Contents

Chapter 1

Multi-dimensional integration in HEP

In this chapter we focus at first on Monte Carlo techniques applied to the problem of multi-dimensional numerical integration. We discuss the two main methods which involve importance sampling and stratified sampling. Secondly we give a brief overview on High Energy Physics (HEP) arguments, with particular attention on the computation of physical observables as a series of perturbative terms which involve high-dimensional integrals. Finally we present the state-of-art of MC integration applied to HEP discussing the current problems and limitations which is facing the High-Luminosity LHC programme.

1.1 Monte Carlo methods for multi-dimensional integration

Monte Carlo (MC) methods are a powerful tool which can provide the answer to a problem by simply running a simulation on the system studied. In the field of multi-dimensional integration techniques MC methods are the solution of choice, since contrary to the standard numerical integration formulas the error on the integral does not scale with the dimension.

1.1.1 A first example of Monte Carlo integration

In the case of multi-dimensional integration we are interested in performing the following integral I:

$$I = \int_V f(\mathbf{x}) d\mathbf{x} \tag{1.1}$$

where V is the domain of the integration and f is a function of n variables $\mathbf{x} = (x_1, x_2, ..., x_n)$.

When performing the integral I the simulation comes down to a sampling of the integrand function. First we need to generate a set of random points $\mathbf{x}_i$ which belong to the integration domain V. The simplest way to performe the sampling is to pick random points uniformly distributed in the volume V.

An estimate of the integral using N random points can be given as:

$$I \approx I_{\mathrm{MC}} = V\frac{1}{N}\sum_{\mathbf{x}_i \in V} f(\mathbf{x}_i) = V\langle f\rangle \tag{1.2}$$

where $\langle f\rangle$ denotes the arithmetic average of the function f.

I_{MC} is a random number, whose value depends on the sampled points, whose mean is given by the exact value of the integral I and the variance is given by:

$$\sigma_I^2 = \frac{1}{N}\Big[V\int_V f^2(\mathbf{x})d\mathbf{x} - I^2\Big] \tag{1.3}$$

This variance is asymptotically related to the variance of the random value I_{MC}, therefore we can estimate the value of σ_I^2 in the limit of large N as:

$$\sigma_I^2 \approx \sigma_{\mathrm{MC}}^2 = \frac{1}{N-1}\Big[V^2\langle f^2\rangle - I_{MC}^2\Big] \tag{1.4}$$

In both cases we can observe that the standard deviation decreases as the sample increased as $N^{-\frac{1}{2}}$ regardless of the dimension of the volume V. This is a remarkable feature typical of MC integration which differs from the standard quadrature techniques where the error increases with the dimension.

MC integration can also deal with another problem of quadrature integration: complicated boundaries. Suppose that we need to integrate a function h over some complicated region H, for which the random sampling becomes challenging. In this particular case we can easily overcome such problem by replacing the region H with a new volume G that includes H more suitable for the process of sampling. After that all we need to do is to replace also the function h with a new function g defined as:

$$g(\mathbf{x}) = \begin{cases} h(\mathbf{x}) & \text{if } \mathbf{x} \in H \\ 0 & \text{if } \mathbf{x} \in G - H \end{cases} \tag{1.5}$$

Obviously one should try to make the new region G not too oversized with respect to H, because every point $\mathbf{x} \in G - H$ will contain no information about the integrand. Therefore the number of effective points used during the sampling N will reduce raising the error in Eq.(1.4).

This first MC integrator has a few disadvantages.

Firstly we have already observed that the error decreases as the square root of the number of sampled points, this implies that if the accuracy requirements are high we will need to increase the size of the sampling. Dealing with huge sizes of sampled points could be, even for a computer with large memory and a fast processor, challenging and we expect that the process of sampling could take a significant amount of time.

Secondly one will probably have to work with large samples, even if the accuracy requirements are modest, when the dimensionality of the integration domain is high. The reason being that if the integrand function is peaked in a small region compared to the volume V, we will need to generate more random points to make sure that the peak is correctly identified; such process

of finding the peaks becomes more challenging in large number of dimensions. This can be seen for example considering the ratio of the volume of a D dimensional hypersphere with unity radius to the D dimensional hypercube with a side of twice the unity radius , which vanishes as the D goes to infinity as:

$$\frac{V_{\text{hypersphere}}}{V_{\text{hypercube}}} = \frac{1}{2^D} \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} \approx \left(\frac{\sqrt{\pi}}{2}\right)^D \xrightarrow{D\to\infty} 0 \tag{1.6}$$

where $\Gamma(x)$ denotes the famous Dirac Gamma function.

In literature this phenomenon known as the *curse of dimensionality*.

1.1.2 Reducing the variance

In the field of MC integrators the main focus is to improve the error estimate in order to achieve more precise results using less number of events. Over the years and also lately with the advent of Machine Learning, several techniques have been proposed, implemented and applied to solve complex multi-dimensional integrals.

In literature there are two main ways of reducing the variance: importance sampling and stratified sampling.

Importance Sampling

In the naive MC integrator the sampling was performed by picking random points uniformly distributed in the integration volume. We have already discussed that if the integrand has a peak in a small region it will be challenging to find it especially because we are selecting points from a uniform distribution.

Suppose that the points $\mathbf{x}_i$ are chosen within the integration volume V with a probability density p correctly normalized:

$$\int_V p(\mathbf{x}) d\mathbf{x} = 1 \tag{1.7}$$

We can calculate the integral I as:

$$I = \int_V f(\mathbf{x}) d\mathbf{x} = \int_V \frac{f(\mathbf{x})}{p(\mathbf{x})} p(\mathbf{x}) d\mathbf{x} = \int_V \frac{f(\mathbf{x})}{p(\mathbf{x})} dP(\mathbf{x}) \tag{1.8}$$

where in the last step we used the transformation $d\mathbf{x} = dP(\mathbf{x})/p(\mathbf{x})$, with $P(\mathbf{x})$ the cumulative distribution of $p(\mathbf{x})$.
From Eq.(1.8) we can deduce that to compute the integral I, instead of using a uniform sampling of the function f, we can also perform a non-uniform sampling of the function f/p in the same integration volume V.
The integral and the relative error will be given by:

$$I \approx I_{\text{MC}} = V\langle f/p \rangle_P \tag{1.9}$$

$$\sigma_I^2 \approx \sigma_{\mathrm{MC}}^2 = \frac{V^2}{N-1}\left[\langle (f/p)^2\rangle_P - \langle f/p\rangle_P^2\right] \tag{1.10}$$

where $\langle\rangle_P$ denotes the average taken with respect to the non-uniform distribution $p(\mathbf{x})$.

This is the concept of *importance sampling*: by changing the distribution of the sampling we can reduce the variance by choosing a suitable $p(\mathbf{x})$.

What is the best choice for the sampling density p? It can be shown that by minimizing the variance, as a functional of sampling density p, the optimal choice for p is to be proportional to $|f|$. This choice is not surprising, in fact by replacing p with $|f|$ in Eq.(1.10) we get a vanishing variance. Moreover in order to satisfy the normalization requirement in Eq.(1.7) the exact solution for p is:

$$p = \frac{|f|}{\int_V |f| d\mathbf{x}} \tag{1.11}$$

As we can see we arrive at a paradox since the optimal sampling density requires the knowledge of $\int_V |f(\mathbf{x})| d\mathbf{x}$ which is the integral that we are trying to compute!
Therefore to minimize the variance the aim is to find a function p that resemble the shape of the integrand function f, this is usually done using adaptive recursive strategies which can provide a better sampling distribution using the sampled points.

Stratified Sampling

Another technique which is a standard one in literature is based on the idea of "stratified sampling". We have already observed that we can estimate the variance of our integral by computing the variance of the random variable I_{MC}; we can exploit this relation by focusing on the average value of the function f over the domain V, denoted by $\langle\!\langle f \rangle\!\rangle$, and the correspoding MC estimate using a uniform sampling $\langle f \rangle$:

$$\langle\!\langle f \rangle\!\rangle = \frac{1}{V}\int_V f(\mathbf{x}) d\mathbf{x} \qquad \langle f \rangle = \frac{1}{N}\sum_{\mathbf{x}_i} f(\mathbf{x}_i) \tag{1.12}$$

We can now rewrite Eq.(1.4) as

$$\mathrm{Var}(f) \equiv \langle\!\langle f^2 \rangle\!\rangle - \langle\!\langle f \rangle\!\rangle^2 = \frac{\langle f^2\rangle - \langle f\rangle^2}{N} \equiv \frac{\mathrm{Var}(\langle f\rangle)}{N} \tag{1.13}$$

Suppose we divide the volume V into two subvolumes V_a and V_b, chosen equal and disjoint, and we sample exactly $N/2$ points in each subvolume. We can formulate another estimator for the mean value of the function, $\langle\!\langle f \rangle\!\rangle$ as the arithmetic mean between the sample average in the two half -regions:

$$\langle f\rangle' \equiv \frac{1}{2}(\langle f\rangle_a + \langle f\rangle_b) \tag{1.14}$$

where $\langle\rangle_a$ denotes the MC estimate of the function for the $N/2$ points belonging to the subvolume V_a and similarly for $\langle\rangle_b$. The variance of the estimator in Eq.(1.14) can be easily computed as:

$$\begin{aligned}\mathrm{Var}(\langle f\rangle') &= \frac{1}{4}\Big[\mathrm{Var}(\langle f\rangle_a) + \mathrm{Var}(\langle f\rangle_b)\Big] && (1.15)\\ &= \frac{1}{4}\left(\frac{\mathrm{Var}_a(f)}{N/2} + \frac{\mathrm{Var}_b(f)}{N/2}\right) && (1.16)\\ &= \frac{1}{2N}\Big[\mathrm{Var}_a(f) + \mathrm{Var}_b(f)\Big] && (1.17)\end{aligned}$$

where $\mathrm{Var}(\langle f\rangle)_i = \langle\!\langle f^2\rangle\!\rangle_i - = \langle\!\langle f\rangle\!\rangle_i^2$ denotes the variance of f limited to the subvolume V_i.

One could ask how the variance in subvolumes V_i is related to the variance of f in the whole integration domain V. This can be easily done using the additivity of the integral, the result is known in literature as the parallel axis theorem:

$$\mathrm{Var}(f) = \frac{1}{2}[\mathrm{Var}_a(f) + \mathrm{Var}_b(f)] + \frac{1}{4}(\langle\!\langle f\rangle\!\rangle_a - \langle\!\langle f\rangle\!\rangle_b)^2 \qquad (1.18)$$

As we can see the stratified sampling gives a variance which is always equal or smaller than the variance of the naive MC integrator. In particular whenever the function f behaves differently in the two regions V_a and V_b, the variance of f increases as $(\langle\!\langle f\rangle\!\rangle_a - \langle\!\langle f\rangle\!\rangle_b)^2$ and such difference doesn't affect the $\mathrm{Var}(\langle f\rangle')$ given by Eq.(1.17).

We can generalize the previous formulas by adding the possibility to sample a specific number of points from each subvolume. Considering again the subvolumes V_a and V_b we can suppose to sample the first one with N_a points and the second one with N_b with the constraint that $N = N_a + N_b$.

It can be shown that in this case the variance is minimized when:

$$\frac{N_a}{N} = \frac{\sqrt{\mathrm{Var}_a(f)}}{\sqrt{\mathrm{Var}_a(f)} + \sqrt{\mathrm{Var}_b(f)}} \qquad (1.19)$$

Consequently we can see that the number of sample in a particular subvolume is proportional to the standard deviation of the samples in that subregion, which is quite expected. If there is a region with high variance we will need more points in order to have a better knowledge of the behaviour of the integrand function.

With the aim of taking full advantage from stratified sampling, the standard procedure is to divide the integration domain in several subvolumes each one with a different number of samples. Again by minimizing the variance we obtain that the number of samples in each subregion must be proportional to the standard deviation in that particular subregion.

Stratified sampling can struggle in higher dimensions due to the large number of subregions. In fact if we divide the integration volume in M subvolumes for each dimension we will end up with a total of M^d total subregions. Considering that in order to compute the variance in each region we need at least two points, we will need a total of $2M^d$ sampled points., which can become challenging.

1.2 HEP and multi-dimensional integration

One of the main fields of physics which requires computing multi-dimensional integrals is High Energy Physics (HEP). In this section we give a brief overview on some aspects of quantum field theory, in particular we show that a prediction for a physical observable can be made as a series of perturbative terms which involve the evaluation of high dimensional integrals.

Finally we present shortly the quantum theory of strong interactions known as Quantum Chromodynamics (QCD).

1.2.1 Cross Section and Decay Rates

The experiments performed at LHC are aimed at probing the behaviour of elementary particles, which can be considered in a relativistic regime due to the high energies required. These are usually scattering experiments in which two beam of particles with well defined momenta collide creating new particles which are detected and measured.

The probability of finding a particular final state can be expressed in terms of a physical observable, the cross section. Suppose that we have two species of particle, $\mathcal{A}$, at rest with density $\rho_{\mathcal{A}}$, and $\mathcal{B}$ moving at velocity v towards $\mathcal{A}$ with density $\rho_{\mathcal{B}}$. Imagine also that we can measure the length of the two beam of particles $l_{\mathcal{A}}$ and $l_{\mathcal{B}}$.

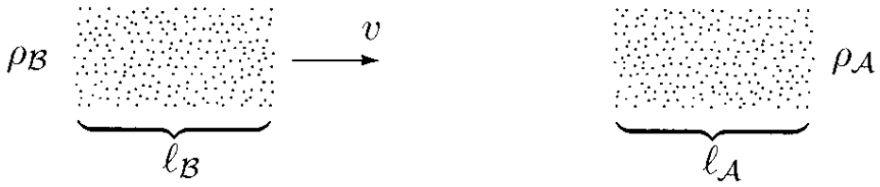

Figure 1.1: Symbolic representation of a scattering process

We expect that the total number of scattering events must be proportional to the cross-sectional area A common to the two beams and all the previous quantities; the cross section, σ, is defined indeed as:

$$\sigma \equiv \frac{\text{Number of scattering events}}{A\rho_\alpha\rho_\beta l_\alpha l_\beta} \tag{1.20}$$

We also wish to measure the momenta of the outgoing particles, which are expected to be infinitesimal. In order to do so we can define a differential cross section $d\sigma/(d^3p_1...d^3p_n)$ such that:

$$\int_\Omega \frac{d\sigma}{d^3p_1...d^3p_n} d^3p_1...d^3p_n = \sigma|_\Omega \tag{1.21}$$

where $\sigma|_\Omega$ gives the cross section for scattering in the region of final-state momentum space Ω. Obviously the final state momenta are not all independent.

Firstly the total four-momentum of the incoming particles must be equal to the total momentum of the outgoing particles by four-momentum conservation. Secondly each particle detected will have a given mass (or zero for massless ones) which fixes the four momentum components, $p_i^2 = m_i^2$.

The majority of the particles produced during the collision are unstable, that is the lifetime τ of such particles is so short that they cannot be detected by the experimental apparatus. Nevertheless we know that an unstable particle decays in other particles species some of which can be detected. For an unstable particle δ we can define a new observable, the decay rate Γ defined as

$$\Gamma = \frac{\text{Number of decays per unit time}}{\text{Number of } \delta \text{ particles present}} \tag{1.22}$$

The lifetime can be computed as the reciprocal of the sum of its decay rates into all possible final states.

This two physical observables can be computed theoretically by using the scattering matrix S firstly introduced by Heisenberg.

1.2.2 The S-matrix formalism

When particles collides during a scattering experiments they interact according to the fundamental interactions describes by the Standard Model (SM). In particular the SM is a quantum field theory that explains how three of the four fundamental forces (strong, weak and electromagnetic) can be described through the exchange of particles called bosons.
The beams of incident particles in a quantum field theory is treated as a quantum state $|\phi_A\phi_B\rangle_{\text{in}}$, in the same way the final state detected will be denoted by $|\phi_1\phi_2\dots\phi_n\rangle_{\text{out}}$. Both the initial and final states can be expressed as linear superpositions of eigenstates of the free theory, i.e. with definite momenta, constructed in the far past and in the far future respectively. For example for the final state we can write:

$$|\phi_1\phi_2...\phi_n\rangle_{\text{out}} = \left(\prod_{i=1}^{n}\int \frac{d^3p_i}{(2\pi)^3}\frac{\phi_i(\mathbf{p}_i)}{\sqrt{2E_i}}\right)|\mathbf{p}_1\mathbf{p}_2...\mathbf{p}_n\rangle_{\text{out}} \tag{1.23}$$

The probability of scattering, which is connected to the overlap between the initial and the final state, can be related to the transition amplitudes between the eigenstates of momenta created in the far past *in* and in the far future *out*

$${}_{\text{out}}\langle\mathbf{p}_1\mathbf{p}_2...\mathbf{p}_n|\mathbf{k}_1\mathbf{k}_2\rangle_{\text{in}} = \lim_{T\to\infty}\langle\mathbf{p}_1\mathbf{p}_2...\mathbf{p}_n|e^{-iHT}|\mathbf{k}_1\mathbf{k}_2\rangle \equiv \langle\mathbf{p}_1\mathbf{p}_2...\mathbf{p}_n|S|\mathbf{k}_1\mathbf{k}_2\rangle \tag{1.24}$$

where we have written explicitly the time evolution. As we can see the braket between the two asymptotic states can be rewritten as the braket between two momenta eigenstates by the limit of a sequence of unitary operators that is called *S-matrix*:

$${}_{\text{out}}\langle\mathbf{p}_1\mathbf{p}_2...\mathbf{p}_n|\mathbf{k}_1\mathbf{k}_2\rangle_{\text{in}} = \langle\mathbf{p}_1\mathbf{p}_2...\mathbf{p}_n|S|\mathbf{k}_1\mathbf{k}_2\rangle \tag{1.25}$$

If we are studying a non-interacting theory the S matrix is simply the identity matrix, since state of definite momentum are eigenstate of the free-field Hamiltonian, which are orthogonal. If we consider an interacting theory we can rewrite S as the identity matrix plus a non-trivial matrix T: $S = \mathbf{1} + iT$. In literature the matrix element of the T matrix is written as:

$$\langle \mathbf{p}_1 \mathbf{p}_2 ... \mathbf{p}_n | iT | \mathbf{k}_1 \mathbf{k}_2 \rangle \equiv (2\pi)^4 \delta^{(4)}\Big(k_1 + k_2 - \sum_{i=1}^{n} p_i\Big) \cdot i\mathcal{M} \tag{1.26}$$

where we introduced the invariant matrix element $\mathcal{M}$ by removing a factor that reflects the four-momentum conservation. The previous separation is useful since $\mathcal{M}$ contains all the informations regarding the interaction, while all the other factors depends merely on the kinematics of the process.

In particular, it can be shown that the differential cross section $d\sigma$ can computed using the square modulus of the invariant matrix element and the kinematic quantities of the particles involved in the collision:

$$d\sigma = \frac{1}{4E_A E_B |v_A - v_B|} \int d\Pi_n \times |\mathcal{M}(k_A, k_B \to p_1, ..., p_n)|^2 \tag{1.27}$$

where $|v_A - v_B|$ is the relative velocity between the two beams and the integral over the final momenta is of the form:

$$\int d\Pi_n = \left(\prod_{i=1}^{n} \int \frac{d^3 p_i}{(2\pi)^3} \frac{1}{2E_i} \right) (2\pi)^4 \delta^{(4)}\Big(k_A + k_B - \sum_{i=1}^{n} p_i\Big) \tag{1.28}$$

which corresponds to an integral involving $3n$ variables where n is the number of particles in the final state.

There is also a formula to compute the differential decay rate $d\Gamma$; it can be obtained considering an initial state with a single unstable particle in his rest frame that decays in n outgoing particles.

$$d\Gamma = \frac{1}{2m_A} \left(\prod_{i=1}^{n} \int \frac{d^3 p_i}{(2\pi)^3} \frac{1}{2E_i} \right) (2\pi)^4 \delta^{(4)}\Big(k_A - \sum_{i=1}^{n} p_i\Big) \times |\mathcal{M}(k_A \to p_1, ..., p_n)|^2 \tag{1.29}$$

We have two equations to compute the differential cross section and the differential decay rate which involve the square modulus of the invariant matrix element $\mathcal{M}$. In the next subsection we show how $\mathcal{M}$ can be calculated pertubatively using Feynman diagrams.

1.2.3 Feynman diagrams

The invariant matrix element $\mathcal{M}$ was firstly introduced in Eq.(1.26) in the overlap between two momenta eigenstates with the S matrix. Such matrix has an exact solution which is expressed as a series of terms, known as Dyson expansion:

$$S = \sum_{n=0}^{\infty} S^{(n)} = \sum_{n=0}^{\infty} = \frac{(-i)^n}{n!} \int \cdots \int d^4 x_1 ... d^4 x_n \mathcal{T}\{\mathcal{H}_I(x_1) \dots \mathcal{H}_I(x_n)\} \tag{1.30}$$

where $\mathcal{T}\{\mathcal{H}_I(x_1)\dots\mathcal{H}_I(x_n)\}$ denotes the normal ordered product of the interacting Hamiltonian densities $\mathcal{H}_I(x_1)\dots\mathcal{H}_I(x_n)$ and the integral is over all space-time.

The expansion is reliable only if we can treat the interacting hamiltonian density, $\mathcal{H}_I$, as a perturbation which is the case for the majority of the theories studied. For example in QED the interaction is proportional to the fine structure constant $\alpha \approx 1/137$, thus the perturbative expansion makes sense.

Thanks to Wick's theorem it is possible to rewrite the time-ordered product in a different way which can be represented graphically using Feynman diagrams. To be more specific Feynman diagrams are graphs in which each part from the lines to the vertices is linked to a mathematical expression. The set of rules that allows us to pass from the pictorial representation to the correct mathematical formula are called Feynman rules, and are easily obtainable from the Lagrangian of the theory.

As for the matrix S also the matrix element $\mathcal{M}$ can be written as a perturbative expansion:

$$\mathcal{M} = \sum_{n=1}^{\infty} \mathcal{M}^{(n)} \tag{1.31}$$

where the contribution $\mathcal{M}^{(n)}$ comes from the nth order perturbation term $S^{(n)}$ and can be obtained by drawing all topologically different, connected Feynman diagrams which contain n vertices and the correct number of external lines.

Among the different diagrams there are some which may contains some loop lines, which correspond to quantum corrections. The Feynman rules tell us that for each loop line we must add an integration over the momentum k of that line, since such momentum is not fixed by the process.

The perturbative expansion of $\mathcal{M}$ can be seen as a sum of integral with increasing dimension. The first contribution comes from the diagrams that contain no loops, which is referred as *tree-level* since such diagram obeys the definition of a tree in graph theory. The second contribution contribution will come from the one-loop diagrams in which the dimension of the integral increases since we are also integrating over the loop momentum. By repeating this process we can express $\mathcal{M}$ as:

$$\mathcal{M} = \mathcal{M}^{\text{tree}} + \mathcal{M}^{\text{1-loop}} + \mathcal{M}^{\text{2-loop}} + \dots \tag{1.32}$$

It is well known in the field of quantum field theories that the diagrams which involves loop corrections can lead to divergent contributions, in the limit where the momenta of the loop particles become large. These divergences are known as ultra-violet (UV) divergences and in renormalizable theories, like QCD, they can be removed by a modification of the continuum limit, at least in perturbation theory. If the considered theory is *renormalizable* there are specific techniques to handle this type of divergences which belong to the field of *renormalization*.

Infra-red IR divergences also appear in perturbation theory for the S-matrix of theories such as QCD and QED that have massless fields. This type of infinities arise when we consider radiative corrections to the tree-level

diagram, i.e. when other particles with vanishing energy are emitted other than the final state particles.

For example in QED we can start from the process $e^-e^+ \to \mu^+\mu^-$ at tree-level and then consider a radiative correction which involves the emission of a photon from one of the two final-state quarks. The second order radiative correction will involve the emission of two photons and so on. Finally we can arrive at an expression to include all this radiative corrections of the form:

$$\mathcal{M}^{\text{inclusive}} = \mathcal{M}^{\text{tree}} + \mathcal{M}^{\text{1-leg}} + \mathcal{M}^{\text{2-leg}} + \dots \tag{1.33}$$

where $\mathcal{M}^{\text{1-leg}}$ corresponds to the emission of a single particle, $\mathcal{M}^{\text{2-legs}}$ emission of two particles and so on. We can observe that also these sequences involves integral of increasing dimensions since by adding one extra particle to the final states we go from an $3n$ dimensional space to a $3n + 3$ space, always considering a final state of n particles.

IR singularities also appear in loop diagrams and thanks to the Kinoshita-Lee-Nauenberg (KLN) theorem it is possible to prove that the singularities of the real emissions and those one of the loop diagrams cancel each other out order by order in perturbation theory.

We have therefore shown that when computing a physical observable in quantum field theory we obtain a perturbative expansion which involves both loop diagrams and radiative correction which is finite. We can denote the first non vanishing term in this series as the *lowest order* (LO) term, the next term will be the *next-to-leading-order* term (NLO) and so no. For a physical observable $\mathcal{O}$ the perturbative series will be of the form:

$$\mathcal{O} = \mathcal{O}^{\text{LO}} + \mathcal{O}^{\text{NLO}} + \mathcal{O}^{\text{NNLO}} \tag{1.34}$$

where $\mathcal{O}^{\text{NNLO}}$ denotes the *next-to-next-to-leading-order* contribution.

1.2.4 Basics of QCD

At the Large Hadron Collider (LHC) we are particularly interested in processes which involves hadronic collisions. In the SM the theory which describe the hadronic interactions is known as Quantum Chromodynamics (QCD), which is a quantum field theory based on the gauge symmetry of the non-Abelian group $SU(3)$.

The hadrons, however, are not the fundamental quanta of the theory; they are described as bound states of subnuclear fermions known as quarks q and the relative anti-particles, the anti-quarks $\bar{q}$. The are two possible bound states observed: mesons, which are made by a quark-anti-quark couple $q\bar{q}$, and the baryons, which are described as a bound state of three quarks qqq.

In order to use the formalism of the S matrix and Feynman diagrams we need to be able to treat the interacting density Hamiltonian of QCD as a perturbation. The interacting term is proportional to the strong coupling constant α_s, thus a perturbative approach is reliable only if we are at scale μ s.t. $\alpha_S(\mu) \ll 1$. It has been proven both theoretically and experimentally

that the strong coupling has the peculiar characteristic of decreasing at UV scales, which is known as asymptotic freedom.

If we consider an hard scattering process between two hadrons usually one compute firstly the differential cross section at a parton level, since due to the asymptotic freedom we can consider the partons almost as free particles. The differential cross section , as we already saw in Eq.(1.27), will be proportional to the phase-space density $d\Phi_n$ and the squared matrix elements $|\mathcal{M}|^2$:

$$d\hat{\sigma}(p_1, \ldots, p_n; Q) \sim |\mathcal{M}(p_1, \ldots, p_n)|^2 d\Phi_n(p_1, \ldots, p_n; Q) \tag{1.35}$$

where Q denotes the renormalisation scale of the hard-process. Since we are at a hard-scale Q we can compute $d\hat{\sigma}$ as a perturbative series in the strong coupling $\alpha_s(Q)$

$$d\hat{\sigma} = d\hat{\sigma}^{\mathrm{LO}} + \alpha_s(Q) d\hat{\sigma}^{\mathrm{NLO}} + \alpha_s^2(Q) d\hat{\sigma}^{\mathrm{NNLO}} + \ldots \tag{1.36}$$

We can then relate the differential partonic cross section with the differential hadronic cross section by using the factorization theorem, when applicable, that allows us to subdivide the calculation of an observable into a short-distance part and an approximately universal long-distance part. The short-distance part in our case is the partonic cross section $d\hat{\sigma}$ while the long-distance part includes the Parton Density Functions (PDF) $f_{a/h}(x_a, \mu_F)$, which can be interpreted, to a first approximation, as the probability density of finding the parton a in the hadron h with a fraction x_a of the hadron's momentum when probed at a scale μ_F which is known as the *factorization scale*.

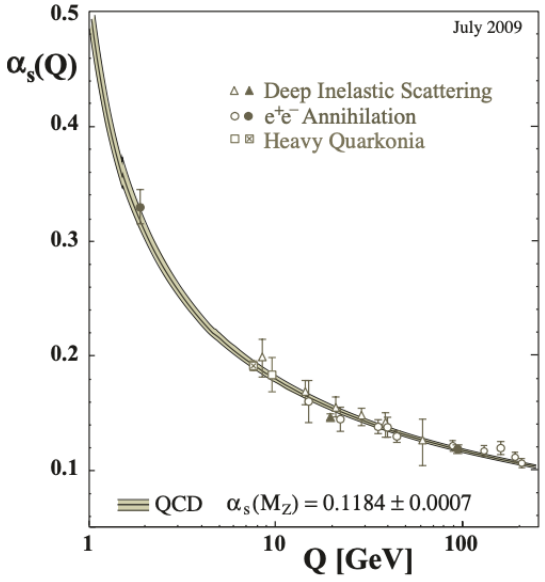

Figure 1.2: Measurements of the strong coupling constant α_s

$$d\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \sum_F \int d\Phi_F f_{a/h_1}(x_a, \mu_F) f_{b/h_2}(x_b, \mu_F) d\hat{\sigma}_{ab \to F} \tag{1.37}$$

where the sum over a and b runs over all the partonic constituents of the two hadrons h_1 and h_2 and the inner sum over F runs over all the possible final states F.

1.3 Modern techniques and limitations

Thanks to the technological development at LHC we are able to obtain experimental data with a very high precision. To test whether the SM can explain and predict correctly these data we need to compare them with theoretical predictions that have the same accuracy. This leads us to the problem of increasing the precision of the theoretical results.

As we saw in the previous section, we can obtain an expression for physical observables such as the differential cross section as a series of terms which involves the calculation of high-dimensional integrals. The most relevant technique used to solve these integrals are MC methods, due to the high-dimensions as expected. The problem of reaching higher accuracy is therefore related to the possibility of reducing the MC estimate of the variance of the considered integral.

In the first section we already discussed some MC techniques that enable us to reduce the variance, however, in the particular case of HEP integrands the task of variance reduction can become problematic.

1.3.1 Problematic of HEP integration

One of the first problems when predicting physical observable based on a quantum field theory, such as QCD, is the fact we need to compute the observable beyond the LO term in order to match the experimental data. Even if we start with a process with a relatively small dimensional phase space at LO, the next terms in the perturbation series will involve the computation of integrals with more complicated functions which are defined in a higher dimensional phase-space.

In particular the dimensionality of the integral will increase s.t.:

- if we consider a real emission the phase-space will change from a $3n$ integration volume to a $3(n+1)$ integration volume
- if we consider a virtual emission, we will need to compute an integral of dimension $3n+4$, where $3n$ comes from the phase-space integration and 4 from the integration over the loop-line

Secondly the squared matrix element $|\mathcal{M}^2|$ can become difficult to sample even for common low-dimensional SM processes, since in general is particularly peaked in smaller region of the integration domain in the vicinity of kinematic divergences. These regions become even smaller for high-dimensional integrands due to the large number of parameters.

This tendency of having sharp peaks in limited regions is the cause of the terrible statistical convergence of naive MC integrators which perform uniform

sampling. Thus importance sampling techniques are the solution of choice, since they enable us to perform a more effective sampling.

Once we have chosen a sampling technique we can reach better accuracies by simply increasing the size of the sample, since, as we have seen, the standard deviation for a MC simulation decreases with the size of the sample as $N^{-1/2}$.

Therefore we expect that for each physical process there will be a number of samples needed in order to reach the target accuracy required. In particular we will need large samples to reach high target accuracies when dealing with NLO terms due to the high-dimension and the complexity of the squared matrix element $|\mathcal{M}|^2$.

However this is only valid from a theoretical point of view, in practice these calculations are performed by computers which can have a hard time in sampling these complicated integrands as well as dealing with huge samples in order to reach the target accuracy. In particular the problem related to the CPU cost and the long computational times has been getting a lot of attention over the last years.

1.3.2 CPU costs and computational times

The CPU cost of MC methods is one of the biggest problem that is driving the budget of big experiments such as ATLAS or CSM. Since 2010, MC integration has gone from being a trivial element of an experiment's CPU budget to, particularly in the case of top quark production processes, an important consumer up to 20% of the experiment's CPU budget. The main driver for this CPU usage has been the availability of the complex multileg and NLO QCD processes.

This trend of high CPU resources is in contrast with the CPU budgets currently available at the LHC experiments. From Figure 1.3 we can see that the annual CPU consumption will overcome the CPU budget especially in Run 4 and Run 5 of the ATLAS experiment. The main problem for the ATLAS experiment is the heavy use of the `SHERPA` event generator which is demanding more CPU by comparison with `MadGraph5` which dominates the CSM simulation budget.

Another problem is related to the long computational times.

During a simulation we expect to perform various iterations of our MC routine and the final result will be expressed as a weighted average of all the previous outputs. In particular in the case of integration we need to repeat the process of sampling and the following instructions to compute the integral according to the technique used. When dealing with complex integrands, such as the NLO contribution for a QCD process, sampling could require long computational times.

The current importance sampling techniques can provide an approximate estimate for the sampling distribution by discarding or re-weighting the samples from such distribution, in particular the aggregation of these imperfect phase-space mappings is one of the major cause for the poor efficiencies of MC event sampling at high fixed order in the strong coupling α_s.

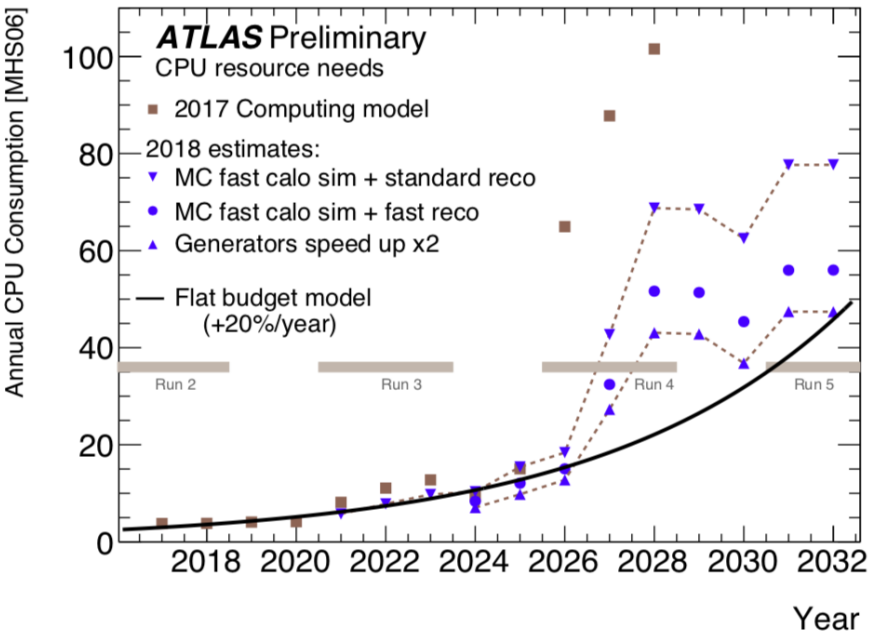

Figure 1.3: ATLAS CPU resource needs 2018 estimates

The ideal distribution would by definition always have unit weights, i.e. if the sampling density matches the optimal one, but in practice the sample weights have a tail to lower values, since the proposal density includes phase-space points which are not relevant for the integration. These lower weights can lead to poor statistical convergence in our computations.

We can also have greater than unit weights , these usually occur when the proposal density underestimated the maximum due to a failure in the previous sampling, with consequently single-event spikes.

In particular this sample rejection from broad distribution of weights combined with an already CPU-intensive computation of the matrix element value for each sample, can explain the huge CPU cost needed in the latest HEP experiments. From Figure 1.4 we can see that the CPU-time per event can become particularly large, for some current processes it can take up to 24 hours for a single event.

1.3.3 Possible solutions and aim of the thesis

The trend of high CPU requirements of the High-Luminosity LHC programme cannot continue, especially because in the future we will need more precision in the form of 1-loop NLO and 2-loop NNLO QCD calculation that may come at unacceptable CPU costs.

The question is therefore: how can we lower the CPU usage while still achieving high accuracy predictions? We can work in two different directions in order to achieve our goal.

Firstly we can develop new algorithms for multi-dimensional integration. In particular new techniques which are able to reach the target accuracy with a lower number of events thus reducing both the computational times and the

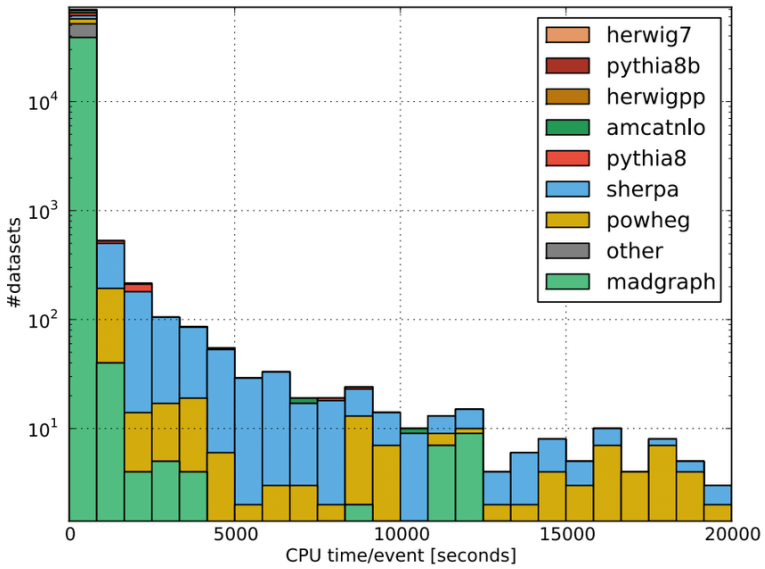

Figure 1.4: CPU time per event using different MC generators

CPU usage. These new techniques may include new numerical MC algorithms as well as ML techniques such as learning the integration phase-space using boosted decision trees or deep neural networks.

Secondly we can lower the CPU usage by looking at new computer architecture such as GPUs or multi-threading CPUs. For the purpose of MC integration of the squared matrix element, which comes down to the event sampling and adaptive strategies to compute the integral, the use of hardware acceleration devices is particularly appealing. In fact the sampling process is embarrassing parallel since we can just use a different random-number generator seeds for each run. Also the other operations used in MC techniques such as stratified sampling may take advantage of parallelizability due to the fact that we can express the integral as a sum of the MC integration in subregions of the integration volume.

The aim of the thesis is to study and implement new MC integration algorithms with the aim of achieving the target accuracies and at the same time overcoming the computational limitations by taking advantage of hardware acceleration devices.