Dynamical Fermions and lattice QCD: Hybrid Monte Carlo Algorithms

Guido Nicotra

July 9, 2020

Contents

1 Introduction

The aim of these pages is to introduce and easily explain the Hybrid Monte Carlo Algorithms to perform numerical simulations with dynamical fermions.
We are going to breefly introduce some basics concepts like Markov chains and the Metropolis Algorithm.
Of course, feel free to skip what you already know, if you want.

2 Markov Chains

Lattice simulations are not a static process. They require a self-interacting dynamical system, similarly to the thermodynamic of a gas in a box. So, we need to evolve the system from some initial conditions to the thermal equilibrium: there, the state variables of the system are constant in time and one can extract measurements. A Markov process satisfies these requirements. The possible configurations of the simulated system in the phase space are the elements of the so called *Markov chain*.

On a lattice, let's call the set of countable states of the system C_i. In a SU(3) pure gauge theory they correspond to the possible configurations of the link variables. Let's consider also a stochastic process that, starting from C_i, generates one after one other sets C_j, following a transition probability $P(C_i \to C_j)$ (shortly P_{ij}). The process is called *"Markov chain"* if the probability P_{ij} depends only on the very last state and not on the entire history that brought that configuration. The way the DNA works with children is an example of Markov process: the birth of a child only depends on the parents, we can completely ignore the other ancestors.
We now require some other stronger hypothesis for the importance sampling simulation: *irreducibility*, *aperiodicity* and the presence of *positive states*.
1) With the irreducibility we are asking that it would be possible to explore the entire phase space. Starting from any configuration i-th, choosen any other configuration j-th and a finite number of steps K, be valid:

$$P_{ij}^{(K)} \neq 0 \tag{1}$$

2) The process is aperiodic if we need at least one step so that $P_{ii}^{(K)} \neq 0$. The change of configuration must give a different state.
3) Finally, given a probability $P_{ii}^{(K)}$ and without reaching this condition in a number of steps lower than K, we define the average recurrent time τ as:

$$\tau = \sum_{K=1}^{\infty} K \cdot P_{ii}^{(K)} \tag{2}$$

If $\tau < \infty$, than C_i is called positive state. Unlike a transient state, the probability that we will return to C_i is not zero.
If the system satisfies all these properties, it can reach the thermal equilibrium: it will be perfect for a Monte Carlo simulation. It's possible to demonstrate the existance of configurations π_j defined by:

$$\lim_{K \to \infty} P_{ij}^{(K)} = \pi_j \tag{3}$$

The limit $K \to \infty$ means that, after a sufficiently long time, π_j are independent from the initial configuration C_i. These π_j satisfy also the usual probability laws:

$$\pi_j > 0 \ ; \ \sum_j \pi_j = 1 \tag{4}$$

and, very importantly:

$$\pi_j = \sum_i \pi_i P_{ij} \tag{5}$$

It means that, applying a transition process P_{ij}, it's not possible to 'leave' the set of $\{\pi\}$. These are exactly the properties of a thermal equilibrium. We can now set the names: $P_{eq}(C_i)$ will be the probability π_i to extract a configuration C_i from the condition of thermal equilibrium.
We only miss a last requirement, the unicity of this equilibrium state. To do this, we introduce the so called *detailed balance equation*:

$$e^{-S(C)} P(C \to C') = e^{-S(C')} P(C' \to C) \tag{6}$$

condition that is sufficient, but not necessary.
Now it's possible to extract the average value of an osservable, generated in agreement with the correct Boltzmann distribution, as:

$$\langle O \rangle = \sum_C P_{eq}(C) \cdot O(C) \tag{7}$$

2.1 Markov Chains, an application: Metropolis Algorithm

In the informatic development, the evolution of the system from a configuration to one other is called *update*. This process is perfomed by some algorithms and we are now going to speak about the Metropolis one.
The Metropolis algorithm is based on the micro-reversibility principle:

$$P(C \to C') = P(C' \to C) \tag{8}$$

To satisfy the detailed balance equation one has to follow some steps. Given an initial configuration C and a proposal C', we can have three different situations: *update*, *fluctuation* and *reject*.
Update. First at all, we can accept the old configuration if it's satisfied:

$$e^{-S(C')} > e^{-S(C)} \tag{9}$$

that follow the classic principle of least action.
Fluctuation. If this condition fails, the algorithm generates a random number $R \in [0, 1)$. The new configuration is accepted only if:

$$e^{-S(C')} \geq R \cdot e^{-S(C)} \tag{10}$$

In this case we obtain an updated action bigger than the old one, in a way similar to the quantum fluctuations.
Reject. If no one of these conditions are verified, we simply reject the proposed new configuration.
Again, it's interesting to notice the dependency of the evolution only on the parents configuration.

3 Dynamical fermions in MC - introduction

To introduce dynamical fermions in Monte Carlo samplings, one can take advantage of the analogy between bosonic and fermionic Gaussian integrals. The main difference, indeed, is the position of $det(D)$ in the Gaussian integral:

$$Z(\chi, \chi^\dagger) = \frac{det(D_{fermionic})}{det(D_{bosonic})} \cdot e^{\sum \chi^\dagger (A^{-1}) \chi} \tag{11}$$

So, the *pseudofermions* are introduced as bosons with the same degree of freedom of fermionic variables, substituting $A = (DD^\dagger)^{-1}$. We replace the integral over fermionic Grassmann variables for two mass-degenerate quarks by an integral over bosonic variables. To ensure the positivity of the determinant, one needs an even number of fermions for a given mass.

If we consider the fermion determinant as a contribution to the gauge action, we obtain the so called *effective fermion action*. With common algebra rules used in QFT, one finds:

$$det(D) = e^{-S_F^{eff}} \;\; with \;\; S_F^{eff} = -Tr(log(D)) \tag{12}$$

For semplicity, we assume that $det(D)$ is real and positive to avoid the sign problem. As well known, this object is highly nonlocal connecting essentially all gauge variables of the system and making its inversion computationally heavy. To include dynamical fermions one can add this fermionic action to generate gauge configurations which follow the appropriate Boltzmann distribution.
In the updating process:

$$e^{-S_{U'} + S_U} = e^{-S_{G'} + S_G} \cdot e^{-S_F'^{eff} + S_F^{eff}} \tag{13}$$

we will need new strategies for this new term.

4 Hybrid Monte Carlo

To treat the nonlocality of the action one could choose two opposite ways to update the system.
A *global change* of the action, where we change all the link variables, can optimize the effort of the algorithm: we maximize the amount of changes after the big calculations with the fermionic determinant. Problem: this solution kills the acceptance rate.
If we *slightly modify* the action updating few links, instead, the acceptance rate should remain stable, but we kill our computer cluster with the computational cost.

A third way that we are going to explore consists into update the fields choosing directions determined by the action. We force the system introducing a bias, of course, but we will also try to eliminate it. To update the field variables, we have to complete two tasks:

1. propose a new configuration with an *a priori selection probability* factor $T_0(U'|U)$;
2. accept/reject this candidate following the acceptance probability $T_A(U'|U)$.

We obtain the total probability $T(U'|U) = T_A(U'|U) \cdot T_0(U'|U)$. Now we look for the best acceptance rate with the lowest volume dependence and the lowest self correlation between consecutive configurations.

4.1 A classical example

Before to speak about dynamical fermions, let's study an easier system in a classic theory with its proper accept/reject mechanism.
Let be Q a real bosonic scalar field and P real conjugate momenta. We want to evaluate the vacuum expectation value of an observable O:

$$\langle O \rangle_Q = \frac{\int DQDP \cdot O(Q) \cdot exp(-\frac{1}{2}P^2 - S_Q)}{\int DQDP \cdot exp(-\frac{1}{2}P^2 - S_Q)} \tag{14}$$

So, our Hamiltonian in this classical system is:

$$H(Q,P) = \frac{1}{2}P^2 + S_Q \tag{15}$$

and we can obtain the classical equations of motion as:

$$\dot{P} = -\frac{\partial H}{\partial Q} = -\frac{\partial S}{\partial Q} \tag{16}$$

$$\dot{Q} = \frac{\partial H}{\partial P} = P \tag{17}$$

These are called *molecular dynamic equations*, since they determine the time evolution of a classical particles system. Because H is a constant of motion ($[H,H] = 0$ and $\dot{H} = 0$), the path of the configurations (P,Q) lies on a hypersurface of a constant energy in the phase space: that means an always accepted path.
Unfortunately, the numerical implementation of these equations requires the introduction of discrete steps like $\epsilon = \Delta\tau$ polluting the path with errors. To solve this problem, one can extract results in the limit of vanishing step, as the lattice continuum limit, or introduce a corrective step, as we are going to see.
We call this sequence of steps *trajectory*. Our Hybrid Monte Carlo algorithm will produce a new configuration following this trajectory and then Metropolis will accept/reject it.

4.1.1 New configuration proposal

We introduce the *leapfrog integration* to build a trajectory. We evolve Q and P in n steps: with Q we simply take n steps of lenght ϵ; in the case of P, we take

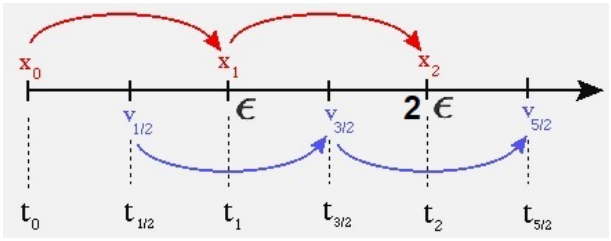

Figure 1: The leapfrog idea

$\epsilon/2$ as first and last step and (n-1) ϵ steps in the middle. If we use this scheme on the molecular dynamic equations, we can evaluate the first steps finding:

$$Q(0) \to Q(\epsilon) = Q_0 + P(\frac{\epsilon}{2})\epsilon \tag{18}$$

and:

$$P(0) \to P(\frac{\epsilon}{2}) = P(0) - \frac{\partial S}{\partial Q}\Big|_{Q(0)} \cdot \frac{\epsilon}{2} \to P(\epsilon) = P(\frac{\epsilon}{2}) - \frac{\partial S}{\partial Q}\Big|_{Q(\epsilon)} \cdot \frac{\epsilon}{2} \tag{19}$$

The first value after the equal sign is always the old value, while the other is the increase.
We now require two hypotesis to satisfy the detailed balance equation: the preservation of the integration measure DQDP and the reversibility of the trajectory $T(P', Q'|P, Q) = T(-P, Q| - P', Q')$.
DIM area preservation. We write a Jacobian chain as:

$$det\frac{\partial(P_1, Q_1)}{\partial(P_0, Q_0)} = det\frac{\partial(P_1, Q_1)}{\partial(P_{1/2}, Q_1)}\frac{\partial(P_{1/2}, Q_1)}{\partial(P_{1/2}, Q_0)}\frac{\partial(P_{1/2}, Q_0)}{\partial(P_0, Q_0)} \tag{20}$$

Because all the right hand side factorizes into three triangular matrices with diagonal of 1s, the integration measure is invariant.
Dim within a dim. Let's evaluate one Jacobian for all:

$$\frac{\partial(P_{1/2}, Q_0)}{\partial(P_0, Q_0)} \tag{21}$$

We find these elements:

$$\frac{\partial P_{1/2}}{\partial P_0} \;\; ; \;\; \frac{\partial P_{1/2}}{\partial Q_0} = 1 \;\; ; \;\; f(P)$$

$$\frac{\partial Q_0}{\partial P_0} \;\; ; \;\; \frac{\partial Q_0}{\partial Q_0} = 0 \;\; ; \;\; 1$$

Apologizing for the ugliness of this matrix, this give $det() = 1$.

□

DIM reversibility. To prove it, we take a step forth and back along the trajectory. Be (Q_0, P_0) an initial value at time 0. From the leapfrog integration, the step forth is:

$$Q_1 = Q_0 + P_0\epsilon - \frac{1}{2}\frac{\partial S}{\partial Q}\Big|_{Q_0}\epsilon^2 \tag{22}$$

$$P_1 = P_0 - \frac{1}{2}\Big(\frac{\partial S}{\partial Q}\Big|_{Q_0} + \frac{\partial S}{\partial Q}\Big|_{Q_1}\Big)\epsilon \tag{23}$$

while the step back starts from an initial value (Q_1, P_1) applying $-\epsilon$.

$$Q(\epsilon - \epsilon) = Q_1 - P_1\epsilon - \frac{1}{2}\frac{\partial S}{\partial Q}\Big|_{Q_1}\epsilon^2 = [...] = Q_0 \tag{24}$$

$$P(\epsilon - \epsilon) = P_1 + \Big(\frac{\partial S}{\partial Q}\Big|_{Q_1} + \frac{1}{2}\frac{\partial S}{\partial Q}\Big|_{Q_0}\Big)\epsilon = P_0 \tag{25}$$

□

Observation. Because time ϵ is always multiplied by P, this is equivalent to say that $T(P', Q'|P, Q) = T(-P, Q|-P', Q')$. Moreover, in our distribution with the action, we have P^2, so one could directly state $T(P', Q'|P, Q) = T(P, Q|P', Q')$.

The new configuration proposal is ready to be accepted/rejected at the end of this leapfrog integration. A common value for the number of steps is almost a hundred.

4.1.2 Accept/reject

To briefly recap what we have done yet, given a configuration Q we generate the conjugate momenta P from a Gaussian distribution $\propto exp(-P^2/2)$. Then, the molecular dynamic trajectory leads from (P,Q) to a new proposal (P',Q'). The last step is to accept/reject the proposal following an *Acceptance probability*:

$$T_A(P', Q'|P, Q) = min\Big(1, \frac{exp(-H[P', Q'])}{exp(-H[P, Q])}\Big) \tag{26}$$

So, if $H > H'$ the new configuration is following the principle of least action and we always accept it ($min = 1$). Otherwise if $H' > H$ we accept the configuration with a probability lower than 1: in this way we admit the quantum fluctuations as the Metropolis algorithm does. It's obvious that, in the limit of exact implementation of the molecular dynamic equations, $H' = H$. In this way it's possible to understand how the acceptance of the configuration is very useful to fix the error introduced by the discretization. Due to this decision process, the HMC algorithm is exact: thay always find the optimal solution to a given optimization problem.
We only miss the detailed balance equation to validate all we have done.

DIM detailed balance equation. The total probability to move from Q to Q' results from integrating over the Ps:

$$T(Q'|Q) = \int DPDP' T_A(P', Q'|P, Q) T(P', Q'|P, Q) e^{-P^2/2} \tag{27}$$

After some algebra and using the parity of P, we rewrite:

$$T_A(P', Q'|P, Q) = [...] = exp(-P'^2/2 + P^2/2 - S_{Q'} + S_Q) \cdot T_A(-P, Q| - P', Q') \tag{28}$$

Now one can work on the integral. Using the reversibility and the area preservation, we find:

$$T(Q'|Q) = [...] = \int DPDP' T_A(P, Q|P', Q') T(P, Q|P', Q') e^{(S_Q - S_{Q'} - P'^2/2)} \tag{29}$$

We can move $exp(S_Q)$ on the left side, finding on the right side the beginning integral but in the opposite direction, namely:

$$e^{-S_Q} T(Q'|Q) = e^{-S_{Q'}} T(Q|Q') \tag{30}$$

This is our detailed balance equation. □

4.2 HMC in QCD

Now it's time to apply our algorithm to lattice QCD, starting with the case of two dynamical, mass-degenerate quarks. Using pseudofermions we get a system made by their fields ϕ and gauge fields U, distributed following the Boltzmann weight factor:

$$exp(-S_U) = exp(-S_G + \phi^\dagger (DD^\dagger)\phi) \tag{31}$$

The update procedure of our system starts from ϕ. One generates complex vectors χ following the Gaussian distribution $exp(-\chi^\dagger \chi)$ and then computes $\phi = D\chi$. When this is done, one updates U fields with the molecular dynamic trajectory, using ϕ as an external constant field. The new configuration can be accepted or rejected.
Since QCD is a gauge theory with SU(3) gauge symmetry, each link variable is:

$$U = exp\Big(i \sum_{j=1}^{8} \omega^j T_j\Big) \tag{32}$$

leading us to traceless hermitian matrices. Then, for each U_μ there are eight P_μ^j coniugate to ω_μ^j. If we combine them with the generators of the algebra T_j as:

$$P_\mu = \sum_{j=1}^{8} P_\mu^j T_j \tag{33}$$

we obtain the same kind of traceless hermitian matrices. We will formulate the algorithm in terms of P_μ and U_μ, substituting in the original Hamiltonian:

$$\frac{1}{2}\sum_j \left(P_\mu^j\right)^2 = Tr\left(P_\mu^2\right) \tag{34}$$

Let's also calculate some derivatives of the action with respect to Q for the molecular dynamic equations.
The leapfrog steps in QCD become:

- Generate pseudofermions;
- Given a gauge configuration U_0, generate P_0 according to the Gaussian distribution $exp(-Tr[P^2])$;
- Update with the first step $P_{1/2} = P_0 - (\epsilon/2)F[U,\phi]|_{U_0}$;
- Iterate for k=1, ..., $n-1$:
  $U_k = exp(i\epsilon P_{k-1/2}) \cdot U_{k-1}$ and $P_{k+1/2} = P_{k-1/2} - \epsilon F[U,\phi]|_{U_k}$;
- Close the steps
  $U_k = exp(i\epsilon P_{n-1/2}) \cdot U_{n-1}$ and $P_n = P_{n-1/2} - (\epsilon/2)F[U,\phi]|_{U_n}$;
- Accept the configuration if a random number $r \in [0,1)$:
  $r < exp(Tr(P^2) - Tr(P'^2) + S_G - S'_G + \phi^\dagger[(DD^\dagger)^{-1} - (D'D'^\dagger)^{-1}]\phi)$.

Talking about F, the so called *driving force*, it stays for the $\partial S/\partial Q$ we saw before. We also know that the derivatives live in the tangent space of the group, so we find that:

$$F[U,\phi] = \sum_{j=1}^{8} T_j \nabla^j (S_G + \phi^\dagger (DD^\dagger)^{-1}\phi) \in su(3) \tag{35}$$

From the derivative of the fermionic contribution, one has to evaluate for each substep of the trajectory:

$$\nabla^j(\phi^\dagger(DD^\dagger)^{-1}\phi) = -((DD^\dagger)^{-1}\phi)^\dagger\Big(\frac{\partial D}{\partial \omega^j}D^\dagger + D\frac{\partial D^\dagger}{\partial \omega^j}\Big)((DD^\dagger)^{-1}\phi) \tag{36}$$

Derivatives of D and inversions of $DD^\dagger$ are, of course, very computationally expensive.

5 References

- Hybrid Monte-Carlo: introduction - Secs. 8.1 and 8.2 in Gattringer, Lang
- Creutz M., Jacobs L., Rebbi C., Monte Carlo computations in lattice gauge theories, in Physics Reports, 1983.
- Rothe H. J., Lattice Gauge Thoeries. An introduction, Singapore, World Scientific Publishing, 2012.