Fourier acceleration, the HMC algorithm and renormalizability

Norman H. Christ[*][†]

Department of Physics, Columbia University, New York, NY 10027, USA
E-mail: nhc@phys.columbia.edu

Evan W. Wickenden

Department of Physics, Columbia University, New York, NY 10027, USA
E-mail: eww2119@columbia.edu

The analysis developed by Lüscher and Schaefer of the Hybrid Monte Carlo (HMC) algorithm is extended to include Fourier acceleration. We show for the ϕ^4 theory that Fourier acceleration substantially changes the structure of the theory for both the Langevin and HMC algorithms. When expanded in perturbation theory, each five-dimensional auto-correlation function of the fields $\phi(x_i, t_i)$, $1 \le i \le N$, corresponding to a specific 4-dimensional Feynman graph separates into two factors: one depending on the Monte-Carlo evolution times t_i and the second depending on the space-time positions x_i. This separation implies that only auto-correlation times at the lattice scale appear, eliminating critical slowing down in perturbation theory.

The 36th Annual International Symposium on Lattice Field Theory - LATTICE2018
22-28 July, 2018
Michigan State University, East Lansing, Michigan, USA.

---

[*]Speaker.

1. Introduction

One of the most challenging aspects of numerical studies using lattice QCD is generating stochastic samples of gauge configurations distributed according to the weight required by the QCD path integral. The current best method for constructing such a sample is the Hybrid Monte Carlo (HMC) algorithm [1]. In this method these samples are obtained as elements of a Markov chain generated by a sequence of molecular dynamics evolution steps followed by a Langevin-style, Gaussian refresh of the evolution momenta. The molecular dynamics follows conventional Hamiltonian mechanics with the gauge matrices treated as classical coordinates and fictitious momenta conjugate to gauge variables defined in the Lie algebra of the gauge group .

These conjugate momenta enter the evolution Hamiltonian through a simple "kinetic" energy term $\frac{1}{2}\sum_{\mu,x,b}(P_\mu^b(x))^2$ where b, $1 \le b \le 8$, labels a generator of the $SU(3)$ gauge group and the position x and direction μ identify a lattice link. The scale of molecular dynamics time is chosen so that the "mass" appearing in this Hamiltonian kinetic energy is unity. As calculations are performed at increasingly small lattice spacing the frequencies present in this classical system extend over a larger range corresponding to an increasingly broad distribution of forces. Since this kinetic energy term implies a single distribution of velocities, the molecular dynamics evolution encounters critical slowing down. The integration of the equations of motion must be carried out with a small step size to allow an accurate evolution for those modes with very large forces, typically ultraviolet, lattice-scale degrees of freedom. This implies that the more physically important long-distance modes which move at the same velocity but for which the potential changes more slowly, must be evolved for an increasing number of time steps to generate an independent sample.

In simpler systems, this critical slowing down of the classical evolution can be mitigated by introducing a mode-dependent mass chosen to give a larger velocity for those modes with smaller frequency, a technique referred to as Fourier acceleration. Such an approach is challenging in a gauge theory where local gauge symmetry implies that simple Fourier transformation cannot be used to distinguish the soft and stiff modes. However, there are a number of interesting proposals that allow Fourier acceleration to be applied to gauge theories [2, 3, 4, 5].

Apart from the difficulty of local gauge symmetry, the asymptotic freedom of QCD might suggest that this is a highly promising application for Fourier acceleration since at short distances in a fixed gauge the modes of the theory become independent harmonic oscillators with known frequencies allowing Fourier acceleration to be applied precisely as the continuum limit is approached. However, this is thrown into doubt because of the complexities of quantum field theory. While the correlation functions in both space-time and Monte Carlo time for the less efficient Langevin algorithm can be analyzed in perturbation theory and shown to possess a well-defined continuum limit [6] the more efficient HMC algorithm contains non-renormalizable singularities implying the presence of non-universal long-time correlations between Monte Carlo samples [7].

As we will show, adding Fourier acceleration to the Langevin or HMC algorithms substantially changes both algorithms and removes correlations between Monte Carlo samples separated by physical times. Thus, Fourier acceleration renders both algorithms "renormalizable" in the trivial sense that there are no Monte Carlo time correlations at a physical-scale in the continuum limit. Since the Langevin algorithm is simpler to analyze than the HMC we will discuss them both with an emphasis on Langevin evolution.

2. Stochastic evolution

Following the treatment of Luscher and Schaefer (L&S) we will examine a ϕ^4 theory. For this case the Langevin equation determines the evolution of the real classical field $\phi(x,t)$ as a function of Monte Carlo time t and as a functional of the noise field $\eta(x,t)$:

$$\partial_t \phi(x,t) = -\frac{\delta}{\delta(x,t)} S[\phi(x,t)] + \eta(x,t) = -(-\partial^2 + m_0^2)\phi(x,t) - \frac{g_0}{3!}\phi(x,t)^3 + \eta(x,t) \qquad (2.1)$$

where x is a space-time coordinate. At large evolution time t, the resulting stochastic field $\phi(x,t)$ will be distributed according to the desired path integral weight $e^{-S[\phi(t)]}$ if the noise field obeys :

$$\langle \eta(x,t)\eta(y,s) \rangle = 2\delta^4(x-y)\delta(t-s). \qquad (2.2)$$

The same distribution results from an ensemble $\phi(x,t_n)$ obtained from a single noise field collected at different sample times t_n. One may view this as a useful algorithm for generating stochastic path integral samples [8] or as a new formulation of quantum mechanics based on Brownian motion [9].

At large time Eq (2.1) can solved as a power series in g_0 by iterating the equation:

$$\phi(x,t) = \int d^4y \int ds K_L(x-y, t-x)\left\{ \eta(y,s) - \frac{g_0}{3!}\phi(x,t)^3 \right\} \qquad (2.3)$$

using the kernel function $K_L(x,t)$ given by

$$K_L(x,t) = \theta(t)\int d^4q \frac{e^{iq \cdot x}}{(2\pi)^4} e^{-(q^2+m_0^2)t}. \qquad (2.4)$$

Iteratively replacing the fields ϕ in the $g_0\phi^3$ term on the right hand of Eq. (2.3) by solutions obtained in lower order will generate a series expansion, giving the field ϕ as a function of η which can be described by a graphical expansion of the sort shown in Fig. 1. The perturbative result for a general Green's function would be obtained by multiplying an appropriate number of such solutions $\phi(x_1,t_1)\phi(x_2,t_2)\ldots\phi(x_N,t_N)$ and then averaging over the noise. The noise functions will combine in pairs. This will join pairs of directed lines ending in a square into what, following the conventions of Ref. [7], we would represent as a single, directionless line, as suggested in Fig. 2.
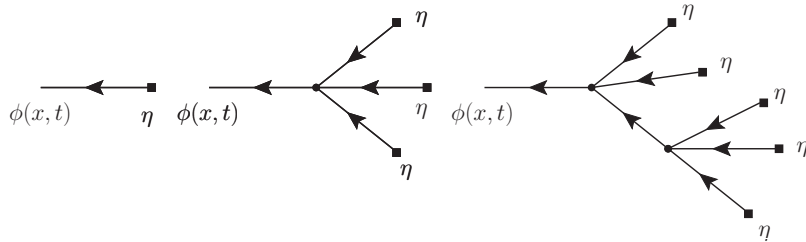

Figure 1: Graphical representation of terms that would appear in a recursive solution of Eq. (2.3).

In contrast to the Langevin algorithm, the HMC involves infrequent refreshes of the molecular dynamics momentum and relies for its efficiency on a Hamiltonian molecular dynamics trajectory during which the system moves deterministically to a new point in which all modes have changed

$$\xrightarrow[\phi(x,t)]{\int d^4z dr \int d^4z' dr' \delta^4(z-z')\delta(r-r')} \overset{\blacksquare}{\underset{\phi(y,s)}{\longrightarrow}} \quad \equiv \quad \overline{\underset{\phi(x,t) \qquad \phi(y,s)}{}}$$

Figure 2: Graphical version of the process of averaging the product of two noise sources and representing the resulting product of directed propagators by a single directionless line.

substantially from their starting point. Ideally, if the initial momenta are chosen randomly, the start and end points of this trajectory will be largely independent. Thus, it is this single-trajectory evolution which we wish to Fourier accelerate.

Following L&S, such a classical evolution can be analyzed if an additional stochastic element is introduced by using the Generalized HMC algorithm (GHMC) [10]. The analogue to the Langevin equation is a pair of Hamiltonian-like equations,

$$\partial_t \pi(x,t) = -\frac{\delta}{\delta(x,t)} S[\phi(x,t)] - 2\mu_0 \pi + \eta'(x,t) \quad \partial_t \phi(x,t) = \pi(x,t), \tag{2.5}$$

which can be rewritten as the second-order evolution equation:

$$\partial_t^2 \phi(x,t) = -\frac{\delta}{\delta(x,t)} S[\phi(x,t)] - 2\mu_0 \partial_t \phi(x,t) + \eta'(x,t) \tag{2.6}$$

where the noise $\eta'(x,t)$ is a simple multiple of the earlier $\eta(x,t)$: $\eta'(x,t) = \sqrt{2\mu_0}\eta(x,t)$.

As discussed by L&S the second order Eq. (2.6) can also be solved using a kernel function:

$$K_{\text{HMC}}(x,t) = \theta(t) \int d^4q \frac{e^{iq\cdot x}}{(2\pi)^4} e^{-\mu_0 t} \frac{\sin(\varepsilon_q t)}{\varepsilon_q} \tag{2.7}$$

where $\varepsilon_q = \sqrt{q^2 + m_0^2 - \mu_0^2}$. Equations (2.4) and (2.7) give insight into the autocorrelation times expected for Langevin and HMC evolutions. As can be seen from the exponential t dependence of $K_{\text{L}}(x,t)$ in Eq. (2.4) the longest autocorrelation times behave as $1/m_0^2$ suggesting a dynamical critical exponent of two. For the HMC algorithm the kernel function $K_{\text{HMC}}(x,t)$ does not fall exponentially as t grows (for $\mu_0 \to 0$ where the GHMC becomes the HMC algorithm) but instead oscillates with a longest period given by $1/m_0$ suggesting a dynamical critical exponent of one.

3. Fourier acceleration

Fourier-accelerated versions of the Langevin algorithm can be obtained by introducing a factor of the Klein-Gordon operator into the left-hand side of Eq. (2.1) [8]. For the HMC algorithm the Klein-Gordon operator is introduced as a mass term in the fictitious HMC kinetic energy: $\frac{1}{2}\pi^\dagger \pi \to \frac{1}{2}\pi^\dagger \frac{1}{(-\partial^2 + m_0^2)}\pi$. For our ϕ^4 example, the resulting stochastic evolution equations become:

$$(-\partial^2 + m_0^2)\partial_t \phi(x,t) = -(-\partial^2 + m_0^2)\phi - \frac{g_0}{3!}\phi^3 + \widetilde{\eta}(x,t) \tag{3.1}$$

$$(-\partial^2 + m_0^2)\partial_t^2 \phi(x,t) = -(-\partial^2 + m_0^2)\phi - \frac{g_0}{3!}\phi^3 - 2\mu_0(-\partial^2 + m_0^2)\partial_t \phi(x,t) + \widetilde{\eta}'(x,t) \tag{3.2}$$

where as before $\widetilde{\eta}' = \sqrt{2\mu_0}\widetilde{\eta}$. However, $\widetilde{\eta}$ obeys:

$$\langle \widetilde{\eta}(x,t)\widetilde{\eta}(y,s) \rangle = 2(-\partial^2 + m_0^2)\delta^4(x-y)\delta(t-s). \tag{3.3}$$

3

The 5-D structure of these equations has been dramatically changed by Fourier acceleration. This can be easily seen if both equations are divided by the factor $(-\partial^2 + m_0^2)$ and written as:

$$(\partial_t + 1)\phi(x,t) = -\frac{g_0}{3!(-\partial^2 + m_0^2)}\phi^3 + \overline{\eta}(x,t) \tag{3.4}$$

$$(\partial_t^2 + 2\mu_0\partial_t + 1)\phi(x,t) = -\frac{g_0}{3!(-\partial^2 + m_0^2)}\phi^3 + \overline{\eta}'(x,t) \tag{3.5}$$

where $\overline{\eta}' = \sqrt{2\mu_0}\overline{\eta}$ and $\overline{\eta}$ obeys

$$\langle\overline{\eta}(x,t)\overline{\eta}(y,s)\rangle = \frac{2}{(-\partial^2 + m_0^2)}\delta^4(x-y)\delta(t-s). \tag{3.6}$$

As can be seen from Eqs. (3.4) and (3.5), Fourier acceleration has effectively separated the dependence on the space-time and stochastic time variables. Now the kernels that can be used to solve these equations perturbatively do not depend on the space-time physics. They are given by:

$$K_{\rm L}^{\rm FA}(x,t) = \delta^4(x)\theta(t)e^{-t} \text{ and } K_{\rm HMC}^{\rm FA}(x,t) = \delta^4(x)e^{-\mu_0 t}\theta(t)\sin\big(t(1-\mu_0^2)^{1/2}\big)/(1-\mu_0^2)^{1/2}. \tag{3.7}$$

All critical slowing down has been removed and both the auto-correlation time for the Langevin algorithm and the oscillation period for the HMC algorithm are now simply one and 2π in lattice units. If we evaluate a particular term in the perturbative solution to either equation corresponding to a given diagram constructed from the ingredients found in Figures 1 and 2 the resulting amplitude will factor into the space-time Green's function that would result if the diagram were viewed as a standard Feynman diagram multiplied by a function of the evolution times t_i which is combination of exponential or trigonometric functions of the t_i multiplied by a finite-order polynomial in the t_i.

This can be expressed algebraically by the relation

$$\Big\langle\phi(x_1,t_1)[\overline{\eta}]\,\phi(x_2,t_2)[\overline{\eta}]\ldots\phi(x_N,t_N)[\overline{\eta}]\Big\rangle_{\overline{\eta}} = \sum_{\{\Gamma_5\}}I_{\Gamma_5}(t_1,t_2,\ldots,t_N)G_{\Gamma_4(\Gamma_5)}(x_1,x_2\ldots,x_N). \tag{3.8}$$

The left-hand side is the product of fields computed using the noise $\overline{\eta}$ and then averaged over $\overline{\eta}$. The right-hand side is a sum over contributions from each 5-D diagram Γ_5, each of which can be written as a product of two factors. The first is $I_{\Gamma_5}(t_1,\ldots,t_N)$ determined by that 5-D diagram and depending on the evolution times $t_1,\ldots,t_N$. The second is the usual amplitude $G_{\Gamma_4(\Gamma_5)}(x_1,x_2\ldots,x_N)$ corresponding to the 4-D Feynman diagram $\Gamma_4(\Gamma_5)$ obtained from the 5-D diagram Γ_5 by treating each line as a 4-D Feynman propagator.

Thus, the t-dependent factors depend only on the stochastic evolution time and will be unaffected by the physical scales in the field theory problem being studied. In this perturbative expansion the scale of the evolution time dependence is that of the lattice cutoff. Possible longer, physical time scales are absent. The non-renormalizable, 5-D light-cone singularity with a logarithmically divergent coefficient found by L&S in second order perturbation theory does not occur in this Fourier-accelerated perturbation theory.

Although we do not thoroughly understand the structure of this perturbative expansion of the 5-D correlation functions, we will comment on two features. First, while individual diagrams that appear in the perturbative expansion of Eqs. (3.4) and (3.5) that determines the N-point function

$$\langle\phi(x_1,t_1)\phi(x_2,t_2)\ldots\phi(x_N,t_N)\rangle \tag{3.9}$$

4

can each be neatly written as the product of a conventional 4-D Feynman diagram multiplied by a factor which depends on the evolution times $t_1, t_2 \ldots t_N$, these evolution-time dependent factors will be different for the different 5-D diagrams. The correctness of the Fourier-accelerated Langevin or GHMC algorithm implies that when evaluated at equal evolution times, $t_i = t_j$ for all i and j these factors must combine to reproduce the required combinatoric factors that appear in the 4-D Dyson-Wick expansion. (Note: given the arrows on the boson lines in the stochastic perturbation theory there will be many 5-D diagrams that will combine to give a single 4-D diagram.) However, when the times t_i are unequal this connection with the 4-D theory is lost. In particular the divergent diagrams requiring renormalization subtractions and the corresponding diagrams which contain the appropriate counter terms will acquire different coefficients and no longer cancel when $t_i \neq t_j$. While this behavior may raise concerns, such potentially strong autocorrelations at short evolution times will have no effect at large evolution times given the rapidly falling or oscillating character of these coefficients.
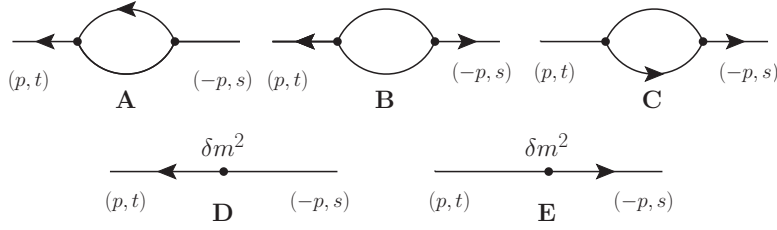

Figure 3: Five type of stochastic diagrams which contribute to the boson self-energy at second order in a ϕ^3 theory.

The diagrams shown in Fig. 3 provide a simple example for Langevin evolution for the case of the second-order self-energy diagrams and counter terms in a ϕ^3 theory. The first three correspond to different factors $I_\Gamma(t-s)$ which will multiply a single conventional divergent 4-D one-loop diagram where $I_\Gamma(t-s)$ depends on the evolution times t and s and $\Gamma = A$, B and C. Similarly, the mass counter terms will be the usual 4-D counter term multiplied by amplitudes $I_D(t-s)$ and $I_E(t-s)$. The five amplitudes I_Γ can be easily worked out. Setting $s = 0$ we find:

$$I_A(t) = I_C(-t) = \frac{2}{3}e^{-|t|}\left[1 + \theta(t)\left(6t - 4(1 - e^{-t})\right)\right] \qquad I_B(t) = \frac{2}{3}e^{-|t|}\left[2 - e^{-|t|}\right]$$

$$I_D(t) = e^{-|t|}\left[\frac{1}{2} + t\theta(t)\right] \qquad I_E(t) = e^{-|t|}\left[\frac{1}{2} - t\theta(-t)\right]. \tag{3.10}$$

At $t = 0$, $I_A(0) + I_B(0) + I_C(0) = 2\left(I_D(0) + I_E(0)\right)$ and the counter term can be chosen to cancel the divergent self-energy as required. However, this will not be true for $t \neq 0$.

The second issue is the dependence of the 5-D auto-correlation functions on the evolution times $t_1, t_2 \ldots t_N$. Order-by-order in perturbation theory all amplitudes decrease exponentially in the separation between the largest and smallest times measured in lattice units or oscillate with odd harmonics of the unit lattice frequency suggesting that critical slowing down has been eliminated. However, the presence of polynomials in these times raises the possibility that when summed to all orders these polynomials may generate additional exponential time dependence that may cancel

the large exponents present in the perturbative theory, replacing them by a weaker time dependence and causing a reemergence of critical slowing down, a superficially unlikely possibility.

Related to this question for our ϕ^4 example, is the choice of the mass used in the Fourier-acceleration factor $(-\partial^2 + \widetilde{m}^2)$ in Eqs. (3.1), (3.2) and (3.3). For simplicity, in the discussion above we have used the bare mass m_0 in these factors. This may be a poor choice since the Fourier-acceleration factor $q^2 + \widetilde{m}^2$ should be small compared to the lattice scale for physical q^2, something that will not be true if the bare mass $m_0 \sim 1/a$ is used in place of $\widetilde{m}$. Apparently using the renormalized mass for $\widetilde{m}$ and in the unperturbed action would be a wiser choice. In a gauge theory where the cut-off scale enters only logarithmically, such an issue may be less important.

4. Conclusion and acknowledgement

We have examined the stochastic evolution equations which can be used to describe the auto-correlation present in an ensemble of Monte Carlo samples generated using the Langevin and GHMC algorithms. In both cases it appears that if Fourier acceleration is introduced the resulting dependence on the Langevin or GHMC simulation time is significantly changed. The kernel used to examine auto-correlations in a perturbative expansion acquires a strong time dependence, decreasing or oscillating with a correlation time or frequency at the cutoff scale. At least in perturbation theory this evolution time scale is unaffected by the longer times scales of the physical theory being studied indicating that critical slowing down has been eliminated.

References

[1] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett.* **B195** (1987) 216.

[2] M. Girolami and B. Calderhead, *Riemann manifold Langevin and Hamiltonian Monte Carlo methods*, *Journal of the Royal Statistical Society: Series B* **73** (2011) 123.

[3] G. Cossu, P. Boyle, N. Christ, C. Jung, A. Jüttner and F. Sanfilippo, *Testing algorithms for critical slowing down*, *EPJ Web Conf.* **175** (2018) 02008 [1710.07036].

[4] X.-Y. Jin, *Ensemble Quasi-Newton HMC*, in *Proceedings, 36th International Symposium on Lattice Field Theory (Lattice 2018)*, vol. LATTICE2018, p. 027, 2018.

[5] Y. Zhao, *Testing a new gauge-fixed Fourier acceleration algorithm*, in *Proceedings, 36th International Symposium on Lattice Field Theory (Lattice 2018)*, vol. LATTICE2018, p. 156, 2018.

[6] J. Zinn-Justin, *Renormalization and Stochastic Quantization*, *Nucl. Phys.* **B275** (1986) 135.

[7] M. Luscher and S. Schaefer, *Non-renormalizability of the HMC algorithm*, *JHEP* **04** (2011) 104 [1103.1810].

[8] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky and K. G. Wilson, *Langevin Simulations of Lattice Field Theories*, *Phys. Rev.* **D32** (1985) 2736.

[9] G. Parisi and Y.-s. Wu, *Perturbation Theory Without Gauge Fixing*, *Sci. Sin.* **24** (1981) 483.

[10] A. M. Horowitz, *A Generalized guided Monte Carlo algorithm*, *Phys. Lett.* **B268** (1991) 247.