PROCEEDINGS OF SCIENCE

Variance reduction via deflation with local coherence

Roman Gruber,[a,*] **Tim Harris**[a] **and Marina Krstić Marinković**[a]

[a] *Institut für Theoretische Physik, ETH Zürich,*
Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland

E-mail: rgruber@ethz.ch, harrist@phys.ethz.ch, marinama@ethz.ch

In large enough volumes, translation-averaging for quark-line connected diagrams reduces the variance inversely proportional to the volume. Stochastic estimators which implement translation averaging however introduce new sources of fluctuations, which in some cases can be relatively large. In this work, we explore whether inexact deflation subspaces can be used to improve the precision of the isovector vector correlators. We perform numerical experiments with $N_{\mathrm{f}} = 2$ non-perturbatively O(a)-improved Wilson fermions and measure the relative contribution from the deflation subspace to the central value and the corresponding variance.

The 40th International Symposium on Lattice Field Theory (Lattice 2023)
July 31st - August 4th, 2023
Fermi National Accelerator Laboratory

*Speaker

PoS(LATTICE2023)153

1. Introduction

When translation averaging is used, the variance of observables with a small footprint is expected to decrease inversely proportional to the volume in a Monte Carlo simulation due to the mass gap of QCD [1]. While this idea certainly works out in the master-field regime [2], it has commonly been put into practice for moderate volumes, too. For quark-line connected correlation functions, such as the isovector vector correlator

$$\langle G(x,y_0)\rangle_{\mathrm{g}} = \langle \mathrm{tr}\{W(x,y_0,x)\}\rangle_{\mathrm{g}}, \qquad W(x,y_0,z) = -\tfrac{1}{3}\sum_k a^3 \sum_{\boldsymbol{y}} \{S(x,y)\gamma_k S(y,z)\gamma_k\}, \tag{1}$$

where $S(x,y)$ denotes the light-quark propagator and $\langle\cdot\rangle_{\mathrm{g}}$ denotes the average with respect to the effective action after integrating out the fermion fields, the variance shows a suppression proportional to the volume. This holds true even when integrating both vertices over the spatial volume L^3 in a box of linear size L. This can be understood from the expression for the variance for its translation-averaged counterpart $G_{\mathrm{vol}}(x_0,y_0) = \frac{a^3}{L^3}\sum_{\boldsymbol{x}} G(x,y_0)$ which is

$$\sigma^2_{\mathrm{vol}}(x_0-y_0) = \frac{a^3}{L^3}\sum_{x'} \delta_{x_0 x_0'}\Big[\langle G(x,y_0)G(x',y_0)\rangle_{\mathrm{g}} - \langle G(x,y_0)\rangle_{\mathrm{g}}\langle G(x',y_0)\rangle_{\mathrm{g}}\Big] \tag{2}$$

and noting the sum converges in infinite volume thanks to the extra suppression due to the quark propagators. Note also that the variance has no power-law divergences for $x_0 \neq y_0$.

Implementing translation averaging is not completely straightforward without incurring a cost which also increases with the volume, for example if propagators would be explicitly computed from every lattice site. One possibility is to use a stochastic estimator for the translation average, such as the one-end trick [3, 4]

$$\mathcal{G}_{\mathrm{vol}}(x_0,y_0) = \frac{1}{N_{\mathrm{src}}}\sum_{i=1}^{N_{\mathrm{src}}}\sum_{\alpha=1}^{4}\frac{a^3}{L^3}\sum_{\boldsymbol{x},z}\eta_i^{(\alpha)\dagger}(x)W(x,y_0,z)\eta_i^{(\alpha)}(z)\delta_{x_0 z_0} \tag{3}$$

where the dimensionless fields $\eta^{(\alpha)}_{i\beta c}(x) = \delta^{\alpha}_{\beta}\xi_{ic}(x)$ are defined in terms of i.i.d. random fields with zero mean and variance

$$\langle \xi_{ic}(x)\xi^*_{jd}(y)\rangle_{\xi} = \delta_{ij}\delta_{cd}\delta_{xy}, \tag{4}$$

and we have chosen the spin-diagonal variant [5] which allows the unbiased estimation of each spin matrix element of W with just $4N_{\mathrm{src}}$ inversions.

The introduction of new auxiliary fields leads to fluctuations, which, in the case of Gaussian-distributed fields, introduces an extra contribution to the variance [6]

$$\sigma^2_{\mathrm{stoch}} = \sigma^2_{\mathcal{G}} - \sigma^2_{\mathrm{vol}} = \frac{1}{N_{\mathrm{src}}}\sum_{\alpha,\beta=1}^{4}\sum_{a,b=1}^{3}\sum_{\boldsymbol{x},z}\delta_{x_0 z_0}\langle W_{\alpha a,\alpha b}(x,y_0,z)W_{b\beta,a\beta}(z,y_0,x)\rangle_{\mathrm{g}}. \tag{5}$$

Although its volume-scaling is the same as the exact average in Eq. (2), it has different matrix elements which can be relatively large compared to σ^2_{vol}, potentially spoiling the variance-reduction from translation averaging. This is illustrated in Fig. 4 where we depict estimators for the variances σ^2_{vol} (red) and $\sigma^2_{\mathrm{stoch}} \approx \sigma^2_{\mathcal{G}}$ (blue) for the isovector vector correlator showing the large hierarchy between them even when N_{src} is large. In the following we investigate improved estimators based on deflation techniques.

2. Variance reduction via deflation

Natural decompositions of traces such as Eq. (1) can be constructed starting from a decomposition of the quark propagator defined in terms of a set of N linear independent fields $\phi_1(x), \ldots, \phi_N(x)$

$$S(x,y) = \sum_{k,l=1}^{N} \phi_k(x)(A^{-1})_{kl}\phi_l^\dagger(y) + P_{\mathrm{R}}S(x,y), \tag{6}$$

where the first term represents the propagator within the subspace of fields, and the matrix elements of the so-called little Dirac operator A are defined explicitly in terms of the Dirac operator D

$$A_{kl} = (\phi_k, D\phi_l). \tag{7}$$

Such operators arise when considering the Schur complement of the propagator in the subspace. The remainder may be written in terms of the projected propagator, namely

$$P_{\mathrm{R}}S(x,y) = S(x,y) - \sum_{k,l=1}^{N} \phi_k(x)(A^{-1})_{kl}\phi_l^\dagger(y). \tag{8}$$

Inserting the representation Eq. (6) into Eq. (1), the trace may be decomposed into terms involving only the little propagator A^{-1} and a remainder. If the fields are chosen to be exact (left) singular vectors of the Dirac operator, the former contribution can be computed exactly with translation averaging. When the remainder is computed stochastically, the extra contribution to the variance has the same form as Eq. (5) with at least two insertions of P_{R}. The efficiency of the deflation for our purposes is thus measured by the reduction in the variance. In practice, we note that the dimension of the little propagator is too large to compute and store directly, and thus its contributions may also need to be estimated stochastically. Nevertheless, here we will assume it is cheap enough that the extra contribution to the variance can be suppressed by a sufficiently large number of samples N_{src}.

Such low-mode averaging and its variants are common variance-reduction strategies in lattice QCD and have been known for a long time [7, 8]. However, given that the spectral density increases with the volume, the cost of achieving a constant efficiency with the volume scales with its square. Volume scaling e.g. achieved in inexact deflation [9] and multigrid solvers [10] profits from the local coherence of the low modes of the Dirac operator. That is, the low modes projected to local domains can be represented by a small basis of fields to a high degree of precision. In this way, only a small number of exact modes are required to span a much larger deflation subspace and the volume-scaling problem is circumvented, as outlined in the following section.

2.1 Domain-decomposed subspaces

A domain-decomposed subspace of dimension N may be constructed from a much smaller set of N_{s} input fields $\psi_1(x), \ldots, \psi_{N_{\mathrm{s}}}(x)$ in the following way. By decomposing the lattice into a set of N_{b} non-overlapping blocks labelled by Λ, the input fields are projected to the blocks

$$\psi_k^\Lambda(x) = \begin{cases} \psi_k(x) & \text{if } x \in \Lambda, \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Name	$T \times L^3$	Pion mass	a [fm]	# configs
D5d	48×24^3	439 MeV	0.0658(10)	100
F7	96×48^3	268 MeV	0.0658(10)	100

Table 1: Ensembles of $N_\mathrm{f} = 2$ non-perturbatively O(a)-improved Wilson fermions used in this work.

and re-orthogonalized afterwards. After relabelling the $N = N_\mathrm{b} N_\mathrm{s}$ fields one implements the deflation exactly as described in the previous subsection. In contrast to the input fields used in the preconditioning of the Dirac equation [9] which may be fairly rough approximations of the low modes, initialized for example through a few inverse iterations, we have in mind to take the input fields as exact (left) singular vectors of the lattice Dirac operator. Such domain decompositions have been utilized as a compression algorithm [11] as well as variance reduction similar to what we investigate here [12].

3. Numerical experiments

In this section, we present some numerical investigations of domain-decomposed deflation subspaces using $N_\mathrm{f} = 2$ non-perturbatively O(a)-improved Wilson fermions generated by the Coordinated Lattice Simulations (CLS) effort [13]. We study two ensembles of representative gauge field configurations separated widely in Monte Carlo time and whose parameters are listed in Tab 1. A small number of input fields $N_\mathrm{s} \sim 10 - 100$ were created on each configuration using the PRIMME library [14] to a relative precision[1] of 10^{-12}. The stochastic sources η were spin-diagonal with $\mathbb{Z}_2 \times \mathbb{Z}_2$ noise. We report results only for the case of the isovector vector current correlator as defined in Eq. (1).

3.1 Efficiency of domain-decomposed subspaces

In order to illustrate one sense in which the enlarged space of fields represents the true low-mode subspace well, one can examine the deficits ϵ_i of the exact low modes χ_i which are not contained in the domain-decomposed subspace

$$\epsilon_i = \left\| \chi_i - \sum_{k=1}^{N} \phi_k (\phi_k, \chi_i) \right\| / \|\chi_i\| \tag{10}$$

In Fig. 1, we display the deficits of the first few hundred exact low modes on the F7 ensemble for various subspaces created from two choices of number of input fields $N_\mathrm{s} = 20, 100$ and two configurations of the domain sizes namely 4^4 and 48×8^3 in lattice units, which correspond to domains in the spatial sizes of around 0.25 and 0.5 fm respectively. The subspace size is thus increased by a factor 41472 or 432 in each case by the domain decomposition. As expected, the deficits are reduced by decreasing the domain size and by increasing the number of input modes. However, as explained previously, the precise measure of the efficiency for our purpose is whether the stochastic variance of the remainder terms is reduced compared to the undeflated version, which we investigate in the following subsection.

[1] $\|D\gamma 5 \chi_i - \lambda_i \chi_i\| / \|\chi_i\| < 10^{-12}$.

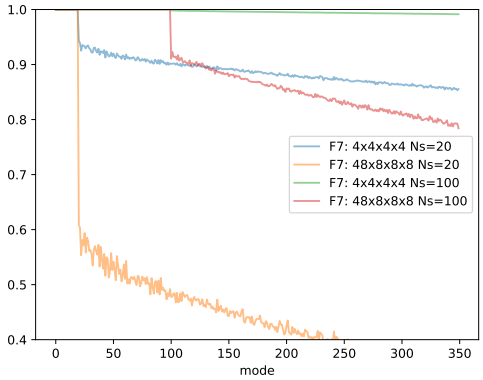

Figure 1: Local coherence of low modes. The deviation of the deficits Eq. (10) from unity $1-\epsilon_i$ versus the mode number. The first N_s modes in each case are exactly contained in the enlarged subspace by construction and have zero deficit. The deficits are reduced by decreasing the domain size or by increasing N_s the number of input modes.

3.2 Efficiency of variance reduction

In this section we test the efficiency of the deflation in reducing the variance on the individual contributions to the trace in Eq. (1) for various domain decompositions. Using the deflation subspace as defined above, we make the decomposition of the stochastic estimator of Eq. (3)

$$\mathcal{G}_{\text{vol}} = \mathcal{G}^{\text{lttl}}_{\text{vol}} + \mathcal{G}^{\text{rem}}_{\text{vol}} \tag{11}$$

where the two estimators are defined as follows. The first contains only the little propagators

$$\mathcal{G}^{\text{lttl}}_{\text{vol}}(x_0,y_0) = \frac{1}{N_{\text{src}}}\sum_{i=1}^{N_{\text{src}}}\sum_{\alpha=1}^{4}\frac{a^6}{L^3}\sum_{\boldsymbol{x},\boldsymbol{y},z}\delta_{x_0z_0}\Big[\sum_{klmn=1}^{N}\eta_i^{(\alpha)\dagger}(x)\phi_k(x)(A^{-1})_{kl}\phi_l^\dagger(y)\gamma_k\phi_m(y)(A^{-1})_{mn}\phi_n^\dagger(z)\gamma_k\eta_i^{(\alpha)}(z)\Big] \tag{12}$$

while the remainder contains a mixed term and a term arising only from the projected propagator

$$\mathcal{G}^{\text{rem}}_{\text{vol}}(x_0,y_0) = \frac{1}{N_{\text{src}}}\sum_{i=1}^{N_{\text{src}}}\sum_{\alpha=1}^{4}\frac{a^6}{L^3}\sum_{\boldsymbol{x},\boldsymbol{y},z}\delta_{x_0z_0}\Big[2\sum_{kl=1}^{N}\eta_i^{(\alpha)\dagger}(x)\phi_k(x)(A^{-1})_{kl}\phi_l^\dagger(y)\gamma_k P_{\text{R}}S(y,z)\gamma_k\eta_i^{(\alpha)}(z) + \eta_i^{(\alpha)\dagger}(x)P_{\text{R}}S(x,y)\gamma_k P_{\text{R}}S(y,z)\gamma_k\eta_i^{(\alpha)}(z)\Big]. \tag{13}$$

As explained earlier, we assume that in practice the little propagator is cheap enough so that N_{src} can be chosen very large for that term. We note, however, that the one-end trick estimator is recovered if N_{src} is chosen to be the same for both terms with the same noise fields. It is therefore interesting to examine each term in turn with a fixed number of N_{src}, keeping in mind that the variance can be effectively eliminated on the little term at no cost.

In Figs. 2 and 3 we show the relative contribution of the two terms to both the central value and the variance as a function of $x_0 - y_0$. In the left panels, the relative contribution to the central

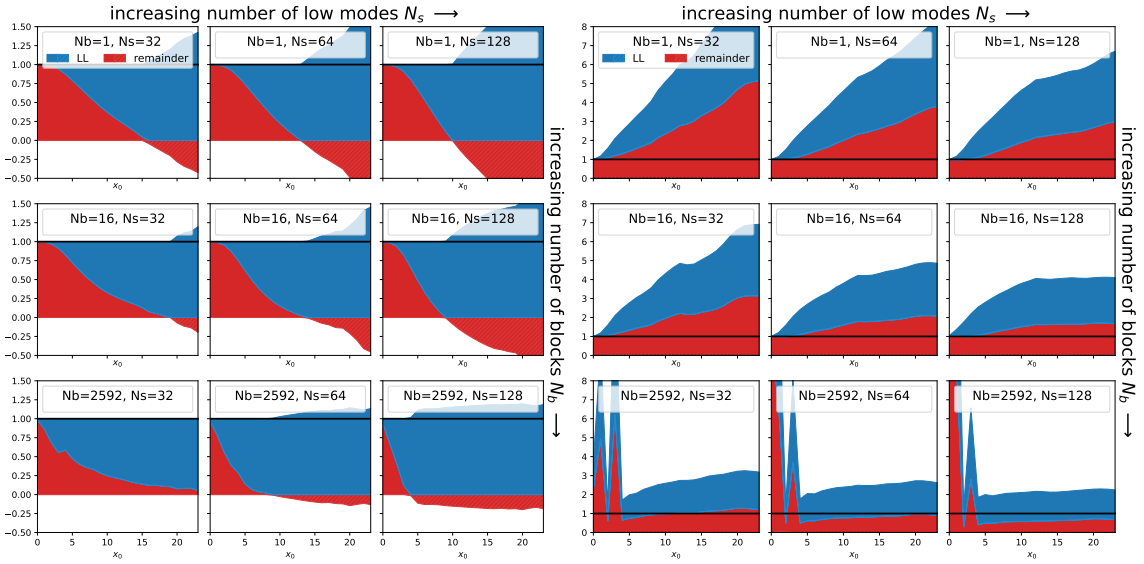

Figure 2: Left: Relative contributions to the value of the correlator Eq. (1), on D5d lattice using different number of low modes N_s and different number of blocks N_b increasing left to right and top to bottom respectively. Right: The variance of the little contribution (blue) and remainder (red) normalized by the variance of the sum. The block sizes ($T \times L^3$) are; 1st row: no blocking (corresponding to traditional LMA), 2nd row: $24 \times 12 \times 12 \times 12$, 3rd row: $4 \times 4 \times 4 \times 4$.

value of the little term (blue) and the remainder term (red) is shown normalized by the sum. Each panel represents a different choice of parameters, with the number of blocks N_b increasing from left to right, and the number of input fields N_s increasing from top to bottom. Note that the remainder term, representing the contribution from the high modes dominates at small separations but even becomes negative at large separations.

In the right panel, we show the variances of the little term (blue) and the remainder (red), normalized by the variance of the sum. The striking feature is the large size of the variance from the remainder term even at the largest separations. Except for the largest deflation subspace sizes (bottom right), this illustrates that the variance on the one-end trick estimator is smaller than the individual contributions, which demonstrates the high degree of covariance between the two pieces. That is, the stochastic variance on the remainder is not guaranteed to be smaller than the original estimator. Even if the variance due to the little propagator is eliminated by increasing the number of samples, the variance on the remainder (red) is larger than the original estimator (black line) except for the smallest domain size of 4^4 for the D5d ensemble (Fig. 2 right, bottom row).

4. Conclusions

In this work we presented a preliminary investigation into variance reduction using deflation with domain-decomposed subspaces for the isovector vector correlator with $N_f = 2$ non-perturbatively O(a)-improved Wilson fermions. Such a domain decomposition exploits the local coherence of the low quark modes of the Dirac operator that almost eliminates the volume-squared problem of generating such subspaces. Variance reduction methods are required for the vector

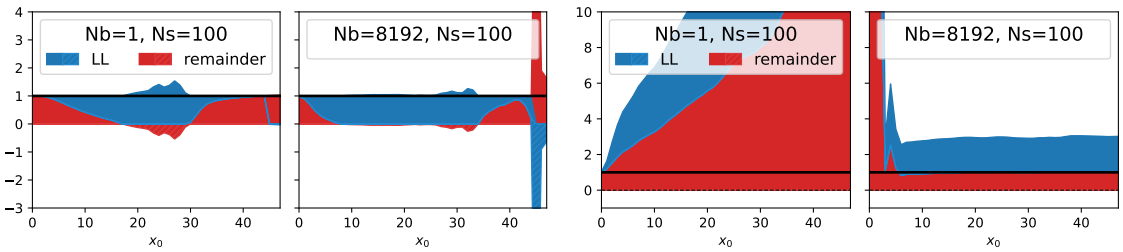

Figure 3: Same as Fig. 2 for the F7 ensemble. The block sizes ($T \times L^3$) are; left: no blocking (LMA) right: $6 \times 6 \times 6 \times 6$.

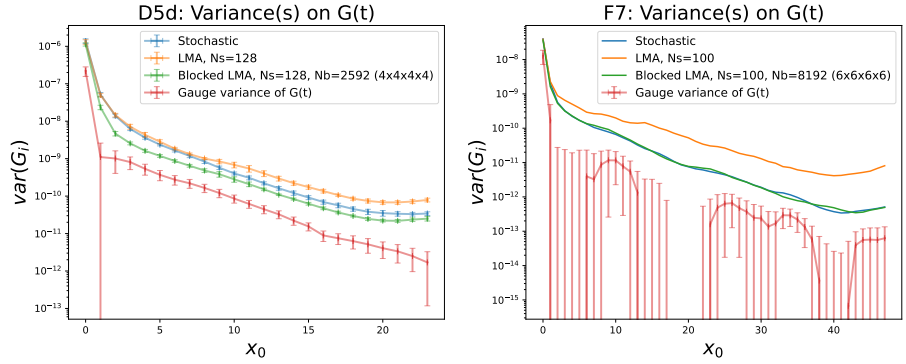

Figure 4: Comparison of variances of the one-end trick (blue), deflated (yellow) and deflation with domain-decomposed subspace estimators (green). An estimator for the variance of the exact translation average Eq. (2) is shown in red. D5d (left) and F7 (right) with $N_s = 128, 100$ low modes and $N_{\text{src}} = 32, 96$ stochastic spin-diagonal sources, respectively.

channel due to the large contribution to the variance from the noise fields, as can be seen in Fig. 4 where we computed an estimator for the variance of the translation-averaged observable Eq. (2).

For the few subspace configurations that we investigated, we observed that in many cases deflating the stochastic estimator may actually increase the variance. In Fig. 4 (left) for the ensemble with larger quark-mass, we see a moderate reduction in the variance with the best parameter choice (green) over the one-end trick (blue). For the smaller quark mass ensemble (right), the variance is at best identical to the case of the one-end trick. We note that in traditional low-mode averaging (yellow), the variance is actually increased with respect to the one-end trick estimator (blue) with these parameter choices and for this observable. Therefore, in order to be safe, a careful analysis of the variances can be is important to gauge the efficiency of deflation. Clearly, there are many possible choices to further improve the estimation of the remainder term (particularly the cross term) which was computed in a simple way in this work, which is ongoing, as well as work towards a more efficient implementation. A more thorough theoretical analysis of low-mode deflation on the variances would be beneficial to better motivate the definition of the subspace.

Acknowledgments

We are grateful to R. Brower, J. Coles, M. Lüscher, and the RC* collaboration members for inspiring discussions the interesting discussions, and the feedback on the initial version of this work.

We acknowledge the access to Piz Daint at the Swiss National Supercomputing Centre, Switzerland under the ETHZ's share with the project IDs go24, eth8 and c21. The supported by the PASC project "Efficient QCD+QED Simulations with openQ*D software" is gratefully acknowledged.

References

[1] M. Lüscher. "Stochastic locality and master-field simulations of very large lattices". In: *EPJ Web Conf.* 175 (2018). Ed. by M. Della Morte et al., p. 01002. arXiv: 1707.09758 [hep-lat].

[2] L. Giusti and M. Lüscher. "Topological susceptibility at $T > T_c$ from master-field simulations of the SU(3) gauge theory". In: *Eur. Phys. J. C* 79.3 (2019), p. 207. arXiv: 1812.02062 [hep-lat].

[3] C. Michael and J. Peisa. "Exotic hadronic states and all to all quark propagators". In: *Nucl. Phys. B Proc. Suppl.* 60 (1998). Ed. by Y. Iwasaki and A. Ukawa, pp. 55–60. arXiv: hep-lat/9705013.

[4] G. M. de Divitiis et al. "Pseudofermion observables for static heavy meson decay constants on the lattice". In: *Phys. Lett. B* 382 (1996), pp. 393–397. arXiv: hep-lat/9603020.

[5] P. Boucaud et al. "Dynamical Twisted Mass Fermions with Light Quarks: Simulation and Analysis Details". In: *Comput. Phys. Commun.* 179 (2008), pp. 695–715. arXiv: 0803.0224 [hep-lat].

[6] M. Lüscher. "Computational Strategies in Lattice QCD". In: *Les Houches Summer School: Session 93: Modern perspectives in lattice QCD: Quantum field theory and high performance computing*. Feb. 2010, pp. 331–399. arXiv: 1002.4232 [hep-lat].

[7] T. DeGrand and S. Schaefer. "Improving meson two-point functions in lattice QCD". In: *Computer physics communications* 159.3 (2004), pp. 185–191.

[8] L. Giusti et al. "Low-energy couplings of QCD from current correlators near the chiral limit". In: *Journal of High Energy Physics* 2004.04 (2004), p. 013.

[9] M. Lüscher. "Local coherence and deflation of the low quark modes in lattice QCD". In: *Journal of High Energy Physics* 2007.07 (2007), p. 081.

[10] J. Brannick et al. "Adaptive multigrid algorithm for lattice QCD". In: *Physical review letters* 100.4 (2008), p. 041601.

[11] M. Clark et al. "Multi-grid lanczos". In: *EPJ Web of Conferences*. Vol. 175. EDP Sciences. 2018, p. 14023.

[12] A. Stathopoulos et al. "Deflation for Monte-Carlo estimation of the trace of a matrix inverse". The Tenth International Workshop on Lattice QFT and Numerical Analysis (QCDNA X). 2017.

[13] "Coordinated Lattice Simulations `https://wiki-zeuthen.desy.de/CLS`".

[14] A. Stathopoulos and J. R. McCombs. "PRIMME: PReconditioned Iterative MultiMethod Eigensolver: Methods and software description". In: *ACM Transactions on Mathematical Software* 37.2 (2010), 21:1–21:30.