Hutch++ and XTrace
to improve stochastic Trace estimation

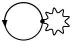
Alessandro Cotellucci, Agostino Patella

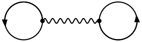
Lattice 23 Fermilab, August 4th 2023

Motivation

Isospin-Breaking effects disconnected loop diagrams:

Strong IB: $\mathrm{Tr}\left[D^{-1}\right] =$ 

Tadpole: $\mathrm{Tr}\left[TD^{-1}\right] =$ 

Bubbles: $\mathrm{Tr}\left[JD^{-1}\right]\mathrm{Tr}\left[JD^{-1}\right] =$ 

Self-Energy: $\mathrm{Tr}\left[JD^{-1}JD^{-1}\right] =$ 

Is there a better stochastic trace estimator than the random sources?

Configurations with $N_f = 3 + 1$ $O(a)$ improved Wilson fermions with C* b. c. in space.

lattice	a [fm]	m_π [MeV]	m_D [MeV]	no. cnfg
64×32^3	0.05393(24)	398.5(4.7)	1912.7(5.7)	1

Hutch++

The Method[a]

$$\langle \text{Tr}\left[D^{-1}\right]\rangle = \underbrace{\sum_i^{nvec} q_i^\dagger D^{-1} q_i}_{\text{Deflated term}} + \underbrace{\frac{1}{nsrc}\sum_i^{nsrc} g_i^\dagger \left(\mathbb{1} - \sum_j q_j q_j^\dagger\right) D^{-1} g_i}_{\text{Stochastic remnant}}$$

- The deflation space is obtained by the orthonormalization of the vectors $D^{-1}w_i$ with w_i U(1) random sources;
- No. of inversions of D: $m = 2nvec + nsrc$;

This method has a better convergence in the asymptotic regime than the Hutchinson[a].

[a] Meyer et al., "Hutch++: Optimal Stochastic Trace Estimation".

Hutch++

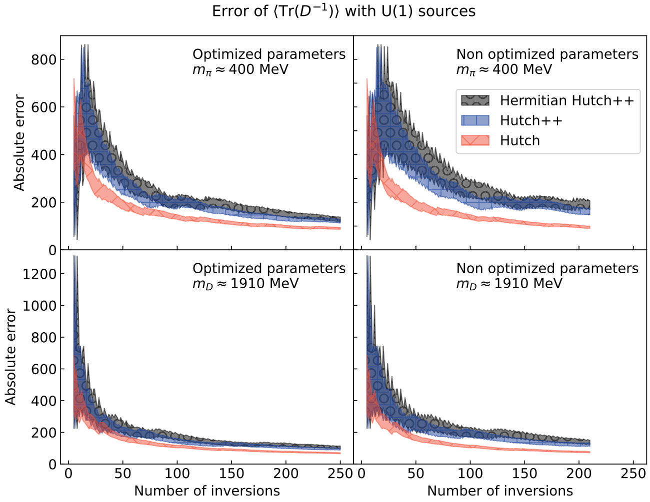
Optimized Parameters

The parameters have been fixed in two ways:

Operator used to compute deflation space	Optimized parameters	Non optimized parameters
D^{-1}	$nvec = \frac{m+2}{4}$ $nsrc = \frac{m-2}{2}$	$nvec = nsrc = \frac{m}{3}$
$\frac{1}{2}\left(D^{-1} + D^{-1\dagger}\right)$	$nvec = \frac{m+2}{5}$ $nsrc = \frac{2m-6}{5}$	$nvec = nsrc = \frac{m}{4}$

Hutch++

Results: scaling with No. of inversions

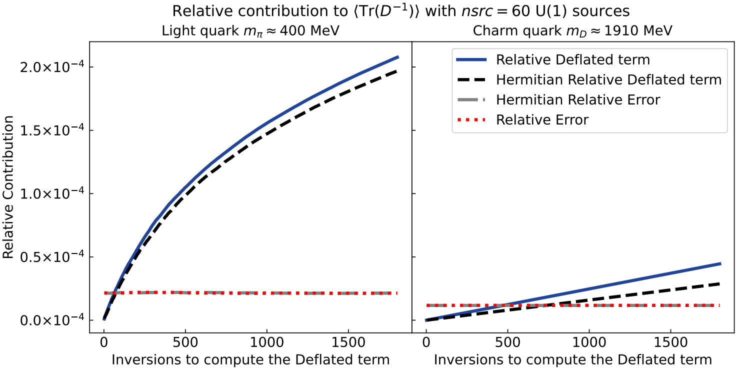


Hutch++

Results: relative contribution of the Deflated term to the trace

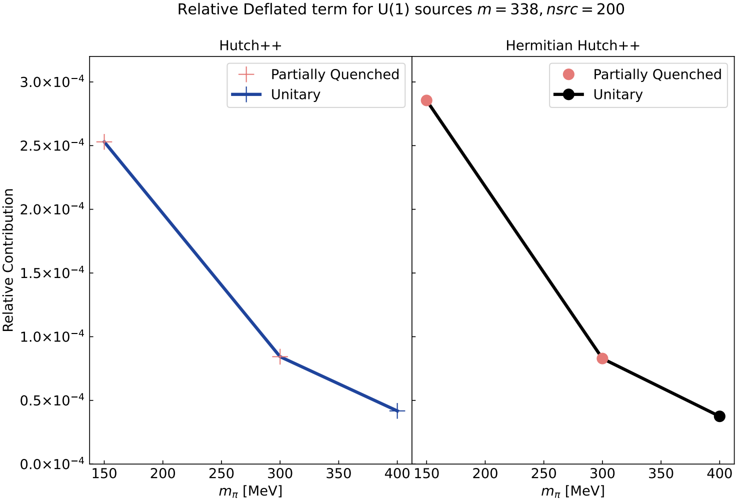
Relative Deflated term: $$\frac{\sum_i^{nvec} q_i^\dagger D^{-1} q_i}{\langle \mathrm{Tr}\left(D^{-1}\right)\rangle}$$

Relative Error: $$\frac{\sigma\left[\mathrm{Tr}\left(PD^{-1}\right)\right]}{\langle \mathrm{Tr}\left(D^{-1}\right)\rangle}$$



Hutch++

Results: dependence with the mass



XTrace

The Method[b]

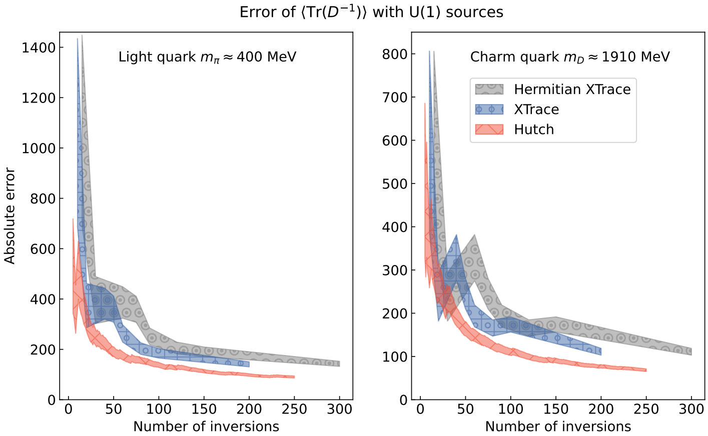
$$\langle \mathrm{Tr}\left[D^{-1}\right]\rangle = \frac{1}{nsrc}\sum_{j}^{nsrc}\left[\sum_i {q_i}^{j\dagger} D^{-1} {q_i}^{j} + w_j^\dagger \left(\mathbb{1} - \sum_k {q_k}^{j} {q_k}^{j\dagger}\right) D^{-1} w_j\right]$$

- The deflation vectors ${q_i}^j$ are obtained by the orthonormalization of $D^{-1}w_i$ with w_i U(1) random sources and $i \neq j$;
- No. of inversions of D: $m = 2nsrc$;

[b] Epperly, Tropp, and Webber, *XTrace: Making the most of every sample in stochastic trace estimation.*

XTrace

Results: scaling with No. of inversions



Conclusions & Outlook

- *Hutch++* and *XTrace* do not improve the stochastic evaluation of the trace of the inverse of the Dirac operator with respect to the *Hutchinson Trace Estimator*.
- For both algorithms the use of the hermitian part of D^{-1} does not provide a better estimation of the deflation space.
- The evaluation of disconnected diagrams requires better techniques which will be part of another study.

Thank you for your attention

References

[1] Ethan N. Epperly, Joel A. Tropp, and Robert J. Webber. *XTrace: Making the most of every sample in stochastic trace estimation*. 2023. arXiv: 2301.07825 [math.NA].

[2] Raphael A. Meyer et al. "Hutch++: Optimal Stochastic Trace Estimation". In: *Symposium on Simplicity in Algorithms (SOSA)*, pp. 142–155. DOI: 10.1137/1.9781611976496.16. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611976496.16. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611976496.16.

Acknowledgments

The work was supported by the North-German Supercomputing Alliance (HLRN) with the project bep00102.

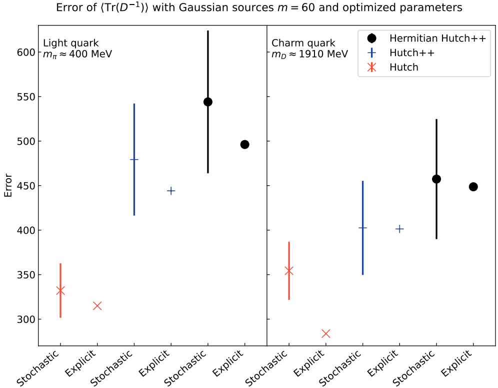
Backup

Test for Hutch++ variance

For Gaussian sources

$$\mathrm{Var}\left\{\mathrm{Re}\left[\mathrm{Tr}\left(D^{-1}\right)\right]\right\} = \frac{1}{2}\left[\mathrm{Tr}\left(D^{-1}D^{-1}\right) + \mathrm{Tr}\left(D^{-1\dagger}D^{-1}\right)\right]$$

$$\mathrm{Var}\left\{\mathrm{Re}\left[\mathrm{Tr}\left(PD^{-1}\right)\right]\right\} = \frac{1}{2}\left[\mathrm{Tr}\left(PD^{-1}PD^{-1}\right) + \mathrm{Tr}\left(PD^{-1\dagger}PD^{-1}\right)\right]$$



Data: D^{-1}
Result: $\langle \mathrm{Tr}\left[D^{-1}\right]\rangle$, $\mathrm{Var}\left\{\mathrm{Tr}\left[D^{-1}\right]\right\}$
Generate n_{vec} random vectors $w_i \in \mathbb{C}^N$;
$y_i = D^{-1} w_i$;
$Q = \mathrm{Ortho}\,(Y)$;
$Proj = \sum_i^{nvec} q_i^\dagger D^{-1} q_i$;
for i *in* $0 : nsrc$ **do**
  Generate g_i random vector;
  $Rem_i = g_i^\dagger \left(\mathbb{1} - \sum_j^{nvec} q_j q_j^\dagger\right) D^{-1} g_i$;
end
$\langle \mathrm{Tr}\left[D^{-1}\right]\rangle = Proj + \frac{1}{nsrc}\sum_i^{nsrc} Rem_i$;
$\mathrm{Var}\left\{\mathrm{Tr}\left[D^{-1}\right]\right\} = \frac{1}{nsrc}\sum_i^{nsrc} Rem_i^2 - \left(\frac{1}{nsrc}\sum_i^{nsrc} Rem_i\right)^2$;

Algorithm 1: Hutch++ Algorithm

XTrace

The Algorithm

Data: D^{-1}
Result: $\langle \mathrm{Tr}\left[D^{-1}\right]\rangle$, $\mathrm{Var}\left\{\mathrm{Tr}\left[D^{-1}\right]\right\}$
Generate n_{src} random vectors $w_i \in \mathbb{C}^N$;
$y_i = D^{-1} w_i$;
$Y = QR$;
$s_i^j = \frac{\left(R^{-1}\right)^*_{ij}}{\sqrt{\left(R^{-1}R^{-1\dagger}\right)_{ii}}}$;
$Proj = \sum_i^{nsrc} q_i^\dagger D^{-1} q_i$;
for i in $0 : nsrc$ **do**

$$Tr_i = -\sum_{jk} s_k^{i*} q_k^\dagger D^{-1} q_j s_j^i + w_i^\dagger \left(\mathbb{1} - \sum_j q_j q_j^\dagger + \sum_{jk} q_j s_j^i s_k^{i*} q_k^\dagger\right) y_i;$$

end
$\langle \mathrm{Tr}\left[D^{-1}\right]\rangle = Proj + \frac{1}{nsrc}\sum_i^{nsrc} Tr_i$;
$\mathrm{Var}\left\{\mathrm{Tr}\left[D^{-1}\right]\right\} = \frac{1}{nsrc}\sum_i^{nsrc} Tr_i^2 - \left(\frac{1}{nsrc}\sum_i^{nsrc} Tr_i\right)^2$;

Algorithm 2: XTrace Algorithm

Backup

The XTrace implementation

Starting from the vectors $Y = D^{-1}W$ we can write the QR decomposition:

$$Y^a = \sum_b Q^b R^{ba}$$

Now we can define the following projector:

$$\begin{aligned} P = & \text{ Orthogonal Projector on the space span}\left\{Y^1, \ldots, Y^m\right\} \\ = & \sum_a Q^a \left[Q^a\right]^\dagger \end{aligned}$$

$$P^{(j)} = \text{ Orthogonal Projector on the space span}\left\{Y^1, \ldots, \cancel{Y^j}, \ldots, Y^m\right\}$$

we can then define the orthogonal complement:

$$P - P^{(j)} = S^j \left[S^j\right]^\dagger$$

And we expand it in terms of the orthogonalized basis of the full space:

$$\tilde{S}^a = \sum_b Q^b \alpha^{ba}.$$

If now we impose the un-normalized vector to be orthogonal to the space of the Ys we get:

$$\left(\tilde{S}^a, Y^b\right) = \delta^{ab} \;\rightarrow\; \alpha = \left(R^{-1}\right)^{\dagger}$$

So now we can rewrite S^a in terms of the Q^as:

$$S^a = \sum_b \frac{\left(R^{-1}\right)^{ab*}}{\sqrt{\sum_c |\left(R^{-1}\right)^{ac}|^2}} Q^b = \sum_c \frac{\left(R^{-1} R^{-1\dagger}\right)^{ca}}{\sqrt{\left(R^{-1} R^{-1\dagger}\right)^{aa}}} Y^c$$

So we can write the projectors as:

$$P^{(j)} = P - S^j \left[S^j\right]^{\dagger}$$

Backup

The XTrace implementation

$P^{(j)}$ is a projector since:

$$P^{(j)} P^{(j)} = \mathbb{1}$$

and $P^{(j)}$ is completely independent on Y^j

$$P^{(j)} = \sum_{bc} \left[\left(R^{-1} R^{-1\dagger} \right)^{bc} - \frac{\left(R^{-1} R^{-1\dagger} \right)^{bj} \left(R^{-1} R^{-1\dagger} \right)^{jc}}{\left(R^{-1} R^{-1\dagger} \right)^{jj}} \right] Y^b \left[Y^c \right]^\dagger$$

in fact whenever $b = j$ or $c = j$ the coefficient is zero so it is independent of it.