

Bayesian Parameter Estimation of the Energy-loss Distribution of QCD Jets in a Quark-Gluon Plasma

Master Thesis in Theoretical Atomic, Nuclear and Particle Physics

by

Mathias Methlie Nilsen

June, 2022



Department of Physics and Technology

University of Bergen

Abstract

In the field of heavy ion collisions, jets are often used as probes of the produced quark-gluon plasma. In this thesis the energy-loss distribution of quark/gluon tagged jets are given a log-normal parametrization. The motivation for this is the study of single parton traveling through a uniform QGP based on the BDMPS spectrum, which is calculated numerically. Bayesian inference is then used to constrain the parameters of the distributions. The data used in the Bayesian analysis are measurements of the nuclear modification factor from ATLAS in lead-lead collisions at $\sqrt{s} = 5.02$ TeV.

The results of the Bayesian analysis fits well with the data, and it predicts well the p_T dependence of the nuclear modification factor when validated on a different dataset from ATLAS at $\sqrt{s} = 2.76$ TeV. However, the model overestimates the amount of jet suppression at this energy. The uncertainty of the R_{AA} data points for p_T above 300 GeV are very large, which makes for a poor constraint on the quark-jet energy-loss distribution. Therefore, constraining the parameters using data for γ -jet events are proposed as alternative approach for further work.

Contents

Introduction	5
1 QCD and Jets	6
1.1 Lie Groups and Color Factors	6
1.2 The QCD Lagrangian	8
1.3 The Gluon Splitting Function	10
1.4 Angular Ordering	13
1.5 Jets in Proton-Proton Collisions	14
1.5.1 Factorization of the Cross Section	15
1.5.2 A Simple Jet Definition	16
1.6 Parton Evolution and Jet Simulation	16
2 Heavy-Ion Collisions and Parton Energy-loss	19
2.1 Basic Terminology and Setup	20
2.2 Signals of Quark-Gluon Plasma	21
2.2.1 Elliptic Flow	21
2.2.2 Jet Quenching	22
2.3 The BDMPS Spectrum	23
2.4 The Poisson Approximation	26
3 Bayesian Inference and MCMC	31
3.1 Bayes Theorem	31
3.2 Posterior, whats the big deal?	32
3.3 Markov Chain Monte Carlo	33
3.4 The Metropolis-Hastings Algorithm	35
3.5 Hamiltonian Monte Carlo	38
3.5.1 Hamiltonian Dynamics	38
3.5.2 The HMC Algorithm	41
4 Jet Quenching Using Bayesian Inference	44
4.1 Re-deriving Earlier Results Using HMC	45
4.2 Quenching of Parton-tagged Jets	49
4.2.1 Classifying Jets in Pythia	49
4.2.2 Nuclear PDFs	51
4.2.3 Log-normal Energy-loss Distributions	52
4.2.4 Model Validation	56

Summary and Outlook	60
Acknowledgement	61
A Numerical Integration of the Convolution-integral	62
A.1 Gamma Distribution	62
A.2 Log-normal Distribution	63
Bibliography	65

Introduction

Quarks and gluons are the constituents of hadrons (protons, neutrons, etc..). In most conditions in the universe they are confined inside of the hadrons. However, at high temperatures or high baryon densities, the partons (quarks and gluons) become deconfined. At these conditions ordinary hadronic matter does not exist, and the quarks and gluons become the main actors of this exotic superdense matter dubbed "quark-gluon plasma" (QGP). This state of plasma is theorized to exist in the interior of neutron stars and at the earliest moments of our universe. In the last decades, evidence from particle colliders such as LHC at CERN and RICH at Brookhaven National Laboratory, suggest the formation of a QGP in high energetic heavy ion collisions [16]. The field of jet quenching uses high energetic objects called jets as probes of the QGP, and this is also the approach taken in this thesis.

Quantum Chromodynamics (QCD) is the theory that describes the interaction of quarks and gluons, and their behavior. Chapter 1 in this thesis contains a short introduction to QCD. It starts by a basic discussion of Lie groups and the fundamental and adjoint representation. Then it describes the QCD Lagrangian and the most relevant Feynman rules. Then a somewhat lengthy derivation of the famous gluon splitting function is shown. The next topic that is covered are jets and how they arise from the underlying theory. The phenomenon of coherent branching is introduced. A short overview of hard proton-proton collisions and the basic dictionary used in jet physics such as transverse momentum and rapidity is given.

In Chapter 2 we take a look at heavy ion collisions and the basic dictionary used in this field. The different observable signals indicating the formation of QGP, such as the nuclear modification factor and elliptic flow are then covered. We then move to the study of a single colored parton traveling through a uniform brick of quark-gluon plasma. The key ingredients of this analysis is the well known BDMPS medium induced gluon spectrum [7]. The low and high energy behavior of this spectrum is looked at to get an intuition of it. We then introduce the energy-loss distribution, and the Poisson approximation which is used to connect the BDMPS spectrum to the overall energy-loss of parton traversing the QGP. We use this result to motivate a parametrization of the energy-loss distribution of a jet. The idea is to use Bayesian inference on the parameters of this distribution to fit it to measured data.

Therefore, Chapter 3 is dedicated to introducing the famous statistical learning method "Bayesian inference". We start by explaining Bayes rule, which is a rule for updating ones belief after taking the experimental data in to consideration. The result of using Bayes rule is the so called "posterior distribution". We demonstrate the importance of the posterior and how it is used to predict new data (or other observables). However, both using and obtaining the posterior is not a trivial task. We will see that a work around way to do Bayesian inference is via MCMC meth-

ods. The class of Metropolis-Hastings algorithms is then introduce, and the common "Random Walk Metropolis" (RWM) method is demonstrated. The weakness of RWM is then discussed, before moving over to a more dimensionally robust method called Hamiltonian Monte Carlo (HMC). Which we will use in our Bayesian analysis of jet quenching (suppression of jets in heavy ion collisions).

Finally in Chapter 4 we apply Bayesian inference on jet quenching using data from the ATLAS collaboration. First we re-derive the result of [17] while demonstrating the efficiency of the HMC method. We then move over to use the parametrization obtained in Chapter 2 to make a model where we distinguish between quark and gluon initiated jets. The spectra of quark and gluon jets are generated in the event generator Pythia, and the process for this is described. The results of the Bayesian analysis is then represented and discussed. The developed model is then validated on a different data set to see how well it generalizes.

All the code developed to produce the results in this thesis is publicly available at GitHub [20]. Interested readers are welcome to download the repository and play around with it. There should be enough info in the GitHub repository for the reader to understand which scripts to run. The repository is divided into chapters corresponding to the same chapters as in this thesis.

Chapter 1

QCD and Jets

Quantum chromodynamics is the theory describing the interaction of quarks and gluons, which can be found inside hadrons. The development of QCD started back in the 1960s when it was theorized that hadrons contains smaller particles. It was proposed that these particles possessed an internal degree of freedom called color. The existence of these so called partons was confirmed experimentally in deep inelastic scattering, and evidence for the color charge of the quarks came from jets of e^+e^- collisions. The goal of this chapter is to introduce the basics of QCD and jets.

1.1 Lie Groups and Color Factors

The QCD Lagrangian (which we will look at in the next section) is invariant under the Lie group $SU(3)$. This begs the question, what is a Lie group? This section gives a short introduction to group theory and the different representations of the group. This leads to the Casimir operators (color factors) which almost always show up in QCD calculations. This small overview over group theory follows from Section 10.1 and 25.1 in [23].

A group is a set of element $\{g_i\}$, and it is defined by a multiplication rule $g_i \times g_j = g_k$, where g_k is also a an element in the group. The rule \times satisfies the following conditions

- Associativity: $(g_i \times g_j) \times g_k = g_i \times (g_j \times g_k)$
- There must exist an identity element \mathbb{I} such that $\mathbb{I} \times g_i = g_i \times \mathbb{I} = g_i$
- For every g_i , an inverse g_i^{-1} exist such that $g_i^{-1} \times g_i = \mathbb{I}$

A Lie group is then a group with an infinite number of elements, which also is a differentiable manifold. Any element in the group can be written as

$$U = e^{i\theta^a T^a} \mathbb{I}, \quad (1.1)$$

where summation over a is implied by the summation convention. The numbers θ^a characterize the given group element. The quantities T^a are called the group generators, and they satisfy the so called Lie algebra

$$[T^a, T^b] = i f^{abc} T^c, \quad (1.2)$$

where the quantity f^{abc} are the so called structure constants, which are anti-symmetric in all indices. One makes the distinction of an Abelian group where $f^{abc} = 0$, and a non-Abelian group where f^{abc} is none zero. The symmetry group SU(3) is an non-Abelian Lie group.

We will now take a look at the two most important representation of the SU(N) groups. A representation of a given Lie group is constructed by embedding the generators into matrices. In the fundamental representation of SU(N) the generators are represented by $N \times N$ Hermitian matrices. The infinitesimal transformation of N fields ψ_i transforming under the fundamental representation is given by

$$\psi_i \rightarrow \psi_i + i\alpha^a(T_{\text{fund}}^a)_{ij}\psi_j. \quad (1.3)$$

In the case of QCD and SU(3) the matrices T_{fund}^a are commonly written in as $T_{\text{fund}}^a = \lambda^a/2$. Where λ^a are the trace-less Gell-Mann matrices, which can be found on page 485 in [23].

$$\begin{aligned} \lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda^5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \\ \lambda^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & \lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda^8 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \frac{1}{\sqrt{3}}. \end{aligned} \quad (1.4)$$

In QCD the quark field transform under the fundamental representation, as opposed to the gluon field A_μ^a which transform under the adjoint representation. The matrices that describe the adjoint representation of SU(N) are $N^2 - 1$ dimensional, and given by the structure constants

$$T_{\text{adj}}^a = -if^{abc}. \quad (1.5)$$

Lets check that T_{adj}^a satisfies the commutation relation in Eq. (1.2).

$$\begin{aligned} [T^a, T^b]_{ce} &= T_{cd}^a T_{de}^b - T_{cd}^b T_{de}^a \\ &= i^2 f^{acd} f^{bde} - i^2 f^{bcd} f^{ade}. \end{aligned} \quad (1.6)$$

We now use the Jacobi identity $f^{bcd} f^{ade} + f^{cad} f^{bde} + f^{abd} f^{cde} = 0$ to get

$$\begin{aligned} [T^a, T^b]_{ce} &= i^2 (f^{bcd} f^{ade} + f^{abd} f^{cde}) - i^2 f^{bcd} f^{ade} \\ &= i^2 f^{abd} f^{cde} \\ &= -i^2 f^{abd} f^{dce} \\ &= if^{abd} T_{ce}^d \longrightarrow [T^a, T^b] = if^{abc} T^c. \end{aligned} \quad (1.7)$$

The infinitesimal transformation of $N^2 - 1$ fields A_μ^a , transforming under the adjoint representation is given by

$$A_\mu^a \rightarrow A_\mu^a - \alpha^b f^{abc} A_\mu^c. \quad (1.8)$$

A basis independent way of characterizing the different representations is through the so called quadratic Casimir $C_2(R)$ defiend by

$$T_R^a T_R^a = C_2(R) \mathbb{I}, \quad (1.9)$$

where R indicate the representation. The quadratic Casimir (often called color factors in QCD) for the fundamental F and adjoint A representation of $SU(N)$ is given by

$$C_F \equiv C_2(\text{fund}) = \frac{N^2 - 1}{2N}, \quad (1.10)$$

$$C_A \equiv C_2(\text{adj}) = N. \quad (1.11)$$

For $SU(3)$, this gives $C_F = 4/3$ and $C_A = 3$. As mentioned, these quantities appear in almost all calculations in QCD.

1.2 The QCD Lagrangian

We start by taking a look at the QCD Lagrangian [14] in all its beauty. The QCD Lagrangian reads

$$\mathcal{L}^{\text{QCD}} = \sum_q \bar{\psi}_q^i (i \not{D}_{ij} - m_q) \psi_q^j - \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu}, \quad (1.12)$$

where the sum over q is a sum over all the quark flavours involved. The index¹ i (and j) is the color index for the quarks, summed from 1 to 3 (or sometimes written as $i \in \{r, g, b\}$). ψ_q^i is then the spinor field for quark q with color index i . We have utilized the slashed notation in Eq. (1.12), $\not{D}_{ij} \equiv \gamma^\mu (D_\mu)_{ij}$, were $(D_\mu)_{ij}$ is the covariant derivative

$$(D_\mu)_{ij} = \delta_{ij} \partial_\mu + i g_s T_{ij}^a A_\mu^a, \quad (1.13)$$

where T^a for $a \in \{1, \dots, 8\}$, are the generators of the $SU(3)$ symmetry group, which satisfying Eq. (1.2). A table of the structure constants f^{abc} is given in Section 1.3 in [25]. The quantity A_μ^a is the gluon field which is a spin-1 boson field. The last term in Eq. (1.12) contains the field strength tensor $F_{\mu\nu}^a$, given by

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_s f^{abc} A_\mu^b A_\nu^c. \quad (1.14)$$

The last term in Eq. (1.14) is the biggest difference between QCD and QED (which does not have this non-Abelian term). It is the source of the famous 3-gluon and 4-gluon interactions. The parameter g_s is of course the strong gauge coupling.

In order to use perturbation theory to make predictions from QCD we need to be able to define Feynman rules. To define the gluon propagator from the Lagrangian of Eq. (1.12) we have to make choice of gauge. We can then add a gauge fixing term with a parameter ξ (see Section 25.4 in [23] for a derivation of the gluon propagator) to the Lagrangian

$$\mathcal{L} = \mathcal{L}^{\text{QCD}} - \frac{1}{2\xi} (\partial^\mu A_\mu^a)^2. \quad (1.15)$$

¹ Not to be confused with the prefactor i in front of \not{D} which is the imaginary number $\sqrt{-1}$

In a non-Abelian gauge theory such as QCD, another term must be added to the the gauge fixing term. This is the famous ghost Lagrangian containing the complex scalar fields called ghost fields c^i , which obeys Fermi statistics. The Lagrangian the becomes

$$\mathcal{L} = \mathcal{L}^{\text{QCD}} - \frac{1}{2\xi} (\partial^\mu A_\mu^a)^2 - \bar{c}^i \not{D}_{ij} c^j. \quad (1.16)$$

The ghost fields only appear in higher order diagrams. They are present to cancel out unphysical degrees of freedom in the covariant gauges parametrized by ξ in Eq. (1.15). Using the Lagrangian in Eq. (1.16) the gluon propagator can be defined,

$$a, \nu \xrightarrow[p]{} b, \mu \longrightarrow \left[-g^{\mu\nu} + (1 - \xi) \frac{p^\mu p^\nu}{p^2} \right] \frac{i\delta^{ab}}{p^2 + i\varepsilon}. \quad (1.17)$$

One of the most common covariant gauges to use is the Feynman gauge, where $\xi = 1$. This choice makes the gluon propagator very simple. However, we still have to take into account ghost in higher order diagrams. There exist a class of non-covariant gauges called axial gauges, where the gauge is fixed in terms of a additional 4-vector n ,

$$\mathcal{L}^{\text{gauge-fixing}} = -\frac{1}{2\xi} (n^\mu A_\mu^a)^2. \quad (1.18)$$

The gluon propagator in this gauge reads

$$\Pi^{\mu\nu} = \left[-g^{\mu\nu} + \frac{n^\mu p^\nu + p^\mu n^\nu}{np} - \frac{(n^2 + \xi p^2)p^\mu p^\nu}{(np)^2} \right] \frac{i\delta^{ab}}{p^2 + i\varepsilon}. \quad (1.19)$$

The axial gauge has the nice property that for on-shell ($p^2 = 0$) gluons, $p_\mu \Pi^{\mu\nu} = 0$. Choosing $\xi = 0$, the ghost fields will decouple from the gluon field, and the need for them vanishes. With the choice of $n^2 = 0$ we arrive at the light-cone gauge. The gluon propagator then only contains the two first terms in Eq. (1.19). In the light-cone gauge only two polarization states propagate, those transverse to n , because $n_\mu \Pi^{\mu\nu} = 0$, and those transverse to p . The numerator of the propagator (which in Feynman gauge can be decomposed into a sum over 4 polarizations) can be written as a sum over 2 physical polarizations for the gluon.

$$d^{\mu\nu} \equiv -g^{\mu\nu} + \frac{n^\mu p^\nu + p^\mu n^\nu}{np} = \sum_{i=1}^2 \varepsilon_i^\mu \varepsilon_i^\nu \quad (1.20)$$

This small overview of gauges is taken from Section 1.5 in [14] and Section 25.4.3 in [23]

As mentioned a big difference of QED and QCD is that QCD has pure bosonic interactions, the 3 and 4 gluon vertices. These appear in the last term of Eq. (1.12) as a consequence of the gluon also carrying color charge (as opposed to the photon who as zero charge). Using Eq. (1.14), we will find that the term in the Lagrangian containing the 3 gluon interaction is

$$\mathcal{L}^{\text{3-gluon}} = g f^{abc} (\partial_\mu A_\nu^a) A_\mu^b A_\nu^c, \quad (1.21)$$

sum over a, b, c and μ, ν implied. The Feynman rule for 3-gluon vertex is [14]

$$a, \mu \quad \quad \quad b, \nu \\ p \quad \quad \quad q \\ \quad \quad \quad r \\ \quad \quad \quad c, \sigma$$

$$\longrightarrow -gf^{abc}[(p-q)^\sigma g^{\mu\nu} + (q-r)^\mu g^{\nu\sigma} + (r-p)^\nu g^{\mu\sigma}], \quad (1.22)$$

where all the momenta are defined as incoming to the vertex. The Feynman rule for the 4-gluon vertex is given by

$$a, \mu \quad \quad \quad b, \nu \\ p \quad \quad \quad q \\ \quad \quad \quad r \\ \quad \quad \quad s \\ c, \rho \quad \quad \quad d, \sigma$$

$$\longrightarrow -ig^2[f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}) \\ \quad \quad f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma}) \\ \quad \quad f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})]. \quad (1.23)$$

Each vertex comes with a conservation of 4-momentum. The 4-gluon vertex do not appear in first order calculations due to the g^2 factor. The rest of the Feynman rules can be found on page 10 of [14]. In the next section we will use the Feynman rules to derive the well known gluon splitting function $\hat{P}_{gg}(z)$.

1.3 The Gluon Splitting Function

Understanding the process of parton branching is crucial for understanding how jets (showers of hadrons) form. In this section we will look at the process of gluon branching, and derive the gluon splitting function $\hat{P}_{gg}(z)$. We will follow closely the approach taking in Section 5.1 in [14]. The Feynman diagram for such a process is

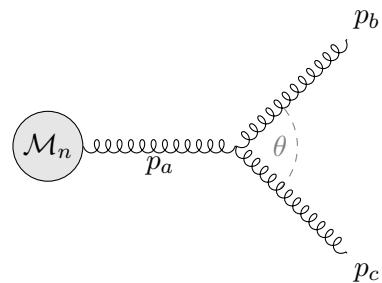


Figure 1.1: Feynman diagram for the process of a gluon branching from \mathcal{M}_n

We start by looking at the kinematics of Fig 1.1, where we assume that $p_c^2, p_b^2 \ll p_a^2 \equiv t > 0$. Defining the energy fraction $z = E_b/E_a = 1 - E_c/E_a$ which is fractional energy carried by gluon

b. All momenta are defined as outgoing from the vertex. Assuming p_b and p_c to be on-shell, we then have

$$\begin{aligned} t &= p_a^2 = (-p_b - p_c)^2 \\ &= p_b^2 + 2p_b \cdot p_c + p_c^2 = 2(p_b^0 p_c^0 - \vec{p}_b \cdot \vec{p}_c) \\ &= 2E_b E_c (1 - \cos \theta) \stackrel{\text{small}\theta}{\approx} z(1-z) E_a^2 \theta^2, \end{aligned} \quad (1.24)$$

where θ is the splitting angle between gluon b and c . We define the coordinate system such that the 4-momentum of the gluon b and c are given by

$$\begin{aligned} p_b &= [1, \sin \theta_b, 0, \cos \theta_b] \cdot z E_a, \\ p_c &= [1, -\sin \theta_c, 0, \cos \theta_c] \cdot (1-z) E_a, \end{aligned} \quad (1.25)$$

where $\theta = \theta_b + \theta_c$. We can now start by looking at the amplitude for this process

$$\mathcal{M}_{n+1} = igf^{ABC} \frac{1}{t} [g_{\mu\nu}(p_a - p_b)_\sigma + g_{\nu\sigma}(p_b - p_c)_\mu + g_{\sigma\mu}(p_c - p_a)_\nu] \varepsilon_a^\mu \varepsilon_b^\nu \varepsilon_c^\sigma \mathcal{M}_n, \quad (1.26)$$

where we change the color index of the gluons to capital letters (A , B and C) to distinguish them from the particle label a , b and c . The factor $\frac{1}{t}$ comes from the propagator, and ε_a comes from the metric tenors in the propagator which can be written as $g^{\mu\rho} = \sum_r \varepsilon_r^\mu \varepsilon_r^\rho$ (for now we will drop the sum). We then then impose that $\varepsilon_i \cdot p_i = 0$, for $i \in \{a, b, c\}$. Combining this with the fact that $p_a = -p_b - p_c$, the amplitude then becomes

$$\mathcal{M}_{n+1} = -2igf^{ABC} \frac{1}{t} [(\varepsilon_a \cdot \varepsilon_b)(\varepsilon_c \cdot p_b) - (\varepsilon_b \cdot \varepsilon_c)(\varepsilon_a \cdot p_b) - (\varepsilon_c \cdot \varepsilon_a)(\varepsilon_b \cdot p_c)] \mathcal{M}_n. \quad (1.27)$$

Assuming on-shell gluons, the polarization vectors are then transverse to the propagation of the gluon. Choosing the polarization's to be either in or out of the plane of the splitting gives

$$\begin{aligned} \varepsilon_i^{\text{in}} \cdot \varepsilon_j^{\text{in}} &= \varepsilon_i^{\text{out}} \cdot \varepsilon_j^{\text{out}} = -1, \\ \varepsilon_i^{\text{in}} \cdot \varepsilon_j^{\text{out}} &= \varepsilon_i^{\text{out}} \cdot p_j = 0. \end{aligned} \quad (1.28)$$

The only non-zero combinations of polarization will be:

ε_a	ε_b	ε_c
in	in	in
in	out	out
out	in	out
out	out	in

Lets look at the first case where all polarization states are in the plane of splitting

$$\begin{aligned} \mathcal{M}_{n+1} &= -2igf^{ABC} \frac{1}{t} [(\varepsilon_a^{\text{in}} \cdot \varepsilon_b^{\text{in}})(\varepsilon_c^{\text{in}} \cdot p_b) - (\varepsilon_b^{\text{in}} \cdot \varepsilon_c^{\text{in}})(\varepsilon_a^{\text{in}} \cdot p_b) - (\varepsilon_c^{\text{in}} \cdot \varepsilon_a^{\text{in}})(\varepsilon_b^{\text{in}} \cdot p_c)] \mathcal{M}_n \\ &= -2igf^{ABC} \frac{1}{t} [-\varepsilon_c^{\text{in}} \cdot p_b + \varepsilon_a^{\text{in}} \cdot p_b + \varepsilon_b^{\text{in}} \cdot p_c] \mathcal{M}_n. \end{aligned} \quad (1.29)$$

Given Eq. (1.28), we have

$$\begin{aligned}\varepsilon_a^{\text{in}} &= [0, 1, 0, 0], \\ \varepsilon_b^{\text{in}} &= [0, \cos \theta_b, 0, -\sin \theta_b], \\ \varepsilon_c^{\text{in}} &= [0, \cos \theta_c, 0, \sin \theta_c].\end{aligned}\tag{1.30}$$

Using the small angle approximation, Eq. (1.29) then becomes

$$\begin{aligned}\mathcal{M}_{n+1} &= -2igf^{ABC} \frac{1}{t} [zE_a\theta - z(1-z)E_a\theta + (1-z)E_a\theta] \mathcal{M}_n \\ &= -2igf^{ABC} \frac{1}{t} [z^2 - z + 1] E_a\theta \mathcal{M}_n.\end{aligned}\tag{1.31}$$

Squaring the amplitude, and summing over gluon colors gives

$$|\mathcal{M}_{n+1}|^2 = \frac{4g^2C_A}{t} \frac{[z^2 - z + 1]^2}{z(1-z)} |\mathcal{M}_n|^2.\tag{1.32}$$

After some algebraic gymnastics this can be rewritten as

$$|\mathcal{M}_{n+1}|^2 = \frac{4g^2C_A}{t} \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right] |\mathcal{M}_n|^2.\tag{1.33}$$

We skip the calculations of the other combinations here because it is only more of the same. The results for all the combinations are:

$$\varepsilon_a^{\text{in}}, \varepsilon_b^{\text{in}}, \varepsilon_c^{\text{in}} : \quad |\mathcal{M}_{n+1}|^2 = \frac{4g^2C_A}{t} \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right] |\mathcal{M}_n|^2\tag{1.34}$$

$$\varepsilon_a^{\text{in}}, \varepsilon_b^{\text{out}}, \varepsilon_c^{\text{out}} : \quad |\mathcal{M}_{n+1}|^2 = \frac{4g^2C_A}{t} [z(1-z)] |\mathcal{M}_n|^2\tag{1.35}$$

$$\varepsilon_a^{\text{out}}, \varepsilon_b^{\text{in}}, \varepsilon_c^{\text{out}} : \quad |\mathcal{M}_{n+1}|^2 = \frac{4g^2C_A}{t} \left[\frac{1-z}{z} \right] |\mathcal{M}_n|^2\tag{1.36}$$

$$\varepsilon_a^{\text{out}}, \varepsilon_b^{\text{out}}, \varepsilon_c^{\text{in}} : \quad |\mathcal{M}_{n+1}|^2 = \frac{4g^2C_A}{t} \left[\frac{z}{1-z} \right] |\mathcal{M}_n|^2.\tag{1.37}$$

To get the un-polarized squared amplitude we now sum over all the polarization vectors. Recall that we already have a sum over ε_a from the propagator.

$$|\mathcal{M}_{n+1}|_{\text{un-pol}}^2 = \sum_{\text{pol}} |\mathcal{M}_{n+1}|^2 = \frac{4g^2C_A}{t} 2 \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right] \equiv \frac{4g^2}{t} \hat{P}_{gg}(z) |\mathcal{M}_n|^2,\tag{1.38}$$

where $\hat{P}_{gg}(z)$ is the well known gluon splitting function,

$$\hat{P}_{gg}(z) = 2C_A \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right] = C_A \frac{[1+z^4+(1-z)^4]}{z(1-z)}.\tag{1.39}$$

We see that there is an enhancement in $\hat{P}_{gg}(z)$ when $z \rightarrow 0$ (gluon b is soft) and $z \rightarrow 1$ (gluon c is soft). Therefore, in the limit of a soft emission, the gluon splitting function can be approximated by $\hat{P}_{gg}(z) = 2C_A/z$. This soft enhancement is when the polarization of the soft gluon is emitted in the splitting plane. One can also derive splitting functions for the following processes

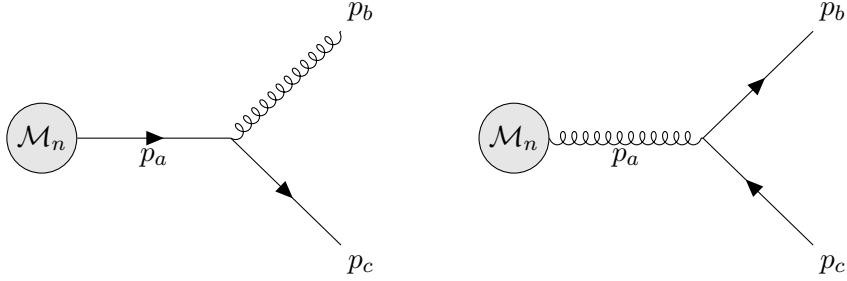


Figure 1.2: Feynmann diagrams for the branching process $q \rightarrow g + q$ (left) and $g \rightarrow \bar{q} + q$ (right)

We will not explicitly do the derivations for the amplitudes in Fig 1.2 here. The main result is that the amplitude squared will follow the same structure as the last expression in Eq. (1.38)

$$|\mathcal{M}_{n+1}|_{\text{un-pol}}^2 = \frac{4g^2}{t} \hat{P}(z) |\mathcal{M}_n|^2, \quad (1.40)$$

where

$$\hat{P}_{gq}(z) = C_F \frac{1 + (1-z)^2}{z}, \quad (1.41)$$

$$\hat{P}_{qg}(z) = T_R [z^2 + (1-z)^2], \quad (1.42)$$

where $T_R = 1/2$. The enhancement in $\hat{P}_{gq}(z)$ is again associated with the emission of soft gluon ($z \rightarrow 0$) in the splitting plane, and the splitting function can again be well approximated (in the soft limit) by $\hat{P}_{gq}(z) = 2C_F/z$. The splitting functions is an important factor in the cross section $d\sigma_{n+1}$ which has the rather simple structure (Section 5.1 in [14])

$$d\sigma_{n+1} = d\sigma_n \frac{dt}{t} dz \frac{\alpha_s}{2\pi} \hat{P}_{ba}(z). \quad (1.43)$$

Using $dt = 2z(1-z)E_a^2\theta d\theta$ for a given z this can then be written as

$$d\sigma_{n+1} = d\sigma_n \frac{\alpha_s}{\pi} \frac{d\theta}{\theta} dz \hat{P}_{ba}(z). \quad (1.44)$$

1.4 Angular Ordering

What is the motivation for the variable change from Eq. (1.43) to Eq. (1.44)? In the collinear splitting functions in the last section we saw that there is an soft enhancement whenever a propagating particle emits a gluon, lets say a quark emits a gluon. However this enhancement is more general than that. The effect we will now present is called angular ordering, and we will follow closely Section 5.5 in [14]. Lets consider the process of $e^+e^- \rightarrow q + \bar{q}$, where one of the quark emit a soft gluon.

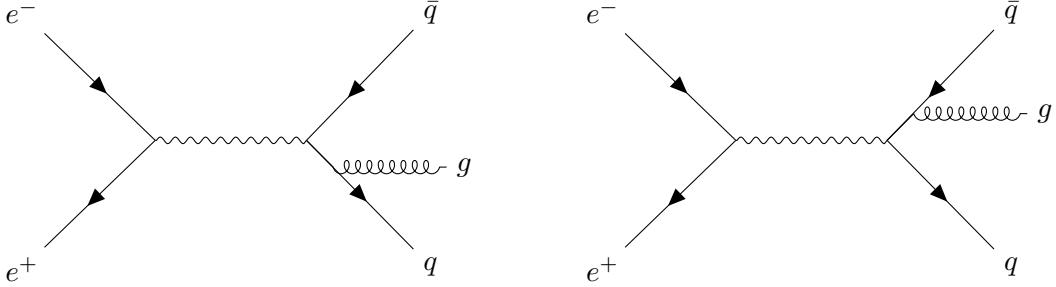


Figure 1.3: $\mathcal{O}(\alpha_s)$ Feynman diagrams for the process $e^+e^- \rightarrow q\bar{q}g$

The cross section $d\sigma_{n+1}$ can then be written as a sum over the two external quark legs

$$d\sigma_{n+1} = d\sigma \frac{d\omega}{\omega} \frac{d\Omega}{2\pi} \frac{\alpha_s}{2\pi} C_F \sum_{i,j \in \{q,\bar{q}\}} W_{i,j}, \quad (1.45)$$

where

$$W_{ij} = \frac{1 - v_i v_j \cos \theta_{ij}}{(1 - v_i \cos \theta_{ig})(1 - v_j \cos \theta_{jg})}, \quad (1.46)$$

where ω and $d\Omega$ is the energy and the solid angle element of the emitted gluon. θ_{ij} is the splitting angle of the quark pair, and θ_{ig} is the gluon emission angle from quark i . v_j is the velocity of the quark i , which we now will approximate to 1 for simplicity. W_{ij} represent interference between the two diagrams, but it can be separated into two parts $W_{ij} = W_{ij}^{(i)} + W_{ij}^{(j)}$ where

$$W_{ij}^{(i)} = \frac{1}{2} \left[W_{ij} + \frac{1}{1 - \cos \theta_{ig}} - \frac{1}{1 - \cos \theta_{jg}} \right]. \quad (1.47)$$

Now lets do the azimuthal integration for the leg i . We then write $d\Omega = d\cos \theta_{ig} d\phi_{ig}$. The details of this integration can be found in [14].

$$\int_0^{2\pi} \frac{d\phi_{ig}}{2\pi} W_{ij}^{(i)} = \frac{1}{2(1 - \cos \theta_{ig})} \left[1 + \frac{\cos \theta_{ig} - \cos \theta_{ij}}{|\cos \theta_{ig} - \cos \theta_{ij}|} \right] = \begin{cases} \frac{1}{1 - \cos \theta_{ig}} & , \text{if } \theta_{iq} < \theta_{ij} \\ 0 & , \text{otherwise} \end{cases} \quad (1.48)$$

This a remarkable result, the gluon emission angle has to be smaller than the splitting angle of the quarks! We will see later that this can be used in branching algorithms. But first, in the next section we will look at proton-proton collisions and how jets can be formed in such events.

1.5 Jets in Proton-Proton Collisions

In this section we will take a brief overview over jet productions in proton-proton collisions. We will first take a look at the factorization of the proton-proton cross section. Then move over to look at the kinematics variables used in collider physics. The discussion of pp-collisions is taken from chapter 7 in [14].

1.5.1 Factorization of the Cross Section

Hard scattering of two protons can be described by hard scattering of quarks and gluons (partons) inside of the protons. The cross section can be written in the following way:

$$\sigma(p_1, p_2) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu^2) f_j(x_2, \mu^2) \hat{\sigma}^{i,j}(\hat{p}_1, \hat{p}_2, \alpha_s(\mu), Q/\mu), \quad (1.49)$$

where p_1 and p_2 are the four-momenta of the colliding protons, x_1 and x_2 are the momentum fractions of the interacting partons defined by $x_1 \equiv \hat{p}_1/p_1$. Q is the characteristic scale of the hard scattering. $f_i(x, \mu^2)$ are the parton distribution functions (PDFs), defined at a scale μ , which can be thought of as the scale separating the long distance physics from the short distance physics. $\hat{\sigma}^{i,j}$ is the scattering cross section of the two interaction partons. The partonic cross section can be calculated perturbatively, which makes the factorization in Eq. (1.49) pretty powerful. The structure of the factorization is visualized in Fig 1.4.

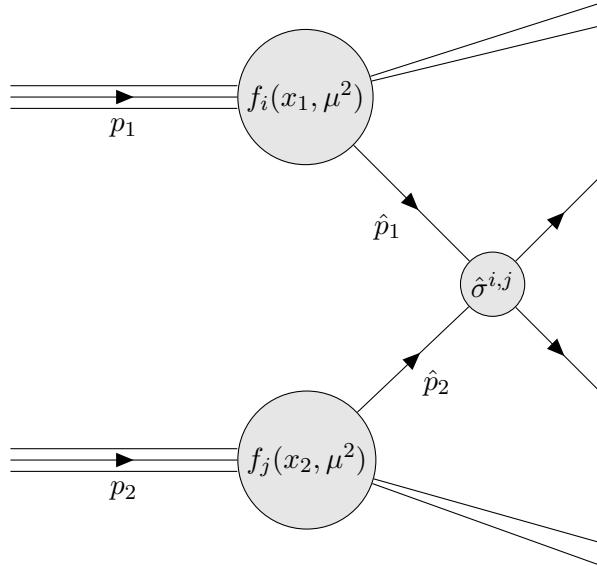


Figure 1.4: Schematic diagram of the QCD factorization given in Eq. (1.49) of a hard scattering of two protons.

It is important to note that Eq. (1.49) do not describe most pp-collisions. There is a lot of soft non-perturbative physics happening in pp-collisions. However Fig 1.4 with Eq. (1.49) is a powerful tool for describe the hard events, which are the most interesting ones.

The usual setup is that the colliding protons travel at almost the speed of light towards each other in the z -direction. This means that in the lab-frame, the protons will be very Lorentz contracted. Therefore, the four-momenta of the outgoing particles from the partonic cross section $\hat{\sigma}^{i,j}$ is often written as

$$\begin{aligned} p^\mu &= (E, p_x, p_y, p_z) \\ &= (m_T \cosh y, p_T \sin \phi, p_T \cos \phi, m_T \sinh y), \end{aligned} \quad (1.50)$$

where p_T is the transverse momentum with respect to the z -axis. $m_T \equiv \sqrt{p_T^2 + m^2}$ is called

the transverse mass. ϕ is the azimuthal angle, and y is the rapidity defined by

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \quad (1.51)$$

The advantage is that the rapidity is additive under a Lorentz boost. A more common use is pseudorapidity η defined as

$$\eta = -\ln \tan \frac{\theta}{2}, \quad (1.52)$$

where θ is the polar angle with respect to the beam. For a massless particle rapidity and pseudorapidity are the same.

1.5.2 A Simple Jet Definition

The partons coming out of the hard vertex in Fig 1.4 will then start to radiate and fragment until they hadronize at some scale. Hadronization is a process where a shower of partons are turned into a shower of hadrons (jets) which is then observed in the detector. Hadronization is not a subject to be discussed in this thesis, for an overview of the different hadronization models see Section 5.6 in [14].

Before we move over talk about parton fragmentation lets first define what we mean by a jet. Qualitatively a jet is a collimated shower of hadrons/particles. There is no single "best" definition of jets, but in hadron collisions it is common to use a definition based on a cone parameter R [14], defined by

$$R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \quad (1.53)$$

A jet is then defined as a concentration of transverse energy $E_T = E \sin \theta$ (or transverse momentum p_T) inside the cone R . Experimentally one constructs a jet by clustering together the observed particles in the detector. One of the most used clustering algorithms is the so called "anti- k_t " algorithm [11]. The Anti- k_t clustering introduces a distance measure d_{ij} , were i and j are particles (or pseudojets),

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta_{ij}^2}{R^2}, \quad (1.54)$$

$$d_{iB} = p_{Ti}^{-2}, \quad (1.55)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and y_i and ϕ_i is the pseudorapidity and azimuthal angle. p_{Ti} is the transverse momentum of particle (or pseudojet) i . The cone parameter R is an input to the algorithm. d_{iB} is the distance measure between i and the beam. The algorithm works by choosing the smallest d_{ij} , then combining i and j into a pseudojet. If d_{iB} is the smallest, i is then classified as a jet, and is taken out of the list of particles and pseudojets. The process is then repeated until there are no particles and pseudojets left, only jets.

1.6 Parton Evolution and Jet Simulation

Consider a timelike parton ($p^2 > 0$), it could for instance be one of the outgoing partons in Fig 1.4. It is likely that it will start radiate, and loose some of its momentum. The average

number of collinear emissions within a cone of angle R is given by

$$N = \frac{\alpha_s}{\pi} \int_0^1 dz \hat{P}(z) \int_0^R \frac{d\theta}{\theta} H(Q_0 - z\theta E), \quad (1.56)$$

where $H(Q_0 - z\theta E)$ is the Heaviside step function, with the perturbative condition that $z\theta E > Q_0$. E is the energy of the parton. Using the approximation that $\hat{P}(z) = 2C_R/z$ we can evaluate the integral and get

$$N = \frac{\alpha_s C_R}{2\pi} \int_0^1 \frac{dz}{z} \int_0^R \frac{d\theta}{\theta} H(Q_0 - z\theta E) = \frac{\alpha_s C_R}{\pi} \log^2 \frac{ER}{Q_0}. \quad (1.57)$$

As an example lets consider the case of a gluon, $C_R = 3$, with energy $E = 200$ GeV. Lets choose $\alpha = 0.2$, $R = 0.4$ and $Q_0 = 1$ GeV. The average number of collinear emissions is then $N \approx 3.67$. This motivates the existence of jets and jet fragmentation. The radiated particles will also probably radiate particles and so on, until we get an avalanche of particles. Recall the principle of angular ordering, radiation will happen at lower and lower angles! Taking this into consideration gives an qualitative picture of the formation of jet, shown in Fig 1.5.

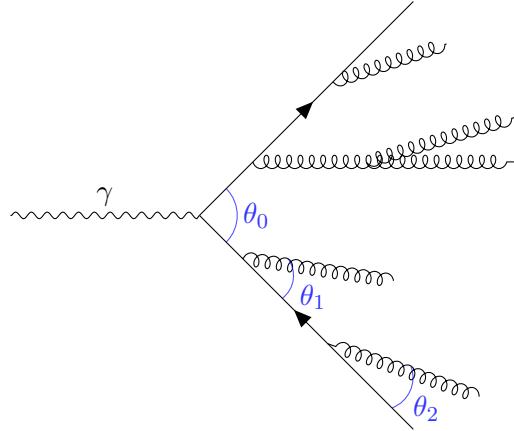


Figure 1.5: Shows a schematic diagram of jet fragmentation form a photon. Angular ordering give $\theta_0 > \theta_1 > \theta_2$.

We can therefore introduce a function $f(x, t)$ which represents the probability of the the particle having momentum fraction x of the original momentum p . t is an arbitrary evolution parameter, for instance the virtual mass or emission angle. Even though we cant calculate $f(x, t)$ from pertubative QCD, we can derive an equation for the t -dependence. It is known as the DGLAP equation, for a derivation see Section 5.2 in [14]. The DGLAP equation reads

$$t \frac{\partial}{\partial t} \left(\frac{f(x, t)}{\Delta(t)} \right) = \frac{1}{\Delta(t)} \int \frac{dz}{z} \frac{\alpha_s}{2\pi} \hat{P}(z) f(x/z, t), \quad (1.58)$$

where $\Delta(t)$ is called the Sudakov form factor, and its defined as

$$\Delta(t) = \exp \left[- \int_{t_0}^t \frac{dt'}{t'} \int dz \frac{\alpha_s}{2\pi} \hat{P}(z) \right]. \quad (1.59)$$

$\hat{P}(z)$ is the relevant splitting function. The Sudakov is the probability of the parton to evolve from t_0 to t without branching. The Sudakov is the basis for jet simulation algorithms.

We will now describe the basics of such branching algorithms, given in Section 5.3 in [14].

Starting with one time-like parton with momentum fraction x_1 and virtual mass t_1 . We first start by generating a new value for t called t_2 . This is done by generating a random number u distributed uniformly between 0 and 1. We then obtain t_2 by solving

$$\frac{\Delta(t_1)}{\Delta(t_2)} = u. \quad (1.60)$$

If the value of t_2 is lower than the scale were we choose to stop the branching, nothing more happens. Else we have to generate a new momentum fraction $z = x_2/x_1$ for the next branching. This is done by solving

$$\int_{\varepsilon}^z dz' \frac{\alpha_s}{2\pi} \hat{P}(z') = u' \int_{\varepsilon}^{1-\varepsilon} dz \frac{\alpha_s}{2\pi} \hat{P}(z) \quad (1.61)$$

for z , where u' is new random number between 0 and 1. ε is a chosen cut-off to eliminate the divergences in the splitting function. The process is then repeated, and the emitted partons can also branch via a similar algorithm. This will then generate a whole shower of partons. We then need to introduce a stopping condition, this will depend on what we are interested in. If we are interested in the distribution of particle in the jet, we would start the evolution at angle R , and evolve down to a minimal angel Q_0/p_T , given by the hadronization scale $Q_0 \sim 0.2 - 1$ GeV. However, if we are interested in distribution of jets originating from a single hard parton, we would start the evolution at a maximal angle $\theta \sim 1$ and evolve down to $\theta = R$. There are of course more subtleties in practice, but the process described above gives the basic ingredients of a branching algorithm.

Chapter 2

Heavy-Ion Collisions and Parton Energy-loss

In particle accelerators such as LHC at CERN, heavy ions are collided at very high centre of mass energies. Typically centre of mass energy per nucleon is in the range $\sqrt{s_{NN}} = 2.76 - 5.02$ TeV. At such energies the heavy ions are Lorentz contracted discs in the lab frame. For a collision at $\sqrt{s_{NN}} = 2.76$ TeV the the Lorentz factor is $\gamma = 1400$ [10], therefore in the lab frame the nuclei are basically two colliding circular pancakes.

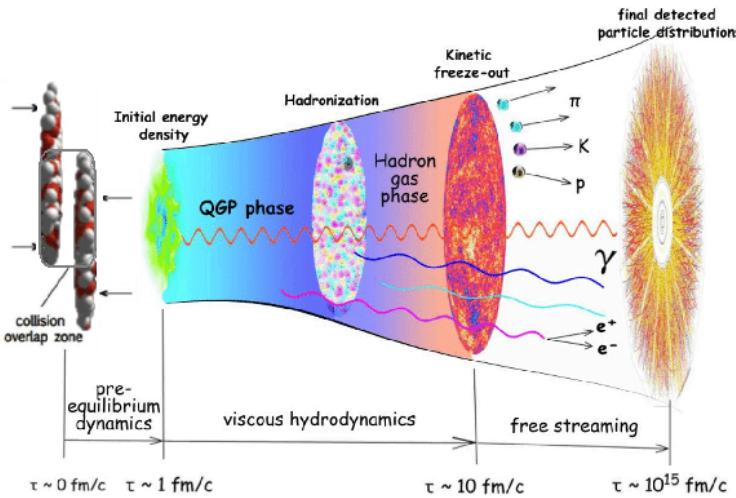


Figure 2.1: Shows the time evolution of a heavy-ion collision. Figure taken from [22] (colors inverted for beauty reasons).

The interdisciplinary study of heavy ion collisions give strong indications of the formation of the quark-gluon plasma. The QGP is a very dense and hot state of strongly interacting quarks and gluons. Its existence was theorized in 1975 (Section 1.4 in [16]) after the discovery of asymptotic freedom in QCD. The earliest moments of our universe and the core of superdense neutron stars are theorized to contain quark-gluon plasma. This is one of the main reasons to understand the QGP and its properties.

Fig 2.1 shows a schematic overview of the time evolution of the collision of two Lorentz con-

tracted nuclei. About 1 fm/c after the collision the Quark-Gluon Plasma (QGP) has formed between the receding discs which is about 2 fm apart at this point. The plasma then expands and cools down, and after about 10 fm/c hadronization happens. In this chapter we will first take a look at some of the observable consequences of the formation of QGP in heavy ion collisions. Then we will study the formalism for a single parton energy-loss in a QGP medium by BDMS [7].

2.1 Basic Terminology and Setup

In high energy heavy ion collisions the following geometrical picture of the collision is often used:

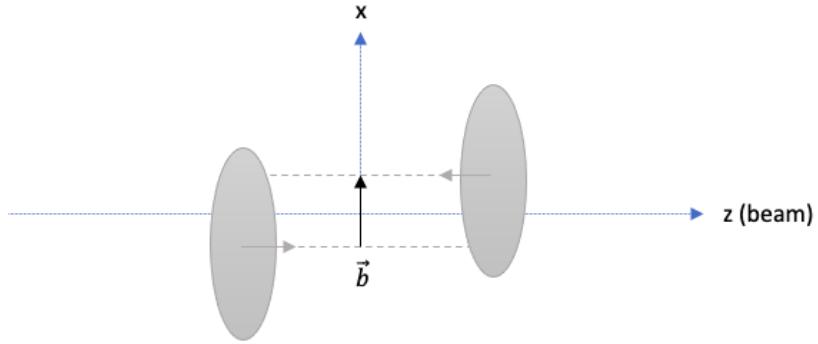


Figure 2.2: Shows the geometrical picture of the collision of two nuclei. \vec{b} is the impact vector.

The vector \vec{b} connecting the centers of the two colliding nuclei in the transverse plane is called the impact vector, and its length $|\vec{b}|$ is the impact parameter. The impact vector defines the x-axis, while the beam defines the z-axis. The x-z plane defines the so called reaction plane. One often distinguishes between the number of spectators N_{spec} and the number of participants N_{part} . Spectators are defined as the nucleons that do not interact with nucleons from the other nuclei, while participants are the nucleons which do interact. The term wounded nucleons are used for the participants which suffer at least one inelastic scattering. N_{part} can be estimated by zero-degree calorimeter, which are calorimeters place near the beam, but far from the collision point.

An import classification of different collisions is the centrality, denoted c . Centrality is defined as the as the percentile of events with the larges number of participants or with the largest particle multiplicity (page 33 in [16]). The centrality class of 0% will the correspond zero impact parameter $|\vec{b}|$. The centrality $c(N)$ of a given event/collision can be well approximated by

$$c(N) = \frac{\pi b^2(N)}{\sigma_{\text{in}}^{AB}}, \quad (2.1)$$

where σ_{in}^{AB} is total inelastic nucleus-nucleus cross section and $b(N)$ is the impact parameter which matches the multiplicity N and the average multiplicity $n(b)$, which can be calculated from the Glauber model. The Glauber model treats the distribution of nucleons in the nucleus as a random distribution given by some density profile. Nucleon-nucleon collisions are given by the total inelastic nucleon-nucleon cross section. Details on the Glauber model is given in

Chapter 3 of [16].

2.2 Signals of Quark-Gluon Plasma

What are the different indicators of the formation of QGP in heavy ion collisions? In this section we will discuss two different observables for the plasma. First we look at elliptic flow and then the suppression of particle spectra quantified by the nuclear modification factor.

2.2.1 Elliptic Flow

In flow analysis of heavy ion collisions one represents the momentum distribution of the observed particles as a Fourier expansion (Section 2.5 in [16])

$$\frac{dN}{dy d^2 p_T} = \frac{dN}{2\pi p_T dp_T dy} \left[1 + 2 \sum_{k=1}^{\infty} v_k \cos(k(\phi - \Psi_{RP})) \right], \quad (2.2)$$

where ϕ is the azimuthal angle, and the angle Ψ_{RP} defines the reaction plane. The coefficients v_k are called the k th harmonic flow, and they are generally functions of rapidity y and transverse momentum p_T . The v_2 coefficient is called "elliptic flow". The reason behind the name "flow" is because v_2 is qualitatively understood as an effect of hydrodynamic-like collective expansion of the matter produced in the collision.

In central collisions of two nuclei the area of particle production is almost circular in the transverse plane. This will result in isotropic expansion in the radial direction. However, in non-central collisions the area of particle production (blue area in Fig 2.3) will have an almond shape as seen from the beam direction.

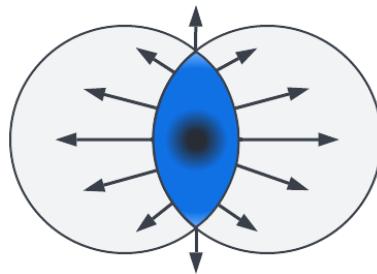


Figure 2.3: Shows the almond shaped overlapping region in a peripheral collisions of two nuclei, as seen from the beam direction.

The particles in the almond will interact with each other. The pressure gradient will be larger in the reaction plane as Fig 2.3 indicates. There will therefore be a more rapid expansion in this direction, which will lead to a momentum anisotropy favoring the momentum to be in the reaction plane. The measurements of v_2 (Figure 2.12 in [16]) match the predictions of ideal hydrodynamics indicating fluid behavior of the medium produced in the almond. The measurements of v_2 indicate a perfect fluid, meaning a very low viscosity to entropy ratio [26]. More details of elliptic flow can be found in [26].

2.2.2 Jet Quenching

One of the main observable of QGP is jet quenching, an extensive overview can be found in [13]. This is where a jet (collimated streams of particles) is modified due to the medium. This manifest itself through the suppression of the high p_T spectrum of jets compared to that in proton-proton collision. This is often quantified by the "nuclear modification factor" R_{AA} , defined experimentally as [16]

$$R_{AA} = \frac{1}{n_{AA}} \frac{dN_{AA}}{dp_T d\eta} / \frac{dN^{pp}}{dp_T d\eta}, \quad (2.3)$$

where n_{AA} (sometimes written as N_{coll}) is the average number of binary nucleon-nucleon collisions. N_{AA} is the average particle multiplicity, and N^{pp} is the multiplicity in proton-proton collisions. Fig 2.4 shows measurements of the nuclear modification factor for jets from ATLAS [1] at different centrality classes with $R = 0.4$ and rapidity range $|y| < 2.8$. The right panel shows R_{AA} for different species of hadrons, as well as direct photons.

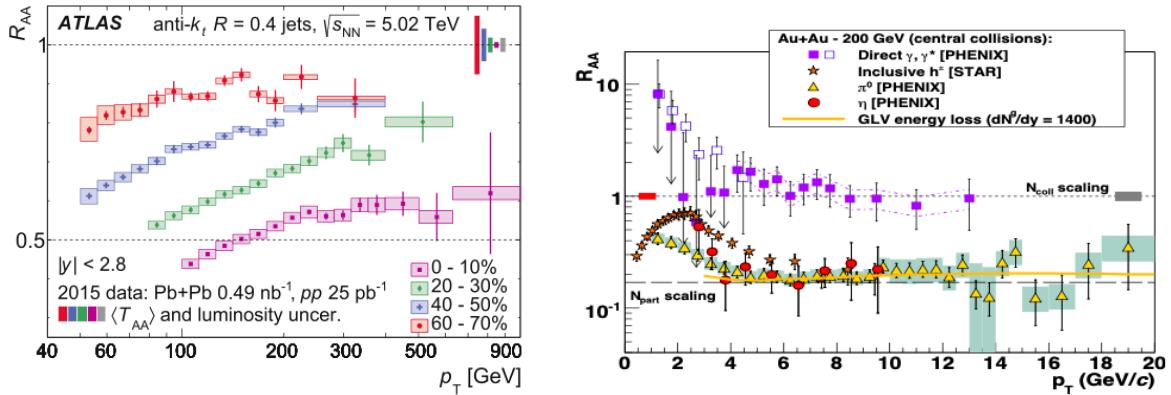


Figure 2.4: Left panel shows R_{AA} measurements for jets with $R = 0.4$ and rapidity range $|y| < 2.8$ at ATLAS [1] as a function of p_T for different centrality classes. Right panel shows R_{AA} for different species of hadrons and direct photons, taken from [13]

If heavy ion collisions were mere superpositions of binary nucleon-nucleon collisions, the scaling with N_{coll} should set the nuclear modification factor to one. However as we can see from Fig 2.4 this is not the case. There is a significant suppression of the jet and hadron spectra in heavy ion collisions. There is a different story for the photons, which after accounting for scaling the R_{AA} is actually one. In the presents of a QCD medium this makes perfect sense. Because photons carry no color charge, they will not interact with the medium, and there will be no suppression.

From now on we will focus on jets. If we wanted to model the spectrum for hadrons, we would need to introduce fragmentation functions. Which are non-perturbative functions describing the momentum fraction distribution of the given hadron. Jets on the other hand are perturbative and infra-red safe observable. The suppression of the jet spectrum will be because of out of cone energy-loss of particles in the medium. Partons produced in the initial collision will fragment into jets, while this is happening the partons will interact with the medium through elastic and inelastic collisions with the medium, as shown in Fig 2.5.

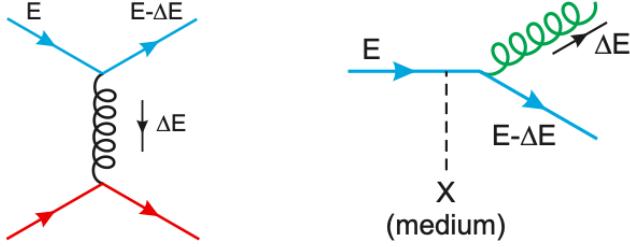


Figure 2.5: (left) Shows the elastic interaction with the medium leading to collisions energy-loss. (right) Shows the inelastic interaction leading to radiative energy-loss. The blue particles represents the transversing quark. Figure taken from [13].

This will lead to a modified fragmentation of the jet, which again will lead to a modification of the jet cross section in heavy ion collisions. An approach we will use later is to encapsulate this modification into an energy-loss distribution $D(\varepsilon)$ [7]. Which represents the probability of loosing energy ε to the QGP. We then relate $\frac{d\sigma^{AA}}{dp_T}$ (the jet cross section in heavy ion collisions, where the medium is present) to $\frac{d\sigma^{pp}}{dp_T}$ (the jet cross section on pp-collisions) via the convolution

$$\frac{d\sigma^{AA}}{dp_T} = \int_0^\infty d\varepsilon D(\varepsilon) \frac{d\sigma^{pp}}{dp_T}(\varepsilon + p_T). \quad (2.4)$$

This has a simple interpretation. Jets in AA collisions are produced, and then they have some probability $D(\varepsilon)$ for loosing energy ε (or transverse momentum) to the medium. We then integrate over all possible values of ε . In the following sections we will study the energy-loss distribution for a single colored particle traveling through a uniform brick of QGP, and we will study this in context of the famous BDMPS medium induced gluon spectrum.

2.3 The BDMPS Spectrum

While a parton moves through a droplet QGP it will undergo collisions with the medium (Fig 2.5). This will lead the parton to emit gluons (medium induced gluons). Each gluon will take away a small amount of energy ω from the parton. Baier, Dokshitzer, Mueller, Peigne and Schiff (BDMPS) calculated the emission probability in such collisions in the medium. Later Arnold [6] showed that the BDMPS result could be cast into a rather simple formalism. He showed that

$$\omega \frac{dI}{d\omega} = \frac{\alpha}{\pi} x P_{gs}(x) \ln |c(t=0)|, \quad (2.5)$$

where $c(t)$ satisfies the following boundary value problem:

$$\frac{d^2c}{dt^2} = \omega_0^2(t)c(t), \quad c(t \rightarrow \infty) = 1, \quad (2.6)$$

where $\omega_0^2(t)$ is a complex number given by

$$\omega_0^2(t) = -i \frac{[(1-x)C_A + x^2 C_R] \hat{q}(t)}{2x(1-x)E}, \quad (2.7)$$

where E is the energy of the hard parton s , $C_R \hat{q}(t) = \hat{q}(t)$ is the average transverse momentum broadening of a hard parton with Casimir representation C_R moving through the medium. I is

the probability of the hard parton to emit a medium induced gluon with energy ω . x the energy fraction ω/E and last but not least $P_{gs}(x)$ is the splitting function for the process $s \rightarrow sg$. Moving forward we will work in the soft gluon approximation, meaning that we will say $x \ll 1$. This means that both P_{gq} and P_{gg} can be approximated by $2C_R/x$, as we saw earlier. The complex number $\omega_0^2(t)$ can also be approximated.

$$\omega_0^2(t) \approx -i \frac{C_R \hat{q}}{2xE} = -i \frac{\hat{q}}{2\omega}. \quad (2.8)$$

We will now solve this for a uniform brick of QGP with length L in the soft gluon approximation. We then have

$$\omega_0^2(t) = \begin{cases} -i \frac{\hat{q}}{2\omega} & , t < L \\ 0 & , t > L. \end{cases} \quad (2.9)$$

It is then fairly easy to show that the solution to the boundary value problem in Eq. (2.6) is

$$c(t) = \begin{cases} \cos(|\omega_0|(L-t)) & , t < L \\ 1 & , t > L. \end{cases} \quad (2.10)$$

where

$$|\omega_0| = \sqrt{\frac{\hat{q}}{4\omega}}(1-i). \quad (2.11)$$

Plugging all of this into Eq. (2.5) gives

$$\omega \frac{dI}{d\omega} = \frac{2\alpha C_R}{\pi} \ln |\cos \sqrt{\frac{\hat{q}}{4\omega}}(1-i)L|. \quad (2.12)$$

And after some mathematical gymnastics we arrive at the BDMPS spectrum

$$\frac{dI}{d\omega} = \frac{\bar{\alpha}}{2\omega} \ln \left[\cosh^2 \sqrt{\frac{\omega_c}{2\omega}} - \sin^2 \sqrt{\frac{\omega_c}{2\omega}} \right], \quad (2.13)$$

where $\bar{\alpha} = 2\alpha_s C_R/\pi$, and ω_c is the characteristic gluon energy given by

$$\omega_c = \frac{1}{2}\hat{q}L^2. \quad (2.14)$$

We will now study the effect of energy-loss of a hard parton using the BDMPS spectrum defined in Eq. (2.13), this analysis is due to BDMS [7] and Arleo [5]. To get a better feeling for the gluon spectrum in Eq. (2.13) we can look at the small and large ω behavior. For $\omega \gg \omega_c$ we can Taylor expand the expression for small $u \equiv \sqrt{\frac{\omega_c}{2\omega}}$ and get:

$$\omega \frac{dI}{d\omega} = \frac{\bar{\alpha}}{12} \left(\frac{\omega_c}{\omega} \right)^2 \quad (2.15)$$

and for small energies, meaning $\omega \ll \omega_c$, we get

$$\begin{aligned}\omega \frac{dI}{d\omega} &= \bar{\alpha} \ln \left[\left(\frac{\exp(\sqrt{\frac{\omega_c}{2\omega}}) + \exp(-\sqrt{\frac{\omega_c}{2\omega}})}{2} \right)^2 - \sin^2 \sqrt{\frac{\omega_c}{2\omega}} \right] \\ &\approx \bar{\alpha} \ln \left[\frac{\exp(\sqrt{\frac{\omega_c}{2\omega}})}{2} \right] = \bar{\alpha} \left[\sqrt{\frac{\omega_c}{2\omega}} - \ln(2) \right].\end{aligned}\quad (2.16)$$

We can also define $N(\omega)$, which is the multiplicity gluons with energy larger than ω

$$N(\omega) \equiv \int_{\omega}^{\infty} d\omega' \frac{dI}{d\omega'}.\quad (2.17)$$

Using the analytical expression in Eq. (2.16), we can obtain

$$\begin{aligned}N(\omega \ll \omega_c) &= \int_{\omega}^{\infty} d\omega' \frac{dI}{d\omega'} = \int_{\omega}^{0.01\omega_c} d\omega' \frac{\bar{\alpha}}{\omega'} \left[\sqrt{\frac{\omega_c}{2\omega}} - \ln(2) \right] + \int_{0.01\omega_c}^{\infty} d\omega' \frac{dI^{\text{Full}}}{d\omega'} \\ &= \bar{\alpha} \left[\sqrt{\frac{2\omega_c}{\omega}} - \ln 2 \ln \frac{\omega}{\omega_c} - 1.44136 \right],\end{aligned}\quad (2.18)$$

where the last integral is calculated numerically with full BDMPS spectrum given in Eq. (2.13). We can also calculate the high energy behavior

$$N(\omega \gg \omega_c) = \int_{\omega}^{\infty} d\omega' \frac{\bar{\alpha}}{12} \frac{\omega_c^2}{\omega'^3} = \frac{\bar{\alpha}}{24} \left(\frac{\omega_c}{\omega} \right)^2.\quad (2.19)$$

The figure below shows the gluon multiplicity of the full spectrum and the analytic low and high energy behavior.

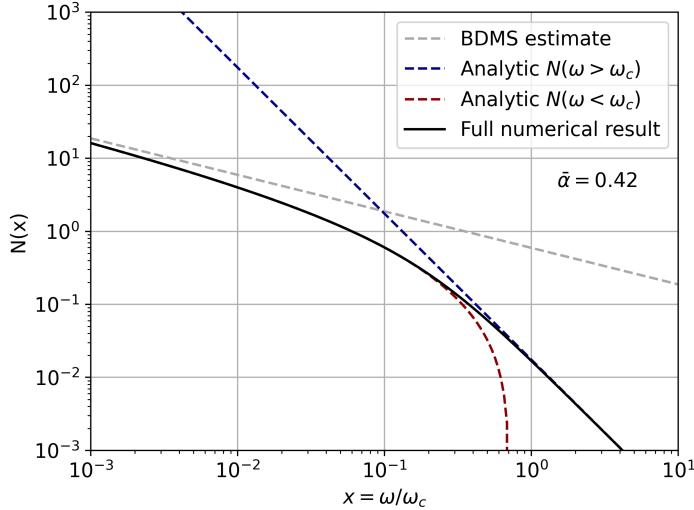


Figure 2.6: $N(\omega)$ of a quark ($C_R = C_F = 4/3$) defined in Eq. (2.17). The black solid curve shows the full numerical result. The red dashed curve shows the analytic IR result from Eq. (2.18). The blue dashed curve shows the high energy behavior given in Eq. (2.19). We have also included the so called analytic BDMS estimate (dashed grey curve), which is given in Eq. (2.20). $\bar{\alpha}$ is set to 0.42

From Fig 2.6 we see that it is extremely unlikely to emit medium induce gluons with energy larger than ω_c . The low and high energy approximations given in Eq. (2.18) and Eq. (2.19) fits very well with the numerical result in their respective regions. While the BDMS estimate, given by

$$N(\omega) = \bar{\alpha} \sqrt{\frac{2\omega_c}{\omega}}, \quad (2.20)$$

only really works in the very low energy regime. Now that we have a better feeling for the medium induce gluon spectrum, we will now look at how this translates into the energy-loss distribution $D(\varepsilon)$ of the parton.

2.4 The Poisson Approximation

Recall that when an hard parton transverses the medium it undergoes several scatterings with the medium, which leads to medium induce gluon emissions. If we assume that each such emission is independent, and neglect interference between the emitted gluons. The energy-loss distribution can take the form of a Poisson distribution [7]

$$D(\varepsilon) = \sum_{n=1}^{\infty} \frac{1}{n!} \left[\prod_{i=1}^n \int d\omega_i \frac{dI(\omega_i)}{d\omega} \right] \delta\left(\varepsilon - \sum_{i=1}^n \omega_i\right) \exp\left(- \int d\omega \frac{dI}{d\omega}\right), \quad (2.21)$$

where n is then the number of emitted gluons, and the n th term in the sum gives the contribution of n emissions, shown in Fig 2.7.

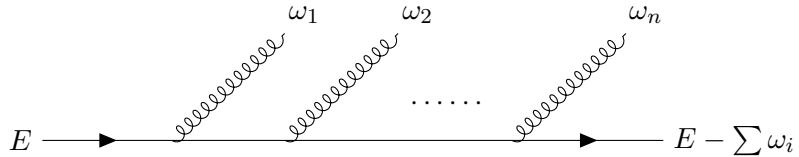


Figure 2.7: Show n medium induced gluon emissions from a quark. The total energy-loss of the quark is the sum of the energies of the emitted gluons.

The delta function gives the constraint that the sum of the energy of the emitted gluons is the energy that the parton loses. The last exponential factor is a normalization factor, to ensure that

$$\int_0^\infty d\varepsilon D(\varepsilon) = 1. \quad (2.22)$$

Now, given an expression for the gluon distribution $\frac{dI}{d\omega}$ it would be very hard to calculate $D(\varepsilon)$ using Eq. (2.21). Luckily we can simplify this procedure by first using the following representation for the delta function

$$\delta\left(\varepsilon - \sum_{i=1}^n \omega_i\right) = \int_C \frac{d\nu}{2\pi i} e^{\nu\varepsilon} \prod_{i=1}^n e^{-\nu\omega_i}, \quad (2.23)$$

where the contour \mathcal{C} runs parallel to the imaginary axis. Plugging this into Eq. (2.21) gives

$$\begin{aligned}
D(\varepsilon) &= \sum_{n=1}^{\infty} \frac{1}{n!} \left[\prod_{i=1}^n \int d\omega_i \frac{dI(\omega_i)}{d\omega} \right] \int_{\mathcal{C}} \frac{d\nu}{2\pi i} e^{\nu\varepsilon} \prod_{i=1}^n e^{-\nu\omega_i} \exp\left(-\int d\omega \frac{dI}{d\omega}\right) \\
&= \int_{\mathcal{C}} \frac{d\nu}{2\pi i} \sum_{n=1}^{\infty} \frac{1}{n!} \left[\prod_{i=1}^n \int d\omega_i \frac{dI(\omega_i)}{d\omega} e^{-\nu\omega_i} \right] e^{\nu\varepsilon} \exp\left(-\int d\omega \frac{dI}{d\omega}\right) \\
&= \int_{\mathcal{C}} \frac{d\nu}{2\pi i} \exp\left(\int d\omega \frac{dI}{d\omega} e^{\nu\omega}\right) \exp\left(-\int d\omega \frac{dI}{d\omega}\right) e^{\nu\varepsilon} \\
&= \int_{\mathcal{C}} \frac{d\nu}{2\pi i} \exp\left[-\int d\omega \frac{dI}{d\omega} (1 - e^{\nu\omega})\right] e^{\nu\varepsilon} \\
&= \int_{\mathcal{C}} \frac{d\nu}{2\pi i} \exp\left[-\nu \int d\omega N(\omega) e^{-\nu\omega}\right] e^{\nu\varepsilon} \\
&\equiv \int_{\mathcal{C}} \frac{d\nu}{2\pi i} \tilde{D}(\nu) e^{\nu\varepsilon},
\end{aligned} \tag{2.24}$$

where

$$\tilde{D}(\nu) = \exp\left[-\int_0^\infty d\omega \frac{dI}{d\omega} (1 - e^{\nu\omega})\right]. \tag{2.25}$$

To evaluate the contour we make the substitution $\nu = ib \rightarrow d\nu = i db$

$$\begin{aligned}
D(\varepsilon) &= \int_{-\infty}^{\infty} \frac{db}{2\pi} \exp\left[-ib \int d\omega N(\omega) e^{-ib\omega}\right] e^{ib\varepsilon} \\
&= \int_{-\infty}^{\infty} \frac{db}{2\pi} \exp\left[-ib \int d\omega N(\omega) (\cos b\omega - i \sin b\omega) + ib\varepsilon\right] \\
&\equiv \int_{-\infty}^{\infty} \frac{db}{2\pi} \exp[ib(\varepsilon - I_c(b))] \exp[-bI_s(b)] \\
&= \int_0^{\infty} \frac{db}{\pi} \exp[-bI_s(b)] \cos[b\varepsilon - bI_c(b)],
\end{aligned} \tag{2.26}$$

where

$$I_s(b) = \int_0^\infty d\omega N(\omega) \sin b\omega, \tag{2.27}$$

$$I_c(b) = \int_0^\infty d\omega N(\omega) \cos b\omega. \tag{2.28}$$

Now given an expression for $\frac{dI}{d\omega}$ we can calculate the energy-loss distribution $D(\varepsilon)$ by doing the following. First calculate $N(\omega)$ by Eq. (2.17), then evaluate the integrals $I_s(b)$ and $I_c(b)$. Once this is done we can finally calculate $D(\varepsilon)$ via Eq. (2.26). We will now do this calculation with the full BDMPS spectrum given in Eq. (2.13), and also using the BDMS estimate given in Eq. (2.20). For the full spectrum we will do it the way suggested by Arleo in [5]. The way he suggests we do it is by splitting the integrals in Eq. (2.27) and Eq. (2.28) into three parts

$$I_s(b) = \left(\int_0^{0.01\omega_c} + \int_{0.01\omega_c}^{2\omega_c} + \int_{2\omega_c}^\infty \right) d\omega N(\omega) \sin b\omega. \tag{2.29}$$

For the first integral term we will use the low energy $N(\omega < \omega_c)$ approximation given in Eq. (2.18), and for the last integral term we will use the approximation given in Eq. (2.19). For the middle part we will use the full BDMS spectrum, and this has to be done numerically. This

works because both the low and high energy approximations of $N(\omega)$ works extremely well in their respective integration limits, as shown in Fig 2.6. For the analytical estimate we have

$$I_s(b) = \bar{\alpha} \int_0^\infty d\omega \sqrt{\frac{2\omega_c}{\omega}} \sin b\omega = \bar{\alpha} \sqrt{\frac{\pi\omega_c}{b}} = I_c(b). \quad (2.30)$$

Inserting this into Eq. (2.26) gives

$$\begin{aligned} D(\varepsilon) &= \int_0^\infty \frac{db}{\pi} \exp[-bI_s(b)] \cos[b\varepsilon - bI_c(b)] \\ &= \int_0^\infty \frac{db}{\pi} \exp[-\bar{\alpha}\sqrt{\pi\omega_c b}] \cos[\varepsilon b - \bar{\alpha}\sqrt{\pi\omega_c b}] \\ &= \bar{\alpha}\sqrt{\frac{\omega_c}{2\varepsilon^3}} \left[\cosh\left(\frac{\pi\bar{\alpha}^2\omega_c}{2\varepsilon}\right) - \sinh\left(\frac{\pi\bar{\alpha}^2\omega_c}{2\varepsilon}\right) \right] \\ &= \bar{\alpha}\sqrt{\frac{\omega_c}{2\varepsilon^3}} \exp\left(-\frac{\pi\bar{\alpha}^2\omega_c}{2\varepsilon}\right). \end{aligned} \quad (2.31)$$

Fig 2.8 shows the full numerical result, and the analytical BDMS estimate, for a quark ($C_R = C_F = 4/3$) and a gluon ($C_R = C_A = 3$). We see that the full numerical result and the analytic

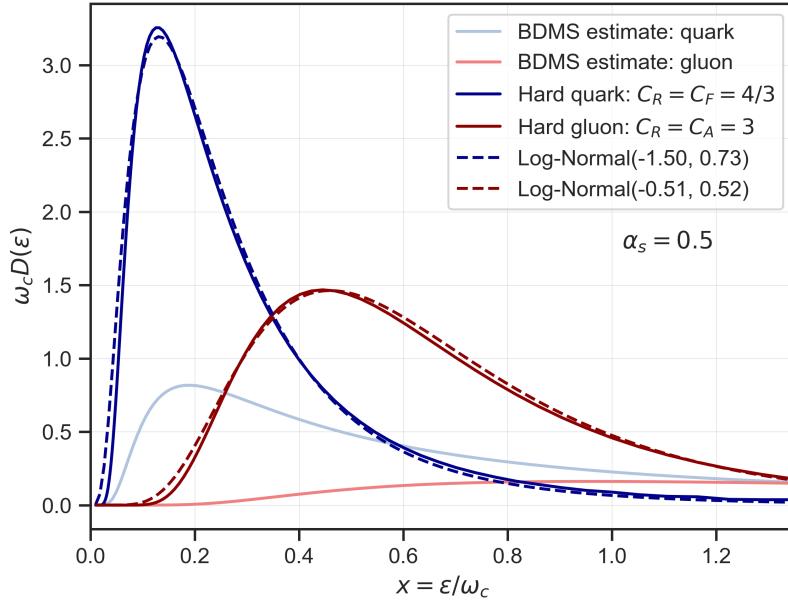


Figure 2.8: Shows $\omega_c D(\varepsilon/\omega_c)$ for the full numerical spectrum for a hard quark (dark-blue solid line), and for an hard gluon (dark-red solid line). While the dark-red and dark-blue dashed lines are Log-Normal distributions fitted to the solid lines. The light-blue and light-red curves show the BDMS analytic estimate Eq. (2.31) for a quark and a gluon

approximation for a quark has its peak at more or less the same location. For a gluon however, the analytical approximation is pretty wide compared to the numerical result. Another thing worth mentioning is that the energy-loss distributions in Fig 2.8 exhibits a tail towards larger ε . This means that the mean energy-loss defined by

$$\langle \varepsilon \rangle \equiv \int_0^\infty d\varepsilon \varepsilon D(\varepsilon), \quad (2.32)$$

will not lie at the peak of the distribution, but at the tail to the right of the peak. We can also conclude that it is not common (and extremely rare for a quark) that the hard parton loses energy larger than ω_c . While the full numerical result in Fig 2.8 is nice, it is not very practical. To use the result to make further predictions for observables such as R_{AA} , it would be convenient to have a parametric expression for the energy-loss distribution. In Fig 2.8 we can see that the log-normal distribution given by

$$\text{LogNormal}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right], \quad (2.33)$$

fits remarkably well to the numerical result with $D_q(x) = \text{LogN}(x|\mu = -1.50, \sigma = 0.73)$ and $D_g(x) = \text{LogN}(x|\mu = -0.50, \sigma = 0.52)$. We can then easily get an estimate for the mean energy-loss for the hard partons. The mean of a log-normal is given by

$$\langle x \rangle = \int_0^\infty dx x \text{LogN}(x|\mu, \sigma) = e^{\mu + \frac{1}{2}\sigma^2}. \quad (2.34)$$

The mean energy-loss will then be $\langle \varepsilon \rangle = \langle x \rangle \omega_c = \exp[\mu + \frac{1}{2}\sigma^2] \omega_c$, which gives

$$\text{quark}(C_R = C_F) : \quad \langle \varepsilon \rangle = \exp\left[-1.51 + \frac{0.73^2}{2}\right] \omega_c \approx 0.32\omega_c, \quad (2.35)$$

$$\text{gluon}(C_R = C_A) : \quad \langle \varepsilon \rangle = \exp\left[-0.51 + \frac{0.52^2}{2}\right] \omega_c \approx 0.78\omega_c. \quad (2.36)$$

As [5] points out, this is pretty close to the analytical estimate of

$$\langle \varepsilon \rangle = \int_0^\infty d\varepsilon \bar{\alpha} \varepsilon \sqrt{\frac{\omega_c}{2\varepsilon^3}} \exp\left(-\frac{\pi\bar{\alpha}^2\omega_c}{2\varepsilon}\right) = \bar{\alpha} \int_0^\infty d\varepsilon \sqrt{\frac{\omega_c}{2\varepsilon}} \exp\left(-\frac{\pi\bar{\alpha}^2\omega_c}{2\varepsilon}\right) = \frac{\alpha_s C_R}{2} \omega_c, \quad (2.37)$$

which is

$$\begin{aligned} \text{quark: } \langle \varepsilon \rangle &= \frac{\alpha_s C_F}{2} \omega_c \approx 0.33\omega_c, \\ \text{gluon: } \langle \varepsilon \rangle &= \frac{\alpha_s C_A}{2} \omega_c = 0.75\omega_c. \end{aligned} \quad (2.38)$$

How well does the energy-loss distributions in Fig 2.8 fit with the R_{AA} data for lead-lead collisions. We can use the two distributions to calculate the nuclear modification factor in the following way

$$R_{AA} = \frac{d\sigma^{AA}}{dp_T} / \frac{d\sigma^{pp}}{dp_T}, \quad (2.39)$$

where $N_{\text{coll}} = 1$ and $\frac{d\sigma^{AA}}{dp_T}$ is modelled by

$$\frac{d\sigma^{AA}}{dp_T} = \int_0^\infty d\varepsilon \sum_{i \in (q,g)} D_i(\varepsilon) \frac{d\sigma^{pp \rightarrow i}}{dp_T}, \quad (2.40)$$

where $D_i(\varepsilon)$ is the energy-loss distribution for parton i , $\frac{d\sigma^{pp \rightarrow i}}{dp_T}$ is cross section for producing parton i (as outgoing from the hard vertex). These cross sections can be generated in Pythia [9]. The $\frac{d\sigma^{pp}}{dp_T}$ is then the sum of the two contributions $\sum_i \frac{d\sigma^{pp \rightarrow i}}{dp_T}$. Lets compare this to the R_{AA}

data in central collisions ($0 - 10\%$) for jets from ATLAS [1] at $\sqrt{s_{NN}} = 5.02$ TeV in Fig 2.4.

The distributions can somewhat explain the data, however it is not very good fit. This might not be so surprising because the log-normal distributions are fitted to the numerical result for partons in an uniform static brick of QGP. However the medium produced in heavy ion collisions is expanding. Another difference is that the data in Fig 2.9 show the suppression of jets (and not single partons). What if we could go the other way? Could we extract the energy-loss distributions from the data? This is topic of the last two chapters. The next chapter will cover the basics of Bayesian inference which we will use. While in the last chapter we will present the results of such an approach.

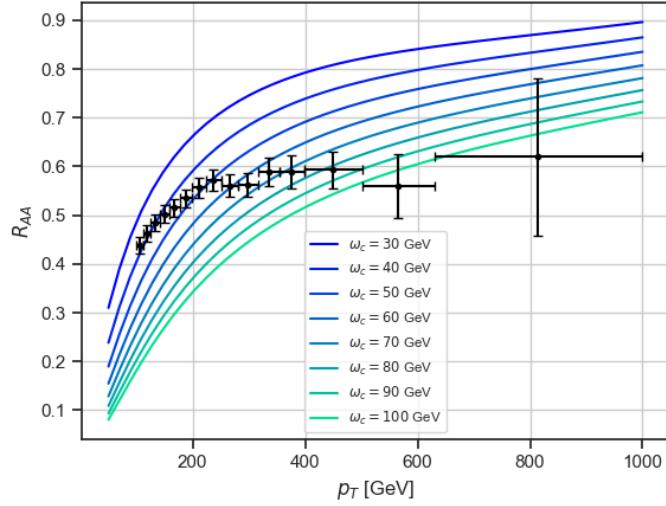


Figure 2.9: Comparison of R_{AA} measurements from ATLAS and the R_{AA} computed from the log-normal estimations of the energy-loss distributions computed numerically from the full BDMPS spectrum for different values of ω_c

Chapter 3

Bayesian Inference and MCMC

Bayesian inference is a powerful tool used in many fields of science. It is a statistical learning method used to update one's belief about some model parameter(s) after taking observed data into account. This is done using the famous Bayes rule. In this chapter we introduce Bayesian inference, and show how this is done in practice. In almost all real life scenarios Markov Chain Monte Carlo (MCMC) methods must be employed. We therefore look at the famous Metropolis-Hastings algorithm, and two different MCMC methods called Random Walk Metropolis and Hamiltonian Monte Carlo, and discuss their strength and weaknesses.

3.1 Bayes Theorem

In scientific endeavors we often want to make a model of how nature works. These models often depend upon an unknown number of parameters, denoted $\boldsymbol{\theta} = [\theta^1, \dots, \theta^n] \in \mathbb{R}^n$. In many cases we have some prior knowledge of what $\boldsymbol{\theta}$ might be, it could for instance be that we know that θ^j (being the j th element of $\boldsymbol{\theta}$) is positive or that it must have a value between 1 and 10. This knowledge could come from some physical constraints, limitations of the model or whatever. At this point it does not really matter! The thing that matters is that Bayesian inference is then a tool we can use to update our belief about $\boldsymbol{\theta}$ in a probabilistic way, by taking data into consideration. So how do we do that? The answer is via Bayes rule (often also called Bayes theorem). Let's say that we have observed some data D , Bayes theorem then reads [27]

$$P(\boldsymbol{\theta}|D) = \frac{P(D|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(D)}. \quad (3.1)$$

The quantity of interest is the so called posterior $P(\boldsymbol{\theta}|D)$, which is the probability distribution of $\boldsymbol{\theta}$ after taking the data D into account. The flipped quantity $P(D|\boldsymbol{\theta})$ is called the likelihood, and it is the likelihood of observing D given that $\boldsymbol{\theta}$ is correct (our given model with the specific parameters in question). If $P(D|\boldsymbol{\theta})$ is very low this will give a low probability for $\boldsymbol{\theta}$ in the posterior.

$P(\boldsymbol{\theta})$ is called the prior, and as the name states this is where we put our prior knowledge of the parameters. If for some reason we know that value of a parameter is positive and lies in the range 0 to 10, our prior could then be $P(\theta^j) = \text{Uniform}[0, 10]$. Finally the denominator $P(D)$

is the marginal likelihood, often referred to as the evidence, given by

$$P(D) = \int d\boldsymbol{\theta} P(D|\boldsymbol{\theta})P(\boldsymbol{\theta}). \quad (3.2)$$

It quantifies how well our model explains the data when averaged over all possible values of $\boldsymbol{\theta}$. When doing Bayesian inference this can be viewed as just a normalization constant to be determined. The problem is that the evidence can actually be very hard to determine, because the integral in Eq. (3.2) can be hard to evaluate (specially with a high dimensional $\boldsymbol{\theta}$).

Moving forward we will use a notation inspired by [27] to easier distinguish the different quantities in Eq. (3.1).

- Posterior: $P(\boldsymbol{\theta}|D) \equiv \mathcal{P}(\boldsymbol{\theta})$
- Likelihood: $P(D|\boldsymbol{\theta}) \equiv \mathcal{L}(D|\boldsymbol{\theta})$
- Prior: $P(\boldsymbol{\theta}) \equiv \pi(\boldsymbol{\theta})$
- Evidence: $P(D) \equiv \mathcal{Z}(D)$

3.2 Posterior, whats the big deal?

We now have some understanding of Bayes rule and what the posterior density is. The question now is how do we use it? Again here we follow closely the approach of [27]. The main points to mention is: making educated guesses, quantifying uncertainty and predicting new data points. Lets have closer look at each of them.

Making Educated Guesses

Often we are interested in making predictions about some other quantity $f(\boldsymbol{\theta})$ that depends on the model parameters. We can do this by marginalizing $f(\boldsymbol{\theta})$ weighted by the posterior distribution over $\boldsymbol{\theta}$.

$$E[f(\boldsymbol{\theta})] = \int d\boldsymbol{\theta} f(\boldsymbol{\theta})\mathcal{P}(\boldsymbol{\theta}). \quad (3.3)$$

An example could be that we want to make an educated guess of what the "true" underlying parameters $\boldsymbol{\theta}'$ could be. This can be done via a loss function $L(\boldsymbol{\theta}|\boldsymbol{\theta}')$ which tells us how far of $\boldsymbol{\theta}$ is from the true value. The problem is that that we don't know the true value $\boldsymbol{\theta}'$ (if we did, this hole endeavour would be pointless). We can instead take a weighted average over the posterior.

$$L_{\mathcal{P}}(\boldsymbol{\theta}) \equiv E[L(\boldsymbol{\theta}|\boldsymbol{\theta}')] = \int d\boldsymbol{\theta}' L(\boldsymbol{\theta}|\boldsymbol{\theta}')\mathcal{P}(\boldsymbol{\theta}'). \quad (3.4)$$

We can then say that our point estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}'$ is the value that minimizes $L_{\mathcal{P}}(\boldsymbol{\theta})$.

Quantifying Uncertainty

Once we have our point estimate $\hat{\boldsymbol{\theta}}$, we are also interested in a region of $\boldsymbol{\theta}$ were we believe there is a $X\%$ chance that the true value $\boldsymbol{\theta}'$ is. One often chooses X to be $1-\sigma$ or $2-\sigma$ credible

intervals. This region Ω_X is defined by the following

$$\int_{\boldsymbol{\theta} \in \Omega_X} \mathcal{P}(\boldsymbol{\theta}) d\boldsymbol{\theta} = \frac{X}{100}, \quad (3.5)$$

where

$$\Omega_X \equiv \{\boldsymbol{\theta} : \mathcal{P}(\boldsymbol{\theta}) \geq \mathcal{P}_X\} \quad (3.6)$$

In practice one often uses a definition for the 1-D marginalized posterior for each parameter θ^i . For further details, the reader is referred to Section 3.2 in [27].

Making Predictions

Once we have used the data D to do inference on $\boldsymbol{\theta}$, it would be nice to predict likelihood of seeing some hypothetical data \tilde{D} . This is what we call the posterior predictive, denoted $P(\tilde{D}|D)$, and its defined by the expectation of the likelihood

$$P(\tilde{D}|D) \equiv E[\mathcal{L}(\tilde{D}|\boldsymbol{\theta})] = \int d\boldsymbol{\theta} \mathcal{L}(\tilde{D}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}). \quad (3.7)$$

We have now seen the importance of the posterior. As the examples above illustrate, we are often more interested in quantities involving integration over the posterior density together with some other function dependent upon the model parameters. The problem is that for most realistic models these integrals are analytically impossible to do. This is actually also a problem when calculating the evidence $\int d\boldsymbol{\theta} \mathcal{L}(D|\boldsymbol{\theta}) \pi(\boldsymbol{\theta})$ which we need for the posterior. Well, no problem you might think, lets just do some good old fashioned numerical integration. Realistic models are often high-dimensional, $\boldsymbol{\theta} \in \mathbb{R}^d$ where d is large. Numerical integration can then be very inefficient! Fortunately there exist a way to get around the problems of computing the posterior and approximating integrals over it. This approach is called Markov Chain Monte Carlo sampling, and is the subject of the next section.

3.3 Markov Chain Monte Carlo

MCMC methods tries to generate a Markov Chain (set of correlated samples) which is then distributed according to the posterior ¹

$$\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n\} \sim \mathcal{P}(\boldsymbol{\theta}). \quad (3.8)$$

At this point a couple of questions needs to be answered. What is a Markov chain? How do we generate it in such a way that the samples mimic the posterior distribution? Last but not least, how do we use the chain to predict expectation values (integrals over the posterior). We can answer the last question fairly easy, given a set of samples $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n\}$, the expectation value

¹ Strictly speaking the samples are distributed according to the posterior mass $\mathcal{P}(\boldsymbol{\theta}) d\boldsymbol{\theta}$ (dubbed "typical set" by [8]), but this distinction is only important for high dimensional models

over some function of the parameters $f(\boldsymbol{\theta})$ can be approximated by the following [27]

$$E[f(\boldsymbol{\theta})] = \int d\boldsymbol{\theta} f(\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}) \approx \frac{1}{n} \sum_{i=1}^n f(\boldsymbol{\theta}_i). \quad (3.9)$$

The expectation value is just an average of $f(\boldsymbol{\theta})$ over all the samples.

We will now look at Markov Chains and their properties. When introducing Markov chains it is easier to get an intuition by first considering a discrete space χ . Following [4], lets consider a space with three states $\chi = x_1, x_2, x_3$. A Markov chain is then a stochastic process $x^1 \rightarrow x^2 \rightarrow \dots \rightarrow x^n$, where the probability of x^i is given by a transition matrix (in the example mentioned above this will be a 3x3 matrix)

$$p(x^i | x^{i-1}, \dots, x^1) = T(x^i | x^{i-1}). \quad (3.10)$$

The chain is called "homogeneous" if $\sum_{x^i} = T(x^i | x^{i-1}) = 1$. Following again the example in [4], consider the following transition matrix

$$T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{bmatrix}. \quad (3.11)$$

This the transition matrix in Eq. (3.11) can be visualized in a transition diagram

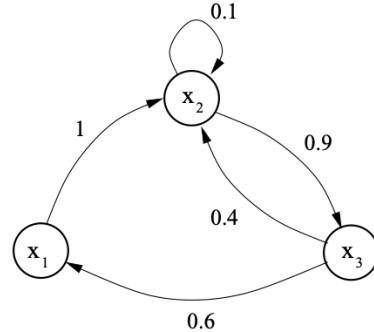


Figure 3.1: Transition graph of Eq. (3.11). The image is taken from [4].

If the current state x^{i-1} of the chain is x_3 , then probability of x^i being x_1 is 0.6 (and the probability of it being x_2 is 0.4). Lets say we start our chain in x_3 , the initial distribution of the visited states is then

$$p_0(x) = (0, 0, 1). \quad (3.12)$$

Lets then apply the transition matrix on this initial distribution once, we then obtain

$$p_0 \cdot T = (0.6, 0.4, 0.0). \quad (3.13)$$

As we expected there is then a 60% chance of moving to x_1 (and 40% to x_2). Now if we apply the transition many times we will see that the distribution will converge to a stable one. This means that after a while we get a distribution $p(x)$ that satisfies

$$p \cdot T = p. \quad (3.14)$$

We recognize $p(x)$ as the left eigenvector of T with eigenvalue 1. $p(x)$ is called the stationary distribution of the Markov chain. The chain will always converge on the stationary distribution (regardless of p_0) if the following two conditions are satisfied [4]:

1. Irreducibility: For any state in χ , there is a non-zero probability of visiting all other states.
2. Aperiodicity: The chain does not get stuck in loops.

Both of these conditions can be met in a concept called "detailed balance", which says that the transition probability is reversible with respect its initial distribution.

$$\begin{aligned} p(x^i)T(x^{i-1}|x^i) &= p(x^{i-1})T(x^i|x^{i-1}) \\ \sum_{x^{i-1}} p(x^i)T(x^{i-1}|x^i) &= \sum_{x^{i-1}} p(x^{i-1})T(x^i|x^{i-1}) \\ p(x^i) &= \sum_{x^{i-1}} p(x^{i-1})T(x^i|x^{i-1}). \end{aligned} \quad (3.15)$$

Eq. (3.15) shows that detailed balance leads to a stationary distribution. Now if we go to a continuous space, the transition matrix becomes a transition kernel and the stationary distribution becomes the eigenfunction of the kernel. Eq. (3.15) then becomes

$$p(x^i) = \int dx^{i-1} p(x^{i-1})T(x^i|x^{i-1}). \quad (3.16)$$

By constructing a Markov process where the stationary distribution is the un-normalized posterior $p(\boldsymbol{\theta}) = \mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ we can generate samples as if they were sampled from the posterior distribution. This looks promising, but it is still quite unclear how this should be done in practice. Let's have a look at how this is done using the Metropolis-Hastings algorithm.

3.4 The Metropolis-Hastings Algorithm

Before moving on let's just take a minute to clearly state what our goal is. We want to draw n samples $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$, such that, as $n \rightarrow \infty$ the distribution of samples converges on the posterior distribution $\mathcal{P}(\boldsymbol{\theta})$. Which is then our stationary target distribution for the Markov chain. We do this by invoking the detailed balance condition

$$P(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)\mathcal{P}(\boldsymbol{\theta}_i) = P(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})\mathcal{P}(\boldsymbol{\theta}_{i+1}), \quad (3.17)$$

where $P(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$ is the probability of moving from $\boldsymbol{\theta}_i$ to $\boldsymbol{\theta}_{i+1}$, and $\mathcal{P}(\boldsymbol{\theta})$ is the posterior. We now introduce some proposal distribution $Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$ which is used to propose new values $\boldsymbol{\theta}_{i+1}$ given $\boldsymbol{\theta}_i$. We then combine this with the probability of accepting the proposal $T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$ in following way

$$P(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) \equiv Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i). \quad (3.18)$$

Putting this into Eq. (3.17) we get the following constraint on the acceptance probability

$$\frac{T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)}{T(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})} = \frac{Q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})\mathcal{P}(\boldsymbol{\theta}_{i+1})}{Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)\mathcal{P}(\boldsymbol{\theta}_i)}. \quad (3.19)$$

The following $T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$ satisfies the constraint above

$$T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = \min\left[1, \frac{Q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})\mathcal{P}(\boldsymbol{\theta}_{i+1})}{Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)\mathcal{P}(\boldsymbol{\theta}_i)}\right]. \quad (3.20)$$

We are now ready to see the Metropolis-Hastings algorithm [27] for generating new samples.

The Metropolis-Hastings algorithm

- Propose a new sample $\boldsymbol{\theta}'_{i+1}$ by drawing a random sample from the proposal distribution:

$$\boldsymbol{\theta}'_{i+1} \sim Q(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)$$

- With probability

$$T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = \min\left[1, \frac{Q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})\mathcal{P}(\boldsymbol{\theta}_{i+1})}{Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)\mathcal{P}(\boldsymbol{\theta}_i)}\right]$$

set $\boldsymbol{\theta}'_{i+1} \rightarrow \boldsymbol{\theta}_{i+1}$. This is done by:

- Computing $T(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)$
- Sample random number $u \sim \text{Uniform}[0, 1]$
- If $u \leq T(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)$: $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}'_{i+1}$
- Else: $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$

- Repeat for i from 1 to n_{samples}

Given some initialization $\boldsymbol{\theta}_0$ the algorithm above is run from $i = 1$ up to $i = n$. At each iteration i , a new sample $\boldsymbol{\theta}'_{i+1}$ is proposed, and then accepted or rejected according to the acceptance probability $T(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)$. The nice thing about this algorithm is that we don't have to know the normalization of the target distribution because we take the ratio $\mathcal{P}(\boldsymbol{\theta}_{i+1})/\mathcal{P}(\boldsymbol{\theta}_i)$. In Bayesian inference this translates to not needing the evidence $\mathcal{Z}(D)$, which is almost impossible to calculate for realistic models.

Random Walk Metropolis

One of the most used MH sampler is the so called Random Walk Metropolis [8], where the proposal distribution is Gaussian

$$Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = \mathcal{N}(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i, \Sigma). \quad (3.21)$$

Now since $Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = Q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})$, the acceptance probability becomes simpler

$$T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = \min\left[1, \frac{\mathcal{P}(\boldsymbol{\theta}_{i+1})}{\mathcal{P}(\boldsymbol{\theta}_i)}\right]. \quad (3.22)$$

This makes the random walk Metropolis extremely easy to implement. Lets consider an 1D example (inspired by the example in [4]) where the target distribution is proportional $0.3e^{-0.2x^2} + 0.7e^{-0.2(x-6)^2}$ and the proposal distribution is $\mathcal{N}(x_{i+1}|\mu = x_i, \sigma = 5)$.

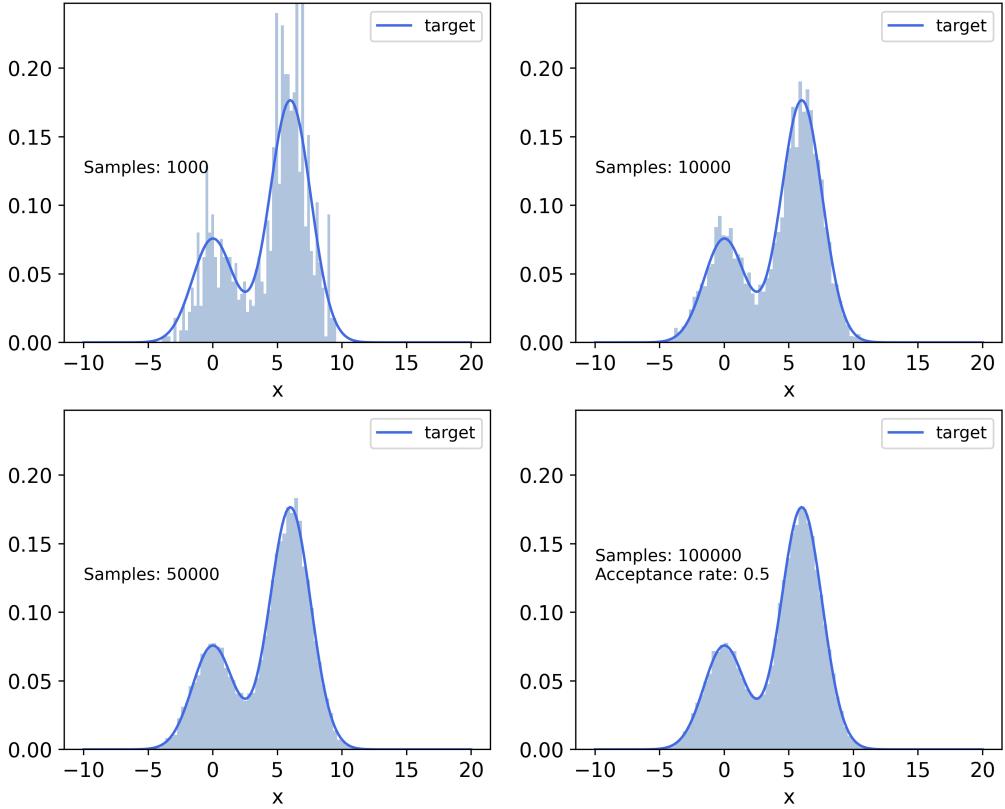


Figure 3.2: Shows an example of running a Random Walk Metropolis algorithm with $\mathcal{N}(x_{i+1}|\mu = x_i, \sigma = 5)$ for 100 000 iterations with a target distribution proportional to $0.3e^{-0.2x^2} + 0.7e^{-0.2(x-6)^2}$. Upper left panel shows the sample distribution after 1000 samples, upper right shows the same for 10 000, bottom left for 50 000 samples and bottom right for 100 000 samples. The target function is shown as the darkblue curve.

Fig 3.2 shows that after 50000 samples the RWM samples has converged pretty good to the target distribution. The figure also shows the distribution after 1000, 10 000, 50 000 and 100 000 samples. It is also quite common to take a look at the traceplot of the chain (Fig 3.3). This is a plot of all the sample values in the chain as function of the iteration. As a rule of thumb this should look like a "fuzzy" caterpillar. If your chain is biased (for example gets stuck in some small region) this will show up pretty clearly in the trace plot.

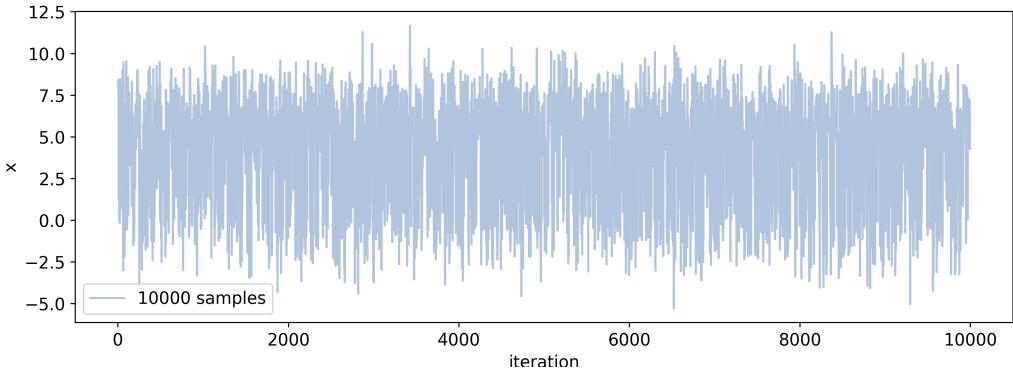


Figure 3.3: Traceplot of the samples in upper right panel in Fig 3.2

At this point a few remarks are in order. Most MCMC algorithms have some hyper-parameter(s) which has to be chosen. For RWM this is the width σ (or the covariance matrix Σ if we are in dimension higher than 1). If σ is too wide the acceptance rate is low and the correlations of the samples are high. This gives a lower effective sample size. On the other hand if the width is too narrow, we run the risk of not visiting all important parts of the target (i.e not all modes)[4]. Our goal is to have an high acceptance rate while also exploring all regions of interest. For the example in Fig 3.2, we used a width of $\sigma = 5$. This seemed to work quite well, with an acceptance rate of 0.5. This essentially means that half of our proposal was accepted.

The price we pay for the simplicity of the RWM becomes apparent when we go two higher dimensions (more complex models). In these cases the RWM is very inefficient, the acceptance rate becomes very low [8]. This means that we need to run the chain with large number of iterations. Even if we choose the "perfect" values for the covariance matrix we will have a problem of inefficient exploration in high dimensions. The sampler will move very slowly and there will be many directions we can randomly move in. The sampler will therefore use a very long time to explore the space. Fortunately there exist more robust MCMC algorithms which are much more clever when it comes to proposing new samples. The algorithm we are going to dive bit deeper into is the Hamiltonian Monte Carlo (HMC)method.

3.5 Hamiltonian Monte Carlo

HMC is a method to propose new samples for the Metropolis Hastings algorithm. It uses the well known dynamics of Hamilton's equation from classical dynamics to simulate trajectories in parameter-space. As we shall see, the potential energy used in the Hamiltonian dynamics is related to the target distribution, which in our case the un-normalized posterior $\mathcal{P}(\boldsymbol{\theta}) = \mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$. Before we discuss the HMC method, lets first quickly review Hamiltonian dynamics and its important properties for MCMC. This small overview of HMC and Hamiltonian dynamics is due to [19] and [8].

3.5.1 Hamiltonian Dynamics

In Hamiltonian dynamics, physical systems are described by the well known function $H(q, p)$, the so called Hamiltonian. q is a D -dimensional position vector and p is the corresponding conjugate momentum vector (also D -dimensional of course). The system evolves according to

Hamilton's equations of motion,

$$\begin{aligned}\frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i}, \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i}.\end{aligned}\tag{3.23}$$

where $i \in (1, \dots, d)$ is the i th element of q and p . The Hamilton's equation defines a mapping T_s from t to $t + s$ for q and p . The Hamiltonian is often written in terms of kinetic K and potential energy U .

$$H(q, p) = K(p) + U(q).\tag{3.24}$$

It is very common to assume that the potential energy depends only on q , and that the kinetic energy only depends on p and is defined as

$$K(p) \equiv \frac{1}{2} p^T M^{-1} p,\tag{3.25}$$

where M is a $d \times d$ "mass matrix" which is symmetric and positive-definite, and often also diagonal. Hamiltonian dynamics has some interesting properties which is important for MCMC. The first one being reversibility. The mapping T_s of $q(t), p(t)$ to $q(t + s), p(t + s)$ is unique (one-to-one). This means that there exists an inverse map T_{-s} , which corresponds to running the dynamics backwards in time. This property is important because for MCMC because it ensures detailed balance for the Markov Chain, (see section 3.2 in [19] for a formal proof).

Another important property is that the time evolution conserves the Hamiltonian (energy). This can be seen fairly easy from Eq. (3.23)

$$\frac{dH}{dt} = \sum_{i=1}^d \left[\frac{\partial H}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} \right] = \sum_{i=1}^d \left[\frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} \right] = 0.\tag{3.26}$$

Once we state the HMC algorithm it will be apparent that conservation of the Hamiltonian will lead to an acceptance probability of 1. This means that all of our HMC proposals will be accepted. However since Eq. (3.23) rarely can be solved analytically, we have to use numerical solutions. In that case the energy will only be approximately conserved.

The last property we will look at is volume preservation. Take a region R in (q, p) space with some volume V . Then the map $T_s(R)$ of R will also have volume V . This result is known as Louisville's theorem, see section 2.2 in [19] for proof.

As mentioned, Eq. (3.23) has to be solved numerically in most cases. There are several approaches for doing this, but the one used in HMC is the leapfrog method. Given a small time-step ε , the leapfrog method for the map $(q(t), p(t))$ to $(q(t + \varepsilon), p(t + \varepsilon))$ is

$$\begin{aligned}p_i(t + \varepsilon/2) &= p_i(t) - \frac{\varepsilon}{2} \frac{\partial U}{\partial q_i}(q(t)), \\ q_i(t + \varepsilon) &= q_i(t) + \varepsilon \frac{\partial K}{\partial p_i}(p_i(t + \varepsilon/2)), \\ p(t + \varepsilon) &= p_i(t + \varepsilon/2) - \frac{\varepsilon}{2} \frac{\partial U}{\partial q_i}(q(t + \varepsilon)).\end{aligned}\tag{3.27}$$

The leapfrog method preserves the phase-space volume. This is fairly easy to show, all we have to do is show that the determinant of the Jacobian is one. So let's consider the first two steps in the transformation given in Eq. (3.27). The Jacobian of the transformation $q(t) \rightarrow q(t + \varepsilon)$ and $p(t) \rightarrow p(t + \varepsilon/2)$ is given by

$$J = \begin{bmatrix} \frac{\partial q(t+\varepsilon)}{\partial q(t)} & \frac{\partial q(t+\varepsilon)}{\partial p(t)} \\ \frac{\partial p(t+\varepsilon/2)}{\partial q(t)} & \frac{\partial p(t+\varepsilon/2)}{\partial p(t)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ f(q(t)) & 1 \end{bmatrix}, \quad (3.28)$$

where $f(q(t)) = -\frac{\varepsilon}{2} \frac{\partial^2 U}{\partial q^2}(q)$. Clearly $\det(J) = 1$ for this transformation. To include the last step we just do the same thing once more, and end up with something similar. Since the determinant of the transformation is 1, the volume is preserved. Trajectories from the leapfrog integrator is also reversible due to the symmetry of the equations. To get the inverse map we just set $p \rightarrow -p$, apply the leapfrog integration with the same amount of steps and same ε , then at the end negate p again. Fig 3.4 shows the effectiveness of the leapfrog method compared to other methods such as the Euler method.

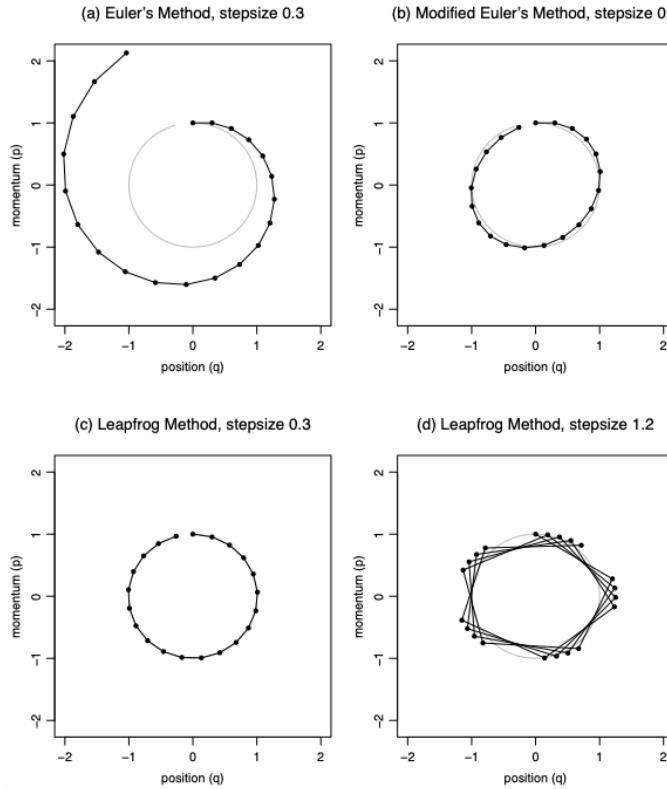


Figure 3.4: Comparison of different numerical methods for a Hamiltonian with $K(p) = p^2/2$ and $U(q) = q^2/2$, with starting point $q = 0$ and $p = 1$. Figure taken from [19]

We see that for even a relatively big leapfrog stepsize, $\varepsilon = 1.2$, the trajectory is stable with some errors of course. If we keep increasing the stepsize more and more the trajectory will diverge away from true path. The stability of the leapfrog method together with its properties of reversibility and volume preservation, makes it ideal for HMC.

3.5.2 The HMC Algorithm

In Bayesian inference with HMC the parameters $\boldsymbol{\theta}$ will play the role of the position vector q . We then introduce corresponding conjugate momenta \mathbf{p} . We then go from a D-dimensional space ($\boldsymbol{\theta}$ space) to a 2D-dimensional space, dubbed phase-space from physics. Recall that our target distribution is the un-normalized posterior $\mathcal{P}(\boldsymbol{\theta})$. To introduce $\mathcal{P}(\boldsymbol{\theta})$ into the phase space we define the so called canonical distribution which is joint distribution of $\boldsymbol{\theta}$ and \mathbf{p} .

$$\mathcal{P}(\boldsymbol{\theta}, \mathbf{p}) = \mathcal{P}(\mathbf{p}|\boldsymbol{\theta})\mathcal{P}(\boldsymbol{\theta}). \quad (3.29)$$

This means that if we marginalize out the momentum, we get the target distribution

$$\int d\mathbf{p} \mathcal{P}(\boldsymbol{\theta}, \mathbf{p}) = \mathcal{P}(\boldsymbol{\theta}) \int d\mathbf{p} \mathcal{P}(\mathbf{p}|\boldsymbol{\theta}) = \mathcal{P}(\boldsymbol{\theta}). \quad (3.30)$$

In a physical system the phase-space distribution $\mathcal{P}(\boldsymbol{\theta}, \mathbf{p})$ is related to the energy of the system (the Hamiltonian) by the following relation

$$\mathcal{P}(\boldsymbol{\theta}, \mathbf{p}) = \frac{1}{Z} \exp\left[-\frac{H(\boldsymbol{\theta}, \mathbf{p})}{T}\right], \quad (3.31)$$

where Z is just an arbitrary normalization constant, which we won't be needing, and T is the temperature of the system, which we will just set to 1. The Hamiltonian can then be written as

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\ln \mathcal{P}(\mathbf{p}|\boldsymbol{\theta}) - \ln \mathcal{P}(\boldsymbol{\theta}) \equiv K(\mathbf{p}, \boldsymbol{\theta}) + U(\boldsymbol{\theta}). \quad (3.32)$$

where the potential is $U(\boldsymbol{\theta}) = -\ln [\mathcal{L}(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})]$, but we are free to choose the kinetic energy by choosing the distribution $\mathcal{P}(\mathbf{p}|\boldsymbol{\theta})$. In HMC the usual choice is a multivariate Gaussian with zero mean and a covariance matrix (mass matrix) M ,

$$\mathcal{P}(\mathbf{p}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{p}|0, M). \quad (3.33)$$

This gives the kinetic energy

$$K(\mathbf{p}) = \frac{1}{2} \mathbf{p}^T M^{-1} \mathbf{p} + C, \quad (3.34)$$

where C is just a constant, which also will cancel out later. The matrix M is often also just set to the identity matrix I .

We are now ready to state the HMC algorithm. Given ε (stepsize) and L (number of leapfrog steps), and an initial sample $\boldsymbol{\theta}_0$, samples are generated by the algorithm below.

The Hamiltonian Monte Carlo Algorithm

```

for  $i = 1$  to  $n_{\text{samples}}$  do
    Sample:  $\mathbf{p}' \sim \mathcal{N}(0, I)$ 
    Set:  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_{i-1}$ ,  $\mathbf{p} \leftarrow \mathbf{p}'$ 
    for  $j=1$  to  $L$  do
         $\mathbf{p} \leftarrow \mathbf{p} - \frac{\varepsilon}{2} \nabla_{\boldsymbol{\theta} U}$ 
         $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \varepsilon \mathbf{p}$ 
         $\mathbf{p} \leftarrow \mathbf{p} - \frac{\varepsilon}{2} \nabla_{\boldsymbol{\theta} U}$ 
    end for
    Set:  $\mathbf{p} \leftarrow -\mathbf{p}$  (need not be done in practice)
    With probability  $T = \min \left[ 1, \frac{\exp[H(\boldsymbol{\theta}, \mathbf{p})]}{\exp[H(\boldsymbol{\theta}_{i-1}, \mathbf{p}')]} \right]$  set  $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}$  by:
    Sampling:  $u \sim \text{Uniform}[0, 1]$ 
    if  $u \leq T$  then
         $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}$ 
    else
         $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_{i-1}$ 
    end if
end for

```

First we (re)sample the momentum variables \mathbf{p}' from multivariate normal distribution (or just normal, if we are in 1D) with zero mean and the identity matrix as a covariance matrix. Then we simulate Hamiltonian dynamics starting from our current position in phase-space $(\boldsymbol{\theta}_{i-1}, \mathbf{p}')$ using the Leapfrog integrator with stepsize ε and L steps. Then we set $\mathbf{p} \rightarrow -\mathbf{p}$ to make the chain reversible. In practice this can be ignored, since we are only interested in Markov chain for $\boldsymbol{\theta}$. After simulating the Hamiltonian dynamics for a time εL , we now have proposal $\boldsymbol{\theta}$, which we then accept with probability T as shown in the algorithm above.

Fig 3.5 shows the same example as in Fig 3.2, but with the HMC algorithm instead. The number of drawn samples goes up to 20 000 (instead of 100 000, as in the example of RWM). Comparing the two examples, it seems that the HMC converges quicker than the RWM. However, to see the true strength of HMC to RWM, we would have to go to higher dimensions. What we can see from this example is that acceptance rate of our proposed samples in HMC is 0.99, which is much higher than the example with RWM.

The HMC methods performance on a given problem depends strongly on the choice of ε and L [18], for the example in Fig 3.5 the values were $\varepsilon = 0.6$ and $L = 15$. If ε is too small then we waste a lot of computational power. If ε is too large, the leapfrog integrator will be too inaccurate, and lead to lower acceptance rates. If L is too small, samples will lie close to each other and the sampler will exhibit random walk behavior (which is the very thing we try to eliminate). If L is too large the Hamiltonian trajectory will loop back towards the start. All in all this means that for a given problem, the two hyper-parameters ε and L must be tuned. Tuning of HMC is beyond the scope of this thesis, but some details can be found in [19]. [18] presents the No-U-Turn Sampler (NUTS) which is an extension to the HMC sampler, which eliminates the need for setting L .

Another problem that can arise when using HMC is that if the true trajectory in the phase-space

exhibits a region with strong curvature. The discretization of time in the leapfrog integrator might not be able to resolve this, and the trajectory will diverge from the true one [8]. In this case the step size ε should be set to a very small value so that the curvature can be resolved.

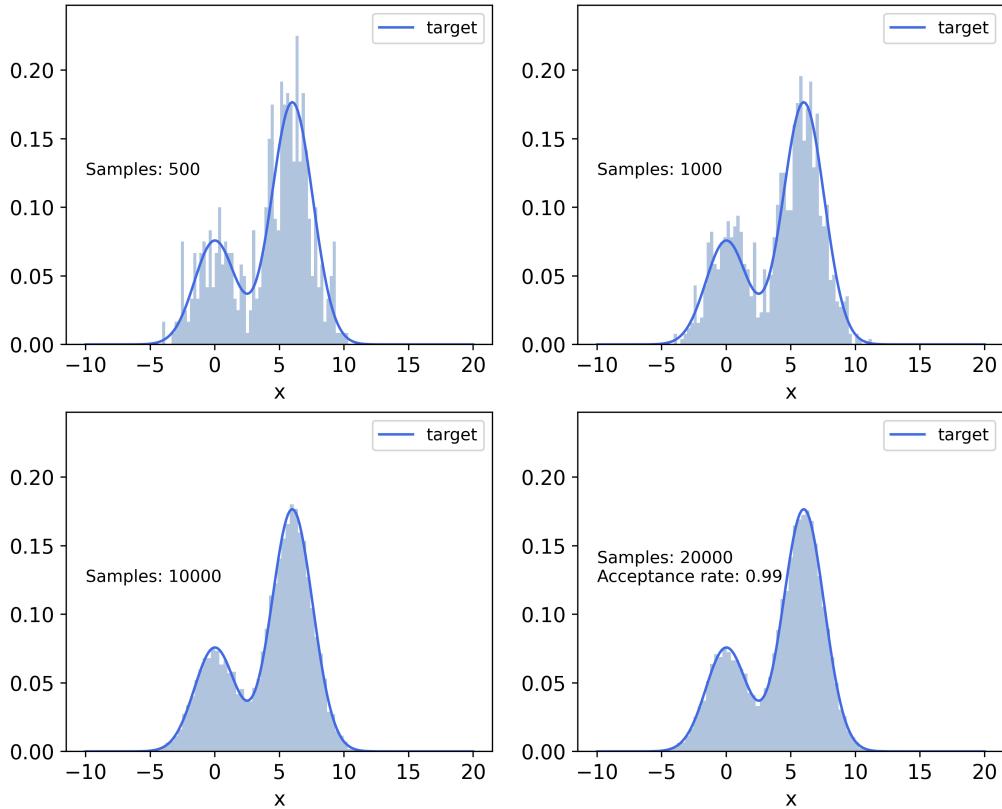


Figure 3.5: Shows an example of the sample distribution using HMC with the same target as in Fig 3.2. $\varepsilon = 0.6$ and $L = 15$ was used in the leapfrog integrator.

Chapter 4

Jet Quenching Using Bayesian Inference

In the Chapter 2 we saw how to apply the formalism of energy-loss of a single hard parton transversing a uniform brick of QGP. We saw that even in this case, to get "realistic" results we had to resort to numerical calculations. Now we would like to extend this to a QCD jet, but we will take a different approach. We take a different approach because it would be even harder to the calculation we did in Chapter 2 for a big shower of particles (a jet). Instead we take a more data driven approach. In practice this means that we will put a parametrization on the energy-loss distribution $D(\varepsilon) \equiv D(\varepsilon, \theta)$. Then use Bayesian inference to constrain the parameters θ . For reasons explained in Chapter 3, this has to be done using MCMC.

Using Bayesian inference to extract the energy-loss distribution $D(\varepsilon)$ of a jet has already been done by Xin-Nian Wang, Long-Gang Pang and Yayun He in [17]. We first take a look at their implementation, and try to replicate it using Hamiltonian Monte Carlo. Next we will build a model for jet quenching based on quark/gluon jets with the goal of generalizability and consistency of the factorization

$$\frac{d\sigma^{AA}}{dp_T}(p_T) = \int_0^\infty d\varepsilon \sum_{i \in (q,g)} D_i(\varepsilon) \frac{d\sigma_i^{\text{vac}}}{dp_T} \quad (4.1)$$

between different data-set. In Eq. (4.1) $D_i(\varepsilon)$ is the energy-loss distribution for a jet fragmented from a parton of type i . $\frac{d\sigma_i^{\text{vac}}}{dp_T}$ is the jet cross section for jets of type i . Quark/gluon tagging of jets is a very active field. Quite often it involves various Machine Learning algorithms. However, in this thesis we take a rather simple approach. We will generate jets in Pythia, and try to match them to the hard partons outgoing from the proton-proton collision. How these quark/gluon jets will be defined and generated in Pythia will be discussed.

4.1 Re-deriving Earlier Results Using HMC

Before we move on lets just quickly recall the definition if the nuclear modification factor. It is a ratio of the jet cross section in a AA collision to that of jet cross section in pp collisions.

$$\begin{aligned} R_{AA} &= \frac{d\sigma^{AA}}{dp_T} / \frac{d\sigma^{pp}}{dp_T} \\ &= \frac{1}{\frac{d\sigma^{pp}}{dp_T}} \int d\varepsilon D(\varepsilon, \theta) \frac{d\sigma^{pp}}{dp_T}(\varepsilon + p_T) \end{aligned} \quad (4.2)$$

Because we model the AA cross section by a convolution with the pp cross section there is no N_{coll} , however it is still present in the experimental data we will use. In [17] the energy-loss distribution is set to be a gamma distribution of $x \equiv \frac{\varepsilon}{\langle \varepsilon \rangle}$

$$D(x, \alpha) = \frac{\alpha^\alpha x^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)} \quad (4.3)$$

where Γ is the normalization factor, given by the gamma function. The mean energy-loss $\langle \varepsilon \rangle$ is a function of p_T

$$\langle \varepsilon \rangle = \beta p_T^\gamma \log p_T \quad (4.4)$$

In this model α can be interpreted as the number of jet-medium scatterings. The task is then reduced to constrain the parameters $\theta = [\alpha, \beta, \gamma]$ using the data in Fig 2.4 via Bayesian inference. We will reproduce the results for the centrality class 0 – 10%. The prior ranges for the parameters was chosen in [17] as

$$\begin{aligned} \alpha &\sim \text{Uniform}[0, 10] \\ \beta &\sim \text{Uniform}[0, 10] \\ \gamma &\sim \text{Uniform}[0, 1]. \end{aligned} \quad (4.5)$$

As explained in the last chapter, the Bayesian analysis has to be done using MCMC sampling. In [17] this is done using the Random Walk Metropolis method with $8 \cdot 10^6$ drawn samples. In our implementation we will use the Hamolition Monte Carlo method with 50 000 drawn samples, and 5000 samples for tuning. This is actually a great example showing how much more sample efficient HMC is compared to RWM. To reduce the risk of divergences in the leapfrog integrator, we set the target acceptance rate for the HMC sampler to be 0.97. This means that we tell the sampler to simulate the Hamiltonian dynamics extremely precise (small time steps). The Bayesian analysis and MCMC sampling is done in the probabilistic programming package PYMC3 in Python [21].

We also need to define a likelihood function. In all of our inferences in this project we will use a Gaussian likelihood, this is also used by [17]. This means that if we have a measurement of $R_{AA}(p_T)$ with uncertainty $\sigma_{R_{AA}}$, the likelihood of $R_{AA}(p_T)$ given θ is

$$\mathcal{L}\left(R_{AA}(p_T) \mid \theta\right) = \mathcal{N}\left(R_{AA}(p_T) \mid \mu = R_{AA}^{\text{model}}(\theta, p_T), \sigma = \sigma_{R_{AA}}\right), \quad (4.6)$$

where $R_{AA}^{\text{model}}(\theta, p_T)$ is the prediction of the model with parameters θ , given by Eq. (4.2). To compute $R_{AA}^{\text{model}}(\theta, p_T)$ we need to be able to evaluate $\frac{d\sigma^{pp}}{dp_T}$ at specific values, especially since the convolution integral in the expression has to be done numerically, info on how this is done

numerically can be found in Sec A.1. We therefore have to be able to interpolate what the values of the cross section is between the data points. Inspired by the parametrization of the cross section given in [28], the cross section will be given by

$$\frac{d\sigma^{pp}}{dp_T}(p_T) = a \left(\frac{p_T^0}{p_T} \right)^{n(p_T)}, \quad n(p_T) \equiv n - b \cdot \log \frac{p_T^0}{p_T} - c \cdot \log^2 \frac{p_T^0}{p_T} - d \cdot \log^3 \frac{p_T^0}{p_T}, \quad (4.7)$$

where a , n , b , c and d are parameters which can be fitted to the pp data from ATLAS in [1] using the method of least squares. p_T^0 is the p_T where the parameter a is the pp cross section. Fig 4.1 shows how well this parameterization fits to the data. The blue curve shows the parametric representation of the jet cross section after fitting it to the data (which are shown with uncertainties in the plot). The parameters of the fit (blue curve) is

$$p_T^0 = 45 \text{ GeV}, \quad a = 133.168342 \text{ nb/GeV}, \quad n = 4.62170656, \\ b = 0.848548201, \quad c = 0.348064297, \quad d = 0.0907623371.$$

It is also worth mentioning that that [17] uses an interpolation algorithm instead of a parametric fit. Ultimately this will not lead to anything different, this is just different "design" choices.

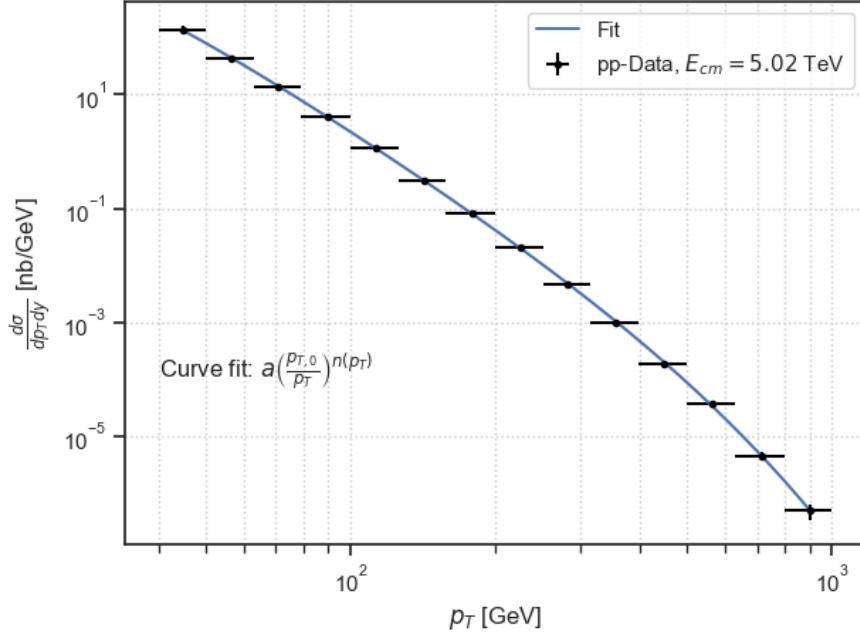


Figure 4.1: Shows the measured jet cross section for proton-proton collisions used for the R_{AA} data in Fig 2.4. The blue curve shows the parametric fit to the data using Eq. (4.7) with the method of least squares.

We are now ready to look at the results of the Bayesian inference. We start by looking at the posterior distributions of the parameters α, β, γ . This is often done using a so called corner plot (Fig 4.2). The upper panels of each column shows the posterior distributions of the individual parameters. While the rest shows the pairwise posterior densities (density is indicated by color).

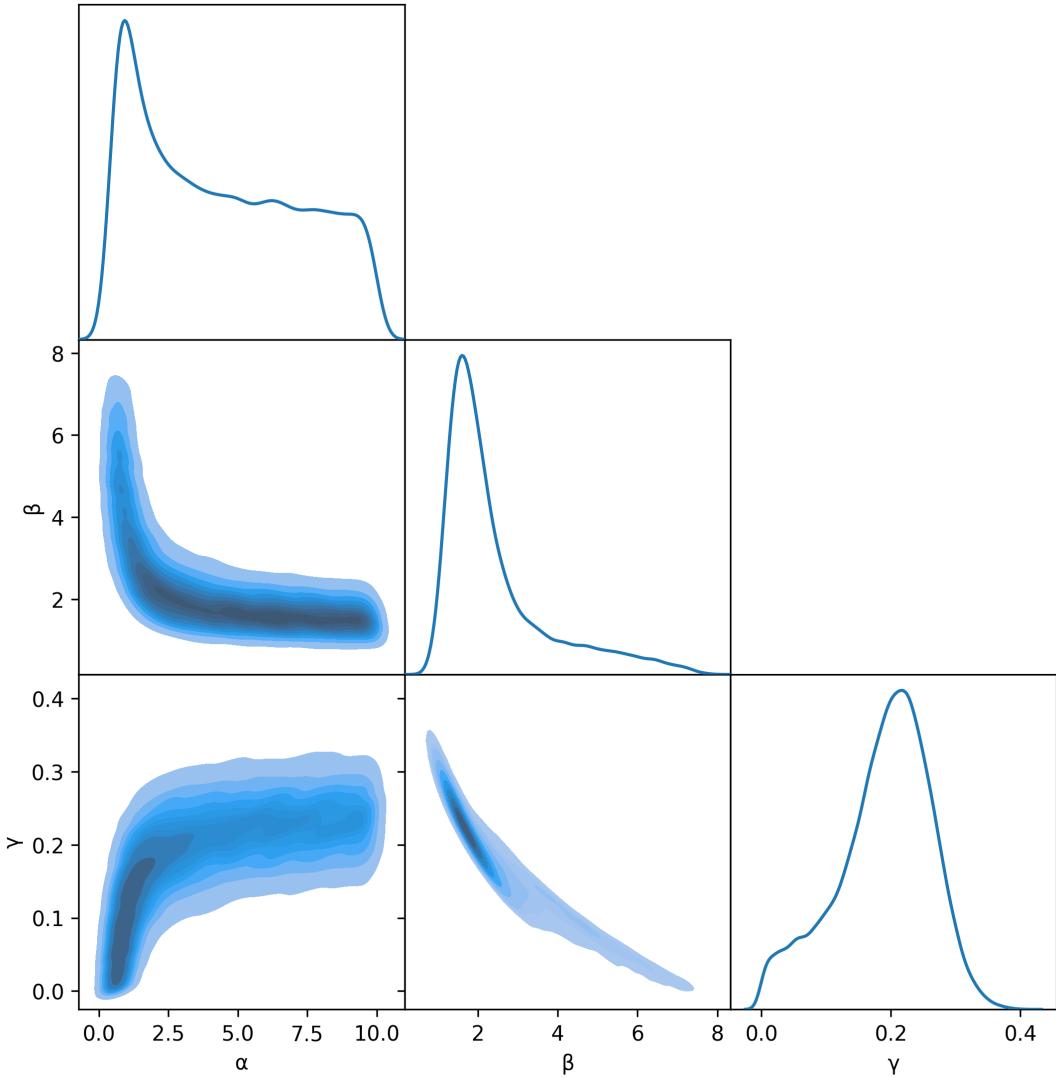


Figure 4.2: Corner plot of the posterior of $[\alpha, \beta, \gamma]$. The upper panels shows the posterior distribution of the individual parameters. The rest of the panels shows the correlations between the parameters.

As they point out, the parameters in Fig 4.2 exhibits clear correlation. Lets compare the parameters of our implementation to that of [17]. The upper row in the table below shoes the mean and 1 std of the posterior of the parameters in our implementation. While the bottom rows shows the result from [17].

α	β	γ
4.40 ± 2.90	2.43 ± 1.34	0.19 ± 0.07
4.41 ± 2.86	1.06 ± 0.97	0.26 ± 0.06

Table 4.1: Shows the mean and standard deviation of the different parameters. The upper values are for our implementation, and the lower values are [17] results.

The parameters values of our implementation, shown in the table above, agree quite well with [17] results, except for the β . This is probably due to the long tail of the β distribution shown in Fig 4.2. Fig 4.3 shows the Bayesian fit to the data for the nuclear modification factor. As

we can see the model fits quite well to the data. However as [17] points out, the experimental data have large uncertainties (especially in the high p_T range). This is most likely the reason for the large uncertainties in the parameters in Table 1.1.

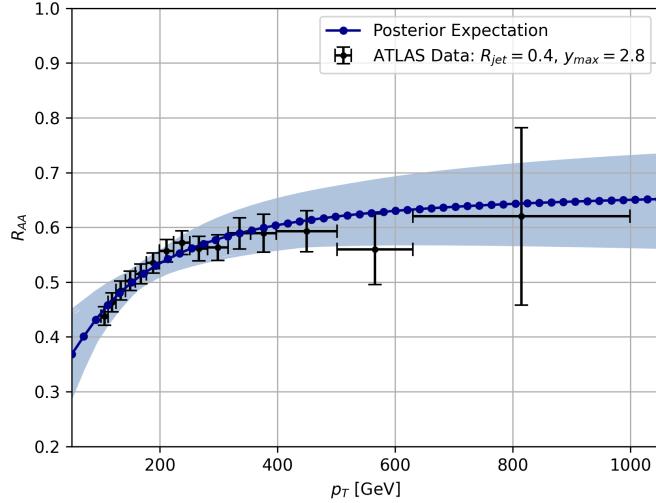


Figure 4.3: Comparison of the fitted model to the R_{AA} data. Here we have sampled 10000 samples from the posterior. The dark-blue curve shows the posterior expectation given by Eq. (3.9), and the light-blue curves show the predictions for the different sampled values for $\theta = [\alpha, \beta, \gamma]$.

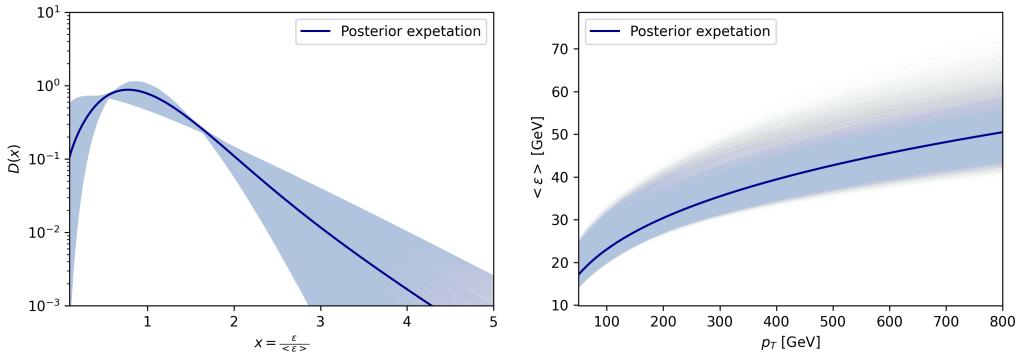


Figure 4.4: Left panel shows the energy-loss distribution $D(x)$, and the right panel shows the mean energy-loss as a function of p_T . Just as [17], we have only used parameter values within 1σ of the mean.

Fig 4.4 shows the energy-loss distribution $D(x, \alpha)$ (left panel) and mean energy-loss $\langle \varepsilon \rangle$ as a function of p_T (right panel). The energy-loss distribution has a tail towards larger values of x . In conclusion we see that we can reproduce the results of [17] well using the HMC method (rather than RWM) with less samples.

4.2 Quenching of Parton-tagged Jets

In the previous section we re-derived the results of [17] using Hamiltonian Monte Carlo. Now we want to take such an analysis to the next step. This would be to distinguish between quark jets and gluon jets. The jet cross section in heavy ion collisions will be modelled by

$$\frac{d\sigma^{AA}}{dp_T}(p_T) = \int_0^\infty d\varepsilon \sum_{i \in (q,g)} D_i(\varepsilon) \frac{d\sigma_i^{\text{vac}}}{dp_T}, \quad (4.8)$$

where $\frac{d\sigma_i^{\text{vac}}}{dp_T}$ is the cross section for producing jet of type i . We define the jet type by the initiator of the jet. This means that if a recorded jet is initiated by a hard (outgoing) quark, it is classified as a quark jet, and vice versa for gluons. Each jet type will have its own energy-loss distribution $D_i(\varepsilon)$. The motivation for this is that we know that quarks and gluons interact a bit different with the medium (different color factor), as we saw in Fig 2.8. However, the spectra in Eq. (4.8) are for jets (not partons). If the spectra was for partons, the interpretation of Eq. (4.8) would be straight forward. We produce a parton, and then it losses energy to the medium (and we sum over the contribution of both quarks and gluons). The difference for jets are that they are not produced, and then quenched. The quenching happens simultaneously as the jet fragments. This makes the factorization in Eq. (4.8) a bit less accurate, but we still expect it to be powerful.

The task will then reduce to putting parametrizations on the distributions $D_i(\varepsilon) = D_i(\varepsilon, \theta_i)$, and then use Bayesian inference to constrain the parameters. Even though we call the jets either a quark or a gluon jet, it is important to remember that the jets will contain both type of particles. Both quarks and gluons will radiate gluons, a gluon will on average emit C_A/C_F more gluons than a quark. A gluon can also split up to a quark and anti-quark pair. All in all this means that we don't expect the energy-loss distributions to be very similar to those we got in Sec 2.4. Before we move on to the Bayesian analysis, we need to generate the parton tagged jets spectra. This will be done in the event generator called Pythia.

4.2.1 Classifying Jets in Pythia

The first step will be to generate the spectra for the quark and gluon jets in Pythia (version 8) [9]. Pythia is an Monte Carlo event generator where we for instance can simulate proton-proton collisions. The nice thing about Pythia is that once a pp-collisions is simulated, we have access to the whole history of the event. We can trace every single final state particles history.

We set Pythia to simulate proton-proton collisions at center of mass energy $\sqrt{s} = 5.02$ TeV. To save time, we turn off hadronization. To reduce noise we also turn off initial state radiation (ISR), multi-parton interactions (MPI) and soft QCD. Using these settings we get a very clean event. We can then start generating the jet spectra using the anti- k_t algorithm described in Sec 1.5. The jet cone parameter R is set to 0.4 and we only record jets within the rapidity range $|y| < 2.8$, this is the same as the data used from ATLAS.

How do we classify a jet as a quark jet or a gluon jet? The approach taken in the thesis is to try matching the jets momentum direction with that of an outgoing hard parton. If the jet and the partons momentum direction match each other within a certain range, the jet will be classified as the same type as the parton. The matching is done in the following way. Once the jet finder

has analyzed the event, and has found n number of jets. We then find the angle θ_{ij} between the momentum of jet i and hard parton j . If this angle is less than the cone parameter R , the jet is classified as initiated by the hard parton.

The upper panel of Fig 4.5 shows the resulting jet spectra after simulating 10 million events in Pythia. As before, the solid lines shows a fit to the generated data with Eq. (4.7). The lower panel shows the fraction of quark and gluon jets $\frac{dN^i}{dp_T} / \frac{dN^{tot}}{dp_T}$ for $i \in (q, g)$. We see that for $p_T < 200$ GeV, the spectrum is dominated by gluon jets. While for $p_T > 200$ GeV, it is dominated by quark jets. We can understand this by looking at the relation of the momentum fractions x_i of the two interacting partons in the colliding protons to the transverse momentum p_T of the two outgoing partons, $x_i \propto p_T$ (Eq. (7.23) in [14]). For high x_i (high p_T), the PDFs of the valence quark dominate [12]. This means that most of the collisions will have quarks as the initial particles in the hard scattering vertex, which again will give a dominance of quarks as the outgoing particles.

The sum of the two contributions should sum up to the total jet spectrum. However, we see (in the lower panel) that there is slight deviation in the low p_T range. This has a simple explanation. The total spectrum will contain small jets that are results of radiation, and will not be matched to the outgoing partons. They will naturally have a small p_T , and its the reason why we see a slight deviation in the lower parts of the spectrum. However this is very small effect.

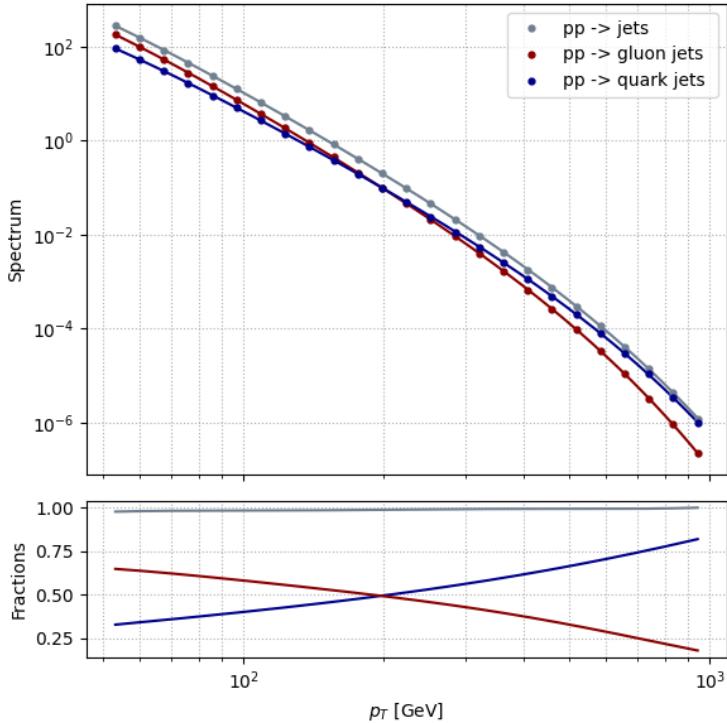


Figure 4.5: Shows the the differnet p_T jet spectra generated Pythia with centre of mass energy $\sqrt{s} = 5.02$ TeV. (upper panel) Red points are the gluon jets. Blue points are the quark jets, and the grey point are the total jet spectrum. The red, blue and grey lines are the fit to the respective data with Eq. (4.7). (lower panel) Shows the fractional contributions. Red line is the gluon spectrum divided by the total, and the blue line show the same for the quark jets. Gray line shows the sum of the quark jets and the gluon jets over the total spectrum.

As a sanity check, Fig 4.6 shows that the jet spectrum generated in Pythia fits pretty well to the measured cross section for proton-proton collisions.

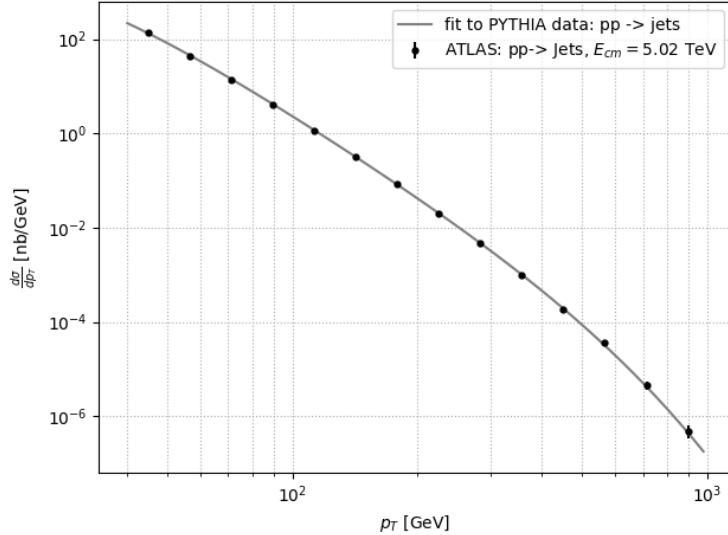


Figure 4.6: Shows how well the Pythia jet spectrum fits (up to constant) with the measured jets cross section.

4.2.2 Nuclear PDFs

An extra piece of physics we want to introduce into our quenching model are nuclear parton density functions (nPDFs). This is the PDF of a bound nucleon in a nucleus. It is defined as [15]

$$f_i^A(x, Q^2) = R_i^A(x, Q^2) f_i(x, Q^2), \quad (4.9)$$

where $f_i(x, Q^2)$ is the free nucleon PDF (those used in Eq. (1.49)), and i indicates the particle flavour. The quantity $R_i^A(x, Q^2)$ is the nuclear modification to $f_i(x, Q^2)$. The left panel of Fig 4.7 shows a typical parametric fit of $R_i^A(x, Q^2)$.

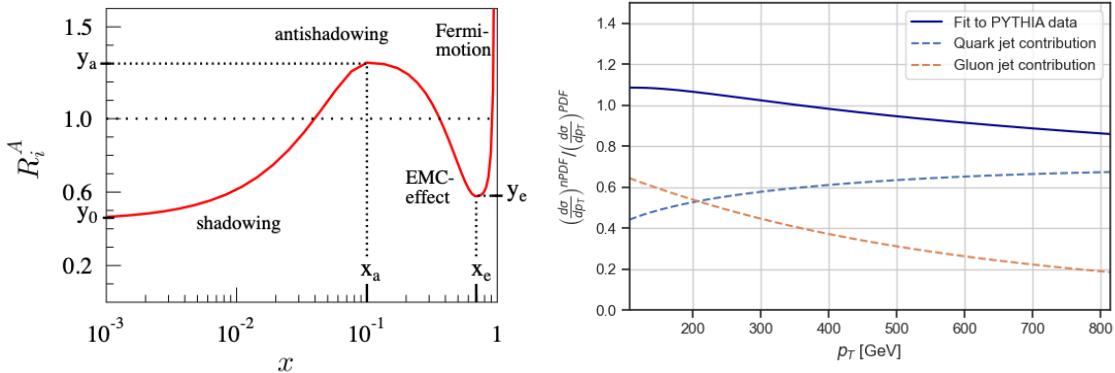


Figure 4.7: Left panel shows a typical parametric fit for the nuclear modification R_i^A in Eq. (4.9), plot taken from [15]. The right panel shows the effect of including nPDFs to the jet spectra in Pythia. The y-axis shows the ratio of the jet spectrum with nPDF over the spectrum with just the regular PDF. It also shows the relative quark and gluon contribution.

Nuclear parton densities can also be included in the Pythia simulation. The right panel of Fig 4.7 shows the effect of including nPDFs. We see that including nPDFs into our model for the medium spectrum will help us with some suppression in the high p_T domain. While the opposite is the case for the low p_T range. This can also be understood by the relation $x \propto p_T$. For high values of x , we see a dip in R_i^A , giving the suppression shown in the right panel in Eq. (4.9). While for x values around 10^{-1} , we have the opposite effect.

4.2.3 Log-normal Energy-loss Distributions

Inspired from Sec 2.4, we will now use log-normal energy-loss distributions for the quark and gluon jets. The log-normal distribution reads:

$$D_i(\varepsilon|\mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i\varepsilon} \exp\left[-\frac{(\log \varepsilon - \mu_i)^2}{2\sigma_i^2}\right]. \quad (4.10)$$

In Sec 2.4 the energy-loss distribution was written as $\omega_c D(x) = \text{LogN}(x|\mu, \sigma)$ where $x = \varepsilon/\omega_c$. We therefore rewrite it as $D(\varepsilon) = \text{LogN}(\varepsilon|\mu, \sigma)$ where $\mu = \mu' + \log \omega_c$. We then recognize that the parameter μ_i in Eq. (4.10) will be $\mu'_i + \log \omega_c$ with the medium parameter ω_c , defined in Eq. (2.14). This means that we know the ω_c dependence of our parameter μ .

We now have a model for $R_{AA}(p_T|\theta)$ with parameters $\theta = [\mu_1, \mu_2, \sigma_1, \sigma_2]$, where 1 is the quark index and 2 is the gluon index. To give the model some extra constraining power (if there is room for it), we introduce an extra parameter δ to scale the $\sigma_{R_{AA}}$ parameter in the likelihood, which again will be chosen to be a Gaussian

$$\mathcal{L}\left(R_{AA}^{\text{data}}(p_T) \middle| \theta, \delta\right) = \mathcal{N}\left(R_{AA}^{\text{data}}(p_T) \middle| \mu = R_{AA}^{\text{model}}(\theta, p_T), \sigma = \delta \cdot \sigma_{R_{AA}}\right). \quad (4.11)$$

Info on how the integral for computing R_{AA} is done numerically can be found in Sec A.2. The priors for θ will be very ignorant. We choose the prior for all the parameters to be uniform, from 0 to 5. While the prior for the δ parameter will be uniform from 0 to 1.

$$\begin{aligned} \mu_i &\sim \text{Uniform}[0, 5], \\ \sigma_i &\sim \text{Uniform}[0, 5], \\ \delta &\sim \text{Uniform}[0, 1]. \end{aligned} \quad (4.12)$$

Fig 4.9 shows the individual posterior distributions (left panels) and the traceplots (right panels). The results are from running the HMC sampling for 20 000 samples (with 10 000 samples for tuning). As we can see, the data struggles to constrain some of the parameters. Specially the parameters μ_1 and σ_1 in the energy-loss distribution $D_1(\varepsilon)$ for the quark jets. This might not be so surprising, because we know from Fig 4.5 that for p_T above 200 GeV the quark jet spectrum is the dominant one. However, at this p_T range the R_{AA} measurements are very uncertain. The parameter δ helps us to constrain the model also, without it the uncertainty in the parameters would be even worse.

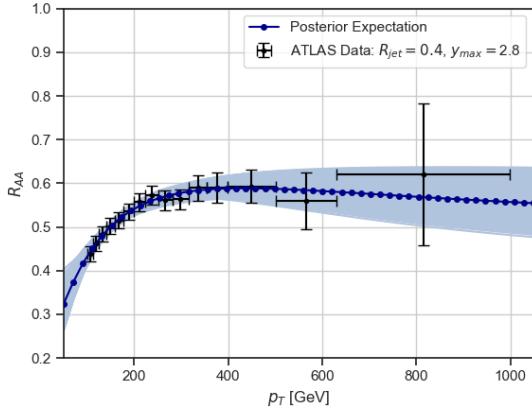


Figure 4.8: Shows the Bayesian fit for the model defined above, together with the R_{AA} data from ATLAS.

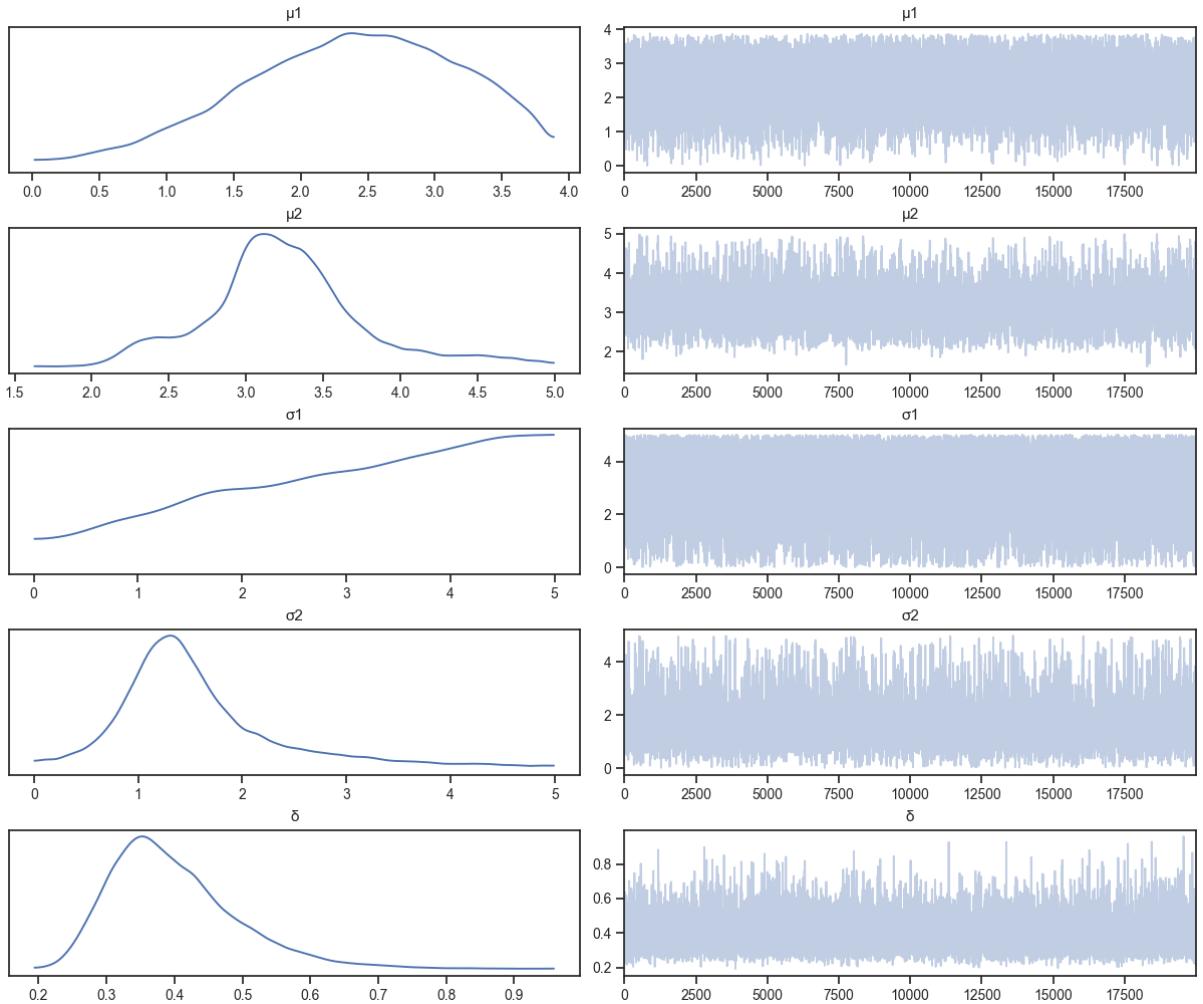


Figure 4.9: The left panels shows the marginalized posterior distributions of the different parameters in the model. The right panels shows the traceplots for the different

Fig 4.8 show the model prediction together with the data used to do the Bayesian fit. We see that the model fits pretty well with the data, even though it struggles to constrain the

parameters. We can therefore conclude that the model contains too many parameters for the available data.

We therefore try to build a model containing more physics. Recall the BDMS estimate of the mean energy-loss of a parton with Casimir C_R traveling through a static QGP.

$$\langle \varepsilon \rangle = \frac{\alpha_s C_R}{2} \omega_c. \quad (4.13)$$

Now we impose that the mean energy-loss of the quark/gluon jets follow the same structure. The mean energy-loss will be proportional to C_R . We then write

$$\exp\left[\mu_i + \frac{1}{2}\sigma_i^2\right] = C_R\omega \quad \rightarrow \quad \mu_i = \log[C_R\omega] - \frac{1}{2}\sigma_i^2, \quad (4.14)$$

where $C_R = C_F$ for quark jets (index $i = 1$), and $C_R = C_A$ for gluon jets (index $i = 2$). This means that we can reduce the number of parameters to 4, including the δ parameter. ω will be a parameter that is shared between the two energy-loss distributions. Because of the expression in Eq. (4.13), it is fair to assume that ω is proportional to the medium parameter ω_c (defined in Eq. (2.14)). Now we have the following parameters to constrain: $\theta = [\omega, \sigma_1, \sigma_2, \delta]$. The priors for the parameters will be

$$\begin{aligned} \omega &\sim \text{Uniform}[0, 100], \\ \sigma_i &\sim \text{Uniform}[0, 5], \\ \delta &\sim \text{Uniform}[0, 1]. \end{aligned} \quad (4.15)$$

We start by looking at the posterior density with the corner plot shown in Fig 4.10. For this MCMC run we used HMC with 80 000 samples and 5000 samples for tuning. The mean and standard deviation of the parameters in the posterior is

ω	σ_1	σ_2	δ
57.75 ± 21.38	1.58 ± 0.51	2.04 ± 0.09	0.41 ± 0.09

We see that the δ parameter exhibits little to none correlation with the other parameters. Which is a good thing, because it is in the model only to give it more constraining power. The parameters σ_1 and σ_2 exhibits some correlation, but not very much. The pairs that show significant correlations are σ_1 and ω , and σ and ω . The mean value of ω is a bit misleading, because the posterior for this parameter has a very long tail. This is an artifact of a very wide prior range. The important thing to notice is that ω values below 20 is completely ruled out by the data, and it has its peak around 30. Lets look at how well this explains the data. By looking at the δ parameter, we can already anticipate that it will be a good fit. The reason for this is that its peak value is well below 1, which means that it managed to "shrink" the errorbars from the experimental data.

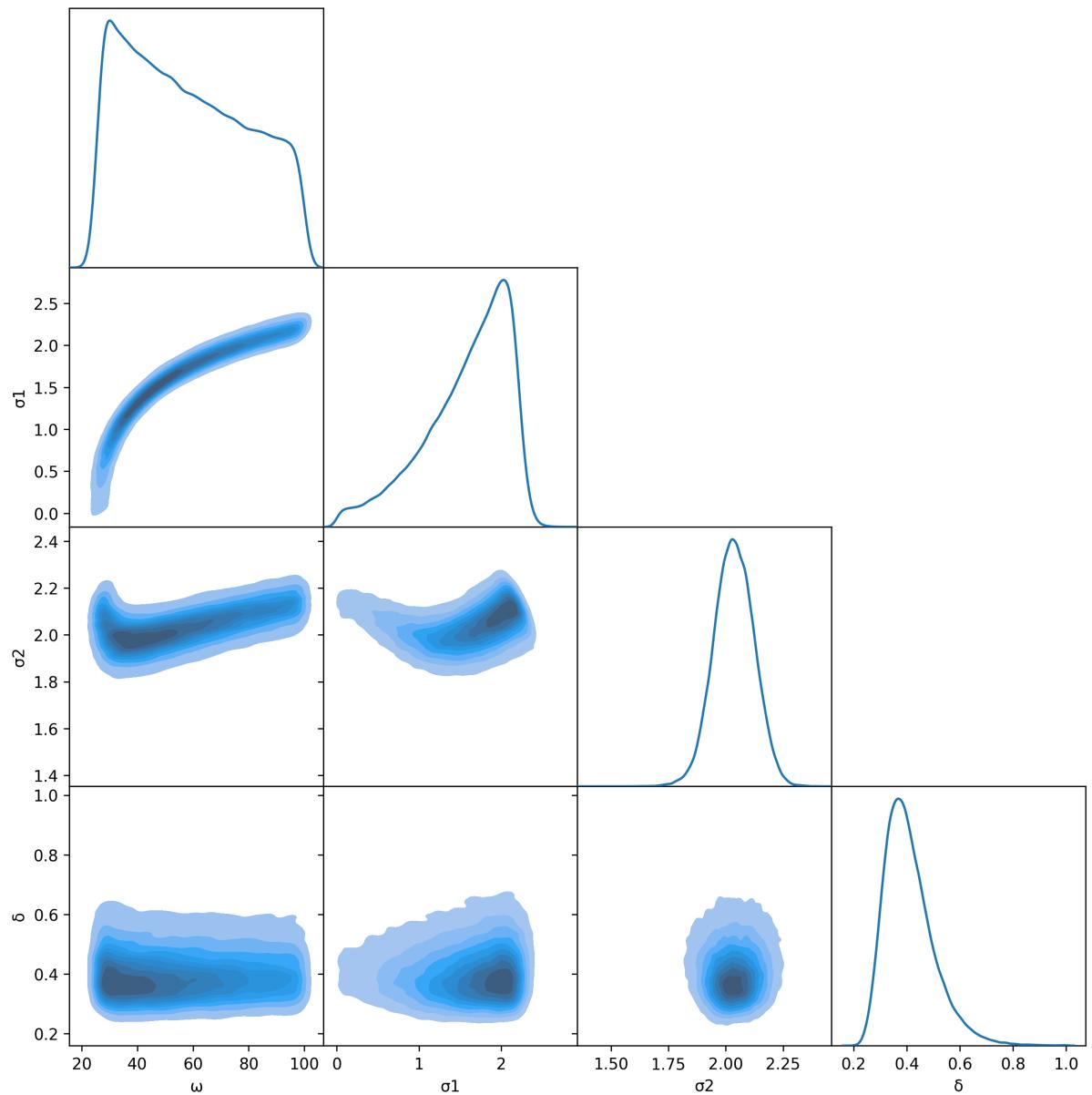


Figure 4.10: Corner plot of the posterior of $\theta = [\omega, \sigma_1, \sigma_2, \delta]$. The upper panels shows the posterior distribution of the individual parameters. The rest of the panels shows the correlations between the parameters

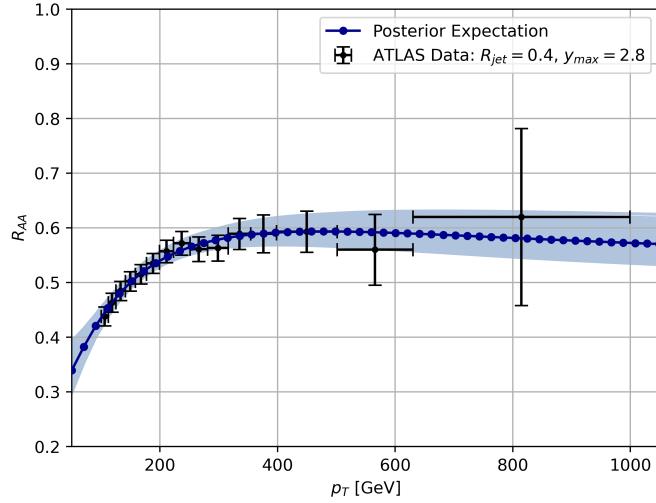


Figure 4.11: Show the Bayesian fit vs the ATLAS data. Dark-blue dotted curve shows the posterior expectation value

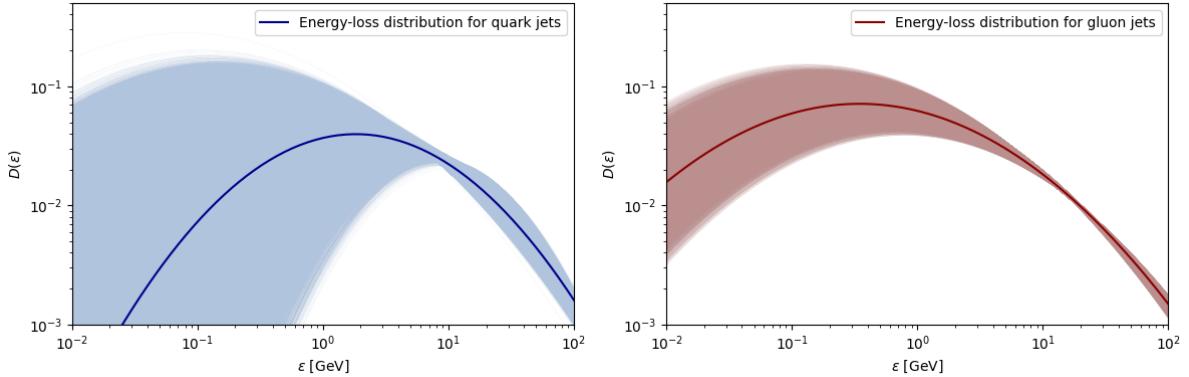


Figure 4.12: Shows the energy-loss distributions $D_i(\varepsilon)$. Blue curve(s) (left panel) is for the quark jets, and red curve(s) (right panel) for the gluon jets. Figure shows all the models within one standard deviation of the parameter means.

Fig 4.11 shows that the model fits very well to the experimental data. However, the energy-loss distributions shown in Fig 4.12 exhibits some unexpected features. $D_g(\varepsilon)$ (dark-red) has a peak that is located below 1 GeV, and even more surprising is that it is lower than the peak for $D_q(\varepsilon)$ (dark-blue). This is the opposite of the result in Sec 2.4. This might be an artifact of the parameter constraint given in Eq. (4.14). Another thing worth mentioning is that the spread (uncertainty) of the quark distributions is much bigger than the gluons. This is again because the uncertainty in the data where the quark spectrum is dominant ($p_T > 200$ GeV).

4.2.4 Model Validation

To truly validate the model, we should see if it can predict a different data set. Lets try to predict the measurements of R_{AA} at the same centrality class 0 – 10% from ATLAS at $\sqrt{s_{NN}} = 2.67$ TeV [2]. First we must generate new spectra for the quark and gluon jets at the correct centre of mass energy and correct rapidity range $|y| < 2.1$. This is done the same way as described

before. Fig 4.13 shows the resulting jet spectra at $\sqrt{s} = 2.67$ TeV. The jet spectra are quite similar, but the quark and gluon fractions are a bit different.

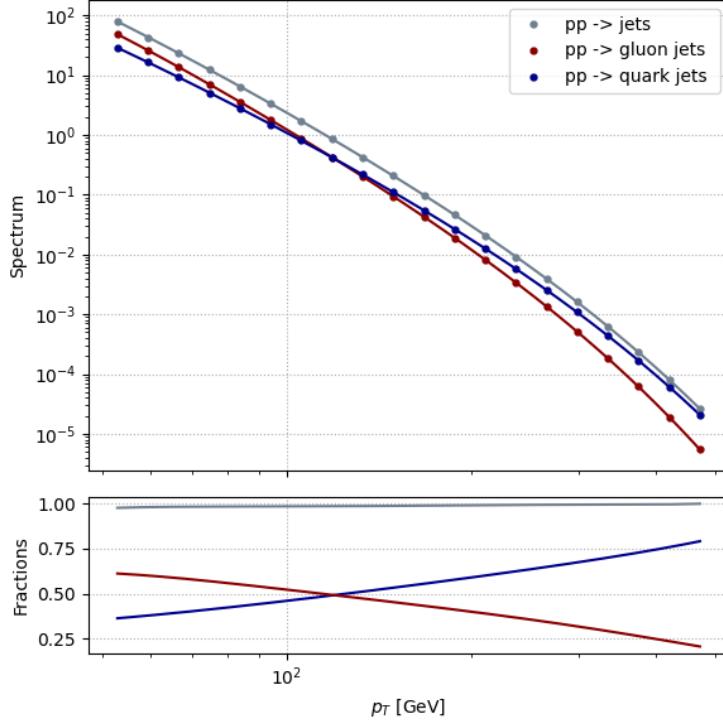


Figure 4.13: Same as Fig 4.5, but with center of mass energy $\sqrt{s} = 2.67$ TeV and rapidity range $|y| < 2.1$.

We then use the model fitted in Fig 4.11 to predict the new data set. The left panel of Fig 4.14 show the model predictions against the data. The p_T dependence of the model seems to be pretty good. However, it looks like the model overestimates the amount of quenching.

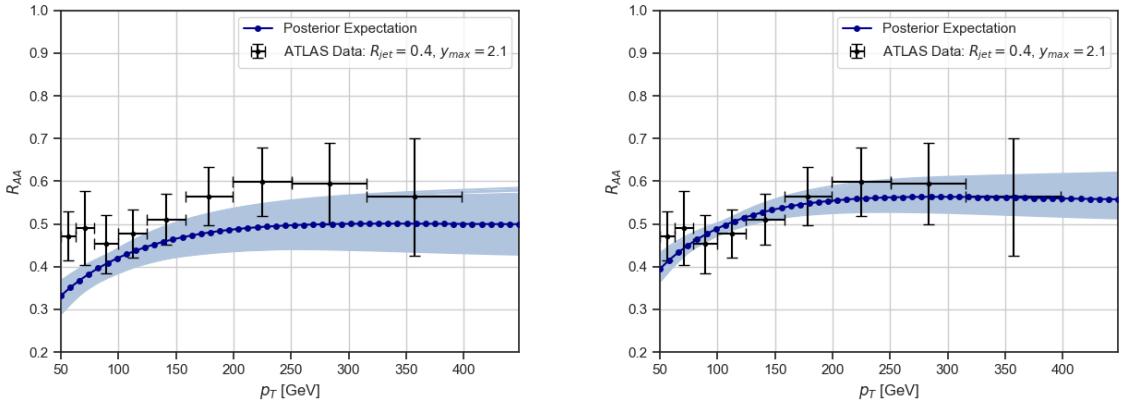


Figure 4.14: (left panel) Shows model prediction vs R_{AA} measurements from ATLAS at $\sqrt{s_{NN}} = 2.67$ TeV with jets with $R = 0.4$ and $|y| < 2.1$. (right panel) Shows the same as the left panel, but with the ω parameter reduced by 30%.

It seems that the model has "overfitted" a bit to the data at $\sqrt{s_{NN}} = 5.02$ TeV. It might be

that at $\sqrt{s_{NN}} = 2.67$ TeV a medium with a different ω_c (which consequently gives a different ω) is produced. To demonstrate that this might be the case, lets do the exact same prediction as in Fig 4.14, but with the parameter ω reduced with 30%. The result for this is given in the right panel of Fig 4.14, the model can now explain the data very well.

We can get a very rough estimate of the difference in temperature by the fact that $\hat{q} \propto T^3$ [13]. Assuming that ω is proportional to ω_c which again is $\propto \hat{q}$,

$$0.7 = \frac{\omega^{2.76\text{GeV}}}{\omega^{5.02\text{GeV}}} = \frac{\omega_c^{2.76\text{GeV}}}{\omega_c^{5.02\text{GeV}}} = \frac{\hat{q}^{2.76\text{GeV}}}{\hat{q}^{5.02\text{GeV}}} = \frac{(T^{2.76\text{GeV}})^3}{(T^{5.02\text{GeV}})^3} \quad \rightarrow \quad \frac{T^{2.76\text{GeV}}}{T^{5.02\text{GeV}}} = 0.7^{1/3} \approx 0.89 \quad (4.16)$$

This means that if the difference of the two centre of mass energies is the QGP (different ω_c) that is produced. The difference in temperature will be about 10%. This might be at the upper end of what one might expect the temperature difference to be. However it is not that big of a difference. It might be that some of the difference in the parameter ω comes from a poor parametrization of the energy-loss distributions.

It is worth mentioning that if we try to predict the R_{AA} data at $\sqrt{s_{NN}} = 2.67$ TeV using the model from [17] in Fig 4.3 we also get a similar overestimation of the amount of jet suppression as shown in Fig 4.15. This may hint that a slightly different medium is produced at $\sqrt{s} = 2.76$ TeV.

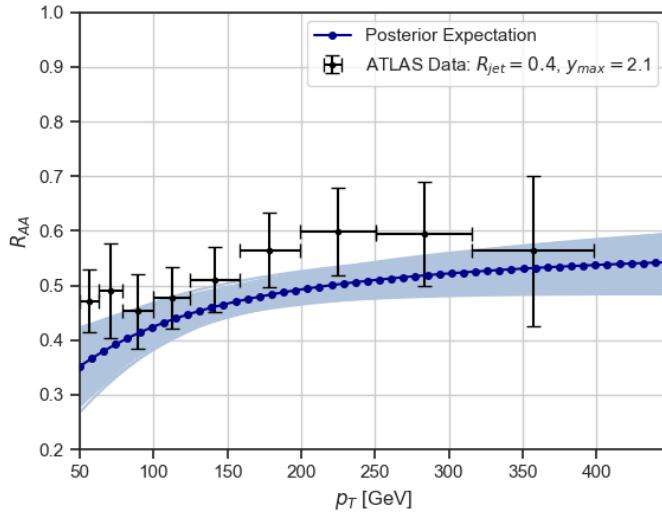


Figure 4.15: Shows the prediction of the model from [17] at $\sqrt{s} = 2.76$ TeV.

Summary and Outlook

In this thesis I have represented a first attempt at extracting the energy-loss distributions for quark and gluon jets in heavy-ion collisions using Bayesian inference. The motivation behind this is that quarks and gluons are fundamental particles appearing the standard model of particle physics. While quarks and gluons are not observable by their own, they are effective probes of the QGP. By connecting the observed jets to the fundamental particles in the underlying theory, one has the potential to learn a great deal about the plasma.

First we took a look at the BDMPS induced gluon spectrum for a hard parton traveling though a uniform brick of QGP. We looked at its behavior in the low and high energy regime. Through the Poisson approximation of the energy-loss distribution we could connect the BDMPS spectrum to the total energy-loss ε . We then saw that the energy-loss distribution $D(\varepsilon)$ for a single parton traveling through the uniform brick exhibits a heavy tail towards higher ε . This means that the mean energy-loss will be much higher than the peak (most probable value). We also saw that the energy-loss distribution of the BDMPS spectrum mimics almost perfect a Log-normal distribution. This gave the motivation for using the same parametrization for the energy-loss distributions for quark and gluon jets.

Since Bayesian inference is the tool we used to constrain the parameters of the energy-loss distributions, we spent a whole chapter discussing the basics of it. The focus of this chapter was to show how this is done in practice using the Markov Chain Monte Carlo methods with the Metropolis-Hastings (MH) algorithm. We looked at two specific step methods for the MH algorithm. First the well known Random Walk Metropolis (RWM), and then the Hamiltonian Monte Carlo (HMC) method. We then argued for why we should use the (HMC) method whenever our models are high-dimensional. The reason for this is that HMC is much more samples efficient than RWM.

As a warm up, we re-derived the results of [17] for the R_{AA} data at $\sqrt{s} = 5.20$ TeV using HMC rather than RWM (which they used). Then we spent some time defining what a quark/gluon jet is, and how they were generated in Pythia. We then went on to define two models, both with log-normal energy-loss distributions for the quark and gluon jets. For the first and most naive model, we saw that the data struggled to constrain the parameters for the quark energy-loss distribution, and argued that this might be because of the high uncertainty in the data where the quark-jet cross section is dominant. We therefor concluded that this model has to many parameters.

We therefore tried to introduce more physics (based on the BDMS estimate) into the second model. This effectively reduced the number of parameters, which gave better constraint on the

model parameters. This model fitted extremely well to the data. However, the resulting energy-loss distributions looked a bit surprising. One would expect the peak of the gluon distribution to be at a higher energy-loss than the quark distribution, just like the BDMS estimate gives, but this was not the case.

The next thing we did was to try and validate the model on a different data set at a lower centre of mass energy $\sqrt{s} = 2.76$ TeV. We saw that the model predicts the p_T dependence of R_{AA} pretty well. However, it overshoots the amount of quenching by a bit. We also saw that this was the case for the model defined by [17]. This might be an indication of over-fitting to the medium produced at a specific centre of mass energy. It is also possible that the both of the models are bad, and don't generalize well. A way to test this would be to evaluate the model on a different observable at the same centre of mass energy.

The ultimate goal would be to get energy-loss distributions $D_i(\varepsilon)$ that would be consistent with all the data sets, at least for the same type of medium (same centrality and centre of mass of the collisions). That is, we want to be able to constrain the parameters on one set, and then be able to predict the rest. This would then show the consistency of the factorization given in Eq. (4.8). Another observable that could be used for either validation or parameter constraining, is the nuclear modification factor for γ -triggered jets. This is when a jet and high energetic photon is created back to back. As described in Chapter 2, since the photon is colorless it will not be affected by the medium¹. Data for γ -triggered jets for PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV from CMS can be found in [24]. To do this analysis one would then have to generate quark and gluon spectra for this process in Pythia, as the jet spectra are treated as an input to the model. Once this is done, we could do an ultimate check of consistency of the given model, and hopefully extract some new information about the QGP.

¹ It might interact with the electrical charge of the medium, but this effect will be much weaker than the QCD interaction

Acknowledgement

I would like to express my deepest appreciation to my supervisor Konrad Tywoniuk, Researcher at Department of Physics and Technology at UiB. He gave great counseling and guidance during the last year, and was always willing to discuss and answer questions about physics. He also taught General Relativity (PHYS301) and Relativistic Heavy Ion Physics (PHYS333), two courses that I very much enjoyed.

I also want to thank PhD student Adam Takacs for helping me install PYTHIA, and PhD student Alexandre Da Fonseca Falcao for usefull discussions about the project. I am also grateful to all my fellow master students for great lunch conversations and good times. A special thanks goes to master students Kristoffer and Sigurd for helping me make Fig 2.3 and for fruitful discussions on the blackboard.

Last but not least, I want to thank Evelinn, the MVP of my life.

Appendix A

Numerical Integration of the Convolution-integral

To do the Bayesian analysis we need to evaluate the convolution

$$\frac{d\sigma^{AA}}{dp_T} = \int d\varepsilon D(\varepsilon) \frac{d\sigma^{pp}}{dp_T}(\varepsilon + p_T), \quad (\text{A.1})$$

which has to be done numerically. We use the package `pymc3` to do the Bayesian inference. This package require us to give an algebraic expression for R_{AA} to do the sampling. We therefore use Gaussian quadrature to approximate the integral in Eq. (A.1).

A.1 Gamma Distribution

To reproduce the results of [17], we have the integral

$$\frac{d\sigma^{AA}}{dp_T} = \int_0^\infty dx \frac{\alpha^\alpha x^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)} \frac{d\sigma^{pp}}{dp_T}(\langle\varepsilon\rangle \cdot x + p_T), \quad (\text{A.2})$$

where $x = \varepsilon / \langle\varepsilon\rangle$, and $\langle\varepsilon\rangle = \beta p_T^\gamma \ln(p_T)$. Introducing the variable change $x' = \alpha x \rightarrow dx = \frac{dx'}{\alpha}$ gives

$$\begin{aligned} \frac{d\sigma^{AA}}{dp_T} &= \int_0^\infty \frac{dx'}{\alpha} \frac{\alpha^\alpha (x'/\alpha)^{\alpha-1} e^{-x'}}{\Gamma(\alpha)} \frac{d\sigma^{pp}}{dp_T} \left(\frac{\langle\varepsilon\rangle}{\alpha} x' + p_T \right) \\ &= \int_0^\infty dx' \frac{x'^{\alpha-1}}{\Gamma(\alpha)} \frac{d\sigma^{pp}}{dp_T} \left(\frac{\langle\varepsilon\rangle}{\alpha} x' + p_T \right) e^{-x'}. \end{aligned} \quad (\text{A.3})$$

This integral can now be approximated using Gauss–Laguerre quadrature, which can approximate integrals on the form (Chapter 25 in [?]):

$$\int_0^\infty dx f(x) e^{-x} \approx \sum_{i=1}^n w_i f(x_i) \quad (\text{A.4})$$

In our analysis we use the same number off integration points as [17] uses, $n = 30$.

A.2 Log-normal Distribution

Lets look at the case inspired by [5], where the energy-loss distributions is given by a log-normal distribution.

$$D(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma\varepsilon} \exp\left[-\frac{(\ln\varepsilon - \mu)^2}{2\sigma^2}\right]. \quad (\text{A.5})$$

The convolution we have to compute is then:

$$\frac{d\sigma^{AA}}{dp_T} = \int_0^\infty d\varepsilon \frac{1}{\sqrt{2\pi}\sigma\varepsilon} \exp\left[-\frac{(\ln\varepsilon - \mu)^2}{2\sigma^2}\right] \frac{d\sigma^{pp}}{dp_T}(\varepsilon + p_T). \quad (\text{A.6})$$

This can be well approximated using Gauss–Hermite quadrature where one can approximate integrals in the form (chapter 25 in [3])

$$\int_{-\infty}^\infty dx f(x)e^{-x^2} \approx \sum_{i=1}^n w_i f(x_i). \quad (\text{A.7})$$

It is quite easy to rewrite eq(A.7) to this form. First lets introduce the variable change $y = \ln(\varepsilon) \rightarrow \varepsilon dy = d\varepsilon$. This gives

$$\frac{d\sigma^{AA}}{dp_T} = \int_{-\infty}^\infty dy \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right] \frac{d\sigma^{pp}}{dp_T}(e^y + p_T) \quad (\text{A.8})$$

Notice that we had to change the lower integration limit. Now we can use the variable change $x = \frac{(y-\mu)^2}{\sqrt{2}\sigma} \rightarrow dy = \sqrt{2}\sigma dx$. The convolution then becomes

$$\frac{d\sigma^{AA}}{dp_T} = \int_{-\infty}^\infty dx \frac{e^{-x^2}}{\sqrt{\pi}} \frac{d\sigma^{pp}}{dp_T}(e^{\sqrt{2}\sigma x + \mu} + p_T) \approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^n w_i \frac{d\sigma^{pp}}{dp_T}(e^{\sqrt{2}\sigma x_i + \mu} + p_T). \quad (\text{A.9})$$

Bibliography

- [1] AABOUD, M., ET AL. Measurement of the nuclear modification factor for inclusive jets in Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV with the ATLAS detector. *Phys. Lett. B* **790** (2019), 108–128.
- [2] AAD, G., ET AL. Measurements of the Nuclear Modification Factor for Jets in Pb+Pb Collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV with the ATLAS Detector. *Phys. Rev. Lett.* **114**, 7 (2015), 072302.
- [3] ABRAMOWITZ, M., AND STEGUN, I. A., Eds. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, tenth printing ed. U.S. Government Printing Office, Washington, DC, USA, 1972.
- [4] ANDRIEU, C., FREITAS, N., DOUCET, A., AND JORDAN, M. An introduction to mcmc for machine learning. *Machine Learning* **50** (01 2003), 5–43.
- [5] ARLEO, F. Tomography of cold and hot qcd matter: tools and diagnosis. *Journal of High Energy Physics* **2002**, 11 (Nov 2002), 044–044.
- [6] ARNOLD, P. Simple formula for high-energy gluon bremsstrahlung in a finite, expanding medium. *Physical Review D* **79**, 6 (Mar 2009).
- [7] BAIER, R., DOKSHITZER, Y. L., MUELLER, A. H., AND SCHIFF, D. Quenching of hadron spectra in media. *Journal of High Energy Physics* **2001**, 09 (Sep 2001), 033–033.
- [8] BETANCOURT, M. A conceptual introduction to hamiltonian monte carlo, 2017.
- [9] BIERLICH, C., ET AL. A comprehensive guide to the physics and usage of PYTHIA 8.3.
- [10] BUSZA, W., RAJAGOPAL, K., AND VAN DER SCHEE, W. Heavy ion collisions: The big picture and the big questions. *Annual Review of Nuclear and Particle Science* **68**, 1 (oct 2018), 339–376.
- [11] CACCIARI, M., SALAM, G. P., AND SOYEZ, G. The anti- k_t jet clustering algorithm. *JHEP* **04** (2008), 063.
- [12] COOPER-SARKAR, A. M. Extraction of the proton parton density functions using a NLO-QCD fit of the combined H1 and ZEUS inclusive DIS cross sections. In *16th International Workshop on Deep Inelastic Scattering and Related Subjects* (8 2008), p. 25.
- [13] D’ENTERRIA, D. 6.4 jet quenching. In *Relativistic Heavy Ion Physics*. Springer Berlin Heidelberg, 2010, pp. 471–520.

- [14] ELLIS, R. K., STIRLING, W. J., AND WEBBER, B. R. *QCD and collider physics*, 1st ed. Cambridge monographs on particle physics, nuclear physics and cosmology. Cambridge University Press, 2003.
- [15] ESKOLA, K. J., PAUKKUNEN, H., AND SALGADO, C. A. EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions. *JHEP* 04 (2009), 065.
- [16] FLORKOWSKI, W. *Phenomenology of ultra-relativistic heavy-ion collisions*. World Scientific Publishing Company, 2010.
- [17] HE, Y., PANG, L.-G., AND WANG, X.-N. Bayesian extraction of jet energy loss distributions in heavy-ion collisions. *Physical Review Letters* 122, 25 (Jun 2019).
- [18] HOMAN, M. D., AND GELMAN, A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.* 15, 1 (jan 2014), 1593–1623.
- [19] NEAL, R. M. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011, ch. 5: MCMC Using Hamiltonian Dynamics.
- [20] NILSEN, M. M. All code can be found on github. <https://github.com/MathiasMNilsen/MasterProject>, 2022.
- [21] SALVATIER, J., WIECKI, T. V., AND FONNESBECK, C. Probabilistic programming in python using pymc3. *PeerJ Computer Science* 2 (2016), e55.
- [22] SCHARENBERG, R. P., SRIVASTAVA, B. K., HIRSCH, A. S., AND PAJARES, C. Hot Dense Matter: Deconfinement and Clustering of Color Sources in Nuclear Collisions. *Universe* 4, 9 (2018), 96.
- [23] SCHWARTZ, M. D. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.
- [24] SIRUNYAN, A. M., ET AL. Study of jet quenching with isolated-photon+jet correlations in PbPb and pp collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *Phys. Lett. B* 785 (2018), 14–39.
- [25] SKANDS, P. Introduction to qcd. *Searching for New Physics at Small and Large Scales* (Sep 2013).
- [26] SNELLINGS, R. Elliptic Flow: A Brief Review. *New J. Phys.* 13 (2011), 055008.
- [27] SPEAGLE, J. S. A conceptual introduction to markov chain monte carlo methods, 2019.
- [28] SPOUSTA, M., AND COLE, B. Interpreting single jet measurements in Pb + Pb collisions at the LHC. *Eur. Phys. J. C* 76, 2 (2016), 50.