Hutch++ and XTrace to Improve Stochastic Trace Estimation

Alessandro Cotellucci[a,*] **and Agostino Patella**[a,b]

[a]*Humboldt Universität zu Berlin, Institut für Physik & IRIS Adlershof,*
Zum Grossen Windkanal 6, 12489 Berlin, Germany

[b]*DESY, Platanenallee 6, D-15738 Zeuthen, Germany*

E-mail: alecote@physik.hu-berlin.de

We present the analysis of two recently proposed noise reduction techniques, Hutch++ and XTrace, both based on inexact deflation. These methods were proven to have a better asymptotic convergence to the solution than the classical Girard-Hutchinson stochastic method. We applied these methods to the computation of the trace of the inverse of the Dirac operator with $O(a)$ improved Wilson fermions on the QCD ensemble generated by the RC* collaboration with $m_\pi \approx 400$ MeV and $V = 64 \times 32^3$. Unfortunately, we see no noise reduction with a moderate number of sources, and we attempt an explanation of why this is the case. This study was part of the effort to evaluate isospin-breaking effects using the RM123 with C* boundary conditions in an unquenched set-up.

*Speaker

1. Motivations

The evaluation of quark disconnected diagrams is a crucial problem in many domains of Lattice QCD. In the context of isospin-breaking effects with the RM123 method [1] a collection of these diagrams arise from the sea-sea contractions. The mentioned diagrams have the following topology:

$$
\begin{aligned}
&\text{Strong IB: Tr}\left[D^{-1}\right] = \\
&\text{Tadpole: Tr}\left[TD^{-1}\right] = \\
&\text{Bubbles: Tr}\left[JD^{-1}\right]\text{Tr}\left[JD^{-1}\right] = \\
&\text{Self-Energy: Tr}\left[JD^{-1}JD^{-1}\right] =
\end{aligned}
\tag{1}
$$

where D^{-1} is the inverse of the Dirac operator, J is the conserved electromagnetic current and T is the Tadpole operator which arises from the lattice discretization of QED. The trace in (1) runs over the spin, colour and spacetime indices of the matrices and due to a large number of entries it can not be evaluated exactly but has to be approximated stochastically.

The straightforward way to compute this quantity is by using the Girard-Hutchinson trace estimator [2, 3]. Recently, two new methods have been proposed: the Hutch++ [4] and the XTrace [5] which modify the Girard-Hutchinson trace estimator to get a better asymptotic scaling. The Hutch++ has already been studied in Lattice QCD simulations by [6] in the context of the MGMLMC++ trace estimator while in this study we focus on the Hutch++ as a substitute candidate for the Girard-Hutchinson. This is the first study involving the XTrace in Lattice QCD simulations.

Since the inversion of the Dirac operator is the most expensive operation in these algorithms, we use the number of inversions as the cost unit to compare the different methods and neglect the impact of all the other operations.

2. Hutch++

The stochastic estimator used to evaluate the trace with this method is:

$$
\langle\text{Tr}\left[D^{-1}\right]\rangle = \underbrace{\sum_i^{n_{vec}} q_i^\dagger D^{-1} q_i}_{\text{Deflated term}} + \underbrace{\frac{1}{n_{src}}\sum_i^{n_{src}} g_i^\dagger\left(\mathbb{1} - \sum_j q_j q_j^\dagger\right)D^{-1} g_i}_{\text{Stochastic remnant}}
\tag{2}
$$

where D^{-1} is the inverse of the Dirac operator, g_i are U(1) random sources and q_i are the vectors belonging to the deflation space. The deflation space is obtained by the orthonormalization of the vectors $D^{-1}w_i$, where w_i U(1) are random sources.

The number of inversions of the Dirac operator is then fixed by $m = 2n_{vec} + n_{src}$. The effect of the deflation space is to give a rough approximation of the lowest eigenmodes of D in the fashion of the low-modes averaging, even with this rough eigenspace the algorithm has been proved to always have a better convergence in the asymptotic regime than the Girard-Hutchinson estimator.

Since we are interested in the real part of $\text{Tr}\left[D^{-1}\right]$ we test also a modification of the method where the deflation space is computed with the hermitian part of the inverse of the Dirac operator $D^{-1} + D^{-1\dagger}$ for which the number of inversions is $m = 3n_{vec} + n_{src}$.

In [4] the authors propose an optimal choice for n_{vec} and n_{src} which minimizes the variance, in our case the optimal parameters are defined by:

Operator used to compute deflation space	Optimized parameters	Non optimized parameters	
D^{-1}	$n_{vec} = \dfrac{m+2}{4}, \ n_{src} = \dfrac{m-2}{2}$	$n_{vec} = n_{src} = \dfrac{m}{3}$	(3)
$\left(D^{-1} + D^{-1\dagger}\right)$	$n_{vec} = \dfrac{m+2}{5}, \ n_{src} = \dfrac{2m-6}{5}$	$n_{vec} = n_{src} = \dfrac{m}{4}.$	

The two methods are tested to evaluate the trace of the inverse of the Dirac operator on an ensemble with $N_f = 3 + 1 \ O(a)$ improved Wilson fermions with C* boundary conditions in space and periodic boundary conditions in time [7].

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

Table 1: Configuration used for this work from [7].

The ensemble characteristics are listed in Table 1, to focus only on the error of the estimator we decide to use only one gauge configuration.

In Figure 1 we show the results of our comparison for a maximum number of inversions of the Dirac operator of $O(200)$.

As one can see the scaling using the optimal parameters is confirmed to be better than the non optimized ones but in all the cases the Hutch++ fails to do better than the Girard-Hutchinson showing a similar scaling.
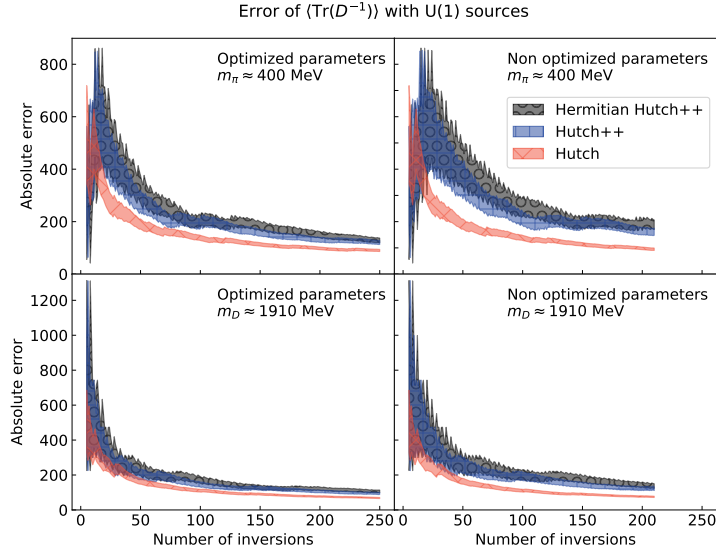

Figure 1: Scaling of the absolute error as a function of the number of inversions of D for the different algorithms using U(1) random sources for both the light quarks (top) and the charm quark (bottom).

3

2.1 The impact of the deflation space

We study the impact of the deflation space by comparing the relative contribution of the deflated term to the total trace and the statistical error of the remnant as defined:

$$
\text{Relative Deflated term:} \quad \frac{\sum_i^{n_{vec}} q_i^\dagger D^{-1} q_i}{\langle \text{Tr}\left(D^{-1}\right)\rangle}
$$

$$
\text{Relative Error:} \quad \frac{\sigma\left[\text{Tr}\left(PD^{-1}\right)\right]}{\langle \text{Tr}\left(D^{-1}\right)\rangle}, \tag{4}
$$

where P is just the projector $\left(\mathbb{1} - \sum_j q_j q_j^\dagger\right)$. In this case, we keep fixed the number of sources used to compute the stochastic remnant and we vary the number of inversions necessary to compute the Deflated term ($2n_{vec}$ for the Hutch++ and $3n_{vec}$ for the Hermitian Hutch++). When the relative deflated term is smaller than the relative error the actual contribution can not be distinguished from a statistical fluctuation.
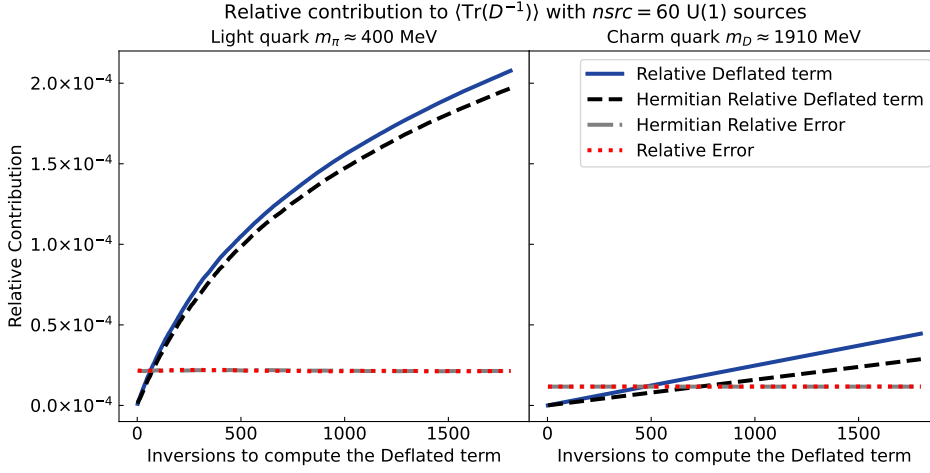
Figure 2: Scaling of the relative contribution of the different Hutch++ methods with the number of inversions to compute the deflation space. The blue solid line represents the Hutch++, the black dashed line the hermitian Hutch++, the red pointed line the relative statistical error in the Hutch++ case and the dashed green line the relative statistical error in the hermitian Hutch++ case.

In Figure 2 one can see that the impact of the deflation space becomes relevant only after a relatively big amount of inversions $O(60)$ for the light quark and over $O(500)$ for the charm quark. There is a clear improvement in the effect going to smaller quark mass but the relative contribution of the deflation space is still of the order $O\left(10^{-4}\right)$ for an extremely high computational effort.

Since the impact of the deflation space is bigger for the light quark than the charm one can expect the method to become more effective close to the physical pion mass like other eigenspace methods, to test this hypothesis we use a partially quenched setup. In Figure 3 we have the results with the relative contribution of the deflation term as a function of the pion mass. One can see that the behaviour of the Hutch++ and the hermitian Hutch++ is similar with a fixed number of inversions m but the effect of the relative term is still irrelevant even at the physical point.

4

Relative Deflated term for U(1) sources $m = 338, nsrc = 200$
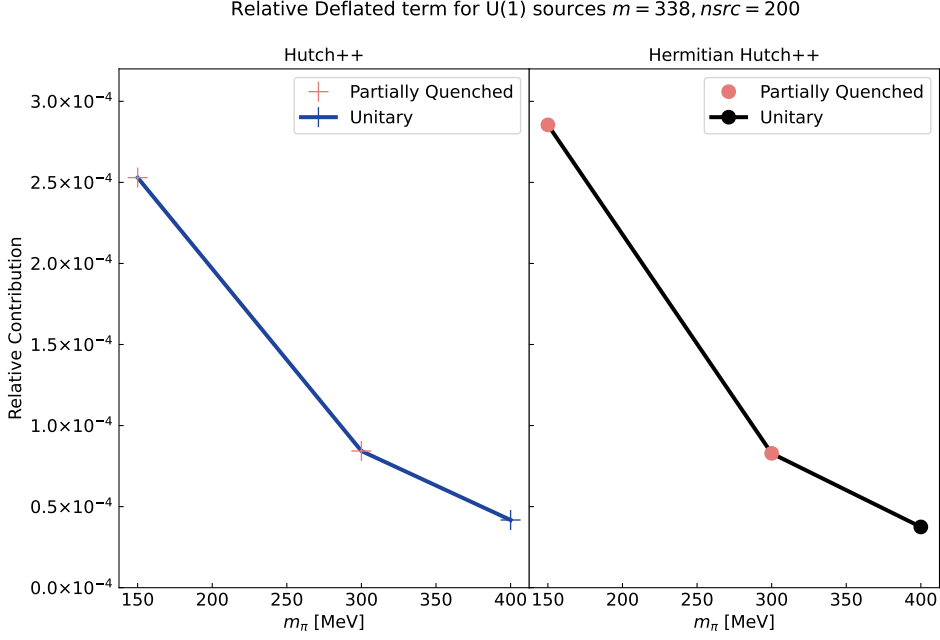
Figure 3: Plot of the Relative contribution of the deflation term as a function of the pion mass for the Hutch++ (left) and the hermitian Hutch++ (right). The red points are the partially quenched setup while others are the unitary points.
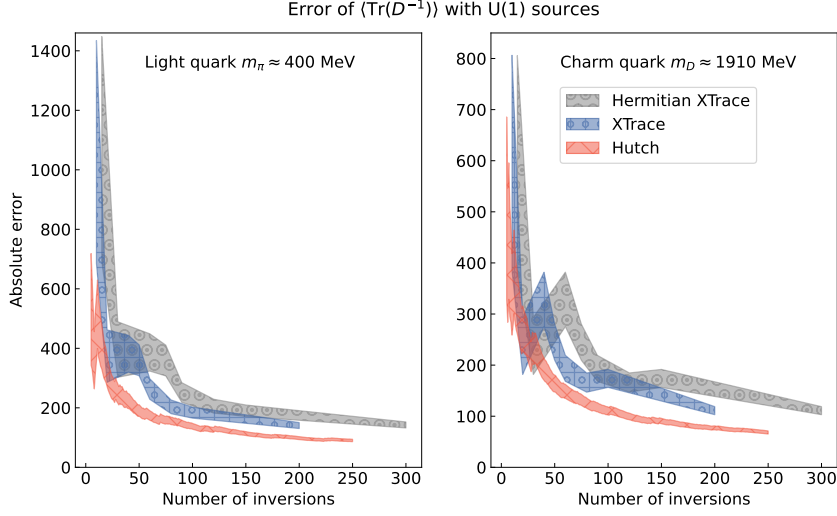
3. XTrace

The other method which was recently proposed, and we decided to test, is the XTrace estimator. In this case, the form of the estimator is:

$$\langle \text{Tr}\left[D^{-1}\right]\rangle = \frac{1}{n_{src}} \sum_{j}^{n_{src}} \left[\sum_{i} q_i^{j\dagger} D^{-1} q_i^{j} + w_j^{\dagger} \left(\mathbb{1} - \sum_{k} q_k^{j} q_k^{j\dagger} \right) D^{-1} w_j \right]$$

where w_i are U(1) random sources and the deflation vectors q_i^{j} are obtained by the orthonormalization of $D^{-1} w_i$ with $i \neq j$ so all the sources are reused to construct the deflation space and the total number of inversions of D is fixed to $m = 2n_{src}$. Also, in this case, we add an implementation of the method where the subspace is generated using the hermitian part of D^{-1} and in that case, the number of inversions is $m = 3n_{src}$. As a modification of the Hutch++, these methods inherit better asymptotic scaling than the Girard-Hutchinson.

We compare the scaling of the XTrace with respect to the Girard-Hutchinson estimator for $O(200)$ number of inversions of the Dirac operator to see which of them performs better. The results are shown in Figure 4 and they clearly show that the method does not overcome the Girard-Hutchinson trace estimator both in the light quark and in the charm quark cases.

Figure 4: Scaling of the absolute error as a function of the number of inversions of D for the different algorithms using U(1) random sources for both the quarks in the ensemble.

4. Conclusions & Outlook

In this study, we tested the Hutch++ and XTrace stochastic trace estimators versus the Girard-Hutchinson method to understand if they can be used as a substitute to speed up the trace computations in Lattice QCD simulations since they have been proven to always have better scaling of the error in the asymptotic regime. We did this test using a single configuration from a QCD ensemble generated by the RC* collaboration. We tested both methods with a variant which we called hermitian that generates the deflation space with the hermitian part of the inverse of the Dirac operator.

Unfortunately, the two methods did not show any improvement in the regime of $O(200)$ inversions of the Dirac operator. We investigated the impact of the deflated term on the total trace for different pion masses in a quenched set-up but this showed that the term is still irrelevant even going to a mass close to the physical one.

Acknowledgments

References

[1] (RM123), G. M. de Divitiis, R. Frezzotti, V. Lubicz, G. Martinelli, R. Petronzio, G. C. Rossi et al., *Leading isospin breaking effects on the lattice*, *Phys. Rev. D* **87** (2013) 114505, [1303.4896].

[2] A. Girard, *A fast 'monte-carlo cross-validation' procedure for large least squares problems with noisy data*, *Numerische Mathematik* **56** (Jan, 1989) 1–23.

[3] M. Hutchinson, *A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines*, *Communications in Statistics - Simulation and Computation* **19** (1990) 433–450, [https://doi.org/10.1080/03610919008812866].

[4] R. A. Meyer, C. Musco, C. Musco and D. P. Woodruff, *Hutch++: Optimal Stochastic Trace Estimation*, pp. 142–155. https://epubs.siam.org/doi/pdf/10.1137/1.9781611976496.16. 10.1137/1.9781611976496.16.

[5] E. N. Epperly, J. A. Tropp and R. J. Webber, *Xtrace: Making the most of every sample in stochastic trace estimation*, 2301.07825.

[6] A. Frommer and M. N. Khalil, *MGMLMC++ as a Variance Reduction Method for Estimating the Trace of a Matrix Inverse*, *PoS* **LATTICE2022** (2023) 017, [2303.11512].

[7] (RCstar), L. Bushnaq, I. Campos, M. Catillo, A. Cotellucci, M. Dale, P. Fritzsch et al., *First results on QCD+QED with C* boundary conditions*, *JHEP* **03** (2023) 012, [2209.13183].