Disconnected Diagrams and High-Degree GMRES Polynomials

Walter Wilcox [a,*], Ronald Morgan[b], Paul Lashomb[a] and Travis Whyte[c,d]

[a] *Department of Physics, Baylor University, Waco, TX 76798-7316, USA*

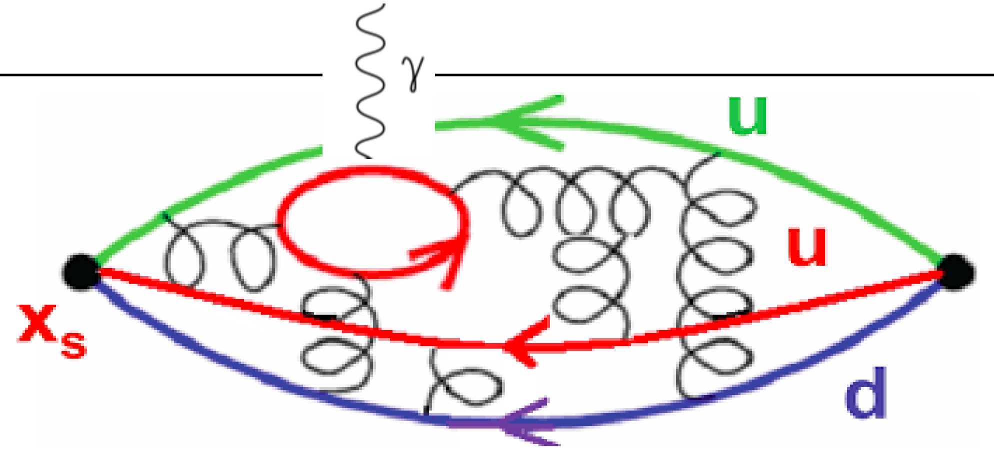
[b] *Department of Mathem… Vaco TX 76798-7316, USA*

[c] *College of William and Mary, D… e, Williamsburg, VA 23187, USA*

[d] *University of Cambric… :, UK CB2 1TN (in Oct.)*

Disconnected Loops

- Disconnected loop effects in many physical quantities
- Hard to evaluate due to many matrix inversions needed to measure all the background fermionic degrees of freedom
- Treat the disconnected quark loops stochastically, through the use of noise vectors to project out operator contributions



Subtraction methods needed in order to reduce the variance of these noisy calculations

Noise Subtraction

The approximate trace of the inverse Wilson matrix can be formed using large N

$$\lim_{N\to\infty} Tr\left(M^{-1}X_{Z(4)}\right) = Tr\left(M^{-1}\right)$$

Expectation value of the trace is invariant under the addition of a traceless matrix

$$\left\langle Tr\left(M^{-1}X_{Z(4)}\right)\right\rangle = \left\langle Tr\left\{\left(M^{-1} - \widetilde{M}^{-1}\right)X_{Z(4)}\right\}\right\rangle$$

The variance of the trace is given by:

$$Var\left[Tr\left(M^{-1}X_{Z(4)}\right)\right] = \frac{1}{N}\sum_{i\neq j}\left|M_{ij}^{-1}\right|^2$$

Variance of the trace is not invariant:

$$Var\left[Tr\left\{\left(M^{-1} - \widetilde{M}^{-1}\right)X_{Z(4)}\right\}\right] = \frac{1}{N}\sum_{i\neq j}\left(\left|M_{ij}^{-1} - \widetilde{M}_{ij}^{-1}\right|^2\right)$$

Goal: Find a traceless matrix $\widetilde{M}^{-1}$ that has off diagonal elements as close to M^{-1} as possible

Subtraction Methods

- Polynomial Subtraction (POLY)

  old (power): $\tilde{M}^{-1}_{poly} = g_0 + g_1 M + g_2 (M)^2 + \ldots g_7 (M)^7$

  new: GMRES polynomial, double GMRES polynomial

- Hermitian Forced Polynomial Subtraction (HFPOLY)

- Eigenvalue Subtraction on Polynomials (ESPOLY)

Solvers

- MINRES-DR(m,k)[1]
  - Calculate the lowest Q eigenpairs of the Hermitian Wilson matrix, $M\gamma_5$, to be used in the HF-type subtraction methods. Used on the first right hand side to produce k eigenvectors.
- GMRES-Proj[2]
  - Uses the k eigenvectors produced from GMRES-DR to accelerate the convergence of the remaining right hand sides.
- PP(d)-GMRES(m), PP-Arnoldi[3]
  - We use unrestarted GMRES and PP-Arnoldi to form the GMRES polynomials, which, at high order also gives eigenvalues and eigenvectors. PP(d)-GMRES then is used to solve for the noises using the GMRES polynomial of degree d.

[1] A. Abdel-Raheim et. al., SIAM J. Sci. Comput. 32 (2010) 129.
[2] D. Darnell, R. B. Morgan, W. Wilcox, Linear Algebra Appl. 429 (2008) 2415.
[3] J. A. Loe and R. B. Morgan, Numer. Linear Algebra Appl. 29 (2022) 1.

The GMRES Polynomial

- GMRES Polynomial
  - Formed by GMRES algorithm,

$$\|r\|_2 = \min_{\pi(A)\in\mathcal{P}} \|\pi(A)b\|_2$$

- "Power" method
  - Polynomial coefficients g solutions to normal equations,

$$(AY)^\dagger AYg = (AY)^\dagger b \;\text{ where }\; Y = [b, Ab, \dots, A^{d-1}b]$$

  - Suffered from instability beyond degree 12 polynomials.

The GMRES Polynomial Cont.

- New GMRES Polynomial used in Morgan and Loe's PP-GMRES:
  - Implementation from factored roots,

$$\pi(\alpha) = \prod_{i=1}^{d}\left(1 - \frac{\alpha}{\tilde{\theta}_i}\right)$$

  - $\tilde{\theta}_i$ are the Leja ordered, harmonic Ritz values of system
  - Much more stable

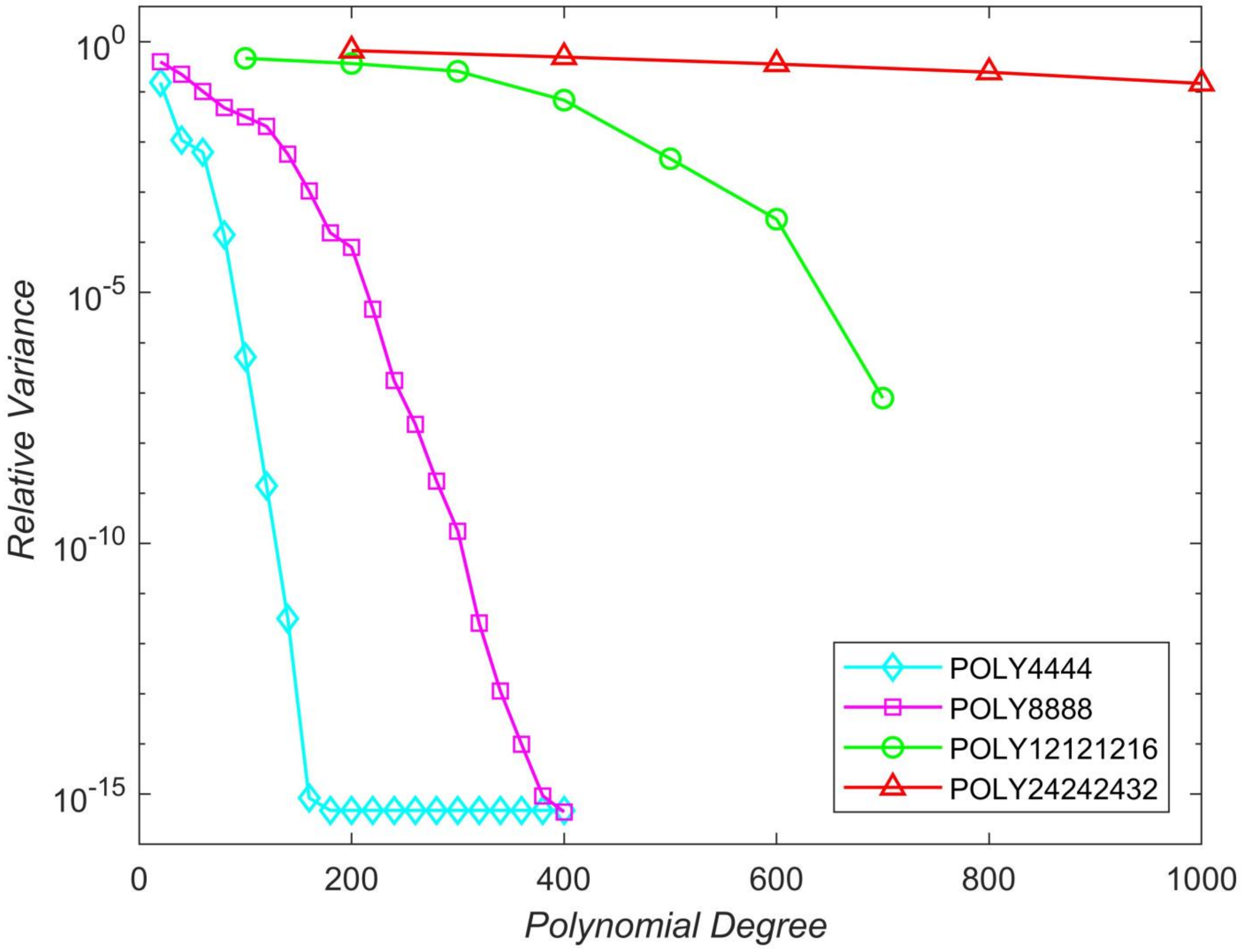
The GMRES Polynomial Cont.

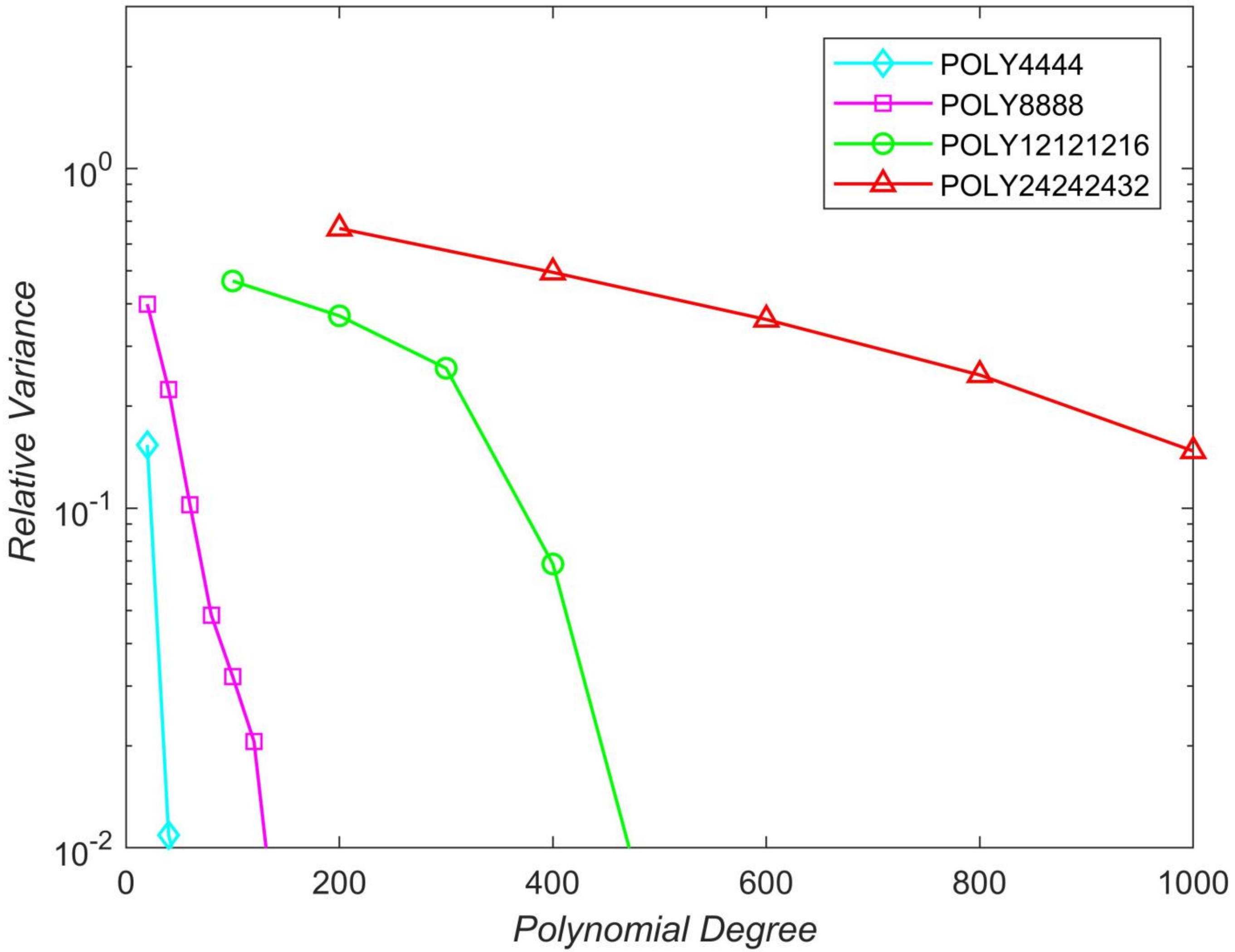
- New GMRES Polynomial
  - GMRES polynomial related to $p(A) \doteq A^{-1}$ by $\pi(\alpha) = 1 - \alpha p(\alpha)$,
  - $p(\alpha)$ reduces to,

$$p(\alpha) = \sum_{k=1}^{d} u_k \text{ where } u_k = \frac{1}{\tilde{\theta}_k}\left(1 - \frac{1}{\tilde{\theta}_1}\right)\left(1 - \frac{1}{\tilde{\theta}_2}\right)\cdots\left(1 - \frac{1}{\tilde{\theta}_{k-1}}\right)$$

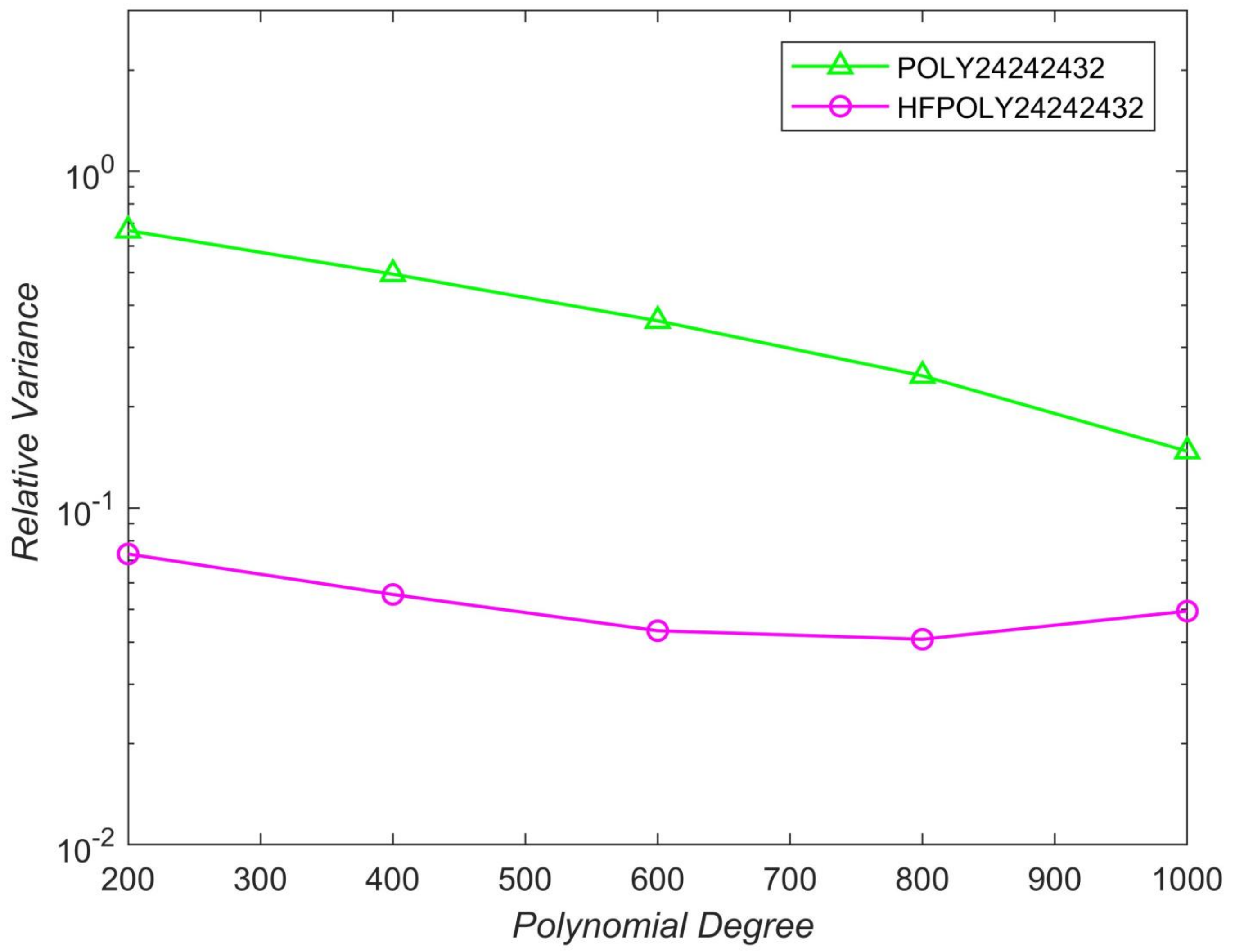
  - For subtraction, $\widetilde{M}^{-1}_{poly} \doteq p(M)$

Wilson fermions,
kappa_crit,
scalar,
unpartitioned noise,
averaged over configs (10),
and noises (10 or 50),
no error bars yet!





POLY4444
POLY8888
POLY12121216
POLY24242432
10^0
10^{-1}
10^{-2}
Relative Variance
0
200
400
600
800
1000
Polynomial Degree


POLY24242432
HFPOLY24242432
Relative Variance
10^0
10^-1
10^-2
200
300
400
500
600
700
800
900
1000
Polynomial Degree

Thanks to Ron Morgan. Uses PP(50)-GMRES(50). Trace(inv(A)-p(A)).

Caveat

Reducing errors on (s=scalar)

$$\sqrt{Re(s)^2 + Im(s)^2}$$

instead of just the real part!! Nevertheless, it demonstrates actual time savings with this method.

Degree	Noises	Time
0	1692	16.82 hours
7	498	4.92 hours
100	158	1.72 hours
200	60	43.0 minutes
300	12	12.4 minutes
400	2	6.80 minutes
500	2	9.60 minutes
600	2	12.5 minutes
700	2	16.4 minutes

Lattice: 12^3 x 16

Tolerance: 0.001 * 12^3 X 16, check every 2 noises

Corrected but different version. Uses Real(Trace(inv(A)-p(A))).

Degree	Variance	Standard Deviation	Relative Variance
No poly	2.6e-4	1.6e-2	1
7	1.9e-4	1.4e-2	7.34e-1
100	8.5e-5	9.2e-3	3.27e-1
200	3.5e-5	5.9e-3	1.34e-1
300	1.5e-5	3.9e-3	5.87e-2
400	4.4e-6	2.1e-3	1.70e-2
500	3.1e-7	5.6e-4	1.2e-3
600	1.4-9	3.8e-5	5.4e-6
700	1.0e-12	1.0e-6	3.8e-6
800	6.2e-16	2.5e-8	2.39e-12

An alternate type of evaluation is to run a fixed number of noises (50 here) and directly measure the relative variance. This establishes directly the connection between variance reduction and cost savings.

Interesting issue: Polynomial effectiveness may be affected by type of starting noise: gaussian, Z2, Z4.

Degree	Relative Variance (Z4)	Relative Variance (Z2)	Relative Variance (Gaussian)
No poly	1	1	1
7	7.34e-1	7.31e-1	7.69e-1
100	3.27e-1	3.27e-1	3.31e-1
200	1.34e-1	1.42e-1	1.77e-1
300	5.87e-2	5.00e-2	6.54e-2
400	1.70e-2	1.23e-2	5.38e-3
500	1.20e-3	1.62e-3	8.08e-4
600	5.40e-6	3.85e-6	3.69e-6
700	3.86e-9	5.38e-10	5.01e-10
800	2.39e-12	5.00e-13	9.62e-13

Multi-Level Trace Cascade

- Single polynomial preconditioning:
  - Right precondition $Ax = b$ using $p_1(A) \doteq A^{-1}$,

$$\phi_1(A)y = b \ \text{ where } \ \phi_1(A) \equiv Ap_1(A)$$

- Double polynomial preconditioning:
  - Preconditioner system a second time with $p_2(A) \doteq \left(\phi_1(A)\right)^{-1}$,

$$\phi_2\left(\phi_1(A)\right)z = b \ \text{ where } \ \phi_2\left(\phi_1(A)\right) \equiv \phi_1(A)p_2\left(\phi_{1(A)}\right)$$

Multi-Level Cascade

- Double polynomial preconditioner:
  - Preconditioners $p_1(A), p_2(\phi_1(A))$ act as a single preconditioner,

$$p(A) \equiv p_1(A)p_2(\phi_1(A))$$

  - Orthonormal basis for Arnoldi is expensive to form and store
  - Much cheaper to form as it requires two smaller Krylov subspaces
- Double polynomial subtraction:
  - For subtraction, we wish to use $\widetilde{M}_{poly}^{-1} \doteq p_1(M)p_2(\phi_1(M))$

Baylor University

FOR 5269

Linear equations use PP(50)-GMRES

Method	Noises	Time in hours
No Poly	1002	9.96
Deg's 700, 7	2 + 662 = 668	3.97
Deg's 700, 300, 7	2 + 86 + 872 = 960	3.15
Deg's 800, 100, 7 with Deflation of 13 eigenvalues (e.vect's computed with the GMRES run out to 800)	2 + 4 + 250 = 256	0.640
Deg's 1019, 100, 7 with Deflation of 13 eigenvalues (e.vect's computed with the PP(30)-GMRES run out to 50 iterations). Note the 1019 = 30*34-1, so 34 iterations of PP(30)-GMRES.	2 + 4 + 258 = 264	0.323

Lattice: 12^3 x 16

Tolerance: 0.0005 * 12^3 X 16, check every 2 noises

Baylor University PHYSICS

Linear equations use PP(50)-GMRES. Larger problem, larger time savings.

Caveat

Reducing errors on (s=scalar)

$\sqrt{Re(s)^2 + Im(s)^2}$

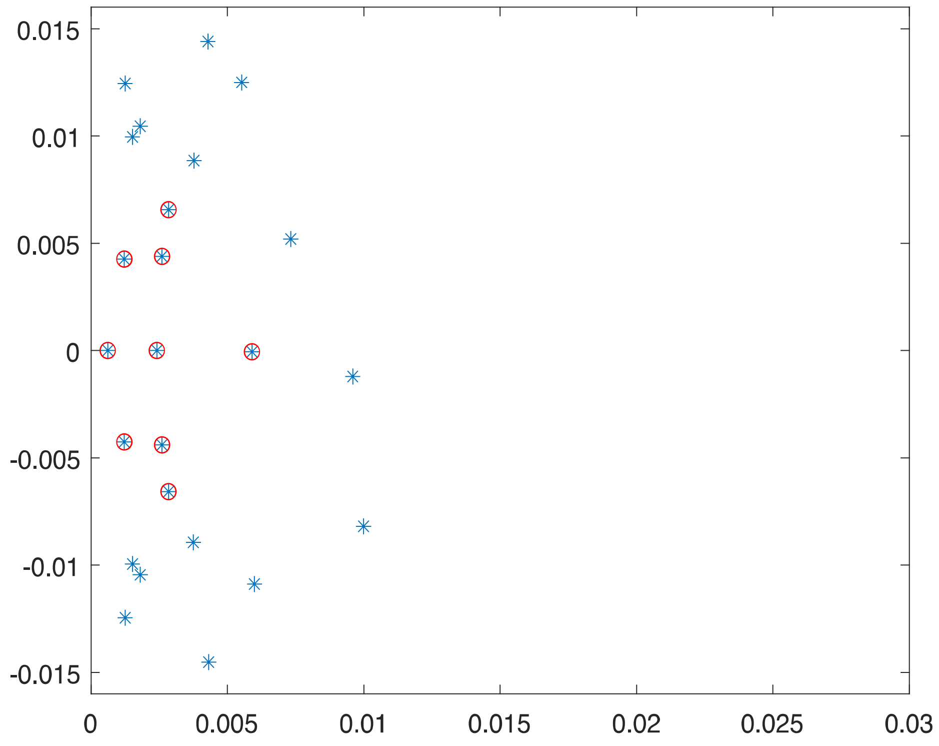
instead of just the real part!!

method	noise vectors	time (hours)
$Tr(A^{-1} - p_{3749}(A)) +$ $Tr(p_{3749}(A) - p_{100}(A)) +$ $Tr(p_{100}(A) - p_4(A))$	$2 + 243 + 40$ $= 285$	86.9
W/ Eigenvalue Deflation, polys 3749, 100, 4	$2 + 3 + 15$ $= 20$	4.1

Table: Trace computation with 24^4 lattice, $n = 3.98$ *million*.
Use Double Poly: degree = 75*50-1 = 3749. Deflation uses 9 eigenvalues.

Tolerance: 0.001* 12^3 X 16, check every 2 noises.
Ron Morgan's statement: "It scares me to think of doing the same problem with no polynomial."

Small eigenvalues of big QCD matrix. Circled ones are deflated.

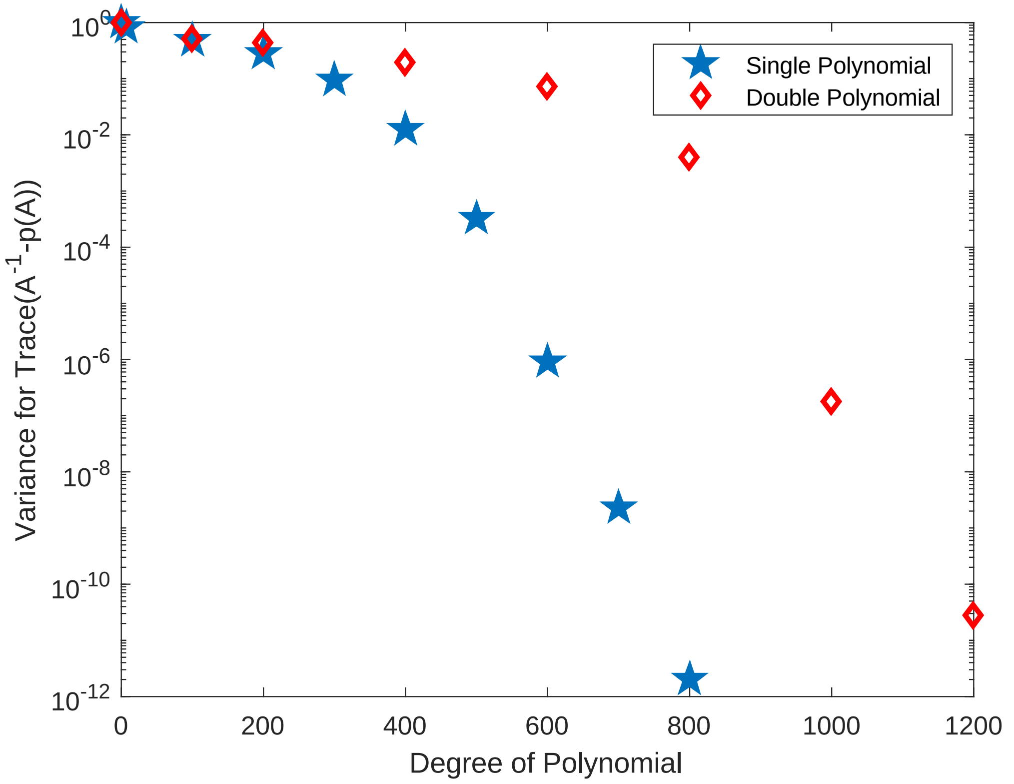


Baylor University

FOR 5269

A Little Double Dealing

12^3x16,
5 configs, 50 noises each



double degree =
[0 99 199 399 599 799 999 1199]

Double poly: Using a p1 polynomial of degree 20. The polynomial used for subtraction is always one degree lower.

Baylor University PHYSICS

Cost Function for Degree Cascade

- Preliminary work done on minimizing the cost function for polynomial degrees d1 and d2. Last polynomial done exactly with probing vectors or point vectors with tricks.
- Adaptation of work by Hallman and Troester, "A Multi-level Approach to Stochastic Trace Estimation".

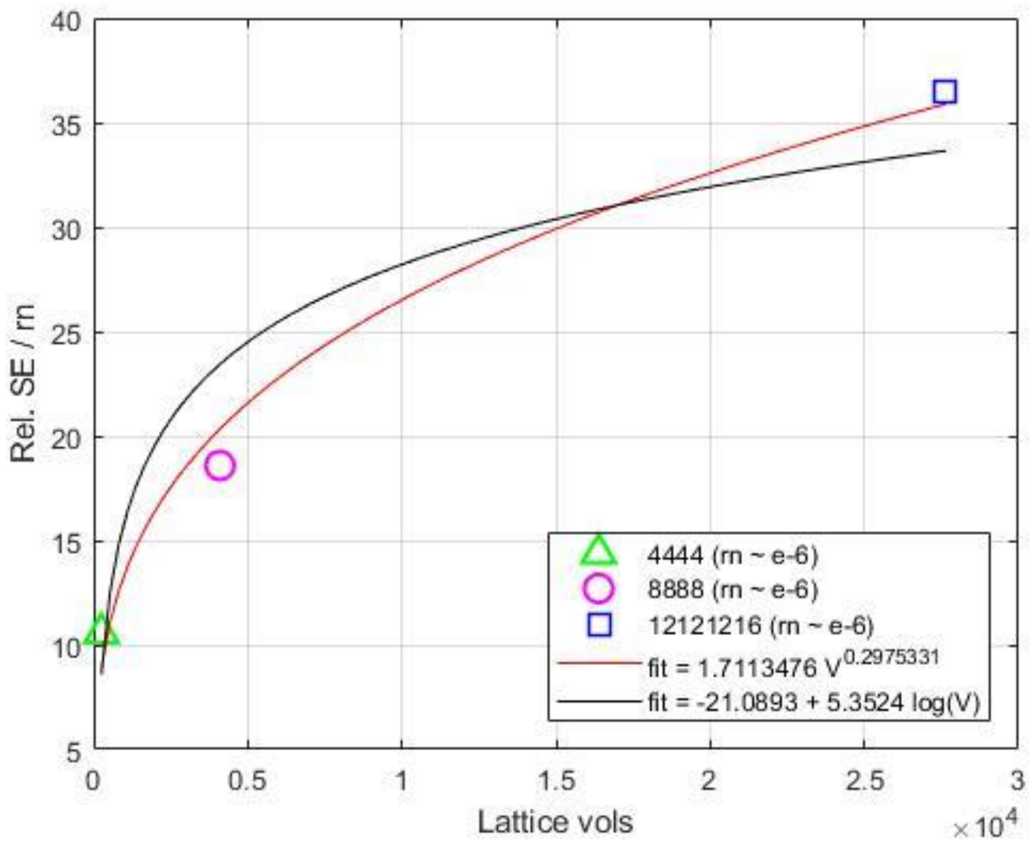
$$C = \sum_{k=1}^{L} m_k C_k = \varepsilon^{-2} \left(\sum_{k=0}^{L} \sqrt{V_k C_k} \right)^2$$

$$m_k = \mu \sqrt{V_k / C_k}, \quad \text{where}$$

$$\mu = \varepsilon^{-2} \sum_{k=1}^{L} \sqrt{V_k C_k}.$$

- Plan to use a scaled GMRES residual norm curve as an estimate or guide of the sample variances.

GMRES res. norm/ variance connection



Summary

- High-degree GMRES polynomials can be efficiently formed and used in lattice QCD for disconnected loop noise subtraction. Our approach combines high-degree polynomials and deflation techniques. HFPOLY is effective, but eventually only POLY helps at very high polynomial order.
- The subtraction algorithm can be completed with a multi-level trace “cascade”. An efficient approach uses a double-polynomial construction to form the GMRES polynomials and PPGMRES and PPArnoldi to solve noises and deflate the polynomial-only steps in the subtraction cascade.
- We can use the eigenvalues generated to both speed up the solution of noise equations as well as to deflate the subtraction algorithm. Thus, deflation plays two roles!
- Our work is ongoing. Current numerical results show that computer time savings for large lattices can be large compared to brute force noise methods.

Acknowledgments

Thanks to the Organizing Committee of the workshop for providing travel and accommodation support. This work was partially supported by a Baylor University Arts and Sciences Summer Research Grant and the Texas Advanced Computing Center (TACC).

For a different approach to noise reduction, please see Travis Whyte's interesting contribution on shift selection noise reduction at 3:10.

Ron Morgan HH XXI 2022 talk. Uses PP(8)-GMRES(50)

Caveat

Reducing errors on (s=scalar)

$$\sqrt{Re(s)^2 + Im(s)^2}$$

instead of just the real part!! Nevertheless, it demonstrates actual time savings with this method.

method	noise vectors	time (hours)
$Tr(A^{-1})$	1700	52.5
$Tr(A^{-1} - p_7(A))$	500	15.4
$Tr(A^{-1} - p_{700}(A)) +$ $Tr(p_{700}(A) - p_7(A))$	$10 + 350 = 360$	2.50
$Tr(A^{-1} - p_{700}(A)) +$ $Tr(p_{700}(A) - p_{300}(A)) +$ $Tr(p_{300}(A) - p_7(A))$	$5 + 40 + 430 = 475$	1.80
W/ Eigenvalue Deflation, polys 800, 100, 7	$5 + 10 + 110 = 125$	0.55
W/ Eigenvalue Deflation, Double 70*15-1=1049, 100, 7	$5 + 10 + 100 = 125$	0.33

Lattice: 12^3 x 16

Tolerance: 0.001 * 12^3 X 16, check every 5 noises