EM sea effects in hadron polarizabilities through reweighting

Walter Freeman, Andrei Alexandru, Michael Lujan, Frank Lee

The George Washington University

July 29, 2013
The background field method

- Expect a mass shift equal to $\alpha_E E^2$ in the presence of a uniform background field.
- Determine polarizability by measuring neutron correlators with $\vec{E} = 0$ and $\vec{E} = \pm iE_0 \hat{x}$, then fitting them to determine the mass shift.
- When fitting correlators, the zero-field and nonzero-field correlators are correlated.
- This results in a much smaller error on $\Delta M$ than on the mass measurements themselves.
Reweighting approach

How do we include the effects of the sea quark charges in the background field approach?

- In principle it’s easy: just generate two otherwise identical ensembles, one with a background field and one without.
- But this requires unaffordably high statistics, since our two mass measurements now no longer have correlated errors.
  - Lose all the information in the “cross-correlation” terms of the covariance matrix.
- Reweighting is a technique for extracting physics from a different action than the one used in generation: “retroactively change the ensemble parameters.”
- We can use it to generate two correlated ensembles, one with and one without the electric field.
Reweighting

Reweighting is just a modification to the standard quantum Monte Carlo, where only a part of the factor $e^{-S}$ is absorbed into a Monte Carlo weight:

**Standard**

$$\frac{\int [dU] O e^{-S_0}}{\int [dU] e^{-S_0}} \rightarrow \sum O_i \sum 1$$

**Reweighted**

$$\frac{\int [dU] O e^{-S_E}}{\int [dU] e^{-S_E}} = \frac{\int [dU] O e^{-(S_E - S_0)} e^{-S_0}}{\int [dU] e^{-(S_E - S_0)} e^{-S_0}} \rightarrow \sum O_i w_i \sum w_i$$

where $w_i = e^{-(S_E - S_0)_i}$.

This will only work well if the two ensembles overlap sufficiently.

- Otherwise, the weight factor will fluctuate wildly, and the ensemble will be dominated by a few configs with large weights.
In order to do reweighting, need the weight factors
\[ w_i = e^{-\Delta S} = \det^{-1} M^{-1}_\eta M_0 \]
- There is a standard stochastic estimator for the inverse determinant
- Several improvement techniques, like low-mode separation and determinant breakup, are very successful in reducing stochastic noise when reweighting in \( m_q \)
- They don’t work at all when reweighting in the background field
- The fluctuations in this standard stochastic estimator are awful (and it is expensive)
  - So long as the estimator is unbiased, the result will be too – just with larger error bars
  - Useful yardstick: ideally stochastic fluctuations (“noise”) less than gauge fluctuations (“signal”)
  - We are so far away from this benchmark that it looks hopeless
A new pseudo-perturbative approach

- The standard improvement techniques (determinant breakup, low mode separation) used for mass reweighting don’t work.
- Can’t distinguish the value of the weight factor from 1 with any sane number of noises on a production lattice.
- Can we make use of the fact that we only need perturbatively small $\eta$?
- Perhaps it is easier to estimate $\frac{\partial w_i}{\partial \eta} \bigg|_{\eta=0}$ and $\frac{\partial^2 w_i}{\partial \eta^2} \bigg|_{\eta=0}$ than $w_i$ itself?
- Expand $w_i$ in a power series in $\eta$ up to second order, about $\eta = 0$
  - Linear term in weight factor can combine with linear dependence of $G_N(t)$ on $\eta$ to give quadratic effect.
  - Quadratic term in weight factor by itself can give quadratic effect.
- If we can estimate these derivatives instead we can evaluate at any $\eta$ we choose to get $w_i(\eta)$.
- Sea contributions taken into account in a way that is similar in practice to the current-insertion approach of Engelhardt.
Derivation of the estimator

For the first derivative, we want \[ \frac{\partial}{\partial \eta} \frac{\det M_\eta}{\det M_0} \bigg|_{\eta=0}. \] Rewrite \( \det M_\eta \) as a Grassman integral:

\[
\frac{\partial}{\partial \eta} \frac{\det M_\eta}{\det M_0} \bigg|_{\eta=0} = \frac{1}{\det M_0} \frac{\partial}{\partial \eta} \int d\psi d\bar{\psi} \ e^{-\bar{\psi} M \psi} \\
= \frac{1}{\det M_0} \int d\psi d\bar{\psi} \ - \bar{\psi} \frac{\partial M_0}{\partial \eta} e^{-\bar{\psi} M_0 \psi} \\
= \text{Tr} \left( \frac{\partial M_0}{\partial \eta} M_0^{-1} \right).
\]

This trace still must be evaluated stochastically.

The second derivative proceeds similarly:

\[
\frac{\partial^2}{\partial \eta^2} \frac{\det M_\eta}{\det M_0} \bigg|_{\eta=0} = -\text{Tr} \frac{\partial^2 M}{\partial \eta^2} M_0^{-1} + \left( \text{Tr} \frac{\partial M}{\partial \eta} M_0^{-1} \right)^2 - \text{Tr} \left( \frac{\partial M}{\partial \eta} M_0^{-1} \right)^2
\]

Unfortunately, stochastic estimators of the traces here are still too noisy.
Hopping-parameter expansion improvement

- We want to estimate $\text{Tr } O$ as $\langle \xi^\dagger O \xi \rangle$, but that estimator is too noisy.
- We can always add and subtract the same thing, so we can also write:

$$\text{Tr } O = \left\langle \xi^\dagger \left( O - \sum_i O'_i \right) \xi \right\rangle + \sum_i \text{Tr } O'_i$$

- Identify other operators $O'_i$ with the following properties:
  - $\text{Tr } O'_i$ can be computed exactly
  - Stochastic estimators of $\text{Tr } O'_i$ have correlated fluctuations with that of $O$
- Construct such operators by making a hopping parameter expansion of each $M^{-1}$ that appears.
- Computing exact traces is possible, but quite messy.
- Do as many orders as you can afford (cost goes as $O(14^n)$).

- Yields substantial improvement.
- Exact traces computed up to $O(\kappa^7)$. 

![Graph showing estimator variance improvement](image)
Calculation parameters, first try

- $24^3 \times 48$ lattice, 2 flavors of dynamical nHYP-clover fermions with $m_\pi \simeq 300 \text{MeV}$, 300 configs
- At least 3000 stochastic estimators of first-order term and 1000 estimators of each second-order term (and often more) per config

- Normally, the effect of reweighting on the statistical power of a calculation can be determined pretty easily:
  - Fluctuations of weight factor decrease effective number of configs: $N_{\text{eff}} = N \frac{\langle w \rangle^2}{\langle w^2 \rangle}$
- But we rely on correlations between zero-field and nonzero-field correlators to reduce error on $\Delta E$
- These correlations are very strong and get stronger for low $\eta$: for $\eta = 10^{-4}$, $\sigma_E = 9 \times 10^{-8}$, $\sigma_{\Delta E} = 6 \times 10^{-3}$
- Reweighting only the nonzero-field correlators may spoil these correlations, even if all the weight factors are $O(1)$
- No way to know how well reweighting will work until the end of the calculation
Even with all of this effort, the first-order estimates are just barely distinguishable from random noise:

To see if we can resolve the gauge fluctuations through the noise, do a zero-parameter fit to \( \text{Tr} \frac{\partial M}{\partial \eta} M^{-1} = 0 \):

This is still dominated by stochastic fluctuations, but the gauge fluctuations (signal) are evident – barely.

What about the second-order term?
Here the gauge average is not zero, so fit to a constant:

This is indistinguishable from pure noise

Ran extra noise sources on the first hundred configs to see if a signal would appear, to no effect.
Weight factors

- Need to choose a particular $\eta$ at which to evaluate the expansion to finish the calculation
- Valence correlators computed mostly with $\eta = 0.0051$
- This value of $\eta$ is too big for a naïve reweighting:
  - Large constant shift in the action from $E$ causes expansion to break down ($\Delta S > 1$)
  - Still small enough that the effects of $E$ on hadron are perturbative (only quadratic valence behavior seen)
- Solve this by using valence correlators “rescaled” to $\eta = 10^{-4}$
  - Effect on valence correlators is nearly perfectly quadratic, so this is okay

Then, everything is nice and perturbative
Combine these with the “rescaled” valence correlators and do the reweighting order by order.

Fitting from \( t = 9 \) to \( t = 21 \) gives the following for \( \Delta E \):

<table>
<thead>
<tr>
<th>Order</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0(^{th})</td>
<td>(6.01(88) \times 10^{-7})</td>
</tr>
<tr>
<td>1(^{st})</td>
<td>(2.86(2.32) \times 10^{-7})</td>
</tr>
<tr>
<td>2(^{nd})</td>
<td>(17.8(10.9) \times 10^{-7})</td>
</tr>
</tbody>
</table>

Clearly we need to beat down the stochastic noise further, especially at second order.

Would like a gain of roughly a factor of ten in error!
Origin of stochastic noise

- Variance of stochastic estimator proportional to sum of off-diagonal matrix elements
- Can’t study them all, but we can map out a representative set grouped by offset

- For most uses of this stochastic estimator (to get, say, $\text{Tr} \, M^{-1}$), the matrix is diagonally dominant
- Not the case here: $\frac{\partial M}{\partial \eta} M^{-1}$ has large offdiagonal elements
- This is a stark depiction of why this problem is so hard
Origin of stochastic noise

- How does hopping parameter expansion help things?
- Improvement operators are ultralocal; only expect reduction near diagonal

Exactly what is expected: improvement up to radius 2

- Suppresses large near-diagonal elements which dominate the noise, but doesn’t eliminate them
Origin of stochastic noise

- To seventh order there is more reduction in offdiagonal elements
- Improvement up to radius 7, as expected

- Shift in value of diagonal elements comes from traces of improvement operators which are added back in the full calculation
- This is as far as the hopping-parameter expansion can realistically take us
- What else can we do?
Dilution separates the matrix dimension into $N$ subsets and stochastically estimates the trace over each separately

- **Advantage:** eliminates noise contributions from offdiagonal terms from different subsets
- **Disadvantage:** Requires $N$ operations to cover the lattice; could have reduced noise by factor of $\sqrt{N}$ by simple repetition
- **Only outperforms simple repetition** if the offdiagonal matrix elements “kept” are lower than the average

We should choose a (four-dimensional generalization) of the rightmost scheme to eliminate the large near-diagonal elements
Dilution, for us

- Spin-color dilution on its own actually makes things worse
- Spatial dilution is somewhat redundant with hopping parameter expansion improvement
- Until you get to very aggressive ($N$ large) dilution schemes, dilution in the presence of HPE improvement makes things worse
- Best dilution schemes are gridding with an additional “8-way hypercubic checkerboard” pattern overlaid, along with standard spin-color dilution
  - Nearest neighbor in a grid with spacing $x$ is Manhattan distance $2x$ away
- Using a $4^4$ grid along with spin-color dilution (24,576 subspaces) breaks even with 24,576 HPE-improved undiluted noises
- Using a $6^4$ grid (124,416 subspaces!) seems to gain roughly a factor of 2 in error per inversion for first-order term
  - One such $6^4$ diluted estimate per config will give us a signal/noise ratio of about 1 on the first-order terms
  - Still not sure what the signal strength for the second-order terms is; haven’t been able to resolve it yet
- How many inversions are we willing to dedicate to each configuration?
Performance improvements and proposed future run

- Have already done some things that will improve performance on a subsequent run
  - Inversion reuse: save result of $M^{-1}\psi$ and use it for all three operators that need estimators (factor of 3)
  - Decreased inverter precision: using $10^{-5}$ rather than $10^{-10}$ gives a three-fold speedup (and negligible penalty)
    - Need to investigate “sloppy CG”: can perhaps wring another factor of 2 or 3 out of it
  - $6^4$ grid + spin-color dilution + 8-way “checkerboard” for the whole ensemble is 75 million inversions = 300k GPU-hours
    - This should buy us the factor of ten we need so that the total error in the polarizability isn’t dominated by the sea contribution
Conclusions

- Stochastic estimates of weight factor expansion terms are very noisy.
- Much, much worse than many common stochastic estimates ($\text{Tr } M^{-1}$) because of strong nondiagonal dominance.
- Hopping-parameter expansion helps some, but not enough.
- Need very strong dilution ($N = 124416$) to show gains over repeated undiluted HPE-improved estimator.
- Combined effect of various optimizations means that such a run is possible, and should give the needed decrease in error.
- If we see a significant shift in the polarizability from the sea effect, we can then attack the ensemble with lighter $m_\pi$ (for which deflation will pay dividends) or the ensemble with a larger $n_x$. 