Interpolation as a means of shift selection in multilevel Monte Carlo with lattice displacements

Travis Whyte[1], Andreas Stathopoulos[1], Eloy Romero[2] , Kostas Orginos[1,2], [1] William & Mary, [2] Jefferson Lab

Problem Description

- Application: lattice QCD disconnected diagrams: estimate the trace of $\Gamma\Pi_p D^{-1}$ for a large, sparse matrix D, where D is the Dirac operator with local rotations (gamma-matrices, Γ) and permutations of the z direction in the lattice with products of gauge links (Π_p); typical matrix size $> 10^8$
- Variance reduction methods include probing and multigrid deflation.
  - Deflation: $\operatorname{tr} D^{-1} \approx \operatorname{tr} D^{-1}Q + t\big(D^{-1}(I-Q)\big)$, using an oblique projector $Q = DV(V^\dagger DV)^{-1}V^\dagger$ that approximately deflates the smallest singular values of D out. (Romero et. al. *J. Comp. Physics.*, 2020)

  But deflation will become prohibitively expensive at larger lattice volumes
- Frequency Splitting (FS) (Giusti et. al. *Eur. Phys. J. C.*, 2019) is a multilevel Monte Carlo method that splits the low and high frequency modes of the propagator:

$$\operatorname{tr}(\Gamma\Pi_p D^{-1}) = \sum_{i=0}^{j-1}(\sigma_{i+1}-\sigma_i)\operatorname{tr}(D+\sigma_i I)^{-1}\Gamma\Pi_p(D+\sigma_{i+1}I)^{-1} + \operatorname{tr}\Gamma\Pi_p(D+\sigma_j I)^{-1}$$

  where each term is computed stochastically with Hutchinson
- Scales with the lattice volume better than deflation; good for exascale regime

Challenges:

- No way to relate the shifts to the variance, so sampling the variance for different shifts must be done
- In general, the shifts for one combination of Γ and Π_p may not be the optimal for a different combination

Sampling and Interpolation

- Sample the variance for a set of initial shifts for a given (Γ,Π_p) pair, then interpolate the variances in logspace using PCHIP/linear interpolation
- Along with the solver cost, C_l, the interpolated variances create a much richer shift space to find shifts that minimize the multilevel cost function: $C_{ML} = \epsilon^{-2}(\sum_{l=0}^{L-1}\sqrt{C_l V_l})^2$ (Giles, M. *Acta Numerica*, 2015)

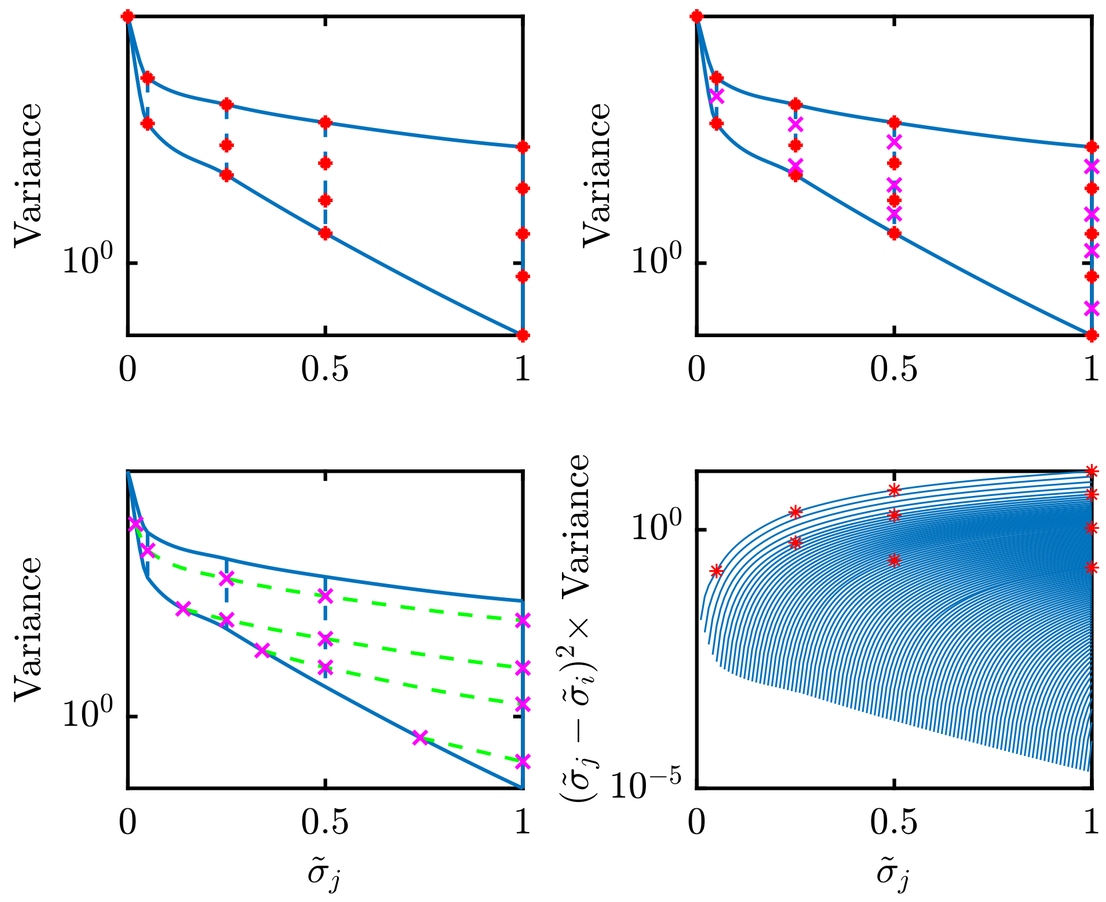

Figure 1. (Upper left) The sampled variances (red bursts), are interpolated along the boundary of the region (solid lines). (Upper right) The dashed lines define the points (purple crosses) resulting from interpolating the sampled interior points. (Lower left) The points populating the interior (green dashes) are created by interpolating the interior points. (Lower Right) The final result after multiplying by $(\tilde{\sigma}_i - \tilde{\sigma}_j)^2$.

Comparison to MG Deflation + Probing

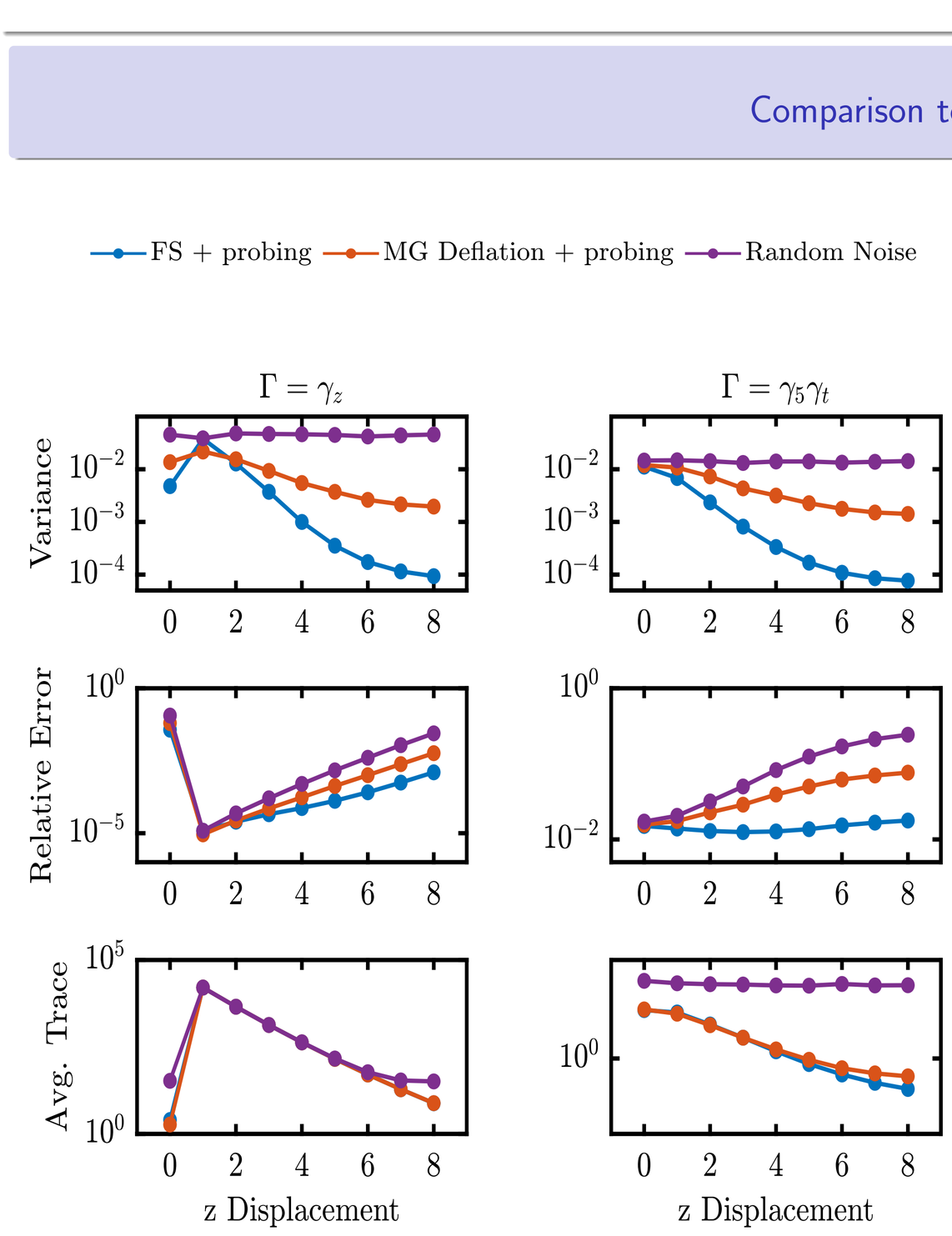

Figure 2. (Left) Comparing FS with p8k7 probing vectors to MG Deflation with p5k8 probing vectors to a target variance of $\epsilon^2 = 0.001$ at equal solver cost. The shifts used in FS were selected from an optimization of $(\Gamma, \Pi_p) = (\gamma_z, \Pi_4)$ on a $32^3 \times 64$ lattice at $m_q = -0.2390$.

	Solver Time ($\times 10^7$)	# of Inversions ($\times 10^6$)	Speedup
FS + Probing	0.0754	2.20	49
MG Deflation + Probing	0.3651	8.09	10
Random Noise	3.6619	68.14	1

Table 1. The estimated solver wallclock time, number of inversions and speedup over random noise to estimate the traces for every (Γ, Π_p) combination for each method while achieving a target variance of $\epsilon^2 = 0.001$ for $(\Gamma, \Pi_p) = (\gamma_z, \Pi_4)$ on 8 KNL nodes. In FS, each $l < L-1$ requires $2\times$ the number of inversions, but the inversions are cheaper as D is shifted.

	Configuration Number				
	1	2	3	4	5
Est. Speedup	4.8436	5.4360	4.8494	4.5541	5.0838
	Configuration Number				
	6	7	8	9	10
Est. Speedup	3.4911	4.9955	4.5245	4.5861	5.7280

Table 2. The estimated speedup of FS + probing over MG deflation + probing for 10 configurations. The set of shifts used in FS are the same set of shifts for all configurations and come from an optimization of $(\Gamma, \Pi_p) = (\gamma_z, \Pi_4)$ for the first configuration.

This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.