

## Error reduction using the covariant approximation averaging

---

**Eigo Shintani\***

*PRISMA Cluster of Excellence, Institut für Kernphysik and Helmholtz Institute Mainz, Johannes Gutenberg-Universität Mainz, D-55099 Mainz, Germany.*

*RIKEN Advanced Institute for Computational Science, Kobe, Hyogo 650-0047, Japan.*

*E-mail: shintani@riken.jp*

In order to enhance the statistics at  $O(10,000)$  measurement, we test the all-mode-averaging technique using the approximation of observable with the optimized size of local deflation field and block size of Schwartz alternative procedure to reduce the computational cost on  $N_f = 2$  Wilson-clover fermion configurations in Mainz-CLS group.

*The European Physical Society Conference on High Energy Physics  
22-29 July 2015  
Vienna, Austria*

---

\*Speaker.

## 1. Introduction

Recent development of the numerical algorithm, software and hardware makes a tremendous progress to be exactly simulating the dynamics of nucleon and nuclei including the dynamical light quarks close to physical pion mass in lattice QCD. On the other hand, aiming for the precise calculation of nucleon structure function in lattice QCD, we confront a dilemma, which is a sort of noise issue pursuing the high precision of Monte-Carlo simulation without the contamination of non-ground state (excited state) effect. According to a naive expectation of signal-to-noise ratio of nucleon propagator given its variance as described by induced pion propagator,

$$S/N \sim \sqrt{N} \exp \left[ - (m_N - 3/2m_\pi)t \right], \quad (1.1)$$

in statistics  $N$  at temporal extension  $t$ , its signal becomes exponentially small rather than statistical noise, even approaching to the light quark, while  $S/N$  is scaled as a function of square of independent statistics. Although there have been several attempts of precise computation of nucleon form factor from several lattice collaborations, those results have not achieved a consensus with each group (see more details, for instance [1] and references therein). Especially for nucleon axial charge, which is even fundamental quantity and precisely measured by experiment, the lattice QCD has not provided an established value, and reported about 10% discrepancy in physical point.

In this proceedings, we first apply the all-mode-averaging (AMA) technique [2, 3, 4] for axial charge computation to make accurate study on  $N_f = 2$  Wilson-clover fermion configurations generated by Mainz-CLS group [5]. As we present in this paper, it also provides quantitative information towards future study of nucleon form factor in the physical point.

## 2. Numerical method

### 2.1 All-mode-averaging

As described in [2, 3, 4], AMA estimator is defined in

$$\mathcal{O}^{\text{AMA}} = \mathcal{O}^{(\text{rest})} + \frac{1}{N_G} \sum_{g \in G}^{N_G} \mathcal{O}^{(\text{appx})g}, \quad \mathcal{O}^{(\text{rest})} = \frac{1}{N_{\text{org}}} \sum_{f \in G}^{N_{\text{org}}} \left[ \mathcal{O}^f - \mathcal{O}^{(\text{appx})f} \right], \quad (2.1)$$

where  $\mathcal{O}^{(\text{appx})}$  is an approximation of observable  $\mathcal{O}$  constructed by the sloppy inversion algorithm of Dirac operator (truncated solver) being around  $10^{-3}$  precision. Using covariant transformation  $g$  in symmetry  $G$  on the lattice, for instance the translational symmetry,  $\mathcal{O}^{(\text{rest})}$  is regarded as the bias correction for low precision in  $\mathcal{O}^{(\text{appx})}$  and because of its covariance, the expectation value of  $\mathcal{O}^{\text{AMA}}$  is consistent with  $\mathcal{O}$  itself. AMA advances that, if  $\mathcal{O}^{(\text{appx})}$  is appropriate observable having strong correlation with original  $\mathcal{O}$ , the statistical quality of  $\mathcal{O}^{\text{AMA}}$  in  $N$  statistics is similar to  $N_G$  times statistics of  $\mathcal{O}$ , even though the computational cost of  $\mathcal{O}^{(\text{appx})}$  is much less than  $\mathcal{O}$ . We notice that in Eq.(2.1) we also include the average over covariant transformation  $f$  with the number of  $N_{\text{org}}$ , for instance, data-set of original observable  $\mathcal{O}$  transformed by  $f$  which has been produced in usual source-shift method. AMA is easily able to reuse the previous data-set in order to enhance the statistical accuracy without any modification or recalculation of  $\mathcal{O}$ .

The formula of the ratio of standard deviation between  $\mathcal{O}$  and  $\mathcal{O}^{\text{AMA}}$  is given as

$$\sigma^{\text{ama}}/\sigma \simeq \sqrt{N_G^{-1} + 2\Delta r + R}, \quad (2.2)$$

$$\Delta r = \left( \frac{1}{N_{\text{org}}} - \frac{1}{N_G} \right) \left( 1 - \frac{1}{N_{\text{org}}} \sum_{f \in G} \frac{\langle \Delta \mathcal{O}^f \Delta \mathcal{O}^{(\text{appx})f} \rangle}{\sigma^f \sigma^{(\text{appx})f}} \right), \quad (2.3)$$

$$R = \frac{1}{N_G^2} \sum_{g, g' \in G} \frac{\langle \Delta \mathcal{O}^{(\text{appx})g} \Delta \mathcal{O}^{(\text{appx})g'} \rangle}{\sigma^{(\text{appx})g} \sigma^{(\text{appx})g'}}, \quad (2.4)$$

where we use  $\Delta \mathcal{O} = \mathcal{O} - \langle \mathcal{O} \rangle$ , and standard deviation  $\sigma = \sqrt{\langle (\Delta \mathcal{O})^2 \rangle}$ . The quantity in Eq. (2.3) represents the magnitude of correlation between  $\mathcal{O}$  and  $\mathcal{O}^{(\text{appx})}$ , *i.e.*  $\mathcal{O}^{(\text{appx})}$  having  $\Delta r \simeq 0$  is appropriate to reduce the error of  $\mathcal{O}^{\text{AMA}}$ . The above equation also represents the dependence of  $N_{\text{org}}$  in which the average of transformed  $\mathcal{O}$  over  $N_{\text{org}}$  much less than  $N_G$  makes the improvement of  $\Delta r$  to being small. The third equation is a summation of correlation between  $\mathcal{O}^{(\text{appx})g}$  with each transformation in  $N_G$ , *i.e.* correlation in correlator with different source location. Here we ignore the correlation between  $\mathcal{O}^f$  since those data is enough separated from each other rather than  $\mathcal{O}^{(\text{appx})}$  (from a point of view of computational cost, AMA should be efficient when computing much large  $N_G$  time approximation than  $N_{\text{org}}$  time original one). Equation (2.2) suggests that, in order to reduce the statistical error of AMA estimator to being close to maximum reduction of  $1/\sqrt{N_G}$ ,  $\mathcal{O}^{(\text{appx})}$  having the small  $\Delta r$  and less correlation between different transformation  $g$  is needed. This relies on a choice of  $g$  and parameter of algorithm in the inversion of Dirac operator with truncated iteration number of conjugate residual (CR) or conjugate gradient (CG) method. In the next subsection, we show our strategy for the implementation of AMA in Wilson-Clover fermion action.

## 2.2 Deflated SAP+GCR in AMA

For Wilson-Clover fermion, it is convenient to adopt the combination of local deflation field with approximation in Schwartz alternative procedure (SAP) [6, 7]. SAP is able to construct the approximation of inverse of Wilson-Dirac operator using domain-decomposition into  $2 \times 2$  matrix of which operator is divided by two local domain  $\Lambda$ ,  $\Lambda^*$  with local boundary fields  $\partial\Lambda$ ,  $\partial\Lambda^*$ . As a consequence of SAP applying into domain-decomposed Dirac operator, the inversion of Wilson-Dirac operator is represented as the polynomial function of Dirac operator in each domain,

$$D^{-1} \simeq M_{\text{sap}} = K \sum_{v=0}^{n_{\text{cy}}-1} (1 - DK)^v, \quad (2.5)$$

$$K = R_{\Lambda}^T D_{\Lambda}^{-1} R_{\Lambda} + R_{\Lambda^*}^T D_{\Lambda^*}^{-1} R_{\Lambda^*} - R_{\Lambda^*}^T D_{\Lambda^*}^{-1} D_{\partial\Lambda^*} D_{\Lambda}^{-1} R_{\Lambda}, \quad (2.6)$$

$$R_{\Lambda}^T = (1, 0), \quad R_{\Lambda^*}^T = (0, 1), \quad (2.7)$$

where  $n_{\text{cy}}$  is the number of cycle of SAP relying on a quality of approximation to inversion. Here  $M_{\text{SAP}}$  is also used in not only preconditioner of generalized conjugate residual (GCR) algorithm but also kernel in smoothing procedure for deflation field. In our set-up, we employ  $n_{\text{cy}} = 5$ .

Deflation field is also efficient to make speed-up the iterative solver. Using the projection with deflation field (see, for example reference [7]), the condition number is improved and then total

iteration number is significantly reduced. Here incorporating SAP into subspace of deflation field [7], the projected Dirac operator, which is called as little Dirac operator, is defined as

$$P = \sum_{\Lambda, \Lambda'}^{N_b} \sum_{k, l}^{N_s} R_{\Lambda}^T \phi_k^{\Lambda} (\phi_l^{\Lambda'})^{\dagger} R_{\Lambda'} (A_{kl}^{\Lambda \Lambda'})^{-1}, \quad A_{kl}^{\Lambda \Lambda'} = (\phi_k^{\Lambda}, R_{\Lambda}^T D R_{\Lambda'} \phi_l^{\Lambda'}), \quad (2.8)$$

and then in this method there are totally  $N_d = N_s N_b$  deflation fields. In order to make deflation method work well, the so-called smoothing process is able to enhance low-mode contribution in deflation subspace. The preconditioner  $M_{\text{SAP}}$  with  $n_{\text{cy}}$  is also useful for this process. The important point is that the performance of preconditioning with deflation field trades off the solving cost of little Dirac operator, *i.e.* large space of deflation field (large number of deflation field and its domain) is required of further cost to compute  $(A_{kl}^{\Lambda \Lambda'})^{-1}$  because of increasing the condition number of little Dirac operator.

In AMA, the number of deflation field and domain size, corresponding to  $N_s$  and  $N_b$  respectively, is regarded as tuning parameter to control the quality of approximation, which is related to the magnitude of  $\Delta r$  in Eq.(2.3), in addition to the number of iteration  $N_{\text{iter}}$  in the inversion algorithm. We notice that the translational symmetry of the approximation  $\mathcal{O}^{(\text{appx})}$  is not trivial as well as even-odd preconditioning [3], since the shift symmetry in inside domain may be broken by inaccuracy of  $\mathcal{O}^{(\text{appx})}$ , besides which is recovered after achieved the enough precision of  $\mathcal{O}$ . To preserve the translational invariance of  $\mathcal{O}^{(\text{appx})g}$  under transformation  $g \in G$ , we use the shift of the same domain  $\Lambda$  (or  $\Lambda^*$ ) and same site with original in local domain. In this case, there is no violation of translational invariance for  $M_{\text{SAP}}$  between the same local site in each  $\Lambda$  (or  $\Lambda^*$ ), and therefore there is limitation of available region of source location relying on the size of domain (this is only case of using covariant symmetry in AMA, and in random source location method [4] it also does not matter). In Section 3.1 we present an example of a choice of source location avoiding such bias.

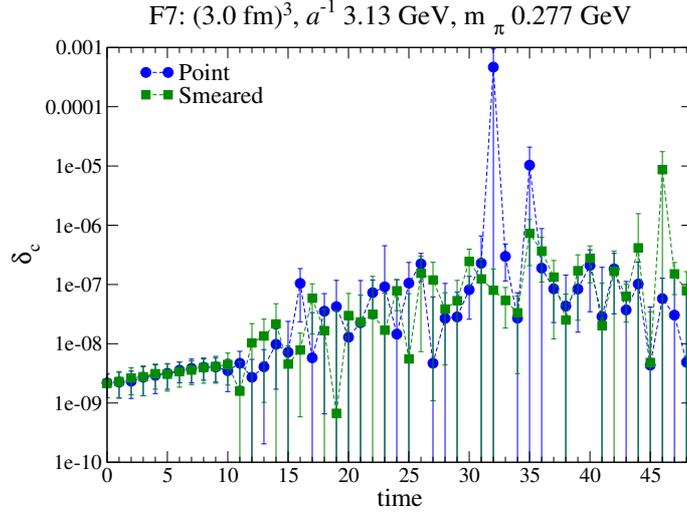
### 3. Performance test of AMA

#### 3.1 Covariance test and tuning parameters

Before going to discussion on the tuning parameter of AMA, we first check the preservation of covariant symmetry of approximation using domain-decomposition for Dirac operator in SAP preconditioning. As tested in [4], the magnitude of the difference between the approximation shifted source and gauge shift, defined as

$$\delta_c = \mathcal{O}^{(\text{appx})g}[U^{\bar{g}}] - \mathcal{O}[U], \quad (3.1)$$

presents the violation of covariance. In the numerical test, we use the domain-size of  $6^4$  and deflation field  $N_s = 30$  on F7 ensemble. We use the transformation  $g$  of the shift of source location  $g = (6, 6, 0, 0)$  and gauge field  $\bar{g} = (-6, -6, 0, 0)$ , where those shifts correspond to even shift of domain-location keeping in the equivalent domain  $\Lambda$  or  $\Lambda^*$  to preserve the covariant symmetry in domain-decomposition as mentioned in Section 2.2. For approximation, we set the threshold to be fixed in 30 iterations of GCR, corresponding to 0.01–0.02 of magnitude of normalized norm of residual vector. In Figure 1 we show  $|\delta_c|$  for nucleon propagator at each time-separation from sink and source point. One sees that  $\delta_c$  increases to the order of  $10^{-3}$ – $10^{-5}$  at large time-separation



**Figure 1:** Error of covariance for the nucleon propagator as a function of time-separation at one configuration in  $96 \times 48^3$  lattice with  $m_\pi = 0.277$  GeV. Different colored lines denote the result in smeared source-sink (blue) and point source-sink (green).

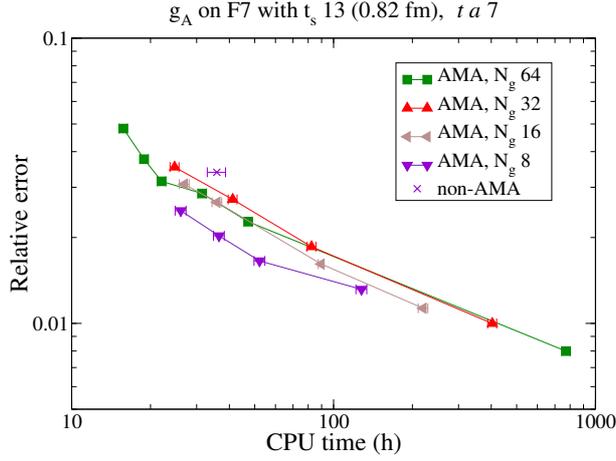
more than 30 in lattice unit due to accumulation of round-off error (although in this region it is statistically noisy), however below 25 time-slice in lattice unit, which is appropriate signal region,  $\delta_c$  is less than  $10^{-6}$ . From the practical point of view, the error of covariant symmetry violation is negligible in our parameter setting.

### 3.2 Error scaling and computational cost

Finally we test the error scaling as a function of total consuming time of CPU and the number of measurement. In this test, we use the cluster machine called as “Wilson” and “Clover” in Institut für Kernphysik and Helmholtz Institute Mainz in Mainz University. Figure 2 plots the relative error of the ratio of three-point and two-point function at time-slice  $t = t_s/2$  with several  $N_G$  on different number of configurations. From  $N_G = 8$  to 64 on F7 ensemble, since the correlators in each  $g$  increase the magnitude of  $R$ , which is correlation between propagator with different  $\mathcal{O}^{(\text{appx})g}$  due to small distance in finite space-time, the computational cost does not ideally decrease. In our test case,  $N_G = 8$  case has about 50% smaller cost than other  $N_G$ , and from  $N_G = 16$  the error scaling is similar to others. Compared to the total cost of conventional method at a point of similar relative error, we have roughly 1.5–3 times gain by use of AMA.

## 4. Summary and discussion

In this proceedings, we apply AMA to Wilson-clover fermion configurations using tuned parameter of SAP+GCR with deflation method, and confirm that our parameter of AMA is able to reduce the computational cost to more than two times smaller than traditional source averaging technique. This is feasible study of determination of physical value of nucleon form factor from lattice QCD, and so that we will perform further study of the chiral behavior of nucleon form fac-



**Figure 2:** The relative error of axial charge at time-slice  $t/a = 7$  in the lattice unit versus CPU time, on which we measure in the cluster machine called as “Wilson” in Mainz University. Different symbols denote the error scaling with different number of  $N_G$  using the several number of configurations. For non-AMA, this is exact result of correlator on the maximum number of configurations with one source

tor including not only  $g_A$  but also iso-vector channel with various lattice parameters [8] in high statistics under way.

## References

- [1] S. Syritsyn, PoS LATTICE **2013**, 009 (2014) [arXiv:1403.4686 [hep-lat]].
- [2] T. Blum, T. Izubuchi and E. Shintani, Phys. Rev. D **88**, no. 9, 094503 (2013) [arXiv:1208.4349 [hep-lat]].
- [3] T. Blum, T. Izubuchi and E. Shintani, PoS LATTICE **2012**, 262 (2012) [arXiv:1212.5542].
- [4] E. Shintani, R. Arthur, T. Blum, T. Izubuchi, C. Jung and C. Lehner, Phys. Rev. D **91**, no. 11, 114511 (2015) [arXiv:1402.0244 [hep-lat]].
- [5] S. Capitani, M. Della Morte, G. von Hippel, B. Jager, A. Juttner, B. Knippschild, H. B. Meyer and H. Wittig, Phys. Rev. D **86**, 074502 (2012) [arXiv:1205.0180 [hep-lat]].
- [6] M. Luscher, Comput. Phys. Commun. **156**, 209 (2004) [hep-lat/0310048].
- [7] M. Luscher, JHEP **0707**, 081 (2007) [arXiv:0706.2298 [hep-lat]].
- [8] S. Capitani *et al.*, Phys. Rev. D **92**, no. 5, 054511 (2015) [arXiv:1504.04628 [hep-lat]].