Effective action for CT-AUX

Kipton Barros

November 13, 2021

1 Perturbation expansion in "vertex interactions"

Our goal is to calculate

$$Z_n \propto \prod_{\sigma=\pm 1} \det N_\sigma^{-1} = e^{-S}, \tag{1}$$

or in terms of the action,

$$S = -\sum_{\sigma=\pm 1} \mathrm{Tr} \, \ln N_\sigma^{-1}. \tag{2}$$

The matrix,

$$N_\sigma^{-1} = e^{V_\sigma} - G_{0\sigma} \left(e^{V_\sigma} - I \right) \tag{3}$$

is defined in terms of the two matrices,

$$V_\sigma = \gamma \, \mathrm{diag} \left[(-1)^\sigma s_1, \dots (-1)^\sigma s_n \right] \tag{4}$$

$$(G_{0\sigma})_{ij} = g_\sigma(\tau_i - \tau_j + 0^+), \tag{5}$$

where $g_\sigma(\tau)$ is the free fermion Green's function. The indices i and j run over all n "vertices". Each vertex i carries spin $s_i = \pm 1$ and imaginary time $\tau_i \in [0, \beta)$. Eventually we want to generalize to cluster problems, which would introduce an additional site index $\mathbf{x}_i$ for the vertex i.

Let us develop a perturbation expansion by decomposing $G_{0\sigma}$ in terms of its diagonal and off-diagonal parts. Every diagonal element of the matrix $G_{0\sigma}$ has the same value, $g_\sigma(0^+)$. Therefore we can write

$$G_{0\sigma} = g_\sigma(0^+) I + G_\sigma^{\mathrm{off}}. \tag{6}$$

where I is the $n \times n$ identity matrix, and G_σ^{off} is what remains of $G_{0\sigma}$.

Similarly, we can decompose

$$N_\sigma^{-1} = D_\sigma - \Gamma_\sigma, \tag{7}$$

where

$$D_\sigma = e^{V_\sigma} - g_\sigma(0^+) \left(e^{V_\sigma} - I \right) \tag{8}$$

$$\Gamma_\sigma = G_\sigma^{\mathrm{off}} \left(e^{V_\sigma} - I \right). \tag{9}$$

are purely diagonal and off-diagonal, respectively. To see this, note that Γ_σ is the product of a purely off-diagonal with a purely diagonal, which produces a purely off-diagonal matrix. Our intention is to expand in powers of the "interaction matrix" Γ_σ, which couples vertices to each other.

The desired action involves the trace of a matrix logarithm,

$$\begin{aligned}\ln N_\sigma^{-1} &= \ln\left(D_\sigma - \Gamma_\sigma\right) \\ &= \ln D_\sigma + \ln(I - D_\sigma^{-1}\Gamma_\sigma).\end{aligned} \tag{10}$$

The magnitude of V_σ can be controlled via the tuneable parameter γ, so e^{V_σ} can be brought arbitrarily close to the identity, I. In this limit, $D_\sigma \approx I$ and $\Gamma_\sigma \approx 0$, which justifies an expansion in small $D_\sigma^{-1}\Gamma_\sigma$. Using the Taylor series,

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots. \tag{11}$$

we find the matrix expansion

$$\ln N_\sigma^{-1} = \ln D_\sigma - D_\sigma^{-1}\Gamma_\sigma - \frac{1}{2}\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 - \ldots. \tag{12}$$

To calculate the action in Eq. (2) we need to evaluate the trace,

$$\mathrm{Tr}\,\ln N_\sigma^{-1} = \mathrm{Tr}\,\ln D_\sigma - \mathrm{Tr}\,D_\sigma^{-1}\Gamma_\sigma - \frac{1}{2}\mathrm{Tr}\,\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 - \ldots \tag{13}$$

The logarithm of the diagonal matrix D_σ, Eq. (8), can be calculated element-by-element. The resulting trace is

$$\mathrm{Tr}\,\ln D_\sigma = \sum_i \left[e^{\gamma(-1)^\sigma s_i} - g_\sigma(0^+)\left(e^{\gamma(-1)^\sigma s_i} - 1\right)\right] \tag{14}$$

Next, observe that $D_\sigma^{-1}\Gamma_\sigma$ is the product of a purely diagonal with a purely off-diagonal, which produces a purely off-diagonal matrix. That is, all diagonal elements of $D_\sigma^{-1}\Gamma_\sigma$ are zero, and therefore its trace is zero,

$$\mathrm{Tr}\,D_\sigma^{-1}\Gamma_\sigma = 0.$$

Finally, we get to the interesting "vertex interaction" terms, associated with

$$\mathrm{Tr}\,\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 = \sum_{ij}\left(D_\sigma^{-1}\Gamma_\sigma\right)_{ij}\left(D_\sigma^{-1}\Gamma_\sigma\right)_{ji}.$$

Since $D_\sigma^{-1}$is diagonal, we can write

$$\left(D_\sigma^{-1}\Gamma_\sigma\right)_{ij} = \left(D_\sigma^{-1}\right)_{ii}\left(\Gamma_\sigma\right)_{ij}.$$

Finally, we arrive at this sum over all vertex pairs (ij),

$$\mathrm{Tr}\,\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 = \sum_{i\neq j}\left(D_\sigma^{-1}\right)_{ii}\left(D_\sigma^{-1}\right)_{jj}\left(\Gamma_\sigma\right)_{ij}\left(\Gamma_\sigma\right)_{ji}. \tag{15}$$

It is valid to restrict attention to $i \neq j$ because Γ_σ is purely off-diagonal. These off-diagonal elements are explicitly,

$$\left(\Gamma_\sigma\right)_{ij} = \left(G_{0\sigma}\right)_{ij} \left(e^{(V_\sigma)_{jj}} - 1\right).$$

To summarize, we can express the target log weight as

$$\begin{aligned} S &= -\sum_{\sigma=\pm 1} \mathrm{Tr}\, \ln N_\sigma^{-1} \\ &= -\sum_{\sigma=\pm 1} \mathrm{Tr}\left[\ln D_\sigma - \frac{1}{2}\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 - \frac{1}{3}\left(D_\sigma^{-1}\Gamma_\sigma\right)^3 - \dots\right]. \end{aligned} \tag{16}$$

The first and second terms have exact closed forms as given in Eqs. (14) and (15).

1.1 Effective Ising interactions

Let us now try to give some interpretation to the "interaction term" of Eq. (15). The perturbation expansion is controlled under the assumption that $V_\sigma \propto \gamma$ is small. In this limit, to leading order we have

$$D_\sigma^{-1} \approx I,$$

$$\left(\Gamma_\sigma\right)_{ij} \approx \left(G_{0\sigma}\right)_{ij} \left(V_\sigma\right)_{jj}$$

such that

$$\mathrm{Tr}\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 \approx \sum_{i\neq j} \left(V_\sigma\right)_{ii} \left(V_\sigma\right)_{jj} \left(G_{0\sigma}\right)_{ij} \left(G_{0\sigma}\right)_{ji}.$$

Recall that

$$\begin{aligned} \left(V_\sigma\right)_{ii} &= \gamma(-1)^\sigma s_i \\ \left(G_{0\sigma}\right)_{ij} &= g_\sigma(\tau_i - \tau_j), \end{aligned}$$

so

$$\mathrm{Tr}\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 \approx \gamma^2 \sum_{i\neq j} s_i s_j \left[g_\sigma(\tau_i - \tau_j) g_\sigma(\tau_i - \tau_j)\right]. \tag{17}$$

This is manifestly a "quadratic interaction between Ising spins", and furthermore it is exactly translation invariant (only depends on the distance $|\tau_i - \tau_j|$ in imaginary time). Note, however, that the "pair potential" is actually associated with the *square* of the free particle Green's function $g_\sigma(\tau_i - \tau_j)$.

It is interesting to examine the next term, of order $\left(D_\sigma^{-1}\Gamma_\sigma\right)^3$. Keeping $V_\sigma \propto \gamma$ small, we get effective three-body interactions at leading order,

$$\mathrm{Tr}\left(D_\sigma^{-1}\Gamma_\sigma\right)^3 \approx \sum_{ijk} \left(\Gamma_\sigma\right)_{ij} \left(\Gamma_\sigma\right)_{jk} \left(\Gamma_\sigma\right)_{ki}, \tag{18}$$

where the summand vanishes unless all i, j, k are distinct. Following the logic from before, we find

$$\text{Tr}\,\left(D_\sigma^{-1}\Gamma_\sigma\right)^3 \approx \gamma^3(-1)^{3\sigma}\sum_{ijk} s_i s_j s_k g_\sigma(\tau_i-\tau_j)g_\sigma(\tau_j-\tau_k)g_\sigma(\tau_k-\tau_i), \qquad (19)$$

an explicit cubic interaction of Ising spins! This term is starting to get expensive to compute explicitly, naively requiring $\mathcal{O}(n^3)$ operations to iterate through all vertex triples. Furthermore, we emphasize that this is not the only term that contributes to the action S at order γ^3. The "lower order" terms Eq. (14) and (15) contribute at all orders in γ, including at order γ^3. Nonetheless, Eq. (19) is interesting because it illustrates a general point: Higher order spin interactions carry additional factors of $g_\sigma(\Delta\tau)$, and therefore tend to be highly localized. Could it be possible to replace *all* these higher order interactions with a local ML model? I'm not sure, because the result would need to depend on $g_\sigma(\Delta\tau)$ in a complicated way.

2 Algorithm proposal

Our goal is to sample vertex configurations x from a distribution

$$P[x] \propto e^{-S[x]}, \qquad (20)$$

where the action is

$$S = -\sum_{\sigma=\pm1} \text{Tr}\,\ln N_\sigma^{-1}. \qquad (21)$$

The matrix N_σ^{-1} is implicitly a function of the entire configuration x. We have developed a perturbation expansion,

$$S = -\sum_{\sigma=\pm1}\text{Tr}\left[\ln D_\sigma - \frac{1}{2}\left(D_\sigma^{-1}\Gamma_\sigma\right)^2 - \frac{1}{3}\left(D_\sigma^{-1}\Gamma_\sigma\right)^3 - \dots\right], \qquad (22)$$

where

$$D_\sigma = e^{V_\sigma} - g_\sigma(0^+)\left(e^{V_\sigma} - I\right) \qquad (23)$$

$$\Gamma_\sigma = G_\sigma^{\text{off}}\left(e^{V_\sigma} - I\right). \qquad (24)$$

Note that $g_\sigma(0^+)$ is a scalar and e^{V_σ} is a diagonal matrix. The tricky part is dealing with the purely off-diagonal matrix G_σ^{off}. The elements

$$\left(G_\sigma^{\text{off}}\right)_{ij} = g_\sigma(\tau_i - \tau_j) \qquad (i \neq j), \qquad (25)$$

can be interpreted as introducing a "coupling" between vertex pairs.

The first term in Eq. (22) is a sum over isolated vertices and can be readily evaluated. The second represents "pair interactions" between vertices. With clever user of the fast Fourier transform (FFT) I believe this second term can be evaluated with cost $\mathcal{O}(n \ln n)$ (see, e.g. particle-particle particle-mesh (P3M) method). The rest of S be interpreted as higher order terms in a many-body expansion. Direct evaluation would require summation over all triples, quadruples, etc., which should be avoided.

2.1 Stochastic approximation of ΔS

Although we cannot readily evaluate all the terms in Eq. (22) exactly, we *can* form an efficient and unbiased stochastic approximator. The core idea is to approximate

$$\operatorname{Tr} A_\sigma^n \approx r^\dagger A_\sigma^n r,$$

where

$$A_\sigma = D_\sigma^{-1} \Gamma_\sigma \tag{26}$$

and r is an appropriately selected random vector (see the literature on "Hutchinson estimation" for example).

It turns out to be best to stochastically approximate the *change* in action,

$$\Delta S = S[x'] - S[x],$$

for a proposed update $x \to x'$. We anticipate significant cancellations in the matrix $A_\sigma^n[x'] - A_\sigma^n[x]$, which should reduce the overall stochastic error.

Our final stochastic approximator has the form

$$\widetilde{\Delta S} = \Delta S_{\text{pair}} + \Delta S_{\text{stoch}}. \tag{27}$$

Single and pair interactions can be evaluated exactly,

$$\Delta S_{\text{pair}} = - \sum_{\sigma = \pm 1} \operatorname{Tr} \left[(\ln D_\sigma[x'] - \ln D_\sigma[x]) - \frac{1}{2} \left(A_\sigma[x']^2 - A_\sigma[x]^2 \right) \right].$$

The higher order terms are (hopefully) much smaller, and here is where we require stochastic approximation,

$$\Delta S_{\text{stoch}} = \sum_{\sigma = \pm 1} \sum_{m=3}^{m_{\max}} \frac{1}{m} r^\dagger \left(A_\sigma[x']^m - A_\sigma[x]^m \right) r.$$

The computational task is to repeatedly apply the matrix $A_\sigma[x']$ or $A_\sigma[x]$ to a random vector r. Unraveling the definitions above, the key difficulty is applying the matrix G_σ^{off} to a vector. Since G_σ^{off} is translation-invariant, I believe we can efficiently apply it in Fourier space using tricks like in P3M. Therefore, all terms in Eq. (27) can be efficiently evaluated up to some cutoff $m_{\max}$.

2.2 Kentucky sampling algorithm

So far, we have been concerned with efficiently approximating $\Delta S = S[x'] - S[x]$ in an unbiased way. Remarkably, this is a sufficient building block to construct a full Monte Carlo method for sampling vertex configurations x from the distribution $\exp(-S[x])$.

The method of Kennedy and Kuti ("Noise without Noise: A new Monte Carlo method") provides a recipe for sampling x when we have available an unbiased estimator $\langle \exp(-\Delta S) \rangle$ of $\exp(-\Delta S)$. The recipe is

$$P(x_1 \to x_2) = \begin{cases} \lambda^+ + \lambda^- \langle \exp(-\Delta S) \rangle & \text{if } f(x_1) > f(x_2) \\ \lambda^- + \lambda^+ \langle \exp(-\Delta S) \rangle & \text{if } f(x_1) \leq f(x_2) \end{cases}, \tag{28}$$

where $\lambda^\pm$ are tunable parameters ranging from 0 to 1, and $f(x)$ is an arbitrary function (ideally, correlated with ΔS) that provides an ordering on x. Statistical correctness can be guaranteed provided that $\langle \exp(-\Delta S) \rangle$ is an unbiased estimator, *and* appropriately bounded (for sure it should be nonnegative, and an upper bound is also required depending on the $\lambda^\pm$).

Given the unbiased estimator for ΔS, it turns out there are various ways to formulate an unbiased estimator for $\exp(-\Delta S)$. One possibility is the series expansion developed by Bhanot and Kennedy ("Bosonic lattice gauge theory with noise")

$$\langle \exp(-\Delta S) \rangle = 1 + x_1(1 + \frac{1}{2} x_2 (1 + \frac{1}{3} x_3 (1 + \dots))),$$

where each $x_1 \dots x_N$ is an independent unbiased estimator of $-\Delta S/N$. Interpreting the above coefficients 1, 1/2, 1/3, ... as probabilities, this infinite sum can be sampled exactly. Implementation details and various improves are described in Lin, Liu, and Sloan ("A Noisy Monte Carlo Algorithm"). A problem is that although the estimator $\langle \exp(-\Delta S) \rangle$ is unbiased, it is not guaranteed to satisfy the bounds required for correctness in Eq. (28). The bounds can be controlled by using more random vectors, but it is argued by Lin et al. that there is an unfavorable scaling with system size. Another possibility is to use the improved "Kentucky sampling" method described in Lin et al., which assuredly gives correct statistics. Reference: `https://arxiv.org/abs/hep-lat/9905033v2`