and short-distance cross sections.

From Eq. (6.103) one can also readily obtain a LL-accurate solution, where at variance with that of Eq. (6.86) the effects due to the running of α are included. Explicitly:

$$\Gamma(z, \mu^2) = \frac{e^{-\gamma_E \xi_0} e^{\hat{\xi}_0}}{\Gamma(1 + \xi_0)} \xi_0 (1 - z)^{-1 + \xi_0} \, , \tag{6.106}$$

where

$$\xi_0 = 2t \, , \qquad \hat{\xi}_0 = \frac{3}{2} t \, ; \tag{6.107}$$

this is again consistent with the findings of Ref. [44]. Finally, the running of α can formally be switched off in the NLL-accurate solution. In order to do so, one must repeat the procedure that leads to Eq. (6.103); however, rather than using the expression of the evolution operator

given in Eq. (6.20), one must use that of Eq. (6.24). By doing so, one arrives at:

$$\Gamma(z,\mu^2) = \frac{e^{-\gamma_E \eta_1} e^{\hat{\eta}_1}}{\Gamma(1+\eta_1)}\, \eta_1 (1-z)^{-1+\eta_1} \tag{6.108}$$

$$\times \left\{ 1 + \frac{\alpha}{\pi}\left[\left(\log\frac{\mu_0^2}{m^2} - 1\right)\left(A(\eta_1) + \frac{3}{4}\right) - 2B(\eta_1) + \frac{7}{4} \right.\right.$$

$$\left.\left. + \left(\log\frac{\mu_0^2}{m^2} - 1 - 2A(\eta_1)\right)\log(1-z) - \log^2(1-z) \right] \right\}.$$

where

$$\eta_1 = \eta_0\left(1 - \frac{5\alpha}{9\pi}\, n_F\right), \tag{6.109}$$

$$\hat{\eta}_1 = \eta_0\left(\lambda_0 + \frac{\alpha}{4\pi}\,\lambda_1\right). \tag{6.110}$$

6.4.3   Alternative z-space derivation of the NLL result

We now show how the large-z solution for the non-singlet can be obtained directly in configuration space, that is without resorting to Mellin-space techniques, and thus providing one with a cross-check. We have considered this alternative procedure by neglecting the running of α. We point out that this method has already been used to obtain the LL solution of Eq. (6.86) – see e.g. Ref. [201]. Here, we extend it to the NLL accuracy.

In essence, the procedure works as follows. One makes an ansatz for the z-space functional form of $\Gamma(z,\mu^2)$, where the μ^2 dependence is parametrised by unknown functions. The PDF evolution equations, simplified in the $z \to 1$ limit, are then turned into differential equations for such unknown functions, where the independent variable is μ^2. By solving these equations, one is left with arbitrary integration constants, whose values are finally determined by matching the solutions to the known PDF initial conditions.

In order to proceed, we start by observing that the assumption of non-running α implies that the dependence on μ^2 can be entirely parametrised by means of the quantity η_0, introduced in Eq. (6.30); thus, we shall use the latter as our independent variable. At the LL, this implies that the evolution equation of Eq. (1.75) reads as follows:

$$\frac{d}{d\eta_0}\Gamma^{\mathrm{LL}}(z,\eta_0) = \frac{1}{2}\, P^{[0]} \otimes_z \Gamma^{\mathrm{LL}}(\eta_0)\,. \tag{6.111}$$

For the computation of the convolution integral on the r.h.s. of Eq. (6.111) we approximate the first-order non-singlet Altarelli-Parisi kernel in the large-z region as follows:

$$P^{[0]}(z) \xrightarrow{z\to 1} 2\left(\frac{1}{1-z}\right)_+ + 2\lambda_0\,\delta(1-z)\,, \tag{6.112}$$

which is the analogue of Eq. (6.81). The parameter λ_0 has been defined in Eq. (6.82), and its value stems from the exact form of the denominator of the splitting kernels, $1+z^2$. We now

make the following ansatz for the functional form of the LL PDF that appear in Eq. (6.111):

$$\Gamma^{\text{LL}}(z, \eta_0) = b(\eta_0) \, (1 - z)^{a(\eta_0)} \, . \tag{6.113}$$

By replacing Eq. (6.113) into Eq. (6.111), and by using Eq. (6.112), the convolution integral has two trivial contributions, induced by the $\delta(1 - z)$ and by the subtraction term of the plus distribution (integrated in the $(0, z)$ range) in Eq. (6.112). The non-trivial part of the convolution integral can also be easily computed in the $z \to 1$ limit, to read:

$$\int_z^1 \frac{dx}{1 - x} \quad \left[\frac{1}{x} \left(1 - \frac{z}{x} \right)^{a(\eta_0)} - (1 - z)^{a(\eta_0)} \right]$$
$$\stackrel{z \to 1}{\simeq} -(1 - z)^{a(\eta_0)} \left[\psi_0(a(\eta_0) + 1) + \gamma_{\text{E}} \right] \, . \tag{6.114}$$

Thus, both sides of Eq. (6.111) are linear combinations of two terms, whose dependence on z is equal to $(1 - z)^{a(\eta_0)}$ and to $(1 - z)^{a(\eta_0)} \log(1 - z)$, respectively. By equating the coefficients of such terms one finally arrives at the sought differential equations:

$$\frac{d}{d\eta_0} a(\eta_0) = 1 \, , \tag{6.115}$$

$$\frac{d}{d\eta_0} b(\eta_0) = b(\eta_0) \left[-\left(\psi_0(a(\eta_0) + 1) + \gamma_{\text{E}} \right) + \lambda_0 \right] \, . \tag{6.116}$$

The solutions of these are:

$$a(\eta_0) = \eta_0 + a_0 \, , \tag{6.117}$$

$$b(\eta_0) = b_0 \frac{e^{(\lambda_0 - \gamma_{\text{E}})\eta_0}}{\Gamma(\eta_0 + a_0 + 1)} \, . \tag{6.118}$$

The quantities a_0 and b_0 are arbitrary integration constants, which can be determined by observing that, in the limit $\eta_0 \to 0$, $\Gamma^{\text{LL}}(z, \eta_0)$ must be equal to the initial condition of Eq. (6.3). By imposing such an equality we obtain:

$$a_0 = -1 \, , \qquad b_0 = 1 \, . \tag{6.119}$$

It then becomes apparent that Eq. (6.113), supplemented with Eqs. (6.117), (6.118), and (6.119), coincides with Eq. (6.86).

The procedure outlined so far can now be extended to the NLL. We write the analogue of Eq. (6.111) as follows:

$$\frac{d}{d\eta_0} \Gamma^{\text{NLL}}(z, \eta_0) = \frac{1}{2} \left(P^{[0]} + \frac{\alpha}{2\pi} P^{[1]} \right) \otimes_z \Gamma^{\text{NLL}}(\eta_0) \, , \tag{6.120}$$

with the second-order non-singlet Altarelli-Parisi kernel approximated in the large-z as follows:

$$P^{[1]}(z) \xrightarrow{z \to 1} -\frac{20}{9} n_F \left(\frac{1}{1 - z} \right)_+ + \lambda_1 \delta(1 - z) \, . \tag{6.121}$$

Equation (6.121) is the z-space analogue of Eq. (6.88), with λ_1 defined in Eq. (6.89). We also

need to replace our LL ansatz of Eq. (6.113) with one that is appropriate at the NLL, namely:

$$
\begin{aligned}
\Gamma^{\mathrm{NLL}}(z, \eta_0) \;=\; & (1 - z)^{a(\eta_0)} \\
& \times \left\{ b(\eta_0) + \frac{\alpha}{\pi} \Big[c(\eta_0) + d(\eta_0) \log(1 - z) + e(\eta_0) \log^2(1 - z) \Big] \right\}.
\end{aligned} \tag{6.122}
$$

The physical motivation of Eq. (6.122) is the following. Firstly, one observes that $P^{[0]}$ and $P^{[1]}$ have the same functional large-z behaviours. Secondly, we have seen that at the LL the convolution of the evolution kernel with the r.h.s. of Eq. (6.113) either leaves the functional form of the latter unchanged, or it multiplies it by a $\log(1 - z)$ term. Therefore, since the $\mathcal{O}(\alpha)$ contribution to the PDF initial condition in the $\overline{\mathrm{MS}}$ scheme, Eq. (6.6), contains logarithmic terms up to the first power, its convolution with the evolution kernel either leave those unchanged, or it increases their powers by one unity.

As was the case at the LL, the convolution of the r.h.s. of Eq. (6.122) with the Altarelli-Parisi kernels features a few trivial contributions, due to the endpoints, and some non-trivial ones, which can nevertheless be readily computed. Among the latter, we find again Eq. (6.114), and:

$$
\begin{aligned}
\int_z^1 & \frac{dx}{1 - x} \left[\frac{1}{x} \left(1 - \frac{z}{x} \right)^{a(\eta_0)} \log \left(1 - \frac{z}{x} \right) - (1 - z)^{a(\eta_0)} \log(1 - z) \right] \\
& \overset{z \to 1}{\simeq} -(1 - z)^{a(\eta_0)} \Big\{ \log(1 - z)[\psi_0(a(\eta_0) + 1) + \gamma_{\mathrm{E}}] + \psi_1(a(\eta_0) + 1) \Big\},
\end{aligned} \tag{6.123}
$$

$$
\begin{aligned}
\int_z^1 & \frac{dx}{1 - x} \left[\frac{1}{x} \left(1 - \frac{z}{x} \right)^{a(\eta_0)} \log^2 \left(1 - \frac{z}{x} \right) - (1 - z)^{a(\eta_0)} \log^2(1 - z) \right] \\
& \overset{z \to 1}{\simeq} -(1 - z)^{a(\eta_0)} \bigg\{ \log(1 - z) \Big\{ \log(1 - z)[\psi_0(a(\eta_0) + 1) + \gamma_{\mathrm{E}}] \\
& \hspace{4cm} + 2\psi_1(a(\eta_0) + 1) \Big\} + \psi_2(a(\eta_0) + 1) \bigg\}.
\end{aligned} \tag{6.124}
$$

Upon using these results, the two sides of Eq. (6.120) become linear combinations of terms proportional to $\log^p(1 - z)$, with $p = 0, 1, 2, 3$. By equating the coefficients of such terms, one finds a system of differential equations:

$$
\frac{d}{d\eta_0} a(\eta_0) = 1 - \frac{5\alpha}{9\pi} n_F, \tag{6.125}
$$

$$
\frac{d}{d\eta_0} b(\eta_0) = b(\eta_0) \left\{ -[\psi_0(a(\eta_0) + 1) + \gamma_{\mathrm{E}}] \left(1 - \frac{5\alpha}{9\pi} n_F \right) + \left(\lambda_0 + \frac{\alpha}{4\pi} \lambda_1 \right) \right\}, \tag{6.126}
$$

$$
\frac{d}{d\eta_0} e(\eta_0) = e(\eta_0) \left\{ -[\psi_0(a(\eta_0) + 1) + \gamma_{\mathrm{E}}] \left(1 - \frac{5\alpha}{9\pi} n_F \right) + \left(\lambda_0 + \frac{\alpha}{4\pi} \lambda_1 \right) \right\}, \tag{6.127}
$$

$$
\begin{aligned}
\frac{d}{d\eta_0} d(\eta_0) = & \, d(\eta_0) \left\{ -[\psi_0(a(\eta_0) + 1) + \gamma_{\mathrm{E}}] \left(1 - \frac{5\alpha}{9\pi} n_F \right) + \left(\lambda_0 + \frac{\alpha}{4\pi} \lambda_1 \right) \right\} \\
& + \left(1 - \frac{5\alpha}{9\pi} n_F \right) \left\{ -2 \, e(\eta_0) \, \psi_1(a(\eta_0) + 1) \right\},
\end{aligned} \tag{6.128}
$$

$$
\frac{d}{d\eta_0} c(\eta_0) = c(\eta_0) \left\{ -[\psi_0(a(\eta_0) + 1) + \gamma_{\mathrm{E}}] \left(1 - \frac{5\alpha}{9\pi} n_F \right) + \left(\lambda_0 + \frac{\alpha}{4\pi} \lambda_1 \right) \right\}
$$

$$+ \left(1 - \frac{5\alpha}{9\pi} n_F\right) \left\{ - d(\eta_0)\,\psi_1(a(\eta_0) + 1) - e(\eta_0)\,\psi_2(a(\eta_0) + 1) \right\}, \qquad (6.129)$$

with solutions:

$$a(\eta_0) = \eta_0 \left(1 - \frac{5\alpha}{9\pi} n_F\right) + a_0 \equiv \eta_1 + a_0\,, \qquad (6.130)$$

$$b(\eta_0) = b_0 \frac{e^{\hat{\eta}_1 - \gamma_E \eta_1}}{\Gamma(\eta_1 + a_0 + 1)}\,, \qquad (6.131)$$

$$e(\eta_0) = e_0 \frac{e^{\hat{\eta}_1 - \gamma_E \eta_1}}{\Gamma(\eta_1 + a_0 + 1)}\,, \qquad (6.132)$$

$$d(\eta_0) = e(\eta_0) \left[d_0 + \left(1 - \frac{5\alpha}{9\pi} n_F\right) \int_{\eta_0}^{1} dt\, 2\,\psi_1(a(t) + 1)\right]$$
$$= e(\eta_0) \left[d_0 - 2\,\psi_0(\eta_1 + a_0 + 1)\right]\,, \qquad (6.133)$$

$$c(\eta_0) = e(\eta_0) \left[c_0 + \left(1 - \frac{5\alpha}{9\pi} n_F\right) \int_{\eta_0}^{1} dt\, d(t)\,\psi_1(a(t) + 1) + \psi_2(a(t) + 1)\right]$$
$$= e(\eta_0) \left[c_0 - d_0\,\psi_0(\eta_1 + a_0 + 1) + \psi_0(\eta_1 + a_0 + 1)^2 - \psi_1(\eta_1 + a_0 + 1)\right]\,, \qquad (6.134)$$

where η_1 and $\hat{\eta}_1$ have been defined in Eqs. (6.109) and (6.110), respectively.

The arbitrary integration constants $a_0\,,\ldots e_0$ can be found by matching with the initial condition. We observe that at $\mu = \mu_0$ the $\alpha \to 0$ NLL result for the PDF must coincide with the LL one; this implies that that Eq. (6.119) must still hold true. Because of this, one can expand Eq. (6.122) by using the techniques employed in Appendix 6.6 (see in particular Eq. (6.234)), to obtain at $\mathcal{O}(\alpha)$ the same functional form as in Eq. (6.6), which leads to the following results:

$$c_0 = -\frac{7}{4} + \gamma_E^2 + \frac{\pi^2}{6} + \left(\gamma_E - \frac{3}{4}\right)\left(\log \frac{\mu_0^2}{m^2} - 1\right)\,, \qquad (6.135)$$

$$d_0 = 1 - 2\gamma_E - \log \frac{\mu_0^2}{m^2}\,, \qquad (6.136)$$

$$e_0 = -1\,. \qquad (6.137)$$

By putting everything back together, one sees that Eq. (6.122) coincides with Eq. (6.108).

6.5   Analytical large-z solution for singlet-photon sector

In this section we compute the asymptotic large-z solution for the singlet-photon sector. The key result relevant to the evolution in the $z \to 1$ region is the following:

$$\mathbb{P}_{\text{S},N} \xrightarrow{N \to \infty} \begin{pmatrix} -2\log \bar{N} + 2\lambda_0 & 0 \\ 0 & -\frac{2}{3} n_F \end{pmatrix} + \frac{\alpha}{2\pi} \begin{pmatrix} \frac{20}{9} n_F \log \bar{N} + \lambda_1 & 0 \\ 0 & -n_F \end{pmatrix} + \mathcal{O}(\alpha^2)\,, \quad (6.138)$$

that is obtained by means of a direct computation starting from the definitions given in Sec. 1.6 and from the explicit expressions for the Altarelli-Parisi splitting kernels provided in Appendix 6.C. Equation (6.138) implies that the singlet and the photon evolve independently from each other in this limit. Since the kernel evolution is a diagonal matrix, so is the evolution operator,

and therefore the solutions for its elements on the diagonal are given by either Eq. (6.20) or Eq. (6.24).

Let us start by considering the singlet. The singlet-singlet elements of the $\mathcal{O}(\alpha^0)$ and $\mathcal{O}(\alpha)$ matrices in Eq. (6.138) are identical to Eqs. (6.81) and (6.88) respectively. Thus, the solutions of Eqs. (6.103), (6.106), and (6.108) are also valid for the singlet.

As far as the photon is concerned, Eq. (1.11) and the photon-photon elements in Eq. (6.138) imply that the second term on the r.h.s. of Eq. (6.20) is equal to zero. Therefore:

$$M^{-1}\big[E_{\gamma\gamma,N}\big] = \frac{\alpha(\mu_0)}{\alpha(\mu)}\,\delta(1-z)\,, \tag{6.139}$$

having used Eq. (6.15). The convolution with the initial conditions of Eqs. (6.3) and (6.7) is thus trivial, and the final result reads as follows:

$$\Gamma_\gamma(z,\mu^2) = \frac{1}{2\pi}\frac{\alpha(\mu_0)^2}{\alpha(\mu)}\frac{1+(1-z)^2}{z}\left(\log\frac{\mu_0^2}{m^2} - 2\log z - 1\right)\,. \tag{6.140}$$

Unfortunately, Eq. (6.140) does not give a good description of the true large-z behaviour of the photon PDF. This is because the off-diagonal terms of the evolution kernel imply that such a PDF receives a contribution that primarily stems from the initial conditions of the electron PDF. As we have seen previously, these are much more peaked towards $z = 1$ than their photon counterparts, so much so that this behaviour compensates the fact that the off-diagonal elements of the evolution kernel are suppressed w.r.t. the diagonal ones, which are the only ones that have been taken into account in Eq. (6.138). It then follows that, in order to improve on the solution in Eq. (6.140), one needs to solve the evolution equations of the singlet-photon system by including those off-diagonal elements.

In order to do this, we start from writing the $\mathcal{O}(\alpha)$ expressions of the Altarelli-Parisi kernels as follows (see Eq. (1.71)):

$$\begin{aligned}
\mathbb{P}_{s,N} &= \mathbb{P}_{s,N}^{[0]} + \frac{\alpha(\mu)}{2\pi}\mathbb{P}_{s,N}^{[1]} + \mathcal{O}(\alpha^2) \tag{6.141}\\
&\equiv \left(\mathbb{P}_{s,N}^{[0,0]} + \frac{1}{N}\mathbb{P}_{s,N}^{[0,1]} + \mathcal{O}\left(N^{-2}\right)\right)\\
&\quad + \frac{\alpha(\mu)}{2\pi}\left(\mathbb{P}_{s,N}^{[1,0]} + \frac{1}{N}\mathbb{P}_{s,N}^{[1,1]} + \mathcal{O}\left(N^{-2}\right)\right) + \mathcal{O}(\alpha^2)\,, \tag{6.142}
\end{aligned}$$

having introduced, at each order in α, the leading- and subleading-N contributions. They read as follows:

$$\mathbb{P}_{s,N}^{[0,0]} = \begin{pmatrix} -2\log\bar{N} + 2\lambda_0 & 0 \\ 0 & -\frac{2}{3}n_F \end{pmatrix}\,, \tag{6.143}$$

$$\mathbb{P}_{s,N}^{[0,1]} = \begin{pmatrix} -1 & 2n_F \\ 1 & 0 \end{pmatrix}\,, \tag{6.144}$$

$$\mathbb{P}_{s,N}^{[1,0]} = \begin{pmatrix} \frac{20}{9}n_F\log\bar{N} + \lambda_1 & 0 \\ 0 & -n_F \end{pmatrix}\,, \tag{6.145}$$

$$
\mathbb{P}_{\mathrm{s},N}^{[1,1]} \;=\; \begin{pmatrix} -4\log\bar{N} + \frac{27+22n_F}{9} & 2n_F\left(\log^2\bar{N} + \frac{15-\pi^2}{6}\right) \\ -\log^2\bar{N} + \frac{15+4n_F}{3}\log\bar{N} - \frac{64n_F+3(36+\pi^2)}{18} & 0 \end{pmatrix} \tag{6.146}
$$

Note that, by considering only Eqs. (6.143) and (6.145), one recovers Eq. (6.138). According to Eq. (6.18), the Altarelli-Parisi kernels define the evolution kernel as follows:

$$
\mathbb{M}_N = \mathbb{P}_{\mathrm{s},N}^{[0]} + \frac{\alpha(\mu)}{2\pi}\left(\mathbb{P}_{\mathrm{s},N}^{[1]} - \frac{2\pi b_1}{b_0}\mathbb{P}_{\mathrm{s},N}^{[0]}\right), \tag{6.147}
$$

whence one can write the evolution equation and its formal solution as follows:

$$
\frac{\partial \mathbb{E}_N(t)}{\partial t} = \mathbb{M}_N(t)\mathbb{E}_N(t) \qquad \implies \qquad \mathbb{E}_N(t) = \exp\left[\sum_{k=1}^{\infty}\Omega_{k,N}(t)\right]. \tag{6.148}
$$

The solution in Eq. (6.148) is based on the so-called Magnus expansion [202] (see also Ref. [203]), which is constructed solely in terms of the evolution kernel:

$$
\Omega_{1,N}(t) = \int_0^t dt_1 \mathbb{M}_N(t_1), \tag{6.149}
$$

$$
\Omega_{2,N}(t) = \frac{1}{2}\int_0^t dt_1 \int_0^{t_1} dt_2 \Big[\mathbb{M}_N(t_1), \mathbb{M}_N(t_2)\Big], \tag{6.150}
$$

$$
\Omega_{3,N}(t) = \dots, \tag{6.151}
$$

with $\Omega_{k,N}(t)$ featuring k instances of $\mathbb{M}_N$, all appearing in commutators. Thus, in the case of a one-dimensional flavour space or of a diagonal evolution kernels, Eq. (6.148) is identical to the solution given in Eq. (6.20). As far as the singlet-photon sector is concerned, we can indeed recover the solutions we have found previously in terms of the quantity introduced in this section. We define the leading-N evolution kernel:

$$
\mathbb{M}_N^{(0)} = \mathbb{P}_{\mathrm{s},N}^{[0,0]} + \frac{\alpha(\mu)}{2\pi}\left(\mathbb{P}_{\mathrm{s},N}^{[1,0]} - \frac{2\pi b_1}{b_0}\mathbb{P}_{\mathrm{s},N}^{[0,0]}\right) \tag{6.152}
$$

and denote by $\mathbb{E}_N^{(0)}(t)$ the corresponding evolution operator. Thus:

$$
\frac{\partial \mathbb{E}_N^{(0)}(t)}{\partial t} = \mathbb{M}_N^{(0)}(t)\mathbb{E}_N^{(0)}(t) \qquad \implies \qquad \mathbb{E}_N^{(0)}(t) = \begin{pmatrix} E_{\Sigma\Sigma,N}^{(0)} & 0 \\ 0 & E_{\gamma\gamma,N}^{(0)} \end{pmatrix}, \tag{6.153}
$$

where:

$$
E_{\Sigma\Sigma,N}^{(0)} = \exp\left[-\xi_1\log\bar{N} + \hat{\xi}_1\right], \tag{6.154}
$$

$$
E_{\gamma\gamma,N}^{(0)} = \exp\left[-\frac{2n_F}{3}t - \frac{\alpha(\mu)-\alpha(\mu_0)}{4\pi^2 b_0}n_F\left(1 - \frac{4\pi b_1}{3b_0}\right)\right] \tag{6.155}
$$

$$
= \left(\frac{\alpha(\mu_0)}{\alpha(\mu)}\right)^{\frac{n_F}{3\pi b_0}}\exp\left[-\frac{\alpha(\mu)-\alpha(\mu_0)}{4\pi^2 b_0}n_F\left(1 - \frac{4\pi b_1}{3b_0}\right)\right] \xrightarrow{\mathrm{QED}} \frac{\alpha(\mu_0)}{\alpha(\mu)}. \tag{6.156}
$$

Equation (6.154) coincides with Eq. (6.90), while Eq. (6.155) coincides with Eq. (6.139), as they

should. This is not immediately apparent in the case of Eq. (6.155) since there, at variance with what has been done in Eq. (6.139), we have not used the simplifications induced by the explicit expressions of the QED β-function coefficients (see Eq. (1.11)). This is useful when one considers the limit of non-running α of the formulae presented here. An expression equivalent to Eq. (6.155), as well as the QED "limit" of both, is given in Eq. (6.156).

We stress that the case of non-running α is problematic, as it might lead to inconsistencies. By switching off the running, one effectively neglects bubble-diagram contributions which are exactly the same as those that lead to the $\gamma\gamma$ entries in Eqs. (6.143) and (6.145). In this context we ignore such potential inconsistencies, but then we need to carefully distinguish the $\gamma\gamma$ contributions to the Altarelli-Parisi kernels (which we always parametrise by means of n_F) from those to the QED β function (which we parametrise by means of the β-function coefficients b_i). We shall return to this point with one explicit example later in this section (see Eqs. (6.215) and (6.216)).

In order to improve on the leading-N results, we shall introduce the subleading-N contributions to the evolution kernel, and treat them as a perturbation to the solution of Eq. (6.153). This entails writing:

$$\mathbb{M}_N = \mathbb{M}_N^{(0)} + \frac{1}{N}\mathbb{M}_N^{(1)} \qquad \Longrightarrow \qquad \mathbb{E}_N(t) = \mathbb{E}_N^{(0)}(t)\mathbb{E}_N^{(1)}(t) , \tag{6.157}$$

having defined:

$$\mathbb{M}_N^{(1)} = \mathbb{P}_{s,N}^{[0,1]} + \frac{\alpha(\mu)}{2\pi}\left(\mathbb{P}_{s,N}^{[1,1]} - \frac{2\pi b_1}{b_0}\,\mathbb{P}_{s,N}^{[0,1]}\right). \tag{6.158}$$

By replacing Eq. (6.157) into Eq. (6.148), one arrives at the evolution equation for the operator $\mathbb{E}_N^{(1)}(t)$:

$$\frac{\partial \mathbb{E}_N^{(1)}(t)}{\partial t} = \widehat{\mathbb{M}}_N^{(1)}(t)\mathbb{E}_N^{(1)}(t) , \qquad \widehat{\mathbb{M}}_N^{(1)}(t) = \frac{1}{N}\left(\mathbb{E}_N^{(0)}(t)\right)^{-1}\mathbb{M}_N^{(1)}(t)\mathbb{E}_N^{(0)}(t) . \tag{6.159}$$

Equation (6.159) can be solved as is written in Eq. (6.148), by constructing the $\Omega_{k,N}(t)$ terms according to Eqs. (6.149)–(6.151) with $\mathbb{M}_N \to \widehat{\mathbb{M}}_N^{(1)}$ there. We then observe that $\Omega_{k,N} \propto 1/N^k$, and thus for consistency with Eq. (6.142) we are allowed to discard all contributions with $k \geq 2$. Therefore:

$$\mathbb{E}_N^{(1)}(t) = \exp\left[\Omega_{1,N}(t)\right] + \mathcal{O}\left(1/N^2\right) = I + \int_0^t dt_1 \widehat{\mathbb{M}}_N^{(1)}(t_1) + \mathcal{O}\left(1/N^2\right) . \tag{6.160}$$

In spite of these simplifications, the integral on the r.h.s. of Eq. (6.160) features contributions of the type $\exp(at_1)\exp(\exp(bt_1))$ for certain a and b, where the functional dependence $\exp(\exp(bt_1))$ stems for the dependence on t_1 of $\alpha(\mu)$ in $\mathbb{E}_N^{(0)}$. Apart from rendering the t_1 integral in Eq. (6.160) non trivial, this will also induce functional forms in the N-space whose analytical inverse Mellin transforms will be extremely hard to compute. We shall therefore resort to simplifying the expression of $\mathbb{E}_N^{(0)}$, by linearising the dependence on t_1 of $\alpha(\mu)$ there. This implies that, as an evolution kernel, we shall use what follows:

$$\widehat{\mathbb{M}}_N^{(1,L)}(t) = \widehat{\mathbb{M}}_N^{(1)}(t)\left[\mathbb{E}_N^{(0)}(t) \longrightarrow \mathbb{E}_N^{(0,L)}(t)\right], \tag{6.161}$$

153

where:

$$\mathbb{E}_N^{(0,L)}(t) = \begin{pmatrix} E_{\Sigma\Sigma,N}^{(0,L)} & 0 \\ 0 & E_{\gamma\gamma,N}^{(0,L)} \end{pmatrix}, \tag{6.162}$$

whose expression can be obtained from Eqs. (6.154) and (6.155) after the linearisation introduced above. Thus:

$$E_{\Sigma\Sigma,N}^{(0,L)} = \exp\left[\left(-\xi_{1,0}\log\bar{N} + \hat{\xi}_{1,0}\right)t\right], \tag{6.163}$$

$$E_{\gamma\gamma,N}^{(0,L)} = \exp\left[-\left(\frac{2n_F}{3} + \chi_{1,0}\right)t\right]. \tag{6.164}$$

In equation (6.163) we have introduced the quantities $\xi_{1,0}$ and $\hat{\xi}_{1,0}$ which we have defined as follows:

$$\xi_1 = \xi_{1,0}\,t + \mathcal{O}(t^2), \qquad \hat{\xi}_1 = \hat{\xi}_{1,0}\,t + \mathcal{O}(t^2), \tag{6.165}$$

with ξ_1 and $\hat{\xi}_1$ given in Eqs. (6.91) and (6.93). By means of an explicit computations from the latter two equations we obtain:

$$\xi_{1,0} = 2\left[1 - \frac{\alpha(\mu_0)}{\pi}\left(\frac{5}{9}n_F + \frac{\pi b_1}{b_0}\right)\right], \tag{6.166}$$

$$\hat{\xi}_{1,0} = \frac{3}{2}\left[1 + \frac{\alpha(\mu_0)}{\pi}\left(\frac{\lambda_1}{3} - \frac{\pi b_1}{b_0}\right)\right]. \tag{6.167}$$

As far as Eq. (6.164) is concerned, its expression stems from that of Eq. (6.155); in particular:

$$-\frac{\alpha(\mu) - \alpha(\mu_0)}{4\pi^2 b_0}n_F\left(1 - \frac{4\pi b_1}{3b_0}\right) = -\chi_{1,0}\,t + \mathcal{O}(t^2), \tag{6.168}$$

from whence:

$$\chi_{1,0} = \frac{\alpha(\mu_0)}{2\pi}n_F\left(1 - \frac{4\pi b_1}{3b_0}\right) \xrightarrow{\text{QED}} 0. \tag{6.169}$$

In summary, the evolution operator we shall use is the following:

$$\mathbb{E}_N(t) = \mathbb{E}_N^{(0,L)}(t)\left(I + \int_0^t dt_1 \widehat{\mathbb{M}}_N^{(1,L)}(t)\right). \tag{6.170}$$

Having established that the asymptotic solutions presented in Sec. 6.4 are perfectly adequate for the case of the singlet, we shall now focus on the implications of Eq. (6.170) on the photon PDF. We obtain:

$$\Gamma_\gamma(z) = M^{-1}\left[\left(\mathbb{E}_N(t)\right)_{\gamma\Sigma}\Gamma_{s,0,N}\right] + M^{-1}\left[\left(\mathbb{E}_N(t)\right)_{\gamma\gamma}\Gamma_{\gamma,0,N}\right], \tag{6.171}$$

with $\Gamma_{s,0,N}$ and $\Gamma_{\gamma,0,N}$ the N-space expressions of the singlet and photon initial conditions, respectively. These can be obtained from Eqs. (6.3)–(6.8):

$$\Gamma_{s,0,N} = 1 + \frac{\alpha(\mu_0)}{2\pi}\left(F_0 + F_1\log\bar{N} + F_2\log^2\bar{N}\right) + \mathcal{O}\left(N^{-1}\right), \tag{6.172}$$

$$\Gamma_{\gamma,0,N} = \mathcal{O}\left(N^{-1}\right), \tag{6.173}$$

where:

$$
\begin{align}
F_0 &= 2 - \frac{\pi^2}{3} + \frac{3}{2}L_0\,, \tag{6.174}\\
F_1 &= 2\,(1 - L_0)\,, \tag{6.175}\\
F_2 &= -2\,. \tag{6.176}
\end{align}
$$

Let us start by considering the contribution of the first term on the r.h.s. of Eq. (6.171). With a straightforward, if tedious, computation we obtain what follows:

$$
\big(\mathbb{E}_N(t)\big)_{\gamma\Sigma}\,\Gamma_{\mathrm{S},0,N} \;\xrightarrow{N\to\infty}\; E^{(0,L)}_{\gamma\gamma,N}\,\frac{1}{N}\sum_{j=1}^{4}\bar{N}^{-\kappa_j}\,\frac{\sum_{i=0}^{4} x_i^{(j)}\log^i \bar{N}}{y_0^{(j)} + y_1^{(j)}\log\bar{N}}\,, \tag{6.177}
$$

with $\{x_0^{(j)},\dots x_4^{(j)}, y_0^{(j)}, y_1^{(j)}\}$ four sets of N-independent quantities, whose specific forms are unimportant here. For any given j, the five terms in the numerators on the r.h.s. of Eq. (6.177) can be re-expressed *algebraically* (i.e. without any approximations) in terms of the corresponding denominators. In this way, one arrives at the following forms (note that $E^{(0,L)}_{\gamma\gamma,N}$ is independent of N):

$$
\begin{align}
\Gamma_{\gamma,j}(z) &= M^{-1}\left[\frac{1}{N}\bar{N}^{-\kappa_j}\,\frac{\sum_{i=0}^{4} x_i^{(j)}\log^i \bar{N}}{y_0^{(j)} + y_1^{(j)}\log\bar{N}}\,,\right]\,, \qquad j = 1,2,3,4\,, \notag\\
&\equiv \sum_{i=1}^{5} R_i\Big(C_{1,j}, C_{2,j}, C_{3,j}, D_{2,j}/D_{1,j}, D_{1,j}\Big)\,\mathfrak{M}_i\Big(z;\kappa_j, D_{1,j}, D_{2,j}\Big)\,. \tag{6.178}
\end{align}
$$

with:

$$
\begin{align}
R_1(C_1, C_2, C_3, C_4, C_5) &= \Big(C_3 - C_4 C_2 + C_4^2 C_1\Big) \notag\\
&\qquad \times \left[1 + \frac{\alpha(\mu_0)}{2\pi}\Big(F_0 - C_4 F_1 + C_4^2 F_2\Big)\right]\,, \tag{6.179}\\
R_2(C_1, C_2, C_3, C_4, C_5) &= \frac{1}{C_5}(C_2 - C_4 C_1) \notag\\
&+ \frac{\alpha(\mu_0)}{2\pi}\frac{1}{C_5}\Big(C_2 F_0 + C_3 F_1 - C_4\big(C_1 F_0 + C_2 F_1 + C_3 F_2\big) \notag\\
&\qquad\qquad + C_4^2\big(C_1 F_1 + C_2 F_2\big) - C_4^3 C_1 F_2\Big)\,, \tag{6.180}\\
R_3(C_1, C_2, C_3, C_4, C_5) &= \frac{C_1}{C_5} + \frac{\alpha(\mu_0)}{2\pi}\frac{1}{C_5}\Big(C_1 F_0 + C_2 F_1 + C_3 F_2 \notag\\
&\qquad\qquad - C_4\big(C_1 F_1 + C_2 F_2\big) + C_4^2 C_1 F_2\Big)\,, \tag{6.181}\\
R_4(C_1, C_2, C_3, C_4, C_5) &= \frac{\alpha(\mu_0)}{2\pi}\frac{1}{C_5}\Big(C_1 F_1 + C_2 F_2 - C_4 C_1 F_2\Big)\,, \tag{6.182}\\
R_5(C_1, C_2, C_3, C_4, C_5) &= \frac{\alpha(\mu_0)}{2\pi}\frac{C_1}{C_5}F_2\,, \tag{6.183}
\end{align}
$$

$$D_{1,1} = \xi_{1,0}\,, \tag{6.184}$$

$$D_{2,1} = -\left(\frac{2n_F}{3} + 2\pi b_0 + \hat{\xi}_{1,0} + \chi_{1,0}\right)\,, \tag{6.185}$$

$$C_{1,1} = \frac{\alpha(\mu_0)}{2\pi} \exp\left(-D_{2,1}t\right)\,, \tag{6.186}$$

$$C_{2,1} = -\frac{\alpha(\mu_0)}{2\pi}\left(5 + \frac{4n_F}{3}\right)\exp\left(-D_{2,1}t\right)\,, \tag{6.187}$$

$$C_{3,1} = \frac{\alpha(\mu_0)}{2\pi}\left(6 + \frac{\pi^2}{6} + \frac{32n_F}{9} + \frac{2\pi b_1}{b_0}\right)\exp\left(-D_{2,1}t\right)\,, \tag{6.188}$$

$$D_{1,2} = D_{1,1}\,, \tag{6.189}$$

$$D_{2,2} = D_{2,1}\,, \tag{6.190}$$

$$C_{1,2} = -\frac{\alpha(\mu_0)}{2\pi}\,, \tag{6.191}$$

$$C_{2,2} = \frac{\alpha(\mu_0)}{2\pi}\left(5 + \frac{4n_F}{3}\right)\,, \tag{6.192}$$

$$C_{3,2} = -\frac{\alpha(\mu_0)}{2\pi}\left(6 + \frac{\pi^2}{6} + \frac{32n_F}{9} + \frac{2\pi b_1}{b_0}\right)\,, \tag{6.193}$$

$$D_{1,3} = D_{1,1}\,, \tag{6.194}$$

$$D_{2,3} = -\left(\frac{2n_F}{3} + \hat{\xi}_{1,0} + \chi_{1,0}\right)\,, \tag{6.195}$$

$$C_{1,3} = 0\,, \tag{6.196}$$

$$C_{2,3} = 0\,, \tag{6.197}$$

$$C_{3,3} = -\exp\left(-D_{2,3}t\right)\,, \tag{6.198}$$

$$D_{1,4} = D_{1,3}\,, \tag{6.199}$$

$$D_{2,4} = D_{2,3}\,, \tag{6.200}$$

$$C_{1,4} = 0\,, \tag{6.201}$$

$$C_{2,4} = 0\,, \tag{6.202}$$

$$C_{3,4} = 1\,, \tag{6.203}$$

and:

$$k_j = \xi_{1,0}\, t\,, \qquad j = 1, 3\,, \tag{6.204}$$

$$k_j = 0\,, \qquad j = 2, 4\,. \tag{6.205}$$

We have introduced the inverse Mellin transforms relevant to Eq. (6.177) which are linearly independent from each other, namely:

$$M^{-1}\left[\frac{\bar{N}^{-\kappa}}{N}\frac{1}{M_2 + M_1 \log \bar{N}}\right] \xrightarrow{z\to 1} \mathfrak{M}_1(z; \kappa, M_1, M_2)\,, \tag{6.206}$$

$$M^{-1}\left[\frac{\bar{N}^{-\kappa}}{N}\log^p \bar{N}\right] \xrightarrow{z\to 1} \mathfrak{M}_{p+2}(z; \kappa, M_1, M_2)\,, \qquad p = 0, 1, 2, 3\,. \tag{6.207}$$

Explicit computations give:

$$\mathfrak{M}_1(z;\kappa,M_1,M_2) = \frac{e^{-\gamma_{\rm E}\kappa}(1-z)^\kappa}{\Gamma(1+\kappa)}\left(\frac{1}{M_2-M_1\log(1-z)} - \frac{(\pi^2\kappa-6\zeta_3\kappa^2)M_1}{6(M_2-M_1\log(1-z))^2}\right.$$

$$\left. - \frac{(30\pi^2-360\zeta_3\kappa+\pi^4\kappa^2)M_1^2}{180(M_2-M_1\log(1-z))^3}\right), \tag{6.208}$$

$$\mathfrak{M}_2(z;\kappa,M_1,M_2) = \frac{e^{-\gamma_{\rm E}\kappa}(1-z)^\kappa}{\Gamma(1+\kappa)}, \tag{6.209}$$

$$\mathfrak{M}_3(z;\kappa,M_1,M_2) = \frac{e^{-\gamma_{\rm E}\kappa}(1-z)^\kappa}{\Gamma(1+\kappa)}\left(-\log(1-z)+\frac{\pi^2\kappa}{6}-\zeta_3\kappa^2\right), \tag{6.210}$$

$$\mathfrak{M}_4(z;\kappa,M_1,M_2) = \frac{e^{-\gamma_{\rm E}\kappa}(1-z)^\kappa}{\Gamma(1+\kappa)}\left(\log(1-z)^2-\frac{\pi^2}{6}+\kappa\left(-\frac{\pi^2}{3}\log(1-z)+2\zeta_3\right)\right.$$

$$\left. + \kappa^2\left(2\zeta_3\log(1-z)-\frac{\pi^4}{180}\right)\right), \tag{6.211}$$

$$\mathfrak{M}_5(z;\kappa,M_1,M_2) = \frac{e^{-\gamma_{\rm E}\kappa}(1-z)^\kappa}{\Gamma(1+\kappa)}\left(-\log(1-z)^3+\frac{\pi^2}{2}\log(1-z)-2\zeta_3\right. \tag{6.212}$$

$$+ \kappa\left(\frac{\pi^2}{2}\log(1-z)^2-6\zeta_3\log(1-z)-\frac{\pi^4}{60}\right)$$

$$\left. + \kappa^2\left(-3\zeta_3\log(1-z)^2+\frac{\pi^4}{60}\log(1-z)+\frac{3}{2}\pi^2\zeta_3-12\zeta_5\right)\right),$$

where, consistently with Eqs. (6.206) and (6.207), in Eqs. (6.208)–(6.212) some terms that vanish at $z\to 1$ have not been included. This is of course arbitrary to some extent, and the logic we have followed is that of keeping those terms which, when expanded in series, either contribute to the same monomials t^n and αt^n as the recursive solutions considered in Sec. 6.3, or have the same power of κ as the former ones. On top of this, one has the special case of Eq. (6.208) which has the structure of a series in $M_1^{k-1}(M_2-M_1\log(1-z))^{-k}$. When $z\to 1$, these terms are progressively more suppressed with increasing k. Unfortunately, this hierarchy is not valid at intermediate z's; in fact, for the values of M_1 and M_2 relevant to our computation there is a singularity at $z\simeq 0.65$ which is dominated by increasingly large values of k. This is what prevents the asymptotic solution of the photon PDF from being well-behaved in all of the z range, at variance with its electron counterpart. This has significant implications for the matching, which are discussed in Sec. 6.7.

We next consider the contribution of the second term on the r.h.s. of Eq. (6.171). Owing to Eq. (6.170), to the $1/N$ suppression implicit in $\widehat{\mathbb{M}}_N^{(1,L)}$, and to Eq. (6.173), it is immediate to see that this contribution, up to terms vanishing in the $z\to 1$ limit, is identical to that of Eq. (6.140), bar for an $\alpha(\mu_0)/\alpha(\mu)$ prefactor that here needs to be written according to Eq. (6.164). Thus, by introducing the quantity:

$$\Gamma_{\gamma,5}(z) = \frac{\alpha(\mu_0)}{2\pi}\frac{1+(1-z)^2}{z}\left(\log\frac{\mu_0^2}{m^2}-2\log z-1\right), \tag{6.213}$$

157

we can write the sought large-z expression of the photon PDFs in a compact form:

$$\Gamma_\gamma(z) \;\; = \;\; \exp\left[-\left(\frac{2n_F}{3} + \chi_{1,0}\right)t\right] \sum_{j=1}^{5} \Gamma_{\gamma,j}(z)\,, \qquad (6.214)$$

with $\Gamma_{\gamma,j}(z)$ given in Eq. (6.178) for $j \leq 4$ and in Eq. (6.213) for $j = 5$.

The results presented above allow one to obtain their counterparts in the case of non-running α, by means of the following formal replacements (see Eq. (6.22)):

$$t \;\longrightarrow\; \frac{\eta_0}{2}\,, \qquad \chi_{1,0}\,t \;\longrightarrow\; \frac{\alpha}{2\pi}\frac{\eta_0}{2}\,n_F\,, \qquad (6.215)$$

$$b_0 \;\longrightarrow\; 0\,, \qquad b_1 \;\longrightarrow\; 0\,, \qquad b_1/b_0 \;\longrightarrow\; 0\,, \qquad (6.216)$$

with η_0 defined in Eq. (6.30). We can also see that, by using the replacements above in the expression for $E^{(0)}_{\gamma\gamma,N}$ given in Eq. (6.155), one obtains the same result as one would have directly read from the solution for the evolution operator relevant to the case of non-running α (Eq. (6.24), with $\alpha(\mu) \to \alpha$ and $b_0 \to 0$ there). We observe that this would not have happened if one had used Eq. (6.156) instead of Eq. (6.155), in spite of these two equations being identical in QED. In other words, the replacements in Eqs. (6.215) and (6.216) might lead to an incorrect result in the limit of non-running α if applied to an expression that contains two values of α computed at different scales; when this is the case, one must first express one of such α values in terms of the other one, and of t. That being said, we point out again that the limit of non-running α must be interpreted with some care — see the comments that follow Eq. (6.156).

When not considering the case of non-running α, one can re-expressed the exponential prefactors in Eq. (6.214) and in Eqs. (6.184)–(6.203), and their combinations, in simpler ways, namely:

$$\exp\left(-D_{2,1}t\right)\exp\left[-\left(\frac{2n_F}{3} + \chi_{1,0}\right)t\right] \;\; = \;\; \frac{\alpha(\mu)}{\alpha(\mu_0)}\,e^{\hat{\xi}_{1,0}t} \;\longrightarrow\; \frac{\alpha(\mu)}{\alpha(\mu_0)}\,e^{\hat{\xi}_1}\,, \qquad (6.217)$$

$$\exp\left(-D_{2,3}t\right)\exp\left[-\left(\frac{2n_F}{3} + \chi_{1,0}\right)t\right] \;\; = \;\; e^{\hat{\xi}_{1,0}t} \;\longrightarrow\; e^{\hat{\xi}_1}\,, \qquad (6.218)$$

$$\exp\left[-\left(\frac{2n_F}{3} + \chi_{1,0}\right)t\right] \;\; = \;\; \frac{\alpha(\mu_0)}{\alpha(\mu)}\,. \qquad (6.219)$$

Two observations are in order. Firstly, the expressions on the r.h.s.'s of Eqs. (6.217) and (6.218) factorise in the functions R_i, owing to the linearity of the latter w.r.t. $C_{1,j}$, $C_{2,j}$, and $C_{3,j}$. Secondly, the replacements on the rightmost sides of Eqs. (6.217) and (6.218) stem from Eq. (6.165); they are not mandatory, but are consistent with the linearisation simplifications made when solving the evolution equations. For scales of the order of up to a few hundred GeV's, in practice they do not induce any significant numerical differences. With the same arguments, in Eq. (6.178) one can also perform the replacements:

$$k_j = \xi_{1,0}\,t \;\longrightarrow\; \xi_1 \qquad j = 1, 3\,, \qquad (6.220)$$

again from Eq. (6.165).

Equation (6.214) is the asymptotic solution that emerges from solving the evolution equation by keeping the dominant off-diagonal terms in the Altarelli-Parisi kernels. As we shall discuss

in Sec. 6.6, it shares with its singlet and non-singlet counterparts the nice property that its perturbative expansion lead to the same coefficients as those of the recursive solutions (for certain classes of basis functions in the z space). However, its functional form is rather involved, but it is fortunately possible to simplify it, by keeping only the truly dominant terms in the $z \to 1$ limit at each order in α. In order to do so one starts by observing that, in such a limit, one has:

$$\mathfrak{M}_i(z; \kappa, M_1, M_2) \quad \overset{z \to 1}{\longrightarrow} \quad 0\,, \tag{6.221}$$

$$\mathfrak{M}_1(z; 0, M_1, M_2) \quad \overset{z \to 1}{\longrightarrow} \quad 0\,, \tag{6.222}$$

$$\mathfrak{M}_2(z; 0, M_1, M_2) \quad \overset{z \to 1}{\longrightarrow} \quad 1\,, \tag{6.223}$$

$$\mathfrak{M}_3(z; 0, M_1, M_2) \quad \overset{z \to 1}{\longrightarrow} \quad -\log(1 - z)\,, \tag{6.224}$$

$$\mathfrak{M}_4(z; 0, M_1, M_2) \quad \overset{z \to 1}{\longrightarrow} \quad \log^2(1 - z)\,, \tag{6.225}$$

$$\mathfrak{M}_5(z; 0, M_1, M_2) \quad \overset{z \to 1}{\longrightarrow} \quad \frac{\pi^2}{2} \log(1 - z) - \log^3(1 - z)\,, \tag{6.226}$$

for any values of M_1 and M_2. Because of Eqs. (6.204) and (6.205), Eq. (6.221) implies that only the $j = 2$ and $j = 4$ contributions to Eq. (6.214) govern the divergent behaviour of $\Gamma_\gamma(z)$ at $z \to 1$. A simple computation then leads to the following result:

$$\Gamma_\gamma(z) \overset{z \to 1}{\longrightarrow} \frac{\alpha(\mu_0)^2}{\alpha(\mu)} \frac{3}{2\pi\xi_{1,0}} \log(1 - z) - \frac{\alpha(\mu_0)^3}{\alpha(\mu)} \frac{1}{2\pi^2\xi_{1,0}} \log^3(1 - z)\,. \tag{6.227}$$

There is a certain similarity between Eq. (6.227) and Eq. (6.103) which is worth stressing. In particular, the dominant term at $z \to 1$ in both equations (proportional to $\log(1 - z)^3$ and $\log(1 - z)^2$, respectively) is suppressed w.r.t. the subdominant one ($\log(1 - z)$ in both cases) by a factor proportional to α (owing to Eq. (6.104) for Eq. (6.103)). This implies that numerically the onset of the behaviour driven by the most divergent terms occurs only at z values which are exceedingly large, and in fact hardly relevant to any phenomenological applications, as we shall see in Sec. 6.9.

Equation (6.214) simplifies considerably when one retains only the LL terms. A direct calculation leads to the following result:

$$\Gamma_\gamma(z) = -e^{\hat{\xi}_0} \mathfrak{M}_1\Big(z; \xi_0, D_1^{(0)}, D_2^{(0)}\Big) + \frac{\alpha(\mu_0)}{\alpha(\mu)} \mathfrak{M}_1\Big(z; 0, D_1^{(0)}, D_2^{(0)}\Big)\,, \tag{6.228}$$

with ξ_0 and $\hat{\xi}_0$ defined in Eq. (6.107), and:

$$D_1^{(0)} = 2\,, \tag{6.229}$$

$$D_2^{(0)} = -\frac{2n_F}{3} - \frac{3}{2}\,. \tag{6.230}$$

We point out that, consistently with the results of Appendix 6.B, the LL photon PDF is of $\mathcal{O}(t)$ (i.e. it does vanish with $\alpha \to 0$): the two terms on the r.h.s. of Eq. (6.228) cancel each other at $t = 0$. From Eq. (6.208), we also see that the LL-accurate photon PDF of Eq. (6.228) vanishes in the $z \to 1$ limit:

$$\Gamma_\gamma(z) \overset{z \to 1}{\longrightarrow} 0\,. \tag{6.231}$$

By comparing Eqs. (6.227) and (6.231) we observe that the photon PDF has a behaviour analogous to that of the electron PDF, namely that its NLL form grows faster than its LL counterpart at $z \to 1$; to a good extent, this is an artifact of the $\overline{\text{MS}}$ scheme.

We finally point out that Eq. (6.228) can be directly obtained from solving the evolution equation of Eq. (6.148), by using there:

$$\mathbb{M}_N = \mathbb{P}_{\text{s},N}^{[0,0]} + \frac{1}{N}\mathbb{P}_{\text{s},N}^{[0,1]}\,. \tag{6.232}$$

Since the kernel of Eq. (6.232) is independent of t, Eq. (6.148) can be simply solved by diagonalisation. After that, one multiplies the results by the LO initial conditions, and performs the inverse Mellin transform. The fact that by doing so one recovers Eq. (6.228) is a rather powerful check on the procedure adopted in this section.

6.6   Expansion of large-z solutions

In view of the matching between the asymptotic large-z solutions and the recursive solutions, it is useful to consider the expansion of the former ones in a series of α; this will also allow us to perform some consistency checks on them. We can formally write the result of such an expansion for the NLL-accurate, running-α solutions of Eqs. (6.103) and (6.214) in the same way as in Eqs. (6.62) and (6.76), namely:

$$\overline{\Gamma}(z,\mu^2) = \sum_{k=0}^{k_{\max}^{\text{LL}}} \frac{t^k}{k!}\,\mathcal{K}_k^{\text{LL}}(z) + \frac{\alpha(t)}{2\pi}\sum_{k=0}^{k_{\max}^{\text{NLL}}} \frac{t^k}{k!}\,\mathcal{K}_k^{\text{NLL}}(z)\,. \tag{6.233}$$

As the notation with an overline suggests, we only take into account contributions that do no vanish at $z \to 1$. We point out that we consider the expansion up to $\mathcal{O}(\alpha^3)$, i.e. we use the values in Eq. (6.63), for the sole reason of consistency with what has been done for the recursive solutions in Sec. 6.3. We shall omit flavour indices here, in order to simplify the notation, since no confusion is possible. In fact, one must bear in mind that the large-z solutions of the singlet and non-singlet PDFs coincide, and that the one of the photon has a functional behaviour significantly different from the former two. Therefore we shall first deal with the singlet non-singlet cases together, and with that of the photon afterwards.

6.6.1   Singlet and non-singlet

When expanding Eq. (6.103) to obtain $\mathcal{K}_k^{\text{LL}}(z)$ and $\mathcal{K}_k^{\text{NLL}}(z)$, one can simply use the explicit expressions of ξ_1 and $\hat{\xi}_1$ in Eqs. (6.91) and (6.93), respectively, and then consider the Taylor series in t and α. However, this procedure cannot possibly give a correct answer at $z = 1$, since $\Gamma(z)$ diverges there order by order, with non-integrable singularities. In order to properly take such an endpoint contribution into account, all z-dependent terms in $\Gamma(z)$ must be regarded as distributions, rather than as regular functions. By doing so, one can exploit the following identities:

$$\frac{\log^p(1-z)}{(1-z)^{1-\kappa}} = \frac{(-1)^p\,\Gamma(1+p)}{\kappa^{1+p}}\,\delta(1-z) + \sum_{i=0}^{\infty} \frac{\kappa^i}{\Gamma(1+\kappa)}\,\mathcal{L}_{i+p}(z)\,, \qquad p \geq 0\,, \tag{6.234}$$

for any κ, and where:

$$\mathcal{L}_i(z) = \Big[\ell_i(z)\Big]_+ \equiv \left[\frac{\log^i(1-z)}{1-z}\right]_+, \qquad i \geq 0\,, \tag{6.235}$$

having introduced $\ell_i(z)$ in Eq. (6.69). By using Eq. (6.234) in Eq. (6.103) with $\kappa = \xi_1$, and by subsequently expanding in t and α, one determines $\mathcal{K}_k^{\mathrm{LL}}(z)$ and $\mathcal{K}_k^{\mathrm{NLL}}(z)$. Because of the structure of Eq. (6.234), it is clear that the latter two quantities can be expressed as linear combinations of the $\mathcal{L}_i(z)$ distributions and of Dirac delta's, namely:

$$\mathcal{K}_k^{\mathrm{LL}}(z) = A_k^{\mathrm{LL}}\delta(1-z) + (1-\delta_{k0})\sum_{i=0}^{i_{\max}^{\mathrm{LL}}(k)} B_{k,i}^{\mathrm{LL}}\,\mathcal{L}_i(z)\,, \qquad k \geq 0\,, \tag{6.236}$$

$$\mathcal{K}_k^{\mathrm{NLL}}(z) = A_k^{\mathrm{NLL}}\delta(1-z) + \sum_{i=0}^{i_{\max}^{\mathrm{NLL}}(k)} B_{k,i}^{\mathrm{NLL}}\,\mathcal{L}_i(z)\,, \qquad k \geq 0\,. \tag{6.237}$$

Equations (6.236) and (6.237) are by construction valid for any z, including $z = 1$, and so is Eq. (6.233). The $z = 1$ contribution will be used in the following, but is not relevant for the matching procedure. For the latter, $\overline{\Gamma}(z)$ will be considered only with $z < 1$, and thus becomes an ordinary function. Its form can be read directly from Eq. (6.233), and is as follows:

$$\overline{\Gamma}(z,\mu^2) = \sum_{k=0}^{k_{\max}^{\mathrm{LL}}} \frac{t^k}{k!}\,K_k^{\mathrm{LL}}(z) + \frac{\alpha(t)}{2\pi}\sum_{k=0}^{k_{\max}^{\mathrm{NLL}}} \frac{t^k}{k!}\,K_k^{\mathrm{NLL}}(z)\,, \tag{6.238}$$

where:

$$K_k^{\mathrm{LL}}(z) = \mathcal{K}_k^{\mathrm{LL}}(z)\Big[A_k^{\mathrm{LL}} \to 0\,,\ \mathcal{L}_i(z) \to \ell_i(z)\Big]\,, \tag{6.239}$$

$$K_k^{\mathrm{NLL}}(z) = \mathcal{K}_k^{\mathrm{NLL}}(z)\Big[A_k^{\mathrm{NLL}} \to 0\,,\ \mathcal{L}_i(z) \to \ell_i(z)\Big]\,. \tag{6.240}$$

Note the strict similarity between Eqs. (6.238) and (6.76). This has to be expected, since both of these expressions are $\mathcal{O}(\alpha^3)$ approximations of the PDF, that retain either some (Eq. (6.238)) or all (Eq. (6.76)) of the terms that are singular for $z \to 1$.

We have determined the coefficients A_k^{LL} and $B_{k,i}^{\mathrm{LL}}$ for $k \leq 3$, and A_k^{NLL} and $B_{k,i}^{\mathrm{NLL}}$ for $k \leq 2$, by means of a direct computation. The results for $k = 0$ are particularly interesting since, in view of Eq. (6.233), they must be related to the initial conditions of Eqs. (6.3) and (6.6). We have obtained:

$$A_0^{\mathrm{LL}} = 1\,, \tag{6.241}$$

$$A_0^{\mathrm{NLL}} = 2 + \frac{3}{2}\,L_0\,, \tag{6.242}$$

$$B_{0,0}^{\mathrm{NLL}} = 2\,(L_0 - 1)\,, \tag{6.243}$$

$$B_{0,1}^{\mathrm{NLL}} = -4\,, \tag{6.244}$$

where L_0 has been defined in Eq. (6.335). With the result of Eq. (6.241), $K_0^{\mathrm{LL}}(z)$ is indeed identical to Eq. (6.3). However, by replacing the results of Eqs. (6.242)–(6.244) into Eq. (6.237),

$K_0^{\text{NLL}}(z)$ turns out *not* to coincide with $\Gamma_{e^-}^{[1]}(z)$ of Eq. (6.6). This is hardly surprising: when working in the large-z region, one is entitled to set $z = 1$ in all of the polynomial terms that appear in the numerators. Therefore, while $K_0^{\text{NLL}}(z)$ should not necessarily be equal to $\Gamma_{e^-}^{[1]}(z)$, it *must* be equal to the $z \to 1$ asymptotic form of the latter – if that were not the case, the large-z solution would not be compatible with the initial conditions from which it supposedly originates. In order to obtain the asymptotic expression of the initial condition, one cannot set $z = 1$ in all of the numerators of the latter right away, since $\Gamma_{e^-}^{[1]}(z)$ is not an ordinary function, but a distribution. Before doing so, one must first pull out the $1 + z^2$ factors from the plus distributions in Eq. (6.6). This can be done by exploiting the following identities:

$$\frac{1+z^2}{(1-z)_+} = \left(\frac{1+z^2}{1-z}\right)_+ - \frac{3}{2}\delta(1-z)\,, \tag{6.245}$$

$$(1+z^2)\left(\frac{\log(1-z)}{1-z}\right)_+ = \left(\frac{1+z^2}{1-z}\log(1-z)\right)_+ + \frac{7}{4}\delta(1-z)\,. \tag{6.246}$$

After having done this, one can finally let $1 + z^2 \to 2$ in the numerators. It is a matter of simple algebra to show that this procedure leads to the expected result:

$$\Gamma_{e^-}^{[1]}(z) \xrightarrow{z\to 1} K_0^{\text{NLL}}(z)\,. \tag{6.247}$$

In summary, we have thus proven that the solution of Eq. (6.103) embeds the initial conditions of Eqs. (6.3) and (6.6).

We conclude this appendix by reporting the results for the coefficients with $k > 0$. We have obtained what follows:

$$A_1^{\text{LL}} = \frac{3}{2}\,, \tag{6.248}$$

$$A_1^{\text{NLL}} = \frac{27}{8} + \frac{\pi^2}{6} - 2\zeta_3 - 4\pi b_0 - \frac{3\pi b_1}{b_0} - \frac{n_F}{18}(3 + 4\pi^2)$$
$$+ \left(\frac{9}{4} - \frac{2\pi^2}{3} - 3\pi b_0\right) L_0 \tag{6.249}$$

$$A_2^{\text{LL}} = \frac{9}{8} - \frac{\pi^2}{3}\,, \tag{6.250}$$

$$A_2^{\text{NLL}} = \frac{45}{16} + \left(4b_0^2 + 3b_1 - \frac{5}{12}\right)\pi^2 + \frac{2\pi^4}{45} - (11 - 10\pi b_0)\zeta_3 - \left(\frac{51}{8} + \frac{5\pi^2}{6}\right)\pi b_0$$
$$- \left(\frac{9}{2} - \frac{4\pi^2}{3}\right)\frac{\pi b_1}{b_0} - n_F\left(\frac{1}{4} - \frac{11\pi^2}{27} - \frac{\pi b_0}{6} - \frac{2\pi^3 b_0}{9}\right)$$
$$+ \left(\frac{27}{16} - \frac{3\pi^2}{2} - \frac{9\pi b_0}{2} + 3\pi^2 b_0^2 + \frac{4\pi^3 b_0}{3} + 8\zeta_3\right) L_0\,, \tag{6.251}$$

$$A_3^{\text{LL}} = \frac{9}{16} - \frac{\pi^2}{2} + \frac{8\zeta_3}{3}\,, \tag{6.252}$$

and:

$$B_{k,i}^{\text{LL}} = b_{\text{S},k,i}^{\text{LL}} = b_{\text{NS},k,i}^{\text{LL}}\,, \tag{6.253}$$

$$B_{k,i}^{\text{NLL}} = b_{\text{S},k,i}^{\text{NLL}} = b_{\text{NS},k,i}^{\text{NLL}}\,, \tag{6.254}$$

with $b^{\text{LL}}_{\text{S},k,i}$, $b^{\text{LL}}_{\text{NS},k,i}$, $b^{\text{NLL}}_{\text{S},k,i}$ and $b^{\text{NLL}}_{\text{NS},k,i}$ given in Appendix 6.B. We point out that Eqs. (6.253) and (6.254) hold for all values of k and i we have considered here. This is remarkable, because it tells one that with the expressions obtained in this chapter all of the $\log^p(1-z)/(1-z)$ terms in the PDF are the same regardless of whether one obtains them from the recursive solution, or by expanding the asymptotic solution. In general, one expects the logarithms from the latter to coincide with those of the former only for the larger values of p at any given k. The result obtained here ultimately stems from keeping some formally subleading contributions in the procedure of Sec. 6.4.2. In particular, it is important that the numerators in Eqs. (6.96) and (6.97) be $1 + x^2$ rather than 2 (which would be a perfectly fine choice in the asymptotic region). Actually, it turns out that the use of $1 + x^2$ is essential in the determination of the endpoint contributions in the plus distributions of Eqs. (6.96) and (6.97), which in turn induce (some of) the z-independent terms in Eqs. (6.98) and (6.99). Conversely, away from the endpoints the replacement of $1 + x^2$ with 2 leads to power-suppressed terms at $z \to 1$.

6.6.2 Photon

In the case of the photon one needs to employ Eq. (6.214). We start by observing that the Taylor series in t and α of such a quantity leads order by order to integrable singularities; as expected, there is therefore no endpoint contribution, and the expansion of the large-z solution can be expressed in terms of ordinary functions. Before turning to the explicit form of the latter, we point out that the t^0 term in the expansion of Eq. (6.214) is equal to $\Gamma_{\gamma,5}(z)$, since the contributions of the $\Gamma_{\gamma,j}(z)$ with $j \leq 4$ terms mutually cancel (that of $j = 1$ ($j = 3$) against that of $j = 2$ ($j = 4$)). One thus recovers the initial conditions of Eqs. (6.3) and (6.7), which is a first consistency check on Eq. (6.214). We now write the expansion of the large-z photon PDF in the same way as was done in Eq. (6.238), but with the K_k functions defined as follows:

$$K^{\text{LL}}_k(z) \;=\; (1 - \delta_{k0}) \sum_{i=0}^{i^{\text{LL}}_{\max}(k)} C^{\text{LL}}_{k,i}\, q_i(z)\,, \qquad k \geq 0\,, \tag{6.255}$$

$$K^{\text{NLL}}_k(z) \;=\; \sum_{i=0}^{i^{\text{NLL}}_{\max}(k)} C^{\text{NLL}}_{k,i}\, q_i(z)\,, \qquad k \geq 0\,. \tag{6.256}$$

having introduced the $q_i(z)$ functions in Eq. (6.70). It is a matter of algebra to arrive at the final results:

$$C^{\text{LL}}_{k,i} \;=\; c^{\text{LL}}_{\gamma,k,i}\,, \tag{6.257}$$

$$C^{\text{NLL}}_{k,i} \;=\; c^{\text{NLL}}_{\gamma,k,i}\,, \tag{6.258}$$

with $c^{\text{LL}}_{\gamma,k,i}$ and $c^{\text{NLL}}_{\gamma,k,i}$ given in Appendix 6.B. As was the case for their singlet and non-singlet counterparts (Eqs. (6.253) and (6.254)), Eqs. (6.257) and (6.258) have the property of holding for all of the k and i values considered here. Thus, the same remarks done previously are valid here as well — with the obvious exception that they apply to the $\log^p(1-z)$ terms rather than to the $\log^p(1-z)/(1-z)$ ones relevant to the singlet and non-singlet cases.

6.7 Additive matching for analytical solutions

The best analytical prediction is obtained by matching the recursive solution of Eq. (6.59), that is valid for all z values but in practice can be computed only up to a certain $\mathcal{O}(\alpha^n)$ (here, $n = 3$), with the solutions of Eqs. (6.103) (for singlet and non-singlet) and (6.214) (for photon), that retain all orders in α but are sensible only when $z \simeq 1$. In order to distinguish these two classes of solutions, we now denote them as follows:

$$\Gamma_{\text{rec}}(z) = \Gamma(z)\big[\text{Eq. (6.59)}\big], \tag{6.259}$$

$$\Gamma_{\text{asy}}(z) = \Gamma(z)\big[\text{Eq. (6.103)}\big] \qquad \text{non}-\text{singlet}, \tag{6.260}$$

$$\Gamma_{\text{asy}}(z) = \begin{pmatrix} \Gamma(z)\big[\text{Eq. (6.103)}\big] \\ \Gamma_\gamma(z)\big[\text{Eq. (6.214)}\big] \end{pmatrix} \qquad \text{singlet}-\text{photon}. \tag{6.261}$$

We shall henceforth consider the case of NLL solutions with running α, which constitutes our most accurate scenario. However, the procedure is unchanged in the case of NLL solutions with fixed α, or in the case of LL solutions. We remind the reader that Eq. (6.59) implicitly encompasses the non-singlet, singlet, and photon cases by means of the J_k^{LL} and J_k^{NLL} functions (see Sec. 6.3).

We define the matched PDFs with the additive formula[7]:

$$\Gamma_{\text{mtc}}(z) = \Gamma_{\text{rec}}(z) + \Big(\Gamma_{\text{asy}}(z) - \Gamma_{\text{subt}}(z)\Big) G(z), \tag{6.262}$$

where $G(z)$ is a largely arbitrary function that must obey the following condition

$$\lim_{z \to 1} G(z) = 1, \tag{6.263}$$

and that might optionally be used to power-suppress at small z the difference in round brackets in Eq. (6.262) – we shall give more details on this point later. The quantity $\Gamma_{\text{subt}}(z)$ (that we call "subtraction term") is responsible for removing the double counting, i.e. the contributions which are present both in the recursive and in the asymptotic solutions. We shall eventually construct it explicitly, but we anticipate the obvious fact that it must feature the dominant $z \to 1$ contributions to the PDF (which, in turn, are present in both the recursive and the asymptotic solutions, as is discussed in Sec. 6.6).

Before proceeding we stress that, although general, the arguments that follow from Eq. (6.262) are best understood if the PDFs are strongly peaked at $z \to 1$, which is indeed what happens for the singlet and non-singlet components, but not for the photon (at least to a certain extent). Thus, we shall first understand the two former cases, and return to the latter one only towards the end of this section.

We want the matched PDF to coincide with the asymptotic or the recursive solution for those z values appropriate for either of the latter two quantities. This is equivalent to requiring:

$$\Gamma_{\text{mtc}}(z) \sim \Gamma_{\text{asy}}(z) \qquad z \simeq 1, \tag{6.264}$$

$$\Gamma_{\text{mtc}}(z) \sim \Gamma_{\text{rec}}(z) \qquad z \text{ elsewhere}. \tag{6.265}$$

---

[7]Additive matching has been considered in Refs. [196, 204, 205]; Refs. [194, 195] adopt a multiplicative one.

Given Eq. (6.263), Eq. (6.264) is satisfied when:

$$\left|\Gamma_{\text{rec}}(z) - \Gamma_{\text{subt}}(z)\right| \ll \left|\Gamma_{\text{asy}}(z)\right|, \qquad z \simeq 1. \tag{6.266}$$

Conversely, there are two ways in which the behaviour in Eq. (6.265) can be achieved.

a) $G(z)$ can be expanded in series around $z = 0$, and is such that:

$$\lim_{z \to 0} G(z) = 0, \tag{6.267}$$

in addition to satisfying Eq. (6.263).

b) One has:

$$\left|\Gamma_{\text{asy}}(z) - \Gamma_{\text{subt}}(z)\right| \ll \left|\Gamma_{\text{rec}}(z)\right|, \tag{6.268}$$

for small and intermediate z values. When Eq. (6.268) holds, one can set:

$$G(z) \equiv 1. \tag{6.269}$$

The option of item *a)* stems from the observation that both $\Gamma_{\text{asy}}(z)$ and $\Gamma_{\text{subt}}(z)$ are only sensible when the $\log^p(1-z)$ terms they contain are large. When this is not the case, i.e. at small- and intermediate-z values, one can suppress them (in fact, one must, if Eq. (6.265) is to be fulfilled) by means of power-suppressed terms, here parametrised by $G(z)$, without affecting the formal accuracy of the matched PDF. However, this has the potential drawback of introducing in $\Gamma_{\text{mtc}}(z)$ a dependence on the arbitrary quantity $G(z)$, which must remain small in order not to lose predictive power. This issue is avoided if the option of item *b)* is viable. This has the drawback that it relies on the condition in Eq. (6.268), that might be problematic since it constrains $\Gamma_{\text{subt}}(z)$ in a z region which is not the natural domain of such a function.

Although there is significant freedom in the construction of the subtraction term, the recursive and asymptotic solutions provide us with two obvious candidates. Namely, we can set either

$$\Gamma_{\text{subt}}(z) \equiv \Gamma_{\text{subt}}^{\text{R}}(z) = \overline{\Gamma}(z)\big[\text{Eq. (6.76)}\big] \tag{6.270}$$

or

$$\Gamma_{\text{subt}}(z) \equiv \Gamma_{\text{subt}}^{\text{A}}(z) = \overline{\Gamma}(z)\big[\text{Eq. (6.238)}\big]. \tag{6.271}$$

In other words: with Eq. (6.270) we use all of the contributions to the recursive solution which are non-vanishing when $z \to 1$, while with Eq. (6.271) we employ the $\mathcal{O}(\alpha^3)$ expansion of the asymptotic solution. Therefore, as it follows from the discussion in appendix 6.6, $\Gamma_{\text{subt}}^{\text{A}}(z)$ essentially contains a subset of the terms present in $\Gamma_{\text{subt}}^{\text{R}}(z)$. More precisely:

$$\Gamma_{\text{subt}}^{\text{R}}(z) \quad \longleftrightarrow \quad \left\{\ell_i(z), q_i(z)\right\} \equiv \left\{\frac{\log^i(1-z)}{1-z}, \log^i(1-z)\right\}, \tag{6.272}$$

$$\Gamma_{\text{subt}}^{\text{A}}(z) \quad \longleftrightarrow \quad \left\{\ell_i(z)\right\} \equiv \left\{\frac{\log^i(1-z)}{1-z}\right\}. \tag{6.273}$$
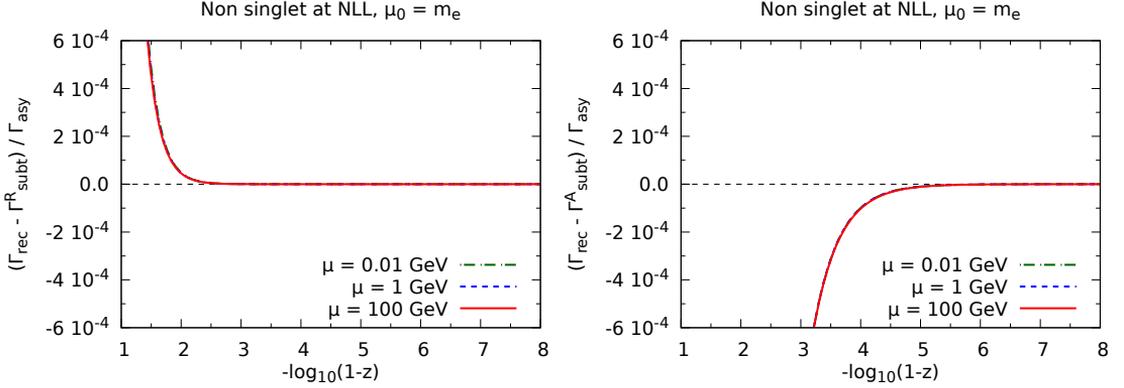
Figure 6.2: Ratio of the l.h.s of Eq. (6.266) over its r.h.s. (without the absolute values), for the two choices of the subtraction term (Eq. (6.270), left panel; Eq. (6.271), right panel), and for three different hard-scale values: $\mu = 0.01$ GeV (dot-dashed green), $\mu = 1$ GeV (dashed blue), and $\mu = 100$ GeV (solid red). As is indicated, the scale on the y axis of these plots is in units of 10^{-4}.

By construction (see Eq. (6.68)), we have

$$\lim_{z \to 1} \left(\Gamma_{\text{rec}}(z) - \Gamma_{\text{subt}}^{\text{R}}(z) \right) = 0 \,, \tag{6.274}$$

and therefore Eq. (6.266) automatically holds when the subtraction term is defined by means of the recursive solution. Conversely,

$$\Gamma_{\text{rec}}(z) - \Gamma_{\text{subt}}^{\text{A}}(z) \;\simeq\; \alpha \, q_2(z) \;\xrightarrow{z \to 1}\; \infty \,. \tag{6.275}$$

However, in spite of this, Eq. (6.266) holds also in this case, since:

$$\frac{\Gamma_{\text{rec}}(z) - \Gamma_{\text{subt}}^{\text{A}}(z)}{\Gamma_{\text{asy}}(z)} \;\simeq\; \frac{q_2(z) + \dots}{\ell_2(z) + \dots} \;\xrightarrow{z \to 1}\; 0 \,. \tag{6.276}$$

The conclusion is that Eq. (6.266) is satisfied with both choices of the subtraction term. The difference between adopting $\Gamma_{\text{subt}}^{\text{R}}(z)$ or $\Gamma_{\text{subt}}^{\text{A}}(z)$ is that with the former function the matched PDF will converge towards the asymptotic solution at z values relatively smaller than those relevant to the latter function. This can be seen in Fig. 6.2, where the ratio of the l.h.s. over the r.h.s. of Eq. (6.266) (without the absolute values) is plotted as a function of $-\log_{10}(1-z)$ for the two choices of the subtraction term considered here, and for three different hard scales μ. Note that the scale on the y axis of the plots in Fig. 6.2 is in units of 10^{-4}. The curves in Fig. 6.2 are relevant to the non-singlet component. We point out that their analogues for the singlet component are qualitatively and quantitatively very similar to those shown here. Apart from being in keeping with the expectations emerging from Eqs. (6.274)–(6.276), Fig. 6.2 shows that, even in the case of Eq. (6.270), the matched PDF will attain its asymptotic form only for values of z which are extremely close to one; in other words, non-logarithmic contributions are almost always important. This being the case, by choosing $\Gamma_{\text{subt}}^{\text{A}}(z)$ as a subtraction term
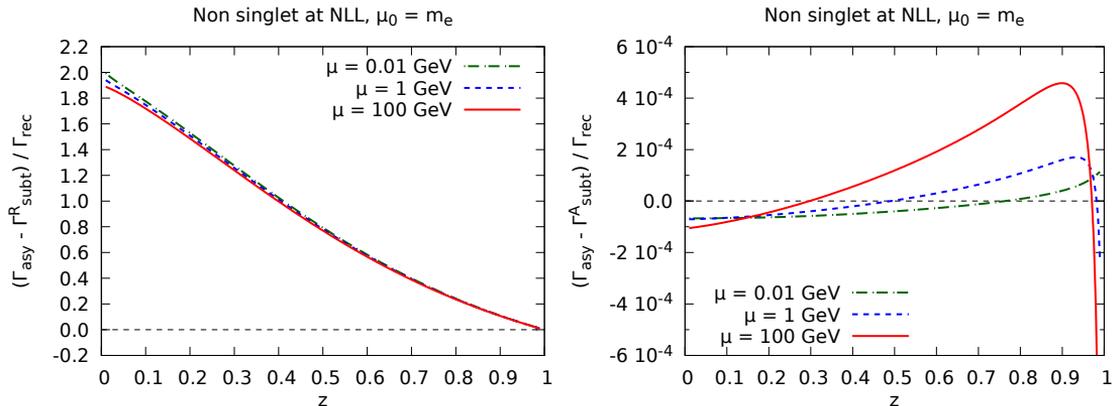
Figure 6.3: Same as in Fig. 6.2, for Eq. (6.268). As is indicated, the scale on the y axis of the plot on the right panel is in units of 10^{-4}.

rather than $\Gamma^{\mathrm{R}}_{\mathrm{subt}}(z)$ (which, as was anticipated, "delays" the onset of the asymptotic regime in the matched PDF) one renders the transition between the asymptotic and recursive solutions less abrupt; this turns out to be beneficial in order to reproduce the results of the numerical evolution.

As far as the small- and intermediate-z region is concerned, we observe that:

$$\Gamma_{\mathrm{asy}}(z) - \Gamma^{\mathrm{R}}_{\mathrm{subt}}(z) = \mathcal{O}(\alpha), \tag{6.277}$$

$$\Gamma_{\mathrm{asy}}(z) - \Gamma^{\mathrm{A}}_{\mathrm{subt}}(z) = \mathcal{O}(\alpha^4). \tag{6.278}$$

Equation (6.277) implies that it is unlikely that, if the subtraction term is defined by means of the recursive solution, one can avoid the use of the $G(z)$ function. Conversely, Eq. (6.278) implies that the definition by means of the asymptotic solution has a better chance of satisfying Eq. (6.268), thus bypassing the need to introduce $G(z)$. Note that the difference in Eq. (6.278) is of $\mathcal{O}(\alpha^4)$ as a direct consequence of the fact that we have computed $\Gamma^{\mathrm{A}}_{\mathrm{subt}}(z)$ to $\mathcal{O}(\alpha^3)$ (see Eq. (6.238)). In Fig. 6.3 we plot the ratio of the l.h.s. over the r.h.s. of Eq. (6.268) (without the absolute values), by using the same layout as in Fig. 6.2. In order to be definite, we have considered again the non-singlet component in Fig. 6.3, and have verified that the singlet one gives results which are extremely similar to those of the non-singlet. Figure 6.3 confirms our expectation based on Eqs. (6.277) and (6.278).

We now turn to discussing the case of the photon PDF, which is quite different from that of the singlet and the non-singlet. The starting point is the same as for the latter PDFs, namely the definition of the subtraction term with either Eq. (6.270) or (6.271), since those formulae are the general parametrisations of the perturbative expansion of the recursive or the asymptotic solutions, respectively, whose actual values are determined by the parameters specific to the particle which is being considered. Indeed, in the case of the photon, the analogues of
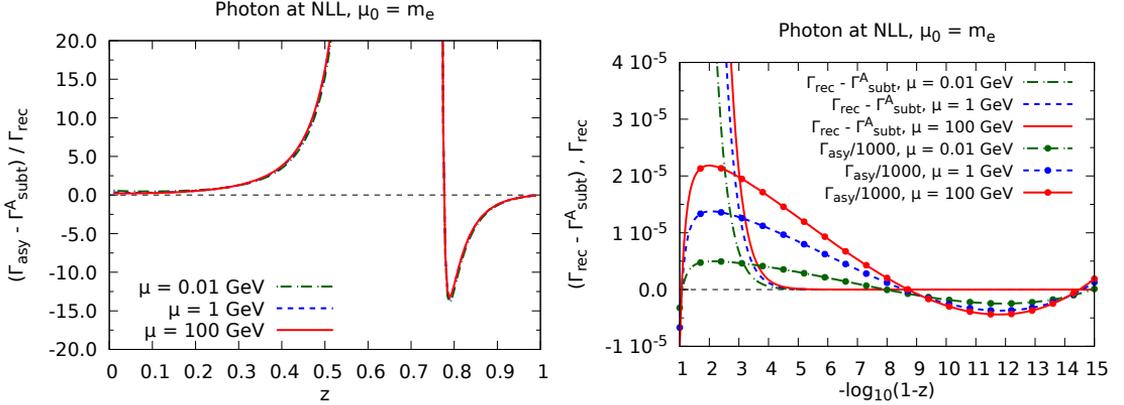
Figure 6.4: Plots assessing the validity of Eq. (6.268) (left panel) and Eq. (6.266) (right panel), in the case of the photon PDF. See the text for details.

Eqs. (6.272) and (6.273) are:

$$\Gamma^{\mathrm{R}}_{\mathrm{subt}}(z) \quad \longleftrightarrow \quad \Big\{ q_i(z) \Big\}, \tag{6.279}$$

$$\Gamma^{\mathrm{A}}_{\mathrm{subt}}(z) \quad \longleftrightarrow \quad \Big\{ q_i(z) \Big\}. \tag{6.280}$$

Actually, because of Eqs. (6.257) and (6.258), one can make a stronger statement, namely:

$$\Gamma^{\mathrm{R}}_{\mathrm{subt}}(z) = \Gamma^{\mathrm{A}}_{\mathrm{subt}}(z). \tag{6.281}$$

We stress that Eq. (6.281) is not a property inherent to the photon PDF, but a consequence of having been able to keep the relevant subleading terms in the computation of its large-z form as carried out in Sec. 6.5. In order to be definite, for consistency with the case of the singlet/non-singlet we shall label the subtraction term with "A" here.

The analogues of the right-hand side panels of Fig. 6.2 and of Fig. 6.3 are presented in the right and left panels of Fig. 6.4, respectively. We start from the right-hand side one, in order to assess the validity of Eq. (6.266). Unfortunately, it turns out that at large z's the NLL photon PDF becomes negative in a certain range, and it crosses twice the zero. For this reason, we cannot consider the ratio of the two sides of Eq. (6.266) as was done in Fig. 6.2, but only plot separately $\Gamma_{\mathrm{rec}}(z) - \Gamma_{\mathrm{subt}}(z)$ and $\Gamma_{\mathrm{asy}}(z)$; these two quantities are displayed in Fig. 6.4 by adopting identical patterns (each associated with a different hard scale μ), with the curves relevant to $\Gamma_{\mathrm{asy}}(z)$ overlaid by full circles. Furthermore, in order for the latter curves to fit into the layout, they have been multiplied by a constant factor equal to 10^{-3}. The plot clearly shows how Eq. (6.266) is safely fulfilled[8].

We now consider the left panel of Fig. 6.4, in order to assess the validity of Eq. (6.268); given that for all of the z values employed in the plot the photon PDFs is positive, we can compute the ratio of the two sides of Eq. (6.268) (without the absolute values) as was done previously in Fig. 6.3. It is immediate to see that the conclusions are the opposite of those valid in the

---

[8]Strictly speaking, no such conclusion is possible in an extremely narrow neighbourhood of the points at which the PDF crosses zero, where it is however not relevant, since all quantities of interest are vanishingly small there.
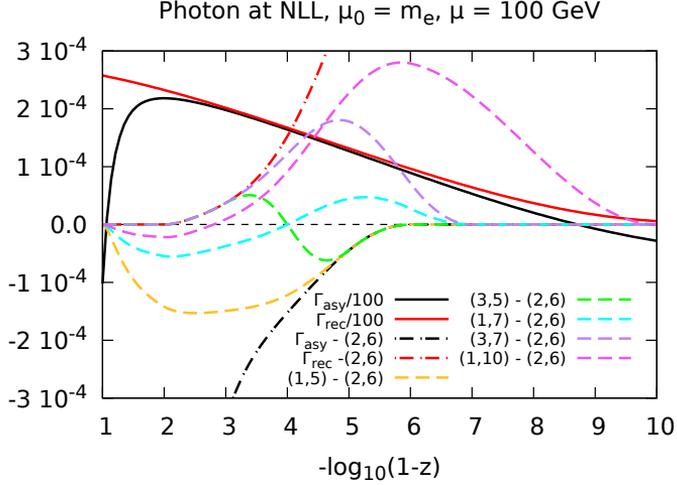
Figure 6.5: Study of the dependence of the matched photon PDF upon the parameters of the matching function $G(z)$, defined in Eq. (6.283). We have set $\mu = 100$ GeV.

cases of the singlet and non-singlet – namely, in a very large range in z the subtraction term and the asymptotic solution do *not* agree with each other[9]. Thus, in the case of the photon the use of a damping function $G(z)$ is unavoidable. In order to define it, it is useful to introduce the function:

$$\hat{z}(z) = -\log_{10}(1 - z) \,, \tag{6.282}$$

by means of which we set:

$$G(z) = \begin{cases} 1 & \hat{z}_1 \leq \hat{z}(z) \,, \\ G_p\left(\frac{\hat{z}(z) - \hat{z}_0}{\hat{z}_1 - \hat{z}_0}\right) & \hat{z}_0 \leq \hat{z}(z) < \hat{z}_1 \,, \\ 0 & \hat{z}(z) < \hat{z}_0 \,, \end{cases} \qquad G_p(v) = \frac{v^p}{v^p + (1 - v)^p} \,. \tag{6.283}$$

This is a smooth function that obeys Eqs. (6.263) and (6.267), and where $\hat{z}_0$, $\hat{z}_1$, and p are free parameters. The physical meaning of the parameters $\hat{z}_0$ and $\hat{z}_1$ is that, for z such that $\hat{z}(z) < \hat{z}_0$ ($\hat{z}(z) > \hat{z}_1$), the matched PDF coincides with the recursive (the asymptotic) solution. As a matter of fact, Eqs. (6.282) and (6.283) stem from the observation that it is $\hat{z}(z)$, and not z, the natural variable to carry out the matching, and this is because the large-z behaviour of the PDFs is achieved when logarithmic terms grow much larger than non-logarithmic ones.

In principle, the parameters $\hat{z}_0$, $\hat{z}_1$, and p are unconstrained. In order to choose them sensibly, we plot in Fig. 6.5 the asymptotic and recursive solutions as solid black and red curves, respectively (both are multiplied by a factor of 10^{-2}, for reasons that will soon become clear). For the matching to work reasonably well, the transition between the recursive and the asymptotic solutions must occur in a region where these two predictions are as close as possible to each other (which we interpret as the signal that both give a reasonable description of the "true" photon PDF). From Fig. 6.5, we gather that such a region is $2 \lesssim \hat{z} \lesssim 6$; this suggests to set $\hat{z}_0 = 2$ and $\hat{z}_1 = 6$. However, it is clear that there is (and there must be) a certain flexibility in

---

[9]See the comments after Eq. (6.212), for a discussion on the origin of this behaviour.

these choices. The agreement between the asymptotic and recursive solutions quickly worsens in the range $z \in (1, 2)$, which implies that $\hat{z}_0 = 1$ must be considered as an extreme choice. On the other hand, for $\hat{z} > 6$ the asymptotic and recursive solutions do tend to stay relatively close to each other, to the extent that even $\hat{z}_1 = 10$ appears to be an acceptable choice. As far as p is concerned, the larger this parameter the more abrupt is the transition between the two regimes. We have therefore opted to employ $p = 2$, which essentially corresponds to the slowest transition compatible with the derivatives of $G(z)$ being continuous. In order to assess the impact of the choices of $\hat{z}_0$ and $\hat{z}_1$ on the matched PDF, we have computed the latter for several values of these parameters. In Fig. 6.5 we display as dashed curves the differences between any of the matched predictions (relevant to $(\hat{z}_0, \hat{z}_1) = (1, 5)$, $(3, 5)$, $(1, 7)$, $(3, 7)$, and $(1, 10)$) minus the one obtained with $(\hat{z}_0, \hat{z}_1) = (2, 6)$. For comparison, we also show the differences between the asymptotic and recursive solutions minus the $(\hat{z}_0, \hat{z}_1) = (2, 6)$ matched PDF as black and red dot-dashed curves, respectively. We see that the differences between any two pairs of matched predictions are roughly in the range $(-2, 3) \cdot 10^{-4}$, i.e. at least a factor 25 smaller than the recursive and the asymptotic solutions. While this statement progressively loses validity when moving towards $\hat{z} = 8$ (where the asymptotic solution, which is the appropriate one in this region, crosses zero), it also becomes less relevant, since indeed all quantities of interest tend to become negligible in absolute value. Having said that, it is important to bear in mind that the dependence on the matching-function parameters is a genuine uncertainty that affects the matched predictions; plots such as those in Fig. 6.5 help assess its size, and should be re-produced whenever new conditions become relevant (specifically, for hard-scale values significantly different w.r.t. those considered in Fig. 6.5). We finally point out that we have repeated the exercise by using $p = 3$ and $p = 4$; the overall uncertainties are similar to those obtained with $p = 2$.

In summary, our best analytical results are obtained with the matching formula of Eq. (6.262). In the case of the singlet and the non-singlet, we employ Eq. (6.271) for the definition of the subtraction term, and a constant $G(z)$ function as in Eq. (6.269). In the case of the photon, the definition of the subtraction term is still given by Eq. (6.271) – however, there are more limited possibilities here, owing to Eq. (6.281). The matched photon PDF does need a non-trivial matching function: we adopt that of Eq. (6.283), with $\hat{z}_0 = 2$, $\hat{z}_1 = 6$, and $p = 2$.

6.8   Numerical solutions for the PDFs

The numerical evolution for the PDFs is achieved by first solving the evolution equation for the evolution operator in Mellin space. As is discussed in Sec. 1.6, the introduction of the singlet and non-singlet combinations, Eq. (1.74), allows for a decoupling of the evolution equations that is well-suited for a numerical implementation. As has been done thus far, in the following we shall implicitly refer to the two-dimensional singlet-photon case, keeping in mind that the non-singlet case is obtained by considering a one-dimensional flavour space.

The numerical solution of Eq. (6.18) for the evolution operator $\mathbb{E}_N$ is obtained by means of a discretised path-ordered product [206]. The evolution range $[0, t]$ is partitioned into n intervals $[t_i, t_{i+1}]$, with $t_0 = 0$ and $t_n = t$, and the evolution operator is written as follows:

$$\mathbb{E}_N(t) = \prod_{i=n-1}^{0} \mathbb{E}_N(t_{i+1}, t_i). \tag{6.284}$$

The product on the r.h.s. of Eq. (6.284) must be understood as a product among matrices. Thus,

by reading the product from left to right one finds decreasing values of the index i. The Magnus expansion of Sec. 6.5 provides a formal solution to the evolution equation, see Eq. (6.148). If the interval $\Delta t_i = t_{i+1} - t_i$ is small enough, we are allowed to discard the $\Omega_{k,N}(t)$ terms with $k > 1$ in the argument of the exponent:

$$\mathbb{E}_N(t_{i+1}, t_i) = \exp\left(\int_{t_i}^{t_{i+1}} \mathbb{M}_N(\alpha(\bar{t}))\, d\bar{t}\right) \mathbb{E}_N(t_i, t_{i-1}).$$
(6.285)

In the same limit, an approximation of the integral appearing in Eq. (6.285) by means of the trapezoidal rule is also justified:

$$\int_{t_i}^{t_{i+1}} \mathbb{M}_N(\alpha(\bar{t}))\, d\bar{t} \simeq \frac{\mathbb{M}_N(\alpha(t_{i+1})) + \mathbb{M}_N(\alpha(t_i))}{2} \Delta t_i.$$
(6.286)

By iteratively replacing Eq. (6.285) in Eq. (6.284), we eventually obtain for the evolution operator:

$$\mathbb{E}_N(t) = \prod_{i=n-1}^{0} \exp\left(\frac{\mathbb{M}_N(\alpha(t_{i+1}) + \mathbb{M}_N(t_i)}{2} \Delta t_i\right).$$
(6.287)

$\mathbb{M}_N$ depends both on the accuracy and on the subtraction scheme: its explicit expression, in $\overline{\mathrm{MS}}$ and at NLO, is implicitly given in Eq. (6.18). The exponential in Eq. (6.287) can be evaluated by means of a diagonalisation procedure. However, in the case of 2×2 matrices, it is possible to write down a closed expression for the exponential, well suited for a numerical implementation:

$$\exp\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{e^{(a+d)/2}}{D} \begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix},$$
(6.288)

with:

$$D = \sqrt{(a-d)^2 + 4bc},$$
(6.289)

$$e_{11} = D \cosh\left(\frac{D}{2}\right) + (a-d) \sinh\left(\frac{D}{2}\right),$$
(6.290)

$$e_{12} = 2b \sinh\left(\frac{D}{2}\right),$$
(6.291)

$$e_{21} = 2c \sinh\left(\frac{D}{2}\right),$$
(6.292)

$$e_{22} = D \cosh\left(\frac{D}{2}\right) + (d-a) \sinh\left(\frac{D}{2}\right).$$
(6.293)

Finally, in order to invert the PDFs from the Mellin to the z space, we employ a numerical algorithm based on the so-called Talbot path. Details on the implementation of this method can be found in Ref. [207].

The computer program that implements what has been described thus far was used to obtain all of the numerical results presented in this chapter. The code can be downloaded at:

$$\texttt{https://github.com/gstagnit/ePDF}$$

In the non-singlet case one can devise an alternative procedure. Namely, one can exploit the analytical N-space solution for the evolution operator, given in Eq. (6.20), multiply it by
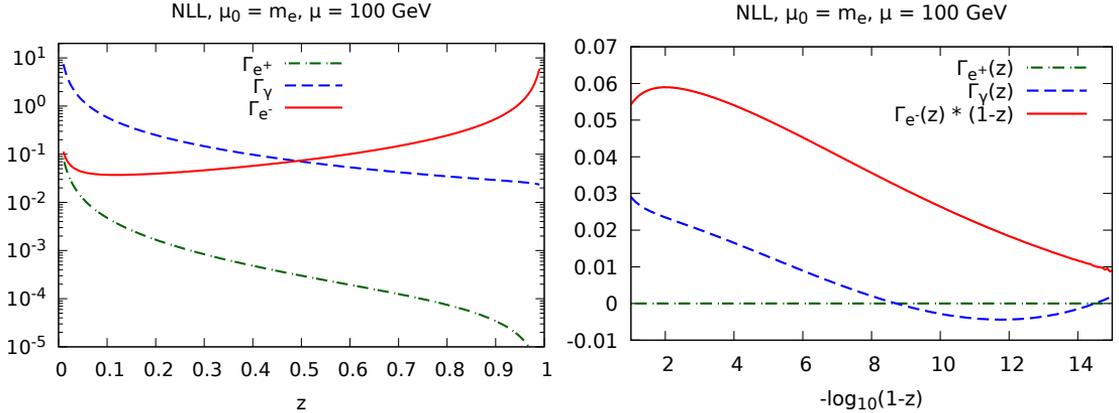
Figure 6.6: Electron (solid red), photon (dashed blue), and positron PDFs (dot-dashed green) PDFs at $\mu = 100$ GeV. The electron PDF is multiplied by a factor $(1 - z)$ in the plot on the right panel.

the Mellin-transformed initial conditions, and then invert the result thus obtained back to the z space by means of a numerical contour integration. We have implemented this strategy in a computer program[10] fully independent from the one described above, and verified that the two are in perfect agreement.

6.9 Results

In this section we present our predictions for the PDFs, by computing them both with the numerical code described in Sec. 6.8, and by evaluating the analytical formulae; these are always the matched ones. We compare these two classes of predictions, mutually validating them in the process. Unless explicitly indicated, all results are NLL-accurate with running α, and all have been obtained by setting $\mu_0 = m$.

We begin by plotting in Fig. 6.6 the electron, photon, and positron PDFs, computed at $\mu = 100$ GeV with the numerical code. The left panel shows these quantities in the full $z \in (0, 1)$ range, while the right panel is a zoom to the large-z region, where we consider $\hat{z} \in (1, 15)$ (see Eq. (6.282) for the definition of $\hat{z}$). Owing to the much faster growth of the electron PDFs in this region w.r.t. that of the other two partons, we have multiplied this PDF by a factor equal to $(1 - z)$, in order for all of the three curves to fit into the same layout. Figure 6.6 renders it manifest that the production of heavy (relative to the collider c.m. energy) objects is dominated by the partonic lepton whose charge is the same as that of the particle lepton it stems from[11] (in Eq. (6.1), one has the implicit constraint $z_+ z_- \geq M^2/S$, with M and $\sqrt{S}$ the mass of the object produced and the collider c.m. energy, respectively). Note that from the right panel of Fig. 6.6, given that the solid-red and dashed-blue curves are roughly of the same order, and that the former includes the $(1 - z)$ factor, one can immediately see that the photon PDF is smaller than the electron PDF by a number of orders of magnitude equal to the value of $\hat{z}$. Conversely,

---

[10]This builds upon the code originally written by the authors of Ref. [198].

[11]The reader must bear in mind that all our results are obtained by assuming an electron particle. In the case of a positron particle, the roles of the electron and positron partons are simply reversed.

by producing lighter objects and/or by increasing the collider energy, the contribution(s) of the incoming photon(s) become(s) important.

In view of the smallness of the positron PDF as is documented in Fig. 6.6, it is more convenient to present our findings in terms of the singlet and the non-singlet PDFs rather than by means of the electron and positron ones. This is what we shall do in the remainder of this section.

In order to establish the level of agreement between our numerical and analytical predictions, we plot in figures 6.7, 6.8, and 6.9 the ratios of the latter over the former, minus one, in the cases of the non-singlet, singlet, and photon, respectively. In each plot, there are three curves, corresponding to three different choices of hard scales: $\mu = 0.01$ GeV (dot-dashed green curves), $\mu = 1$ GeV (dashed blue curves), and $\mu = 100$ GeV (solid red curves). An overarching observation is that, in all of the cases bar for the photon at large z's (an exception to which we shall return later), the $\mu = 100$ GeV results are those which display the largest analytical-numerical disagreements. However, even in this worst-case scenario, the level of agreement is excellent, being typically of the order of 10^{-5}–10^{-4} (relative); the largest disagreements are to be found at small z's in the case of the singlet (because of the presence of un-resummed $\log z$ terms[12]). In keeping with the previous remark relevant to the hard-scale dependence, the case of the photon at $z \simeq 1$ constitutes an exception: from the right panel of Fig. 6.9 we see that the analytical and numerical predictions agree at the level of 10^{-3} (10^{-2}) at $\mu = 100$ GeV ($\mu = 0.01$ GeV) for $2 \lesssim \hat{z} \lesssim 6$; furthermore, the behaviour at $\hat{z} > 6$ might seem to suggest that the $z \to 1$ limits of the analytical and numerical computations are different. We shall show in the following (see Fig. 6.13) that this is in fact *not* the case. For the time being, the crucial thing to bear in mind is that, in the z region we are discussing, the photon PDF is very small in absolute value and, more importantly, smaller than the electron PDF by several orders of magnitude. Thus, even a relatively large discrepancy of 0.1–1% between the numerical and analytical photon PDFs will be quite irrelevant. The general conclusion, which applies to all partons, is therefore that the analytical formulae appear to be perfectly adequate, and can be employed in calculations of cross sections for phenomenological purposes.

We now turn to assessing the effects on the PDFs of the NLL corrections, by comparing the NLL results with their LL counterparts. While this will fully account for the predictions obtained here for the first time, it is important to bear in mind that the PDFs are unphysical quantities, and that beyond LL cancellations do occur (in particular, in $\overline{\text{MS}}$) between them and the short-distance cross sections. Thus, an increase or decrease by a factor X of an NLL PDF w.r.t. an LL one will most definitely not translate into a corresponding increase or decrease of the NLO physical cross section w.r.t. its LO counterpart. In the main frames of Figs. 6.10, 6.11, and 6.12, we plot the ratios of the NLL PDFs over the LL ones, both computed with the numerical code. As was done previously, all figures feature three curves, that correspond to different choices of hard scales; the same scale values, and the same graphical patterns, are used here as in Figs. 6.7–6.9. All the figures have an inset, which displays the double ratio (minus one):

$$\frac{\text{PDF}_{\text{NLL}}}{\text{PDF}_{\text{LL}}}\bigg|_{\text{an}} \bigg/ \frac{\text{PDF}_{\text{NLL}}}{\text{PDF}_{\text{LL}}}\bigg|_{\text{num}} - 1 \,. \tag{6.294}$$

The agreement between the numerical and analytical predictions is again extremely good, especially at large z's; once more, the photon in this region is the (relative) exception to that

---

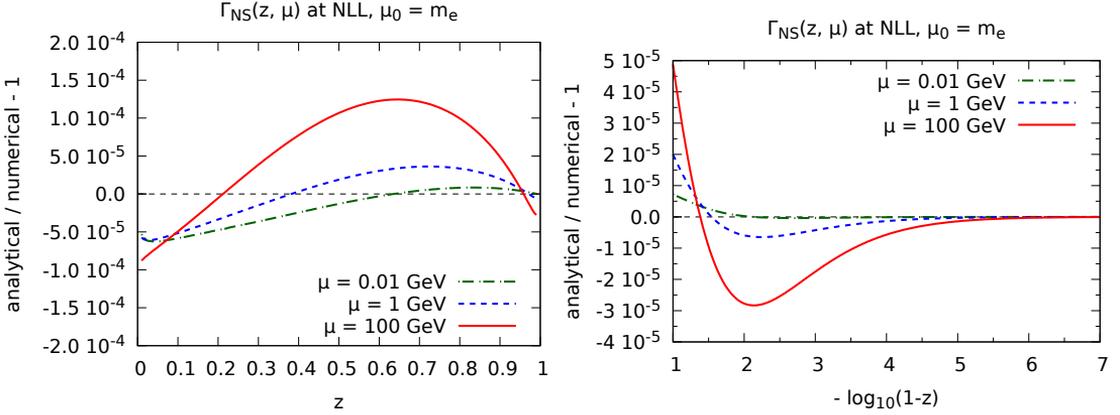[12]Techniques to resum such logarithms exist, see e.g. Refs. [208, 209].

Figure 6.7:   Comparison between the numerical and analytical predictions for the non-singlet, for three different hard-scale choices.
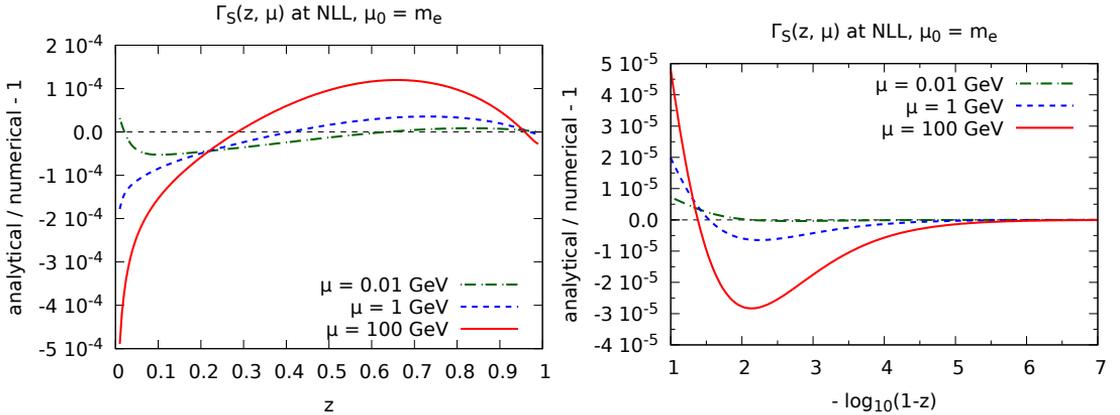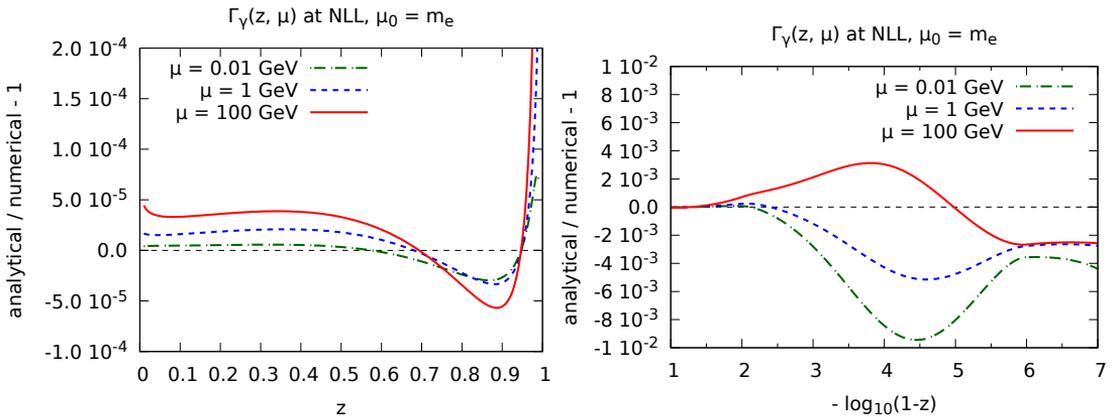

Figure 6.8:   As in Fig. 6.7, for the singlet.

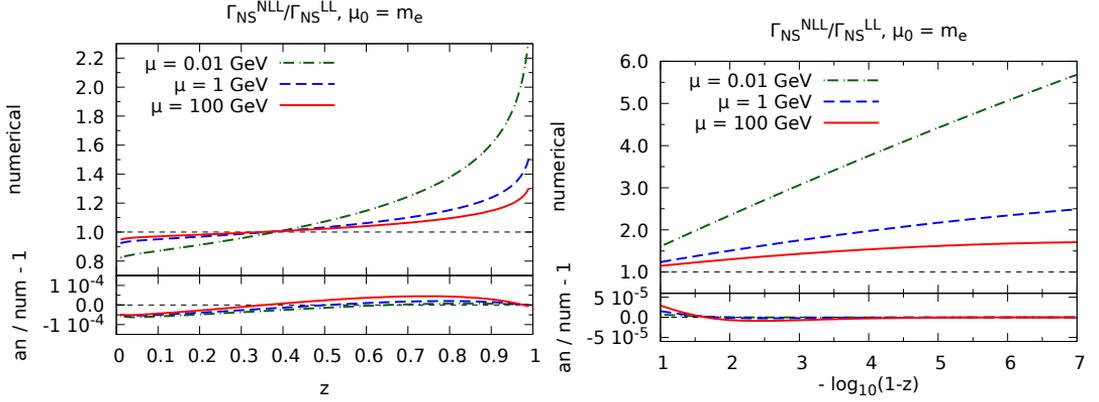**Figure 6.9:**   As in Fig. 6.7, for the photon.

Figure 6.10: Main frames: ratios of NLL over LL PDF, as computed with the numerical code, for the non-singlet, and for three different hard-scale choices. Insets: ratio of the ratio shown in the main frame, over the same quantity computed analytically, minus one.
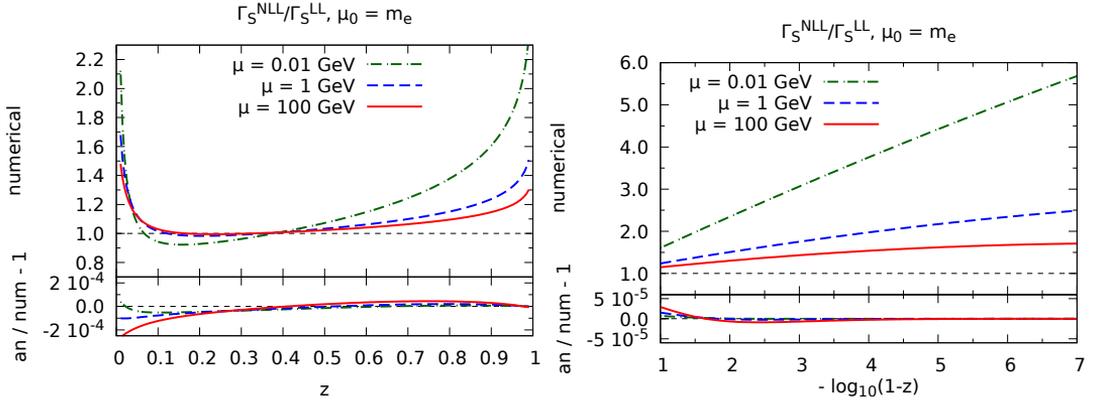

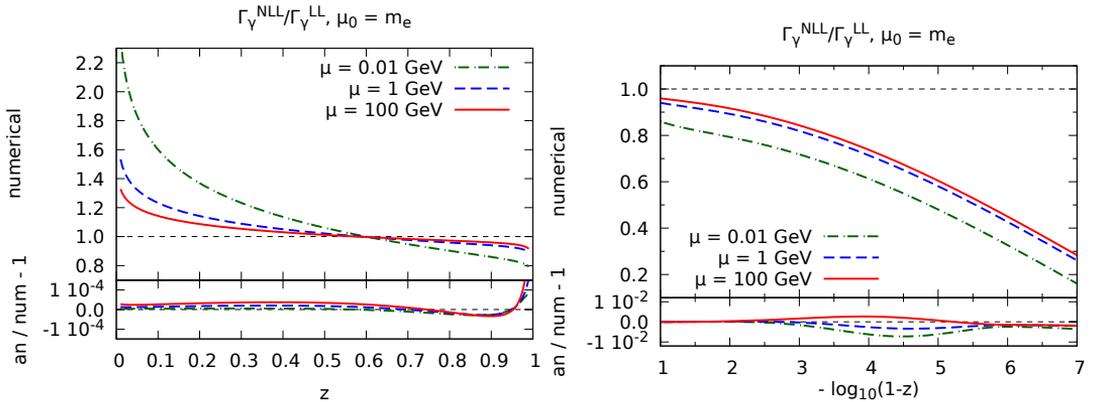**Figure 6.11:** As in Fig. 6.10, for the singlet.

Figure 6.12: As in Fig. 6.10, for the photon.

general rule, on which we shall comment later. The agreement becomes marginally worse with increasing μ, but this effect is less evident w.r.t. that in the case of the absolute predictions of Figs. 6.7–6.9. Interestingly, the size of the NLL effects *decreases* with the hard scale. This is particularly easy to understand in the large-z region in the case of the singlet (or non-singlet), since it can be directly read from Eq. (6.103). As was already remarked there, this behaviour is driven by Eq. (6.104), which implies that: *a)* the coefficient of the $\log(1-z)$ term is much larger than that of the $\log^2(1-z)$ term up to extremely large values of z; *b)* such a coefficient, being proportional to $1/\alpha(\mu)$, decreases with μ. These two effects can clearly be seen in the main frames of the right panels of Figs. 6.10 and 6.11, where the various lines are almost straight ones, but relatively less so at larger values of the hard scales. Keeping in mind the general observation made above on the unphysical nature of the PDFs, we point out that the natural applications of the quantities computed here involve scales that are large.
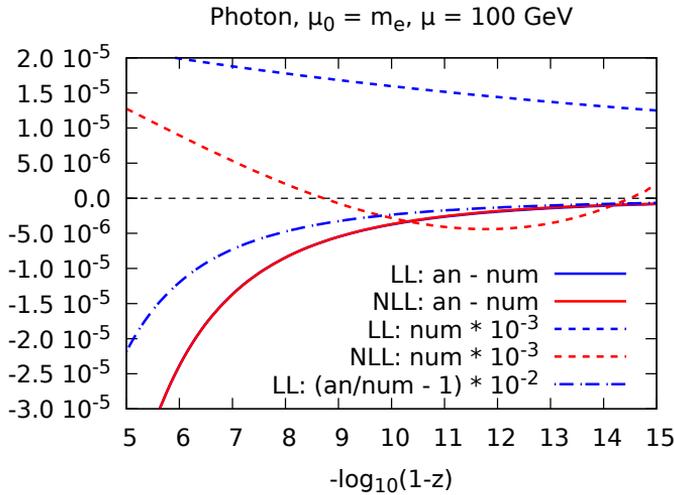

Figure 6.13:   Behaviour of the photon PDF at very large z values, where the analytical and numerical predictions are compared. We have set $\mu = 100$ GeV.

We now go back to commenting on the large-z behaviour of the photon PDF. We have already remarked in Fig. 6.9 that such a PDF in this region is close to zero in absolute value, and orders of magnitude smaller than its electron counterpart. On top of that, for the specific issue of the NLL vs LL results, the comparison between Eqs. (6.214) and (6.228) (or between their over-simplified forms of Eqs. (6.227) and (6.231)) shows that, at variance with the case of the electron (Eqs. (6.103) and (6.106)), the NLL asymptotic photon PDF does not factorise the functional form relevant to its LL version. Hence, larger differences in the matching region have to be expected between the NLL and LL photon PDF, which are larger than those for the electron.

At the right end of the z range in Fig. 6.12 we see again the kind of pattern as in the same region of Fig. 6.9, which might cast doubts on the agreement between the $z \to 1$ limits of the analytical and numerical predictions. In order to address this concern, in Fig. 6.13 we plot the photon PDF in a much more extended z range w.r.t. what was done so far. The blue and red solid curves are the differences between the analytical and numerical results computed at the

LL and NLL, respectively; the dashed curves of the same colours are the corresponding PDFs, multiplied by an overall constant factor equal to 10^{-3}; finally, the blue dot-dashed curve is the rescaled ratio of the analytical over the numerical LL results, minus one, which can be sensibly computed owing to the fact that the LL PDF does not vanish for values of $z \neq 1$. Apart from the similarity between the LL and NLL differences, the key message of Fig. 6.13 is that at $z \to 1$ the analytical and numerical predictions do tend to the same value, but in a much slower way w.r.t. the case of the singlet/non-singlet. In other words, the onset of the true asymptotic regime occurs at much larger z values for the photon than for the singlet or non-singlet. This needs not be surprising, owing to the mechanism that governs the asymptotic photon behaviour, as is documented in Sec. 6.5. An improvement of the analytical large-z PDF computed here would require keeping all terms suppressed by powers of N^{-2} in Mellin space, an extremely involved computation which is not justified in view of the smallness of the photon PDF in this region.

6.A   Numerical integrals in the recursive solutions

In the course of the recursive procedure, we have found that some integrals relevant to $\mathcal{J}_2^{\mathrm{NLL}}$ (i.e. the function associated with the $\mathcal{O}(\alpha t^2)$ term in the representation of the PDFs) are not easily computable analytically. We have therefore opted to limit ourselves to obtaining their $z \to 1$ leading terms analytically, while evaluating all of the remaining terms numerically, so that the latter contribute only to $\hat{J}_2^{\mathrm{NLL}}$ (we point out that an analogous strategy has already been adopted in Ref. [196]). More precisely, let us consider the generic modified-convolution integral of Eq. (6.27). We distinguish two possibilities: either $g(x)$ is a plus distribution, or it is an ordinary function. Notation-wise, these two cases are written as follows:

$$\text{plus distribution}: \qquad g(x) = \Big[\hat{g}(x)\Big]_+ \, , \tag{6.295}$$

$$\text{ordinary function}: \qquad g(x) \, . \tag{6.296}$$

In the case of Eq. (6.295), we have:

$$\Big[\hat{g}\Big]_+ \overline{\otimes}_z h = \Big[\hat{g}\Big]_+ \overline{\otimes}_z h\Big|_{\mathrm{end}} + \Big[\hat{g}\Big]_+ \overline{\otimes}_z h\Big|_{\mathrm{bulk}} \, , \tag{6.297}$$

where we have defined the endpoint and bulk contributions, respectively, as follows:

$$\Big[\hat{g}\Big]_+ \overline{\otimes}_z h\Big|_{\mathrm{end}} \;\; = \;\; -h(z) \int_0^z dx\, \hat{g}(x) \, , \tag{6.298}$$

$$
\begin{aligned}
\Big[\hat{g}\Big]_+ \overline{\otimes}_z h\Big|_{\mathrm{bulk}} \;\; &= \;\; \int_z^1 dx\, \hat{g}(x) \left[h\left(\frac{z}{x}\right) - h(z) \right] \\
&= \;\; \int_0^1 dy\, (1-z)\, \hat{g}\Big(1 - (1-z)y\Big) \left[h\left(\frac{z}{1-(1-z)y}\right) - h(z) \right] \, .
\end{aligned}
\tag{6.299}
$$

These equations can also be used in the simpler case of Eq. (6.296): one simply sets the endpoint contribution equal to zero, and computes Eq. (6.299) by removing the subtraction term $h(z)$ and with the formal replacement $\hat{g} \to g$ there.

The endpoint contribution of Eq. (6.298) is always computed analytically, and its results

are included in $\bar{J}_2^{\text{NLL}}(z)$ and/or $\hat{J}_k^{\text{NLL}}(z)$, according to the behaviour of $h(z)$ at $z \to 1$. As far as Eq. (6.299) is concerned, for the sake of the forthcoming discussion let us re-write it more compactly as follows:

$$F(z) = \int_0^1 dy\, f(y, z)\,. \tag{6.300}$$

If the integral in Eq. (6.300) were strongly convergent, then we might obtain its contribution to the PDF (see Eq. (6.26)) by means of a derivation under the integral sign, namely:

$$-\frac{\partial F(z)}{\partial z} = -\int_0^1 dy\, \frac{\partial f(y, z)}{\partial z}\,. \tag{6.301}$$

Unfortunately, the strong convergence of the integral is not guaranteed, given that $F(z)$ in general is logarithmically divergent at $z \to 1$. However, the contributions that are non vanishing at $z = 1$ are also easy to compute analytically; such computation can be carried out directly at the differential level of Eq. (6.301), and stems from expanding the integrand on the r.h.s. of that equation in a series of z around 1. The latter must include all terms that result in either a logarithmically-divergent or a constant non-null term at $z \to 1$, which typically implies up to $(1-z)^0$ contributions. In this way we arrive at the following identity:

$$-\frac{\partial F(z)}{\partial z} = -\left[\frac{\partial F(z)}{\partial z} - \frac{\partial F(z)}{\partial z}\bigg|_{\text{asy}}\right] - \frac{\partial F(z)}{\partial z}\bigg|_{\text{asy}}\,, \tag{6.302}$$

with:

$$\frac{\partial F(z)}{\partial z}\bigg|_{\text{asy}} = \int_0^1 dy\, \frac{\partial f(y, z)}{\partial z}\bigg|_{\text{exp}}\,, \tag{6.303}$$

having denoted by $\partial f/\partial z|_{\text{exp}}$ the aforementioned series expansion. The integral in Eq. (6.303) is computed analytically, and its result added to $\bar{J}_2^{\text{NLL}}$ (thus, given Eq. (6.72), it contributes to $c_{2,i}^{\text{NLL}}$ for some i, depending on $h(z)$; there are no contributions to $b_{2,i}^{\text{NLL}}$):

$$-\frac{\partial F(z)}{\partial z}\bigg|_{\text{asy}} \longrightarrow \bar{J}_2^{\text{NLL}}(z)\,. \tag{6.304}$$

Conversely, the quantity in square brackets in Eq. (6.302), where the rightmost term is regarded as a regularising factor, is computed numerically[13], and eventually included in $\hat{J}_2^{\text{NLL}}$:

$$-\left[\frac{\partial F(z)}{\partial z} - \frac{\partial F(z)}{\partial z}\bigg|_{\text{asy}}\right] \equiv -\int_0^1 dy\, \left(\frac{\partial f(y, z)}{\partial z} - \frac{\partial f(y, z)}{\partial z}\bigg|_{\text{exp}}\right) \longrightarrow \hat{J}_2^{\text{NLL}}(z)\,. \tag{6.305}$$

We list here the pairs $\hat{g}$ (or g) and h that we handle in the way we have just described:

$$\hat{g}_a(v) = \frac{1 + v^2}{1 - v}\,, \qquad\qquad h_a(v) = \log^2(1 - v)\log v\,, \tag{6.306}$$

$$\hat{g}_b(v) = \frac{1 + v^2}{1 - v}\,, \qquad\qquad h_b(v) = \log(1 - v)\,\text{Li}_2(v)\,, \tag{6.307}$$

---

[13]These are one-dimensional integrations of regularised integrals: the routine `gsl_integration_qag` of the GSL library is employed, which guarantees a fast and reliable convergence.

$$\hat{g}_c(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_c(v) = \log^2(v)\,\log(1+v)\,, \tag{6.308}$$

$$\hat{g}_d(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_d(v) = \log(v)\,\log^2(1+v)\,, \tag{6.309}$$

$$\hat{g}_e(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_e(v) = \log(v)\,\mathrm{Li}_2(-v)\,, \tag{6.310}$$

$$\hat{g}_f(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_f(v) = \log(1+v)\,\mathrm{Li}_2(-v)\,, \tag{6.311}$$

$$\hat{g}_g(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_g(v) = \log(1+v)\,\mathrm{Li}_2\left(\frac{1}{1+v}\right)\,, \tag{6.312}$$

$$\hat{g}_h(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_h(v) = \mathrm{Li}_3(1-v)\,, \tag{6.313}$$

$$\hat{g}_i(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_i(v) = \mathrm{Li}_3(-v)\,, \tag{6.314}$$

$$\hat{g}_j(v) = \frac{1+v^2}{1-v}\,, \qquad\qquad h_j(v) = \mathrm{Li}_3\left(\frac{1}{1+v}\right)\,, \tag{6.315}$$

$$\hat{g}_k(v) = \frac{1+v^2}{1-v}\,\log(1-v)\,\log v\,, \qquad\qquad h_k(v) = \log(1-v)\,, \tag{6.316}$$

$$g_l(v) = \frac{1+v^2}{1+v}\,\log^2 v\,, \qquad\qquad h_l(v) = \log(1-v)\,, \tag{6.317}$$

$$g_m(v) = \frac{1+v^2}{1+v}\,\log v\,\log(1+v)\,, \qquad\qquad h_m(v) = \log(1-v)\,, \tag{6.318}$$

$$g_n(v) = \frac{1+v^2}{1+v}\,\mathrm{Li}_2(-v)\,, \qquad\qquad h_n(v) = \log(1-v)\,, \tag{6.319}$$

$$g_o(v) = \frac{1}{v}\,, \qquad\qquad h_o(v) = \log v\,\log^2(1+v)\,, \tag{6.320}$$

$$g_p(v) = \frac{1}{v}\,, \qquad\qquad h_p(v) = \log(1+v)\,\mathrm{Li}_2(-v)\,, \tag{6.321}$$

$$g_q(v) = \frac{1}{v}\,, \qquad\qquad h_q(v) = \log(1+v)\,\mathrm{Li}_2\left(\frac{1}{1+v}\right)\,, \tag{6.322}$$

$$g_r(v) = 1\,, \qquad\qquad h_r(v) = \mathrm{Li}_3(1-v)\,, \tag{6.323}$$

$$g_s(v) = \frac{1}{v}\,, \qquad\qquad h_s(v) = \mathrm{Li}_3(1-v)\,, \tag{6.324}$$

$$g_t(v) = 1\,, \qquad\qquad h_t(v) = \mathrm{Li}_3\left(\frac{1}{1+v}\right)\,, \tag{6.325}$$

$$g_u(v) = \frac{1}{v}\,, \qquad\qquad h_u(v) = \mathrm{Li}_3\left(\frac{1}{1+v}\right)\,, \tag{6.326}$$

$$g_v(v) = \frac{1}{v}\,\log^2(1-v)\,, \qquad\qquad h_v(v) = \log(1-v)\,. \tag{6.327}$$

We stress again that each of these pairs will contribute to both Eq. (6.304) and (6.305). We denote generically either of these contributions as follows:

$$J_\rho^{\mathrm{num}}(z) \;\longleftrightarrow\; \left(\hat{g}_\rho, h_\rho\right) \quad \text{or} \quad \left(g_\rho, h_\rho\right)\,, \qquad \rho = a, \ldots v\,. \tag{6.328}$$

179

These will enter $\bar{J}_2^{\text{NLL}}(z)$ and $\hat{J}_2^{\text{NLL}}(z)$ as linear combinations with identical coefficients (owing to Eq. (6.302)), which however do depend on the flavour structure. Explicitly:

$$\text{non} - \text{singlet}: \qquad \sum_\rho w_{\text{NS},\rho}\, J_\rho^{\text{num}}(z) = \qquad\qquad\qquad\qquad\qquad (6.329)$$

$$4\, J_a^{\text{num}} + 4\, J_b^{\text{num}} + 4\, J_h^{\text{num}} + 2\, J_c^{\text{num}} + 4\, J_d^{\text{num}} + 4\, J_e^{\text{num}} + 4\, J_f^{\text{num}}$$
$$- 4\, J_g^{\text{num}} - 4\, J_i^{\text{num}} + 8\, J_j^{\text{num}} - 4\, J_k^{\text{num}} - 2\, J_l^{\text{num}} + 8\, J_m^{\text{num}} + 8\, J_n^{\text{num}}\,,$$

$$\text{singlet}: \qquad \sum_\rho w_{\text{S},\rho}\, J_\rho^{\text{num}}(z) = \qquad\qquad\qquad\qquad\qquad (6.330)$$

$$4\, J_a^{\text{num}} + 4\, J_b^{\text{num}} + 4\, J_h^{\text{num}} - 2\, J_c^{\text{num}} - 4\, J_d^{\text{num}} - 4\, J_e^{\text{num}} - 4\, J_f^{\text{num}}$$
$$+ 4\, J_g^{\text{num}} + 4\, J_i^{\text{num}} - 8\, J_j^{\text{num}} - 4\, J_k^{\text{num}} + 2\, J_l^{\text{num}} - 8\, J_m^{\text{num}}$$
$$- 8\, J_n^{\text{num}} - 24\, n_F\, J_r^{\text{num}}\,,$$

$$\text{photon}: \qquad \sum_\rho w_{\gamma,\rho}\, J_\rho^{\text{num}}(z) = \qquad\qquad\qquad\qquad\qquad (6.331)$$

$$- 8\, J_o^{\text{num}} - 8\, J_p^{\text{num}} + 8\, J_q^{\text{num}} + 8\, J_s^{\text{num}} + 16\, J_t^{\text{num}} - 16\, J_u^{\text{num}} - 4\, J_v^{\text{num}}\,.$$

The results of these linear combinations when the J_ρ^{num} contributions are computed analytically as in Eq. (6.304) are the following:

$$\sum_\rho w_{\text{NS},\rho}\, J_\rho^{\text{num}}(z) = -\frac{2}{3}\pi^2 \log(1-z) + \frac{4}{3}\pi^2 + 10\log(2)^2\,, \qquad (6.332)$$

$$\sum_\rho w_{\text{S},\rho}\, J_\rho^{\text{num}}(z) = \frac{2}{3}\pi^2 \log(1-z) + 4\pi^2 - 10\log(2)^2\,, \qquad (6.333)$$

$$\sum_\rho w_{\gamma,\rho}\, J_\rho^{\text{num}}(z) = -4\log^3(1-z) + \frac{4}{3}\pi^2 \log(1-z) + \frac{4}{3}\pi^2 \log(2) - 4\log(2)^3 - 8\zeta_3\,. \qquad (6.334)$$

As was anticipated, the results on the r.h.s. of Eqs. (6.332)–(6.334) do not contribute to any of the $b_{2,i}^{\text{NLL}}$ coefficients, while they enter the coefficients $c_{2,1}^{\text{NLL}}$ and $c_{2,0}^{\text{NLL}}$ (singlet and non-singlet), and $c_{2,3}^{\text{NLL}}$, $c_{2,1}^{\text{NLL}}$, and $c_{2,0}^{\text{NLL}}$ (photon).

6.B   Coefficients of recursive solutions

In this appendix we report the results for the coefficients that enter Eq. (6.71) and Eq. (6.72). Note that the NLL coefficients do already include the r.h.s. of Eqs. (6.332)–(6.334). All of the coefficients that do not appear below are understood to be equal to zero. In the NLL coefficients, we employ the following shorthand notation:

$$L_0 = \log \frac{\mu_0^2}{m^2}\,. \qquad (6.335)$$

6.B.1   LL coefficients

- Non-singlet:

$$b_{\text{NS},\,1,0}^{\text{LL}} = 2\,, \tag{6.336}$$

$$c_{\text{NS},\,1,0}^{\text{LL}} = -2\,, \tag{6.337}$$

$$b_{\text{NS},\,2,1}^{\text{LL}} = 8\,, \tag{6.338}$$

$$c_{\text{NS},\,2,1}^{\text{LL}} = -8\,, \tag{6.339}$$

$$b_{\text{NS},\,2,0}^{\text{LL}} = 6\,, \tag{6.340}$$

$$c_{\text{NS},\,2,0}^{\text{LL}} = -2\,, \tag{6.341}$$

$$b_{\text{NS},\,3,2}^{\text{LL}} = 24\,, \tag{6.342}$$

$$c_{\text{NS},\,3,2}^{\text{LL}} = -24\,, \tag{6.343}$$

$$b_{\text{NS},\,3,1}^{\text{LL}} = 36\,, \tag{6.344}$$

$$c_{\text{NS},\,3,1}^{\text{LL}} = -12\,, \tag{6.345}$$

$$b_{\text{NS},\,3,0}^{\text{LL}} = \frac{27}{2} - 4\pi^2\,, \tag{6.346}$$

$$c_{\text{NS},\,3,0}^{\text{LL}} = \frac{9}{2} + 4\pi^2\,. \tag{6.347}$$

- Singlet:

$$
\begin{aligned}
b_{\text{S},\,k,i}^{\text{LL}} &= b_{\text{NS},\,k,i}^{\text{LL}} & \forall\,k\,,i\,, && \text{(6.348)}\\
c_{\text{S},\,k,i}^{\text{LL}} &= c_{\text{NS},\,k,i}^{\text{LL}} & \forall\,k\,,i\,. && \text{(6.349)}
\end{aligned}
$$

- Photon:

$$c_{\gamma,\,1,0}^{\text{LL}} = 1\,, \tag{6.350}$$

$$c_{\gamma,\,2,1}^{\text{LL}} = 2\,, \tag{6.351}$$

$$c_{\gamma,\,2,0}^{\text{LL}} = \frac{3}{2} - \frac{2}{3} n_F\,, \tag{6.352}$$

$$c_{\gamma,3,2}^{\text{LL}} = 4 \,, \tag{6.353}$$

$$c_{\gamma,3,1}^{\text{LL}} = 6 - \frac{4}{3} n_F \,, \tag{6.354}$$

$$c_{\gamma,3,0}^{\text{LL}} = \frac{9}{4} - \frac{2}{3} \pi^2 - n_F + \frac{4}{9} n_F^2 \,. \tag{6.355}$$

6.B.2 NLL coefficients

• Non-singlet:

$$b_{\text{NS},0,1}^{\text{NLL}} = -4 \,, \tag{6.356}$$

$$c_{\text{NS},0,1}^{\text{NLL}} = 4 \,, \tag{6.357}$$

$$b_{\text{NS},0,0}^{\text{NLL}} = 2 \left(L_0 - 1 \right) \,, \tag{6.358}$$

$$c_{\text{NS},0,0}^{\text{NLL}} = -2 \left(L_0 - 1 \right) \,, \tag{6.359}$$

$$b_{\text{NS},1,2}^{\text{NLL}} = -12 \,, \tag{6.360}$$

$$c_{\text{NS},1,2}^{\text{NLL}} = 12 \,, \tag{6.361}$$

$$b_{\text{NS},1,1}^{\text{NLL}} = -14 + 8 L_0 + 8 \pi b_0 \,, \tag{6.362}$$

$$c_{\text{NS},1,1}^{\text{NLL}} = 10 - 8 L_0 - 8 \pi b_0 \,, \tag{6.363}$$

$$b_{\text{NS},1,0}^{\text{NLL}} = 1 - \frac{20}{9} n_F + 4 \pi b_0 - \frac{4 \pi b_1}{b_0} + \frac{4}{3} \pi^2 + L_0 (6 - 4 \pi b_0) \,, \tag{6.364}$$

$$c_{\text{NS},1,0}^{\text{NLL}} = -2 + \frac{32}{9} n_F - 4 \pi b_0 + \frac{4 \pi b_1}{b_0} - \frac{4}{3} \pi^2 + L_0 (-2 + 4 \pi b_0) \,, \tag{6.365}$$

$$b_{\text{NS},2,3}^{\text{NLL}} = -32 \,, \tag{6.366}$$

$$c_{\text{NS},2,3}^{\text{NLL}} = 32 \,, \tag{6.367}$$

$$b_{\text{NS},2,2}^{\text{NLL}} = 12(-5 + 2 L_0 + 4 \pi b_0) \,, \tag{6.368}$$

$$c_{\text{NS},2,2}^{\text{NLL}} = -12(-3 + 2 L_0 + 4 \pi b_0) \,, \tag{6.369}$$

$$b_{\text{NS},2,1}^{\text{NLL}} = -17 - \frac{160}{9}n_F + 56\pi b_0 - \frac{32\pi b_1}{b_0} + \frac{40}{3}\pi^2 - 16\pi^2 b_0^2 - 4L_0(-9 + 8\pi b_0), \qquad (6.370)$$

$$c_{\text{NS},2,1}^{\text{NLL}} = -7 + \frac{208}{9}n_F - 32\pi b_0 + \frac{32\pi b_1}{b_0} - \frac{40}{3}\pi^2 + 16\pi^2 b_0^2 + 4L_0(-3 + 8\pi b_0), \qquad (6.371)$$

$$
\begin{aligned}
b_{\text{NS},2,0}^{\text{NLL}} &= 9 - \frac{24\pi b_1}{b_0} - 4\pi b_0 + 6\pi^2 + 8\pi^2 b_1 - 8\pi^2 b_0^2 - \frac{16}{3}\pi^3 b_0 - 40\zeta_3 \\
&\quad + L_0\left(\frac{27}{2} - 24\pi b_0 - 4\pi^2 + 8\pi^2 b_0^2\right) + n_F\left(\frac{40\pi b_0}{9} - \frac{2}{9}(33 + 4\pi^2)\right), \quad (6.372)
\end{aligned}
$$

$$
\begin{aligned}
c_{\text{NS},2,0}^{\text{NLL}} &= -4 - \frac{10}{3}\pi^2 + 8\pi^2 b_0^2 + \frac{8\pi b_1}{b_0} - 8\pi^2 b_1 + 14\pi b_0 + \frac{16}{3}\pi^3 b_0 \\
&\quad + n_F\left(\frac{22}{9} - \frac{64}{9}\pi b_0 + \frac{8}{9}\pi^2\right) + L_0\left(\frac{9}{2} + 8\pi b_0 + 4\pi^2 - 8\pi^2 b_0^2\right) + 40\zeta_3. \quad (6.373)
\end{aligned}
$$

• Singlet:

$$
\begin{aligned}
b_{\text{S},k,i}^{\text{NLL}} &= b_{\text{NS},k,i}^{\text{NLL}} && \forall\, k\,, i\,, & (6.374) \\
c_{\text{S},k,i}^{\text{NLL}} &= c_{\text{NS},k,i}^{\text{NLL}} && \forall\, k\,, i\,. & (6.375)
\end{aligned}
$$

• Photon:

$$c_{\gamma,0,0}^{\text{NLL}} = (L_0 - 1), \qquad (6.376)$$

$$c_{\gamma,1,2}^{\text{NLL}} = -3, \qquad (6.377)$$

$$c_{\gamma,1,1}^{\text{NLL}} = -7 + 2L_0 - \frac{4}{3}n_F, \qquad (6.378)$$

$$c_{\gamma,1,0}^{\text{NLL}} = -4 + n_F\left(-\frac{26}{9} - \frac{2}{3}L_0\right) + 2\pi b_0 - \frac{2\pi b_1}{b_0} + L_0\left(\frac{3}{2} - 2\pi b_0\right), \qquad (6.379)$$

$$c_{\gamma,2,3}^{\text{NLL}} = -6, \qquad (6.380)$$

$$c_{\gamma,2,2}^{\text{NLL}} = -\frac{37}{2} + 4L_0 - \frac{2}{3}n_F + 10\pi b_0, \qquad (6.381)$$

$$
\begin{aligned}
c_{\gamma,2,1}^{\text{NLL}} &= -\frac{37}{2} + \frac{8}{9}n_F^2 + 18\pi b_0 - \frac{8\pi b_1}{b_0} + 2\pi^2 \\
&\quad + L_0(6 - 8\pi b_0) - \frac{4}{3}n_F(5 + L_0 - 2\pi b_0), \quad (6.382)
\end{aligned}
$$

183

$$
\begin{aligned}
c_{\gamma,2,0}^{\mathrm{NLL}} &= -\frac{45}{8} + \left(\frac{52}{27} + \frac{4}{9}L_0\right) n_F^2 + 4\pi b_0 + \frac{11}{6}\pi^2 - 4\pi^2 b_0^2 - \frac{6\pi b_1}{b_0} + 4\pi^2 b_1 \\
&\quad + n_F \left(-\frac{23}{6} + \frac{40\pi b_0}{9} + \frac{8\pi b_1}{3b_0} + \frac{2}{9}\pi^2 - L_0 + L_0\frac{8\pi b_0}{3}\right) \\
&\quad + L_0 \left(\frac{9}{4} - 6\pi b_0 - \frac{2}{3}\pi^2 + 4\pi^2 b_0^2\right) - 6\zeta_3 \,.
\end{aligned}
\tag{6.383}
$$

6.C   Splitting functions for QED

In this appendix we report explicit expressions for the LO and NLO splitting functions in QED [210]. The analogous LO and NLO splitting functions in QCD can be found on standard textbooks [14]. The notation adopted here is the same of Sec. 1.6. Consistently with the rest of the chapter, n_F denotes the number of lepton families. Note that each charged lepton has a charge squared equal to unity.

6.C.1   Leading order

The LO splitting functions are:

$$
P_{\mathrm{NS}}^{[0]} = P_{\Sigma\Sigma}^{[0]} = \left(\frac{1+x^2}{1-x}\right)_+ = -(1+x) + \frac{2}{(1-x)_+} + \frac{3}{2}\delta(1-x)\,,
\tag{6.384}
$$

$$
P_{\Sigma\gamma}^{[0]} = 2n_F[x^2 + (1-x)^2]\,,
\tag{6.385}
$$

$$
P_{\gamma\Sigma}^{[0]} = \frac{1+(1-x)^2}{x}\,,
\tag{6.386}
$$

$$
P_{\gamma\gamma}^{[0]} = -\frac{2}{3}n_F\delta(1-x)\,.
\tag{6.387}
$$

6.C.2   Next-to-leading order

At NLO the lepton-lepton valence, lepton-antilepton valence and singlet splitting functions are:

$$
\begin{aligned}
P_{ee}^{\mathrm{v},[1]} &= \left\{ -\left[\frac{3}{2}\ln x + 2\ln x \ln(1-x)\right]p_{ee}(x) - \left(\frac{3}{2} + \frac{7}{2}x\right)\ln x \right. \\
&\quad \left. -\frac{1}{2}(1+x)\ln^2 x - 5(1-x)\right\} + n_F\left\{ -\left[\frac{10}{9} + \frac{2}{3}\ln x\right]p_{qq}(x) - \frac{4}{3}(1-x)\right\} \\
&\quad + \delta(1-x)\left\{ \left(\frac{3}{8} - \frac{\pi^2}{2} + 6\zeta_3\right) - n_F\left(\frac{1}{6} + \frac{2\pi^2}{9}\right)\right\}\,,
\end{aligned}
\tag{6.388}
$$

$$
P_{e\bar{e}}^{\mathrm{v},[1]} = 4(1-x) + 2(1+x)\ln x + 2p_{ee}(-x)S_2(x)\,,
\tag{6.389}
$$

$$
P_{ee}^{\mathrm{s},[1]} = \frac{20}{9x} - 2 + 6x - \frac{56}{9}x^2 + \left(1 + 5x + \frac{8}{3}x^2\right)\ln x - (1+x)\ln^2 x\,,
\tag{6.390}
$$

with:

$$p_{ee}(x) = \frac{1 + x^2}{(1 - x)_+}, \tag{6.391}$$

$$S_2(x) = -2\text{Li}_2(-x) + \frac{1}{2}\ln^2 x - 2\ln x \ln(1 + x) - \frac{\pi^2}{6}. \tag{6.392}$$

Note that at NLO $P^{\text{s}}_{ee} = P^{\text{s}}_{e\bar{e}}$. Thus the non-singlet and the singlet combinations of splitting functions, Eq. (1.76) and Eq. (1.79) respectively, read:

$$P^{[1]}_{\text{NS}} = P^{\text{v},[1]}_{ee} - P^{\text{v},[1]}_{e\bar{e}}, \tag{6.393}$$

$$P^{[1]}_{\Sigma\Sigma} = P^{\text{v},[1]}_{ee} + P^{\text{v},[1]}_{e\bar{e}} + 2n_F P^{\text{s},[1]}_{ee}. \tag{6.394}$$

Eqs. (6.393)-(6.394) can be cast in the following form:

$$P^{[1]} = P^{[1]}_R + P^{[1]}_+ \frac{2}{(1 - x)_+} + P^{[1]}_\delta \delta(1 - x) \tag{6.395}$$

with a the regular term, a the term proportional to a plus prescription and a term proportional to a delta function. By an explicit computation, for the regular parts we obtain:

$$P^{[1]}_{\text{NS},R} = \left\{ -\left[\frac{3}{2}\ln x + 2\ln x \ln(1 - x)\right](-1 - x) - \left(\frac{3}{2} + \frac{7}{2}x\right)\ln x \right.$$
$$\left. -\frac{1}{2}(1 + x)\ln^2 x - 5(1 - x) - 4(1 - x) - 2(1 + x)\ln x - 2\left(-1 + x + \frac{2}{1 + x}\right)S_2(x) \right\}$$
$$+ n_F\left\{ -\left[\frac{10}{9} + \frac{2}{3}\ln x\right](-1 - x) - \frac{4}{3}(1 - x) \right\}, \tag{6.396}$$

$$P^{[1]}_{\Sigma\Sigma,R} = \left\{ -1 + x + \left(\frac{1}{2} - \frac{3}{2}x\right)\ln x - \frac{1}{2}(1 + x)\ln^2 x \right.$$
$$\left. -\left[\frac{3}{2}\ln x + 2\ln x \ln(1 - x)\right](-1 - x) + 2\left(\frac{2}{1 + x} - 1 + x\right)S_2(x) \right\}$$
$$+ n_F\left\{ -\frac{16}{3} + \frac{40}{3}x + \left(10x + \frac{16}{3}x^2 + 2\right)\ln x \right.$$
$$\left. -\frac{112}{9}x^2 + \frac{40}{9x} - 2(1 + x)\ln^2 x - \left[\frac{10}{9} + \frac{2}{3}\ln x\right](-1 - x) \right\}, \tag{6.397}$$

while the distributional parts are the same for the two functions:

$$P^{[1]}_{\text{NS},+} = P^{[1]}_{\Sigma\Sigma,+} = -\left[\frac{3}{2}\ln x + 2\ln x \ln(1 - x)\right] - n_F\left[\frac{10}{9} + \frac{2}{3}\ln x\right], \tag{6.398}$$

$$P^{[1]}_{\text{NS},\delta} = P^{[1]}_{\Sigma\Sigma,\delta} = \left(\frac{3}{8} - \frac{\pi^2}{2} + 6\zeta_3\right) - n_F\left(\frac{1}{6} + \frac{2\pi^2}{9}\right). \tag{6.399}$$

185

The remaining NLO splitting functions are:

$$
P_{\Sigma\gamma}^{[1]} = n_F \left\{ 4 - 9x - (1 - 4x)\ln x - (1 - 2x)\ln^2 x + 4\ln(1 - x) \right.
$$

$$
\left. + \left[2\ln^2\left(\frac{1 - x}{x}\right) - 4\ln\left(\frac{1 - x}{x}\right) - \frac{2}{3}\pi^2 + 10 \right] p_{e\gamma}(x) \right\}, \tag{6.400}
$$

$$
P_{\gamma\Sigma}^{[1]} = \left\{ -\frac{5}{2} - \frac{7}{2}x + \left(2 + \frac{7}{2}x\right)\ln x - \left(1 - \frac{1}{2}x\right)\ln^2 x - 2x\ln(1 - x) \right.
$$

$$
\left. - \left[3\ln(1 - x) + \ln^2(1 - x) \right] p_{\gamma e}(x) \right\} + n_F \left\{ -\frac{4}{3}x - \left[\frac{20}{9} + \frac{4}{3}\ln(1 - x) \right] p_{\gamma e}(x) \right\}, \tag{6.401}
$$

$$
P_{\gamma\gamma}^{[1]} = n_F \left\{ -16 + 8x + \frac{20}{3}x^2 + \frac{4}{3x} - (6 + 10x)\ln x - (2 + 2x)\ln^2 x \right\} - n_F \delta(1 - x), \tag{6.402}
$$

with:

$$
p_{e\gamma}(x) = x^2 + (1 - x)^2, \tag{6.403}
$$

$$
p_{\gamma e}(x) = \frac{1 + (1 - x)^2}{x}. \tag{6.404}
$$

Conclusions

In the thesis, we have investigated several topics in Standard Model precision phenomenology for LHC and future colliders. The concept of *substructure* has been the main thread running through the various chapters of the thesis. Though the individual topics discussed differ significantly from each other, they are intimately connected. We has seen how jet measurements provide unique constraints on PDFs, and how more precise PDFs provide better predictions for jet processes. Moreover, the formalism of PDFs usually adopted in QCD has been successfully transplanted into QED, thereby providing a systematic improvement on fixed-order predictions. In these Conclusions, we will recall the main outcomes of our studies while also broadening our discussion towards possible directions for future analyses.

First, we concentrated on jet observables at LHC. In Chapter 2 we discussed the theoretical predictions at various degrees of accuracy for QCD processes involving jets. In particular, we focused on the single-jet inclusive cross section. In Chapter 3 we went on to study this observable in great detail, concentrating on its definition. In an attempt to correct for its non-unitarity, we proposed a set of alternative unitary weighted definitions [1]. Our study revealed unknown features of the standard definition. It turned out that, thanks to its own non-unitarity, the standard definition does not depend in a problematical way upon the value of the momentum cutoff used in the jet definition. In addition, we found that the apparent perturbative instability of the standard definition is the manifestation of an unnatural smallness of the NLO K-factor for a value of the jet radius $R \sim 0.4$, similar to the one usually used by the experimental collaborations at the LHC. This unnatural smallness is a consequence of an accidental cancellation between the individual contributions of the leading jet and of the subleading jet to the total cross-section. Even if weighted definitions do not show obvious benefits over the standard definition, those which include only the two leading jets show interesting features. This finding indirectly confirms the observed better perturbative stability of dijet observables.

We then moved to hadronic PDFs. In Chapter 4, within the NNPDF framework, we studied in a systematic way the inclusion of jet measurements in a PDF fit [3]. We have found full consistency among the constraints imposed by single-inclusive jets and dijets on PDFs, specifically the gluon PDF. We have shown that NNLO corrections are crucial in order to ensure compatibility of the jet observables with the rest of the global dataset. In addition, at NLO the choice of central scale has a significant impact (with the scale choice $\hat{H}_T$ for single-inclusive jets better behaved at NLO), while at NNLO perturbative stability appears to be achieved. In a comparative assessment of single-inclusive jets vs. dijets, we found that the dijet observable has a more marked impact on the gluon central value. Moreover, dijet observables behave better

perturbatively: as the perturbative order of the theory prediction increases, the data-theory agreement clearly improves. However, the single-inclusive jet observable leads to a more significant reduction of the gluon uncertainty. These observations could be due to theoretical reasons but also to the nature of the current data. We expect that these issues will be settled with the availability of more precise data, possibly for a greater variety of kinematic observables, including more differential ones. We are waiting for the analysed data from LHC Run-II at 13 TeV, as well as for data coming from future LHC Run-III and HL-LHC runs. Chapter 4 has been a first step towards widening the set of jet observables used in precision PDF studies, which include not only multi-differential jet cross-sections, but also jet substructure observables.

Chapter 5 focused on an analytical understanding of machine learning techniques used for quarks versus gluon discrimination [4], which is a hot topic in jet substructure studies. First, we proposed a variant of the standard N-subjettiness observable, the *primary N-subjettiness* $\mathcal{T}_N$, which is only sensitive, at LL accuracy, to emissions of soft gluons off the original hard parton. Thanks to the simple all-order behaviour of the observables $\{\mathcal{T}_1 \ldots \mathcal{T}_n\}$, we have been able to determine that the optimal discriminant at LL is just a cut on $\mathcal{T}_n$. The central part of Chapter 5 was the analytical study of the LL behaviour of a perceptron that takes primary N-subjettiness variables as inputs. We found that the ability of the neural network to find the correct minimum actually depends on the functional form of the inputs. Namely, it succeeds if logarithms (or square logarithms) of the N-subjettiness are passed, but it fails with linear inputs. Furthermore, we have also found that, in the case of linear inputs, the learning rate of the perceptron is slower than with other forms of inputs. Finally, we have considered a more realistic framework for our analysis, with a full neutral network trained on pseudodata generated with a general-purpose Monte Carlo parton shower. We have obtained qualitative agreement with the perceptron analysis. The findings in Chapter 5 highlight, in a quantitative way, how having expert-knowledge of the underlying physical phenomena helps when dealing with classification problems that employ ML techniques. In addition, from this preliminary analysis, primary N-subjettiness appears to be an intriguing observable worthy of further investigations; going beyond LL accuracy for primary N-subjettiness is not only possible but desirable. Whether or not it is possible to study more complicated network architectures analytically is an open question. We have just begun scratching the surface of these types of calculations, and we look forward to future work in this direction.

Finally, in Chapter 6, we devoted our attention to the computation of the electron, positron, and photon PDFs of an unpolarised electron. QED PDFs are crucial ingredients for the high-precision predictions needed for future e^+e^- colliders. We improved the accuracy of the existing LL results to NLL accuracy, solving the DGLAP evolution by means of both numerical and analytical methods [2]. The two solutions were shown to agree extremely well (typically, at the 10^{-4} level). The analytical results stem from an additive matching formula which combines a prediction that is accurate to all orders in α for $z \to 1$, with a prediction that is accurate up to $\mathcal{O}(\alpha^3)$ in the whole z range. We found the NLL large-z asymptotic result of the electron PDF to possess the same qualitative behaviour as that of the LL result, with the $z \to 1$ singularity even more pronounced than at LL because of the presence of additional $\log(1-z)$ terms. As for the photon PDF, which at LL vanishes at $z \to 1$, at NLL it grows logarithmically, thus exhibiting, similarly to the electron PDF, an enhanced growth at higher orders. In view of the rapid growth of the electron PDF at $z \to 1$, the analytical knowledge of the PDFs is crucial in the context of numerical computations, because it allows to adjust the integration procedure, which would otherwise be hardly converging.

Work is still ongoing in the continuation of Chapter 6. In Chapter 6 we worked in the $\overline{\text{MS}}$ subtraction scheme. Some of the explicit $\log(1-z)$ logarithms in the NLL PDFs mentioned above stem directly from this scheme choice. In an alternative factorization scheme, similar to the DIS scheme usually adopted in QCD, in which one maximally simplifies the PDFs' initial conditions, such logarithms are absent. Moreover, the phenomenological impact of NLL PDFs on physical observables is still undetermined, and work is ongoing in this direction too. NLL electron PDFs not only provide a NLL correction to processes with incoming electrons but also allow to treat photon-initiated hard processes in the same framework.

We summarized the main results of each individual chapter in this thesis, and we now bring the reader's attention to some transversal topics which pervaded the thesis.

One such recurring topic was machine learning. We met neural networks in two different contexts: in Chapter 4, when discussing the NNPDF methodology, and in Chapter 5, which is entirely devoted to the application of ML to jet physics. Indeed, machine learning is nowadays becoming increasingly used in the field of particle physics phenomenology. However, despite its rapid development and the unquestionable improvements that ML brings to particle physics, it is often met with a certain degree of suspicion, given its intrinsic black box nature. Hence, it is extremely important — and will be even more so in the future. — to be able to test the ML framework in environments we can control. Chapter 5 took a first step in this direction: we anchored the applicability of ML to the solid ground of perturbative QCD. We shed a new light on the possibility of validating the neural network by means of analytical calculations. In spirit, the approach adopted in Chapter 5 shares similarities with closure tests like those the NNPDF collaboration routinely performs to assess the robustness of the PDF fitting methodology.

The importance of an analytical approach was another recurring topic throughout the thesis. Chapter 3 used an analytical approximation to the full cross section in order to highlight features of the various definitions for the single-jet inclusive cross section. Chapter 6 found important differences in the large-z region between the electron PDFs at NLL as compared to those at LL. Analytic calculations have allowed us to understand these behaviours beyond what would have been achievable numerically. As already remarked, such analytical knowledge is also crucial when performing convolutions. Furthermore, the recursive analytical solutions allowed us to cross-check the numerical solutions. This is vital for the degree of accuracy we aimed to reach. Finally, the good analytical properties of the primary N-subjettiness definition paved the way for the studies of Chapter 5.

Another leitmotiv of this thesis has been the exploration of new definitions of well-known observables, as developed both in Chapter 3 and Chapter 5. Chapter 3 introduced alternative unitary definitions of the single-jet inclusive cross section. Regrettably, in the end we found no notable improvement over the standard definition; however, as a by-product of this investigation, we revealed new features of the standard definition. In Chapter 5 we introduced the primary N-subjettiness, a variant of the standard N-subjettiness. The new observable is more amenable to analytical calculations and it maintains, if not exceeds, the discriminating power of the standard definition, leading to an improvement in term of performance.

The huge amount of experimental data LHC will have collected at the end of its life will drive the experimental community towards tremendously precise measurements. Equally precise theoretical predictions within the Standard Model framework will then be an indispensable prerequisite for a proper comparison. Wise use of machine learning technologies, analytical control over numerical results, and flexibility in the choice of observables will all play a key role in the near future of Standard Model phenomenology.

Résumé substantiel (Français)

Dans cette thèse, nous avons étudié plusieurs sujets de phénoménologie du Modèle Standard au LHC et aux futurs collisionneurs. Le concept de *sous-structure* a été le principal fil conducteur des différents chapitres de la thèse. Bien que les thèmes abordés diffèrent sensiblement les uns des autres, ils sont intimement liés. Nous avons vu comment les mesures des jets fournissent des contraintes uniques sur les PDF, et comment des PDF plus précises permettent de mieux prévoir les processus des jets. En outre, le formalisme des PDF généralement adopté dans la QCD a été utilisé avec succès en QED, ce qui permet une amélioration systématique des prévisions d'ordre fixe.

Tout d'abord, nous nous sommes concentrés sur les observables de jets au LHC. Dans le chapitre 2, nous avons discuté de prédictions théoriques à divers degrés de précision pour les processus de QCD impliquant des jets. En particulier, nous nous sommes concentré sur la section transversale inclusive d'un seul jet. Dans le chapitre 3, nous avons étudié ce observable dans de détail, en se concentrant sur sa définition. Dans le but de corriger son non-unitarieté, nous avons proposé des définitions alternatives unitaires [1]. Notre étude a révélé des caractéristiques inconnues de la définition standard. Nous avons constaté que c'est grâce à sa propre non-unitarieté, que la definition standard ne dépend pas de manière problématique de la valeur de la coupure sur l'impulsion utilisée dans la définition du jet. En outre, nous avons constaté que l'instabilité perturbrive apparente de la definition standard est la manifestation de la petite taille accidentelle du facteur K du NLO pour une valeur du rayon du jet $R \sim 0,4$, similaire à celle habituellement utilisée par les collaborations expérimentales au LHC. Cette petite taille accidentelle est une conséquence d'une annulation entre les contributions individuelles du jet de tête et du jet secondaire à la section transversale totale. Même si les définitions pondérées unitaires ne présentent pas d'avantages évidents par rapport à la définition standard, celles qui n'incluent que les deux principaux jets montrent des aspects interessants. Ce résultat confirme indirectement les meilleures caractéristiques de stabilité perturbative des observables de dijet.

Nous sommes ensuite passés aux PDF hadroniques. Au chapitre 4, dans le cadre de la collaboration NNPDF, nous avons étudié de manière systématique l'inclusion des mesures des jets dans un fit de PDF [3]. Nous avons constaté une cohérence totale entre les contraintes imposées par des mesures de dijet et d'un seul jet sur les PDF, en particulier la PDF du gluon. Nous avons montré que les corrections NNLO sont cruciales pour assurer la compatibilité des observables de jet avec le reste de l'ensemble des données. En outre, au NLO, le choix de l'échelle centrale a un impact significatif (avec le choix d'échelle choix $\hat{H}_T$ pour les observables d'un seul jet donnant un meilleur comportement au NLO), tandis qu'au NNLO une stabilité perturbatrive semble être atteinte. Dans une évaluation comparative des jets et observables

de dijet, nous avons constaté que le dijet a un impact plus marqué sur le valeur centrale du gluon. De plus, les observables de dijet se comportent mieux au niveau perturbatif: à mesure que l'ordre perturbatif de la prédiction de la théorie augmente, l'accord des données avec la theorie s'améliore nettement. Cependant, l'observable d'un seul jet conduit à une réduction plus significative de l'incertitude sur la PDF du gluon. Ces observations pourraient être dues à des raisons théoriques mais aussi à la nature des données actuelles. Nous espérons que ces questions seront réglées grâce à la disponibilité de données plus précises, éventuellement pour une plus grande variété d'observables cinématiques, notamment plus différentielles. Le chapitre 4 a été un premier pas vers l'élargissement de l'ensemble des observables de jets utilisées dans les études de PDF de précision.

Le chapitre 5 est focalisé sur une compréhension analytique des techniques de machine learning utilisées pour la discrimination de quark contre gluon jet [4], qui est un sujet d'actualité dans les études sur la sous-structure des jets. Tout d'abord, nous avons proposé une variante de l'observable N-subjettiness, le *primary N-subjettiness* $\mathcal{T}_N$, qui est seulement sensible, à précision logarithmique dominante, aux émissions de gluons soft à partir de la partie hard d'origine. Grâce au comportement simple de tous les ordres des observables $\{\mathcal{T}_1 \ldots \mathcal{T}_n\}$, nous avons pu déterminer que le discriminant optimal à LL n'est qu'une coupure $\mathcal{T}_n$. La partie centrale du chapitre 5 était l'étude analytique du comportement d'un perceptron qui prend des variables primaires comme inputs. Nous avons trouvé que la capacité du réseau neuronal à trouver le bon minimum en fait dépend de la forme fonctionnelle des inputs. En d'autres termes, il réussit si les logarithmes (ou logarithmes carrés) du N-subjettiness sont donnés, mais il échoue avec les inputs linéaires. De plus, nous avons également constaté que, dans le cas des inputs linéaires, le taux d'apprentissage du perceptron est plus lent que avec d'autres formes d'inputs. Enfin, nous avons envisagé un cadre plus réaliste pour notre analyse, avec une réseau de neurons complet formé sur les pseudodonnées générées avec un Monte Carlo parton shower. Nous avons obtenu un accord qualitatif avec l'analyse du perceptron. Les conclusions du chapitre 5 mettent en évidence, de manière quantitative, comment la connaissance des phénomènes physiques sous-jacents est utile pour traiter les problèmes de classification qui font appel à des techniques de ML. En outre, d'après cette analyse préliminaire, la primary N-subjettiness semble être une observable intéressante digne d'être approfondie; aller au-delà de la précision logarithmique dominante est non seulement possible mais souhaitable. S'il est possible ou non d'étudier analytiquement des architectures de réseau plus complexes est une question ouverte. Nous venons de commencer à gratter la surface de ces types de calculs, et nous attendons avec impatience les travaux futurs dans ce direction.

Enfin, au chapitre 6, nous avons consacré notre attention au calcul de la PDF de l'électron, du positron et du photon PDF non polarisés. Les PDF en QED sont des ingrédients essentiels pour les prévisions de haute précision nécessaires pour les futurs collisioneurs e^+e^-. Nous avons amélioré la précision des résultats au LL existants pour atteindre la précision du NLL, et résolu l'évolution des équations DGLAP par des moyens à la fois numériques et analytiques [2]. Les deux solutions se sont révélées très concordantes (généralement, au niveau de 10^{-4}). Ces résultats proviennent d'une formule d'appariement qui combine un terme exacte à toutes les ordres en α pour $z \to 1$, avec un qui est précis jusqu'à $\mathcal{O}(\alpha^3)$ dans toute la gamme des z. Nous avons constaté que le résultat asymptotique de la PDF électronique au NLL à grande échelle (z) a un comportement similaire a celui du résultat LL, la singularité $z \to 1$ étant encore plus prononcée que au LL en raison de la présence de termes supplémentaires $\log(1-z)$. Quant à la PDF du photon, qui est nulle à LL à $z \to 1$, à NLL elle croît de manière logarithmique,

exposant ainsi, à l'instar de la PDF électronique, une croissance accrue aux ordres supérieures. Compte tenu de la croissance rapide de la PDF électronique à $z \rightarrow 1$, la connaissance analytique des PDF est cruciale dans un contexte de haute precision car ça permet d'ajuster la procédure d'intégration, qui, autrement, ne convergerait pas. L'impact phénoménologique des PDF NLL sur les observables physiques est encore indéterminé, et des travaux sont en cours dans ce sens.

Bibliography

[1] M. Cacciari, S. Forte, D. Napoletano, G. Soyez and G. Stagnitto, *Single-jet inclusive cross section and its definition*, *Phys. Rev.* **D100** (2019) 114015 [1906.11850].

[2] V. Bertone, M. Cacciari, S. Frixione and G. Stagnitto, *The partonic structure of the electron at the next-to-leading logarithmic accuracy in QED*, *JHEP* **03** (2020) 135 [1911.12040].

[3] R. Abdul Khalek, S. Forte, T. Gehrmann, A. Gehrmann-De Ridder, T. Giani, E. W. N. Glover, A. Huss, E. R. Nocera, J. Pires, J. Rojo and G. Stagnitto, *Phenomenology of NNLO jet production at the LHC and its impact on parton distributions*, *Eur. Phys. J. C* **80** (2020) 797 [2005.11327].

[4] G. Kasieczka, S. Marzani, G. Soyez and G. Stagnitto, *Towards Machine Learning Analytics for Jet Substructure*, 2007.04319, accepted by JHEP.

[5] G. F. Giudice, *The Dawn of the Post-Naturalness Era*, pp. 267–292. World Scientific, 2019. 1710.07663. 10.1142/9789813238053_0013.

[6] European Strategy Group collaboration, *2020 Update of the European Strategy for Particle Physics*. CERN Council, Geneva, 2020, 10.17181/ESU2020.

[7] ATLAS collaboration, *Standard Model Summary Plots Summer 2019*, .

[8] LCC Physics Working Group collaboration, *Tests of the Standard Model at the International Linear Collider*, 1908.11299.

[9] G. Soyez, *Event generators*, Talk given at LHCP 2020 conference, 2020.

[10] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[11] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2014.

[12] M. Srednicki, *Quantum field theory*. Cambridge University Press, 2007.

[13] G. F. Sterman, *An Introduction to quantum field theory*. Cambridge University Press, 1993.

[14] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and collider physics*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1.

[15] T. Muta, *Foundations of quantum chromodynamics. Second edition*, *World Sci. Lect. Notes Phys.* **57** (1998) 1.

[16] R. D. Field, *Applications of Perturbative QCD*, *Front. Phys.* **77** (1989) 1.

[17] G. Dissertori, I. G. Knowles and M. Schmelling, *High energy experiments and theory*. Oxford Science Publications, 2003.

[18] M. Bohm, A. Denner and H. Joos, *Gauge theories of the strong and electroweak interaction*. Springer, 2001.

[19] M. H. Seymour, *Quantum chromodynamics*, in *2004 European School of High-Energy Physics, Sant Feliu de Guixols, Spain, 30 May - 12 June 2004*, pp. 49–94, 2005, hep-ph/0505192, http://doc.cern.ch/yellowrep/CERN-PH-TH-2005-083.

[20] P. Nason, *Introduction to perturbative QCD*, in *11th Jorge Andre Swieca Summer School on Particle and Fields*, pp. 409–486, 1, 2001.

[21] G. P. Salam, *Elements of QCD for hadron colliders*, in *High-energy physics. Proceedings, 17th European School, ESHEP 2009, Bautzen, Germany, June 14-27, 2009*, 2010, 1011.5131, http://inspirehep.net/record/880643/files/arXiv:1011.5131.pdf.

[22] M. L. Mangano, *Introduction to QCD*, in *1998 European School of High-Energy Physics*, pp. 53–97, 1998.

[23] A. Bacchetta, *Notes on QED*, Lectures, http://www2.pv.infn.it/~bacchett/teaching.html.

[24] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *Phys. Rev.* **D98** (2018) 030001.

[25] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) 1343.

[26] H. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346.

[27] T. Appelquist and J. Carazzone, *Infrared Singularities and Massive Fields*, *Phys. Rev. D* **11** (1975) 2856.

[28] R. P. Feynman, *Very high-energy collisions of hadrons*, *Phys. Rev. Lett.* **23** (1969) 1415.

[29] T. C. Rogers, *An overview of transverse-momentum–dependent factorization and evolution*, *Eur. Phys. J.* **A52** (2016) 153 [1509.04766].

[30] R. Angeles-Martinez et al., *Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects*, *Acta Phys. Polon.* **B46** (2015) 2501 [1507.05267].

[31] J. C. Collins, D. E. Soper and G. F. Sterman, *Factorization of Hard Processes in QCD*, *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1 [hep-ph/0409313].

[32] J. D. Jackson, *Classical Electrodynamics*. Wiley, 1998.

[33] S. Catani and M. H. Seymour, *A General algorithm for calculating jet cross-sections in NLO QCD*, *Nucl. Phys.* **B485** (1997) 291 [`hep-ph/9605323`], [Erratum: Nucl. Phys.B510,503(1998)].

[34] G. Luisoni and S. Marzani, *QCD resummation for hadronic final states*, *J. Phys.* **G42** (2015) 103101 [`1505.04084`].

[35] M. Cacciari and S. Catani, *Soft gluon resummation for the fragmentation of light and heavy quarks at large x*, *Nucl. Phys.* **B617** (2001) 253 [`hep-ph/0107138`].

[36] F. Bloch and A. Nordsieck, *Note on the Radiation Field of the electron*, *Phys. Rev.* **52** (1937) 54.

[37] D. Yennie, S. C. Frautschi and H. Suura, *The infrared divergence phenomena and high-energy processes*, *Annals Phys.* **13** (1961) 379.

[38] T. Kinoshita, *Mass singularities of Feynman amplitudes*, *J. Math. Phys.* **3** (1962) 650.

[39] T. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, *Phys. Rev.* **133** (1964) B1549.

[40] G. Curci, W. Furmanski and R. Petronzio, *Evolution of Parton Densities Beyond Leading Order: The Nonsinglet Case*, *Nucl. Phys.* **B175** (1980) 27.

[41] C. Muselli, *Resummations of Transverse Momentum Distributions*, Ph.D. thesis, Università degli Studi di Milano, 2017. `https://air.unimi.it/handle/2434/530136`.

[42] S. Frixione, *Initial conditions for electron and photon structure and fragmentation functions*, *JHEP* **11** (2019) 158 [`1909.03886`].

[43] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, *Sov. Phys. JETP* **46** (1977) 641 [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[44] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438 [Yad. Fiz.15,781(1972)].

[45] L. N. Lipatov, *The parton model and perturbation theory*, *Sov. J. Nucl. Phys.* **20** (1975) 94 [Yad. Fiz.20,181(1974)].

[46] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys.* **B126** (1977) 298.

[47] W. Furmanski and R. Petronzio, *Singlet Parton Densities Beyond Leading Order*, *Phys. Lett. B* **97** (1980) 437.

[48] S. Moch, J. A. M. Vermaseren and A. Vogt, *The Three loop splitting functions in QCD: The Nonsinglet case*, *Nucl. Phys.* **B688** (2004) 101 [`hep-ph/0403192`].

[49] A. Vogt, S. Moch and J. A. M. Vermaseren, *The Three-loop splitting functions in QCD: The Singlet case*, *Nucl. Phys.* **B691** (2004) 129 [`hep-ph/0404111`].

[50] G. P. Salam and J. Rojo, *A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)*, *Comput. Phys. Commun.* **180** (2009) 120 [0804.3755].

[51] M. Botje, *QCDNUM: Fast QCD Evolution and Convolution*, *Comput.Phys.Commun.* **182** (2011) 490 [1005.1481].

[52] V. Bertone, S. Carrazza and J. Rojo, *APFEL: A PDF Evolution Library with QED corrections*, *Comput.Phys.Commun.* **185** (2014) 1647 [1310.1394].

[53] A. Vogt, *Efficient evolution of unpolarized and polarized parton distributions with qcd-pegasus*, *Comput. Phys. Commun.* **170** (2005) 65 [hep-ph/0408244].

[54] S. Weinzierl, *Fast evolution of parton distributions*, *Comput.Phys.Commun.* **148** (2002) 314 [hep-ph/0203112].

[55] S. Forte, *Parton distributions at the dawn of the LHC*, *Acta Phys.Polon.* **B41** (2010) 2859 [1011.5247].

[56] 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.* **C75** (2015) 204 [1412.3989].

[57] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, 1912.10053.

[58] NNPDF collaboration, *Parton distributions from high-precision collider data*, *Eur. Phys. J.* **C77** (2017) 663 [1706.00428].

[59] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, vol. 958. Springer, 2019, 10.1007/978-3-030-15709-8, [1901.10342].

[60] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145 [1101.2599].

[61] G. F. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, *Phys.Rev.Lett.* **39** (1977) 1436.

[62] G. P. Salam, *Towards Jetography*, *Eur. Phys. J. C* **67** (2010) 637 [0906.1833].

[63] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock and B. R. Webber, *New clustering algorithm for multi - jet cross-sections in e+ e- annihilation*, *Phys. Lett.* **B269** (1991) 432.

[64] S. Catani, Y. L. Dokshitzer, M. Seymour and B. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187.

[65] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160 [hep-ph/9305266].

[66] Y. L. Dokshitzer, G. Leder, S. Moretti and B. Webber, *Better jet clustering algorithms*, *JHEP* **9708** (1997) 001 [hep-ph/9707323].

[67] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, `hep-ph/9907280`.

[68] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [`0802.1189`].

[69] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys. Lett. B* **641** (2006) 57 [`hep-ph/0512210`].

[70] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) 1896 [`1111.6097`].

[71] 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 (PhysTeV 2015) Les Houches, France, June 1-19, 2015*, 2016, `1605.04692`, `http://lss.fnal.gov/archive/2016/conf/fermilab-conf-16-175-ppd-t.pdf`.

[72] P. Gras, S. Hoeche, D. Kar, A. Larkoski, L. Lönnblad, S. Platzer, A. Siódmok, P. Skands, G. Soyez and J. Thaler, *Systematics of quark/gluon tagging*, *JHEP* **07** (2017) 091 [`1704.03878`].

[73] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [`1011.2268`].

[74] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, *N-Jettiness: An Inclusive Event Shape to Veto Jets*, *Phys. Rev. Lett.* **105** (2010) 092002 [`1004.2489`].

[75] M. Dasgupta, L. Schunk and G. Soyez, *Jet shapes for boosted jet two-prong decays from first-principles*, *JHEP* **04** (2016) 166 [`1512.00516`].

[76] G. P. Salam, L. Schunk and G. Soyez, *Dichroic subjettiness ratios to distinguish colour flows in boosted boson tagging*, *JHEP* **03** (2017) 022 [`1612.03917`].

[77] D. Napoletano and G. Soyez, *Computing N-subjettiness for boosted jets*, *JHEP* **12** (2018) 031 [`1809.04602`].

[78] A. J. Larkoski and E. M. Metodiev, *A Theory of Quark vs. Gluon Discrimination*, *JHEP* **10** (2019) 014 [`1906.01639`].

[79] A. J. Larkoski, *An Unorthodox Introduction to QCD*, `1709.06195`.

[80] M. Cacciari and N. Houdeau, *Meaningful characterisation of perturbative theoretical uncertainties*, *JHEP* **1109** (2011) 039 [`1105.5152`].

[81] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, *The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence*, *JHEP* **04** (2004) 068 [`hep-ph/0303085`].

[82] B. L. Combridge, J. Kripfganz and J. Ranft, *Hadron Production at Large Transverse Momentum and QCD*, *Phys. Lett.* **70B** (1977) 234.

[83] G. Soyez, *QCD for the LHC*, Lectures, `http://www.lpthe.jussieu.fr/~soyez/lectures/plan.html`.

[84] V. Shtabovenko, R. Mertig and F. Orellana, *New Developments in FeynCalc 9.0*, *Comput. Phys. Commun.* **207** (2016) 432 [1601.01167].

[85] R. Mertig, M. Bohm and A. Denner, *FEYN CALC: Computer algebraic calculation of Feynman amplitudes*, *Comput. Phys. Commun.* **64** (1991) 345.

[86] ATLAS collaboration, *Measurement of the inclusive jet cross-sections in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **09** (2017) 020 [1706.03192].

[87] CMS collaboration, *Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV*, *JHEP* **03** (2017) 156 [1609.05331].

[88] R. Ellis and J. Sexton, *QCD Radiative Corrections to Parton Parton Scattering*, *Nucl. Phys. B* **269** (1986) 445.

[89] Z. Kunszt and D. E. Soper, *Calculation of jet cross-sections in hadron collisions at order α_s^3*, *Phys. Rev.* **D46** (1992) 192.

[90] F. Aversa, P. Chiappetta, M. Greco and J. Guillet, *QCD Corrections to Parton-Parton Scattering Processes*, *Nucl. Phys. B* **327** (1989) 105.

[91] F. Aversa, L. Gonzales, M. Greco, P. Chiappetta and J. Guillet, *Jet cone size dependence of inclusive cross-sections to O $(alpha - s^3)$*, *Z. Phys. C* **49** (1991) 459.

[92] W. Giele, E. Glover and D. A. Kosower, *Higher order corrections to jet cross-sections in hadron colliders*, *Nucl. Phys. B* **403** (1993) 633 [hep-ph/9302225].

[93] W. T. Giele, E. W. N. Glover and D. A. Kosower, *The inclusive two jet triply differential cross-section*, *Phys. Rev.* **D52** (1995) 1486 [hep-ph/9412338].

[94] R. Ellis, D. Ross and A. Terrano, *The Perturbative Calculation of Jet Structure in e+ e- Annihilation*, *Nucl. Phys. B* **178** (1981) 421.

[95] S. D. Ellis, Z. Kunszt and D. E. Soper, *The One Jet Inclusive Cross-section at Order α_s^3: Gluons Only*, *Phys. Rev. Lett.* **62** (1989) 726.

[96] S. D. Ellis, Z. Kunszt and D. E. Soper, *The One Jet Inclusive Cross-section at Order α_s^3 Quarks and Gluons*, *Phys. Rev. Lett.* **64** (1990) 2121.

[97] S. D. Ellis, Z. Kunszt and D. E. Soper, *Two jet production in hadron collisions at order α_s^3 in QCD*, *Phys.Rev.Lett.* **69** (1992) 1496.

[98] S. Frixione, Z. Kunszt and A. Signer, *Three jet cross-sections to next-to-leading order*, *Nucl. Phys.* **B467** (1996) 399 [hep-ph/9512328].

[99] Z. Nagy, *Three jet cross-sections in hadron hadron collisions at next-to-leading order*, *Phys. Rev. Lett.* **88** (2002) 122003 [hep-ph/0110315].

[100] Z. Nagy, *Next-to-leading order calculation of three jet observables in hadron hadron collision*, *Phys. Rev.* **D68** (2003) 094002 [hep-ph/0307268].

[101] E. Glover, C. Oleari and M. Tejeda-Yeomans, *Two loop QCD corrections to gluon-gluon scattering*, *Nucl. Phys. B* **605** (2001) 467 [`hep-ph/0102201`].

[102] E. Glover and M. Tejeda-Yeomans, *One loop QCD corrections to gluon-gluon scattering at NNLO*, *JHEP* **05** (2001) 010 [`hep-ph/0104178`].

[103] Z. Bern, A. De Freitas and L. J. Dixon, *Two loop helicity amplitudes for gluon-gluon scattering in QCD and supersymmetric Yang-Mills theory*, *JHEP* **03** (2002) 018 [`hep-ph/0201161`].

[104] Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, *One loop n point gauge theory amplitudes, unitarity and collinear limits*, *Nucl. Phys. B* **425** (1994) 217 [`hep-ph/9403226`].

[105] M. L. Mangano and S. J. Parke, *Multiparton amplitudes in gauge theories*, *Phys. Rept.* **200** (1991) 301 [`hep-th/0509223`].

[106] S. Amoroso et al., *Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*, in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 3, 2020, `2003.01700`.

[107] A. Gehrmann-De Ridder, T. Gehrmann and E. Glover, *Antenna subtraction at NNLO*, *JHEP* **09** (2005) 056 [`hep-ph/0505111`].

[108] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover and G. Heinrich, *Infrared structure of e+ e- —> 3 jets at NNLO*, *JHEP* **11** (2007) 058 [`0710.0346`].

[109] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover and G. Heinrich, *Jet rates in electron-positron annihilation at O(alpha(s)**3) in QCD*, *Phys. Rev. Lett.* **100** (2008) 172001 [`0802.0813`].

[110] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover and G. Heinrich, *NNLO corrections to event shapes in e+ e- annihilation*, *JHEP* **12** (2007) 094 [`0711.4711`].

[111] A. Daleo, T. Gehrmann and D. Maitre, *Antenna subtraction with hadronic initial states*, *JHEP* **04** (2007) 016 [`hep-ph/0612257`].

[112] J. Currie, E. Glover and S. Wells, *Infrared Structure at NNLO Using Antenna Subtraction*, *JHEP* **04** (2013) 066 [`1301.4693`].

[113] J. Currie, E. W. N. Glover and J. Pires, *NNLO QCD predictions for single jet inclusive production at the LHC*, *Phys. Rev. Lett.* **118** (2017) 072002 [`1611.01460`].

[114] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover and J. Pires, *Second order QCD corrections to jet production at hadron colliders: the all-gluon contribution*, *Phys.Rev.Lett.* **110** (2013) 162003 [`1301.7310`].

[115] J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss and J. Pires, *Precise predictions for dijet production at the LHC*, *Phys. Rev. Lett.* **119** (2017) 152001 [`1705.10271`].

[116] M. Czakon, A. van Hameren, A. Mitov and R. Poncelet, *Single-jet inclusive rates with exact color at $\mathcal{O}\ (\alpha_s^4)$*, *JHEP* **10** (2019) 262 [1907.12911].

[117] M. Czakon, *A novel subtraction scheme for double-real radiation at NNLO*, *Phys. Lett. B* **693** (2010) 259 [1005.0274].

[118] M. Czakon and D. Heymes, *Four-dimensional formulation of the sector-improved residue subtraction scheme*, *Nucl. Phys. B* **890** (2014) 152 [1408.2500].

[119] ATLAS collaboration, *Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s}$=2.76 TeV and comparison to the inclusive jet cross section at $\sqrt{s}$=7 TeV using the ATLAS detector*, *Eur.Phys.J.* **C73** (2013) 2509 [1304.4739].

[120] J. Currie, E. W. N. Glover, A. Gehrmann-De Ridder, T. Gehrmann, A. Huss and J. Pires, *Single jet inclusive production for the individual jet p_T scale choice at the LHC*, in *23rd Cracow Epiphany Conference on Particle Theory Meets the First Data from LHC Run 2 Cracow, Poland, January 9-12, 2017*, 2017, 1704.00923, http://inspirehep.net/record/1589454/files/arXiv:1704.00923.pdf.

[121] J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss and J. Pires, *Infrared sensitivity of single jet inclusive production at hadron colliders*, *JHEP* **10** (2018) 155 [1807.03692].

[122] J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, N. Glover, A. Huss and J. Pires, *Jet cross sections at the LHC with NNLOJET*, *PoS* **LL2018** (2018) 001 [1807.06057].

[123] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr and G. Watt, *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J.* **C75** (2015) 132 [1412.7420].

[124] M. Dasgupta, F. A. Dreyer, G. P. Salam and G. Soyez, *Inclusive jet spectrum for small-radius jets*, *JHEP* **06** (2016) 057 [1602.01110].

[125] J. Bellm et al., *Jet cross sections at the LHC and the quest for higher precision*, 1903.12563.

[126] Z.-B. Kang, F. Ringer and I. Vitev, *The semi-inclusive jet function in SCET and small radius resummation for inclusive jet production*, *JHEP* **10** (2016) 125 [1606.06732].

[127] X. Liu, S.-O. Moch and F. Ringer, *Threshold and jet radius joint resummation for single-inclusive jet production*, *Phys. Rev. Lett.* **119** (2017) 212001 [1708.04641].

[128] I. W. Stewart and F. J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, *Phys. Rev.* **D85** (2012) 034011 [1107.2117].

[129] A. D. Martin, R. G. Roberts and W. J. Stirling, *Structure Function Analysis and psi, Jet, W, Z Production: Pinning Down the Gluon*, *Phys. Rev.* **D37** (1988) 1161.

[130] ATLAS collaboration, *Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s}$ = 7 TeV using 4.5 fb^{-1} of data with the ATLAS detector*, *JHEP* **02** (2015) 153 [1410.8857].

[131] CMS collaboration, *Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector, Phys.Rev.* **D87** (2013) 112002 [1212.6660].

[132] ATLAS COLLABORATION collaboration, *Measurement of dijet cross sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector, JHEP* **1405** (2014) 059 [1312.3524].

[133] CMS collaboration, *Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV and constraints on parton distribution functions, Eur. Phys. J.* **C77** (2017) 746 [1705.02628].

[134] T.-J. Hou et al., *Progress in the CTEQ-TEA NNLO global QCD analysis*, 1908.11394.

[135] CMS collaboration, *Constraints on parton distribution functions and extraction of the strong coupling constant from the inclusive jet cross section in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J.* **C75** (2015) 288 [1410.6765].

[136] ATLAS collaboration, *Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP* **05** (2018) 195 [1711.02692].

[137] CMS collaboration, *Measurement of the double-differential inclusive jet cross section in proton–proton collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J.* **C76** (2016) 451 [1605.04436].

[138] CMS collaboration, *Measurements of the differential jet cross section as a function of the jet mass in dijet events from proton-proton collisions at $\sqrt{s} = 13$ TeV, JHEP* **11** (2018) 113 [1807.05974].

[139] CMS collaboration, *Dependence of inclusive jet production on the anti-k_T distance parameter in pp collisions at $\sqrt{s} = 13$ TeV*, 2005.05159.

[140] ATLAS collaboration, *Measurement of three-jet production cross-sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector, Eur. Phys. J.* **C75** (2015) 228 [1411.1855].

[141] ATLAS collaboration, *Measurement of four-jet differential cross sections in $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS detector, JHEP* **12** (2015) 105 [1509.07335].

[142] CMS collaboration, *Measurement of the inclusive 3-jet production differential cross section in proton–proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range, Eur. Phys. J.* **C75** (2015) 186 [1412.1633].

[143] E. R. Nocera and M. Ubiali, *Constraining the gluon PDF at large x with LHC data, PoS* **DIS2017** (2018) 008 [1709.09690].

[144] D. Britzger et al., *Calculations for deep inelastic scattering using fast interpolation grid techniques at NNLO in QCD and the extraction of α_s from HERA data, Eur. Phys. J.* **C79** (2019) 845 [1906.05303].

[145] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, *Eur.Phys.J.* **C66** (2010) 503 [0911.2985].

[146] FASTNLO collaboration, *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*, 1109.1310.

[147] V. Bertone, S. Carrazza and N. P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, *Comput. Phys. Commun.* **212** (2017) 205 [1605.02070].

[148] 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** (2019) 102001 [1905.09047].

[149] A. Gehrmann-De Ridder, T. Gehrmann, N. Glover, A. Huss and T. A. Morgan, *NNLO QCD corrections for Z boson plus jet production*, *PoS* **RADCOR2015** (2016) 075 [1601.04569].

[150] S. Dittmaier, A. Huss and C. Speckner, *Weak radiative corrections to dijet production at hadron colliders*, *JHEP* **1211** (2012) 095 [1210.0438].

[151] NNPDF collaboration, *A first determination of parton distributions with theoretical uncertainties*, *Eur. Phys. J.* **C** (2019) 79:838 [1905.04311].

[152] NNPDF collaboration, *Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies*, *Eur. Phys. J. C* **79** (2019) 931 [1906.10698].

[153] THE NNPDF collaboration, *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* **B809** (2009) 1 [0808.1231].

[154] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [1410.8849].

[155] S. Carrazza, *Parton distribution functions with QED corrections*, Ph.D. thesis, Milan U., 2015. 1509.00209.

[156] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015.

[157] THE NNPDF collaboration, *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, *JHEP* **05** (2010) 075 [0912.2276].

[158] S. Carrazza and J. Cruz-Martinez, *Towards a new generation of parton densities with deep learning models*, *Eur. Phys. J. C* **79** (2019) 676 [1907.05075].

[159] F. Chollet et al., "Keras." https://keras.io, 2015.

[160] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015.

[161] S. Alekhin et al., *The PDF4LHC Working Group Interim Report*, 1101.0536.

[162] Z. Kassabov, "Reportengine: A framework for declarative data analysis."
https://doi.org/10.5281/zenodo.2571601, Feb., 2019. 10.5281/zenodo.2571601.

[163] FASTNLO collaboration, *New features in version 2 of the fastNLO project*, in *20th International Workshop on Deep-Inelastic Scattering and Related Subjects*, pp. 217–221, 2012, 1208.3641, DOI.

[164] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images – deep learning edition*, *JHEP* **1607** (2016) 069 [1511.05190].

[165] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet substructure classification in high-energy physics with deep neural networks*, *Phys. Rev. D* **93** (2016) 094034 [1603.09349].

[166] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **1701** (2017) 110 [1612.01551].

[167] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **1507** (2015) 086 [1501.05968].

[168] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning Top Taggers or The End of QCD?*, *JHEP* **1705** (2017) 006 [1701.08784].

[169] G. Kasieczka, T. Plehn et al., *The Machine Learning Landscape of Top Taggers*, *SciPost Phys.* **7** (2019) 014 [1902.09914].

[170] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Physical Review D* **101** (2020) [1902.08570].

[171] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Periwal, M. Pierini, A. Serikova, M. Spiropulu and J.-R. Vlimant, *JEDI-net: a jet identification algorithm based on interaction networks*, *Eur. Phys. J.* **C80** (2020) 58 [1908.05318].

[172] CMS Collaboration, *Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector*, *CMS-DP-2018-058* (2018) .

[173] G. Louppe, M. Kagan and K. Cranmer, *Learning to Pivot with Adversarial Networks*, in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds., pp. 981–990, Curran Associates, Inc., (2017), 1611.01046, http://papers.nips.cc/paper/6699-learning-to-pivot-with-adversarial-networks.pdf.

[174] G. Kasieczka and D. Shih, *DisCo Fever: Robust Networks Through Distance Correlation*, 2001.05310.

[175] S. Bollweg, M. Haußmann, G. Kasieczka, M. Luchmann, T. Plehn and J. Thompson, *Deep-Learning Jets with Uncertainties and More*, *SciPost Phys.* **8** (2020) 006 [1904.10004].

[176] G. Kasieczka, M. Luchmann, F. Otterpohl and T. Plehn, *Per-Object Systematics using Deep-Learned Calibration*, 2003.11099.

[177] B. Nachman, *A guide for deploying Deep Learning in LHC searches: How to achieve optimality and account for uncertainty*, 1909.03081.

[178] M. Dasgupta, A. Fregoso, S. Marzani and G. P. Salam, *Towards an understanding of jet substructure*, *JHEP* **1309** (2013) 029 [1307.0007].

[179] M. Dasgupta, A. Fregoso, S. Marzani and A. Powling, *Jet substructure with analytical methods*, *Eur.Phys.J.* **C73** (2013) 2623 [1307.0013].

[180] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **1405** (2014) 146 [1402.2657].

[181] M. Dasgupta, A. Powling, L. Schunk and G. Soyez, *Improved jet substructure methods: Y-splitter and variants with grooming*, *JHEP* **12** (2016) 079 [1609.07149].

[182] M. Dasgupta, A. Powling and A. Siodmok, *On jet substructure methods for signal jets*, *JHEP* **08** (2015) 079 [1503.01088].

[183] A. J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, *JHEP* **12** (2014) 009 [1409.6298].

[184] A. J. Larkoski, I. Moult and D. Neill, *Analytic Boosted Boson Discrimination*, *JHEP* **05** (2016) 117 [1507.03018].

[185] A. J. Larkoski, *Improving the Understanding of Jet Grooming in Perturbation Theory*, 2006.14680.

[186] A. J. Larkoski, G. P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [1305.0007].

[187] Z.-B. Kang, K. Lee, X. Liu, D. Neill and F. Ringer, *The soft drop groomed jet radius at NLL*, *JHEP* **02** (2020) 054 [1908.01783].

[188] P. Cal, D. Neill, F. Ringer and W. J. Waalewijn, *Calculating the angle between jet axes*, *JHEP* **04** (2020) 211 [1911.06840].

[189] F. A. Dreyer, G. P. Salam and G. Soyez, *The Lund Jet Plane*, *JHEP* **12** (2018) 064 [1807.04758].

[190] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, *Phil. Trans. Roy. Soc. Lond.* **A231** (1933) 289.

[191] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

[192] A. J. Larkoski, D. Neill and J. Thaler, *Jet Shapes with the Broadening Axis*, *JHEP* **1404** (2014) 017 [1401.2158].

[193] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [`1108.2701`].

[194] M. Skrzypek and S. Jadach, *Exact and approximate solutions for the electron nonsinglet structure function in QED*, *Z. Phys.* **C49** (1991) 577.

[195] M. Skrzypek, *Leading logarithmic calculations of QED corrections at LEP*, *Acta Phys. Polon.* **B23** (1992) 135.

[196] M. Cacciari, A. Deandrea, G. Montagna and O. Nicrosini, *QED structure functions: A Systematic approach*, *Europhys. Lett.* **17** (1992) 123.

[197] S. Frixione, *A General approach to jet cross-sections in QCD*, *Nucl. Phys. B* **507** (1997) 295 [`hep-ph/9706545`].

[198] B. Mele and P. Nason, *The Fragmentation function for heavy quarks in QCD*, *Nucl. Phys.* **B361** (1991) 626 [Erratum: Nucl. Phys.B921,841(2017)].

[199] W. Furmanski and R. Petronzio, *Lepton - Hadron Processes Beyond Leading Order in Quantum Chromodynamics*, *Z. Phys.* **C11** (1982) 293.

[200] W. Beenakker et al., *WW cross-sections and distributions*, in *CERN Workshop on LEP2 Physics (followed by 2nd meeting, 15-16 Jun 1995 and 3rd meeting 2-3 Nov 1995) Geneva, Switzerland, February 2-3, 1995*, pp. 79–139, 1996, `hep-ph/9602351`.

[201] G. Altarelli, *Partons in Quantum Chromodynamics*, *Phys. Rept.* **81** (1982) 1.

[202] W. Magnus, *On the exponential solution of differential equations for a linear operator*, *Commun. Pure Appl. Math.* **7** (1954) 649.

[203] S. Blanes, F. Casas, J. Oteo and J. Ros, *The magnus expansion and some of its applications*, *Physics Reports* **470** (2009) 151–238.

[204] E. A. Kuraev and V. S. Fadin, *On Radiative Corrections to e^+e^- Single Photon Annihilation at High-Energy*, *Sov. J. Nucl. Phys.* **41** (1985) 466 [Yad. Fiz.41,733(1985)].

[205] O. Nicrosini and L. Trentadue, *Soft Photons and Second Order Radiative Corrections to $e^+e^- \to Z^0$*, *Phys. Lett.* **B196** (1987) 551.

[206] M. Bonvini, *Resummation of soft and hard gluon radiation in perturbative QCD*, Ph.D. thesis, Genoa U., 2012. `1212.0480`. `http://www.infn.it/thesis/thesis_dettaglio.php?tid=6613`.

[207] NNPDF collaboration, *Neural network determination of parton distributions: The Nonsinglet case*, *JHEP* **03** (2007) 039 [`hep-ph/0701127`].

[208] G. Altarelli, R. D. Ball and S. Forte, *Small x Resummation with Quarks: Deep-Inelastic Scattering*, *Nucl.Phys.* **B799** (2008) 199 [`0802.0032`].

[209] M. Ciafaloni, D. Colferai, G. Salam and A. Stasto, *Renormalization group improved small x Green's function*, *Phys.Rev.* **D68** (2003) 114003 [`hep-ph/0307188`].

[210] D. de Florian, G. F. R. Sborlini and G. Rodrigo, *Two-loop QED corrections to the Altarelli-Parisi splitting functions*, *JHEP* **10** (2016) 056 [`1606.02887`].