Multigrid Multilevel Monte Carlo

Gustavo Ramirez-Hidalgo
(work with prof. Andreas Frommer)

Bergische Universität Wuppertal
Department of Applied Mathematics

June 26, 2023

Seminar @ DESY/Berlin

BERGISCHE
UNIVERSITÄT
WUPPERTAL

Table of Contents

BERGISCHE UNIVERSITÄT WUPPERTAL

Some Essentials

Some Essentials

The problem at hand: compute $\mathrm{tr}(\Gamma D^{-1})$. We focus on $\Gamma = I$ here.

- D the Dirac operator
- do it stochastically

We focus here on the Wilson discretization:

$$D_W = \underbrace{m_0 I}_{\text{mass shift}} + \frac{1}{2}\sum_{\mu=1}^{4}\left(\gamma_\mu \otimes \underbrace{(\partial_\mu + \partial^\mu)}_{\text{centralized fd}} - \underbrace{a I_4 \otimes \partial_\mu \partial^\mu}_{\text{stabilization term}}\right)$$

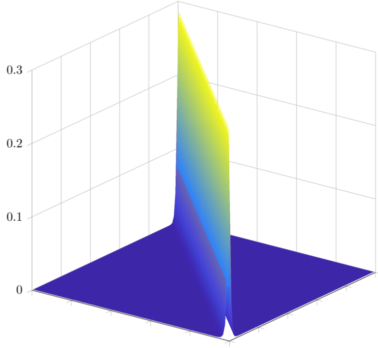
(**note**: $(\Gamma_5 D_W)^H = \Gamma_5 D_W$).

BERGISCHE UNIVERSITÄT WUPPERTAL

Some Essentials

Another problem at hand:

compute $\mathrm{tr}(\widetilde{P}^H D^{-1})$, where D is the Dirac matrix and $\widetilde{P}$ is a shift in the rows (which comes from a displacement on the lattice in one of the four spacetime directions).



- to illustrate how this can be difficult, note here a plot of the entries of A^{-1}, with A the 2D Laplace problem
- in LQCD, we see a rapid decay in the entries of D^{-1}

BERGISCHE UNIVERSITÄT WUPPERTAL

(taken from Claudia Schimmel's PhD thesis)

Some Essentials

How do we compute the trace stochastically?

- use the Hutchinson estimator:

$$\mathrm{tr}(A) \approx \frac{1}{N}\sum_{i=1}^{N}(x^{(i)})^H A x^{(i)},$$

- in such a Monte Carlo computation, the variance determines the work, and in this particular case the variance takes the form (if we choose Rademacher vectors):

$$\begin{aligned}\mathbb{V}[x^H A x] &= \frac{1}{2}\|\mathrm{offdiag}(A + A^T)\|_F^2 \\ &= \frac{1}{2}\left(\|A + A^T\|_F^2 - \sum_{i=1}^{n}(A_{ii} + (A^T)_{ii})\right)\end{aligned}$$

BERGISCHE UNIVERSITÄT WUPPERTAL

Some Essentials

How do we compute the trace stochastically?

- the SVD of a matrix $A \in \mathbb{C}^{n \times n}$ has the form: $A = U \Sigma V^H$
- there is a relation between the Frobenius norm and the singular values of a matrix:

$$\|A\|_F^2 = \sum_{i=1}^{n} \sigma_i^2$$

- so, **to reduce the variance** of the Hutchinson estimator we need to **remove the largest singular values of** A. In the case of $\mathrm{tr}(D^{-1})$, this means removing the smallest singular values of D

BERGISCHE UNIVERSITÄT WUPPERTAL

Some Essentials

How do we deflate in the Hutchinson estimator?

- set the orthogonal projector: $\Pi = U_k U_k^H$
- if the columns of U_k are left singular vectors of D to very high accuracy:

$$
\begin{aligned}
\mathrm{tr}(D^{-1}) &= \mathrm{tr}(D^{-1}(I - \Pi)) + \mathrm{tr}(D^{-1}\Pi) \\
&= \mathrm{tr}(D^{-1}(I - \Pi)) + \mathrm{tr}(U_k^H V_k \Sigma_k^{-1})
\end{aligned}
$$

BERGISCHE UNIVERSITÄT WUPPERTAL

Some Essentials

How do we deflate in the Hutchinson estimator?

- note (remember: $D = U\Sigma V^H$):

$$D^{-1}(I - \Pi) = (V\Sigma^{-1}U^H)(I - U_k U_k^H)$$

- therefore, when computing $\mathrm{tr}(D^{-1}(I - \Pi))$, this deflated form will have a smaller variance compared to $\mathrm{tr}(D^{-1})$

BERGISCHE UNIVERSITÄT WUPPERTAL

Some Essentials

How do we deflate in the Hutchinson estimator?

- BUT: computing singular vectors of D to high accuracy can be very expensive
- if we compute the vectors in U_k to low accuracy, then:

$$\begin{aligned}\mathrm{tr}(D^{-1}) &= \mathrm{tr}(D^{-1}(I-\Pi)) + \mathrm{tr}(D^{-1}\Pi) \\ &= \mathrm{tr}(D^{-1}(I-\Pi)) + \mathrm{tr}(U_k^H D^{-1} U_k)\end{aligned}$$

which requires extra k inversions!

BERGISCHE UNIVERSITÄT WUPPERTAL

Some Essentials

How do we deflate in the Hutchinson estimator?

- in this case of low accuracy for the vectors in U_k, we might want to switch from an orthogonal to an oblique projector:

$$\mathrm{tr}(D^{-1}) = \mathrm{tr}(D^{-1} - U_k(V_k^H D U_k)^{-1} V_k^H) + \mathrm{tr}((V_k^H D U_k)^{-1} V_k^H U_k)$$

BERGISCHE
UNIVERSITÄT
WUPPERTAL

Multigrid and DD-αAMG:
DD-αAMG = Domain Decomposition Aggregation-Based αdaptive Algebraic Multigrid

Multigrid and DD-αAMG

DD-αAMG = Domain Decomposition Aggregation-Based αdaptive Algebraic Multigrid.

The main problem with iterative solvers when dealing with $Ax = b$

- condition number of a matrix A, $\kappa(A)$, gives a measure on how difficult it is to solve a system of equations of the form $Ax = b$
- $\kappa_p(A) = \|A\|_p \|A^{-1}\|_p$
- $\kappa_2(A) = \frac{\sigma_1}{\sigma_n}$

iter. count	$\kappa(A)$
37	100
592	400
9,472	1,600
151,552	6,400
...	...

BERGISCHE UNIVERSITÄT WUPPERTAL

Multigrid and DD-αAMG

Iterative solvers as smoothers: Gauss-Seidel applied to the Laplace 2D problem

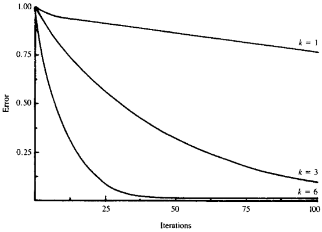
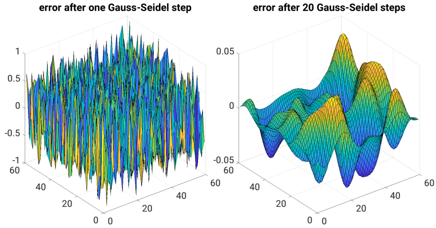


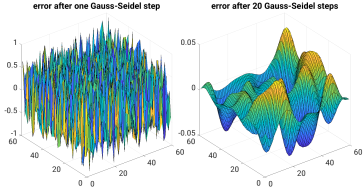
Image taken from *A Multigrid Tutorial* by Briggs et.al. Three different right hand sides.

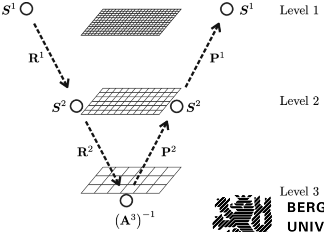


Starting with a random initial guess.

Multigrid and DD-αAMG

- the error can be split as: $e = e_{low} + e_{high}$.
- a good smoother eliminates e_{high} effectively.
- multigrid solvers can lead to convergence independent of $\kappa(A)$.





BERGISCHE UNIVERSITÄT WUPPERTAL

Multigrid and DD-αAMG

Algebraic multigrid: non-geometric.

Two-level AMG:

- Petrov-Galerkin projection: $D_c = RDP$
- coarse-grid correction: $\psi \leftarrow \psi + PD_c^{-1}Rr$, with $r = \eta - D\psi$
- error propagation: $E = I - PD_c^{-1}RD$
- $\Rightarrow$ **choose** $\mathrm{range}(P)$ **to contain the near kernel**

Algorithm 1 Two-level V-cycle

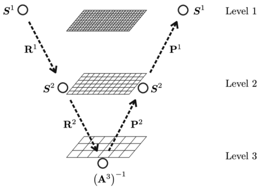
Input : ψ, η, ν
Output : ψ

$r \leftarrow \eta - D\psi$
$\psi \leftarrow \psi + PD_c^{-1}Rr$
$r \leftarrow \eta - D\psi$
$\psi \leftarrow \psi + M_{SAP}^{(\nu)}r$

BERGISCHE UNIVERSITÄT WUPPERTAL

Multilevel AMG:

this extension is possible by decomposing $e_{low} = e_{c,low} + e_{c,high}$ into the second level.



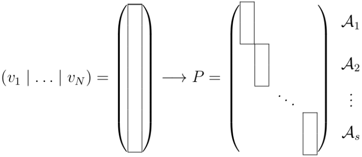
Algorithm $\boxed{\text{Two-level V-cycle}}$

Input : ψ, η, ν
Output : ψ

$r \leftarrow \eta - D\psi$
$\psi \leftarrow \psi + P\boxed{D_c^{-1}Rr}$
$r \leftarrow \eta - D\psi$
$\psi \leftarrow \psi + M_{SAP}^{(\nu)} r$

BERGISCHE UNIVERSITÄT WUPPERTAL

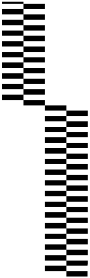
Multigrid and DD-αAMG



- $\mathcal{A}_i := \mathcal{L}_{j(i)} \times \mathcal{W}_i$
- **standard aggregation:** $\{\mathcal{A}_{j,\tau} \; : \; j = 1, ..., n_{\mathcal{L}_c}, \tau = 0, 1\}$, such that:

  $$\mathcal{A}_{j,0} := \mathcal{L}_j \times \{0, 1\} \times \mathcal{C}$$
  $$\mathcal{A}_{j,1} := \mathcal{L}_j \times \{2, 3\} \times \mathcal{C}$$
- when using the standard aggregation: $\Gamma_5 P = P\Gamma_5^c$
- also, it makes sense to choose: $D_c = P^H DP$

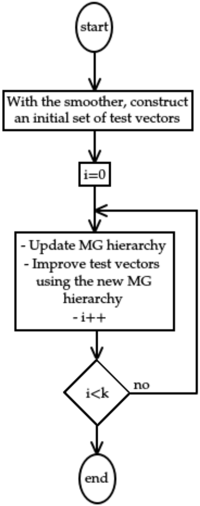
A section of P:



BERGISCHE UNIVERSITÄT WUPPERTAL

Multigrid and DD-αAMG

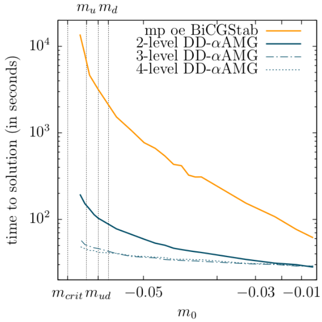
- **test vectors**: the approximate eigenvectors used in the construction of P, i.e. $\{v_i : i = 1, ..., N\}$
- the setup phase in DD-αAMG builds these test vectors
- the method updates the multigrid hierarchy as the test vectors improve



BERGISCHE UNIVERSITÄT WUPPERTAL

Multigrid and DD-αAMG

Comparison of computational cost for solving linear systems with a configuration from a BMW collaboration configuration using DD-αAMG and a Krylov subspace method. The left plot reports on timings for the solve only, whereas the right plot includes the multigrid setup time. Both plots were generated on the JUROPA high performance computer from the Jülich Supercomputing Centre (taken from Matthias Rottmann's PhD thesis).



MGMLMC:

Multigrid Multilevel Monte Carlo

MGMLMC

Assumming a multigrid construction such that $R_\ell = P_\ell^H$ and $P_\ell^H P_\ell = I$, then (Frommer, Nasr, R-H, 2022):

- two levels:

$$\operatorname{tr}(A_0^{-1}) = \operatorname{tr}(A_0^{-1}) - \operatorname{tr}(P_0 A_1^{-1} R_0) + \operatorname{tr}(P_0 A_1^{-1} R_0) \tag{1}$$

$$= \operatorname{tr}(A_0^{-1} - P_0 A_1^{-1} P_0^H) + \operatorname{tr}(A_1^{-1}) \tag{2}$$

BERGISCHE
UNIVERSITÄT
WUPPERTAL

MGMLMC

Assumming a multigrid construction such that $R_\ell = P_\ell^H$ and $P_\ell^H P_\ell = I$, then (Frommer, Nasr, R-H, 2022):

- two levels:

$$\begin{aligned} \mathrm{tr}(A_0^{-1}) &= \mathrm{tr}(A_0^{-1}) - \mathrm{tr}(P_0 A_1^{-1} R_0) + \mathrm{tr}(P_0 A_1^{-1} R_0) && (1) \\ &= \mathrm{tr}(A_0^{-1} - P_0 A_1^{-1} P_0^H) + \mathrm{tr}(A_1^{-1}) && (2) \end{aligned}$$

- we can connect this to inexact/approximate deflation, which is discussed in *Multigrid deflation for Lattice QCD* (Romero, Stathopoulos and Orginos), and already in Lüscher's local coherence paper

BERGISCHE UNIVERSITÄT WUPPERTAL

MGMLMC

We move from a two-level to a multilevel method by means of Multilevel Monte Carlo (Giles, 2008), and then write (Frommer, Nasr, R-H, 2022):

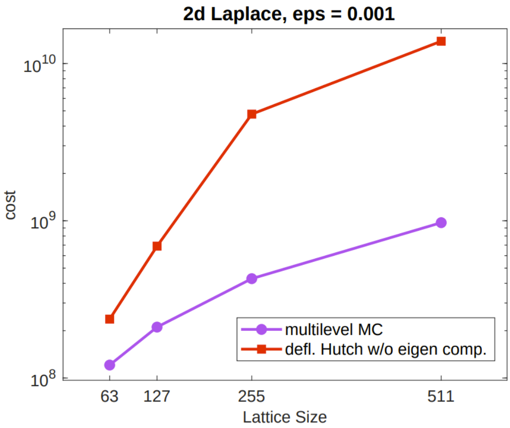
- multilevel:

$$\mathrm{tr}(A_0^{-1}) = \sum_{\ell=0}^{L-2} \mathrm{tr}\left(A_\ell^{-1} - P_\ell A_{\ell+1}^{-1} R_\ell\right) + \mathrm{tr}(A_{L-1}^{-1}) \tag{3}$$

$(\mathrm{cost}(D_1^{-1} b_1) \sim 0.1 \mathrm{cost}(D_0^{-1} b_0))$

MGMLMC

MGMLMC vs exactly deflated Hutchinson (Frommer, Nasr, R-H, 2022):



BERGISCHE UNIVERSITÄT WUPPERTAL

A short note on deflation in LQCD:

if we want to deflate as $\mathrm{tr}((I-\Pi)D^{-1})$, we can use the following to compute approximate deflation vectors:

- the Dirac operator fulfills Γ_5-hermiticity: $(\Gamma_5 D)^H = \Gamma_5 D$

BERGISCHE UNIVERSITÄT WUPPERTAL

A short note on deflation in LQCD:

if we want to deflate as $\mathrm{tr}((I-\Pi)D^{-1})$, we can use the following to compute approximate deflation vectors:

- the Dirac operator fulfills Γ_5-hermiticity: $(\Gamma_5 D)^H = \Gamma_5 D$
- this gives us access to the SVD:
  $\Gamma_5 D = V \Sigma V^H \Rightarrow D = (\Gamma_5 V \mathrm{sign}(\Sigma))\mathrm{abs}(\Sigma) V^H$
  $\Rightarrow D^{-1} = V \mathrm{abs}(\Sigma^{-1})(\Gamma_5 V \mathrm{sign}(\Sigma))^H$

BERGISCHE UNIVERSITÄT WUPPERTAL

A short note on deflation in LQCD:

if we want to deflate as $\mathrm{tr}((I - \Pi)D^{-1})$, we can use the following to compute approximate deflation vectors:

- the Dirac operator fulfills Γ_5-hermiticity: $(\Gamma_5 D)^H = \Gamma_5 D$
- this gives us access to the SVD:
  $\Gamma_5 D = V \Sigma V^H \Rightarrow D = (\Gamma_5 V \mathrm{sign}(\Sigma)) \mathrm{abs}(\Sigma) V^H$
  $\Rightarrow D^{-1} = V \mathrm{abs}(\Sigma^{-1})(\Gamma_5 V \mathrm{sign}(\Sigma))^H$
- $\Rightarrow$ block pow. it. on $(\Gamma_5 D)^{-1}$ gives columns of V, ready for deflation

BERGISCHE
UNIVERSITÄT
WUPPERTAL

MGMLMC

A short note on deflation in LQCD:

if we want to deflate in MGMLMC as e.g. $\mathrm{tr}((I-\Pi)(D^{-1}-P_1 D_2^{-1} P_1^H))$, do we have something similar to the previous slide to compute the deflation vectors?

- in our multigrid construction:

$$\Gamma_5^\ell P_\ell = P_\ell \Gamma_5^{\ell+1},\ (\Gamma_5^\ell D_\ell)^H = \Gamma_5^\ell D_\ell \tag{4}$$

  which implies: $((D_\ell^{-1} - P_\ell D_{\ell+1}^{-1} P_\ell^H)\Gamma_5^\ell)^H = (D_\ell^{-1} - P_\ell D_{\ell+1}^{-1} P_\ell^H)\Gamma_5^\ell$

BERGISCHE UNIVERSITÄT WUPPERTAL

MGMLMC for a lattice with a displacement
(i.e. $\mathrm{tr}(\widetilde{P}^H D_0^{-1})$)

MGMLMC for a lattice with a displacement

Two-level decomposition:

$$\begin{aligned}\operatorname{tr}(\widetilde{P}^H D^{-1}) &= \operatorname{tr}\left(\widetilde{P}^H (D_0^{-1} - P_0 D_1^{-1} P_0^H)\right) + \operatorname{tr}(\widetilde{P}^H P_0 D_1^{-1} P_0^H) \\ &= \operatorname{tr}\left(\widetilde{P}^H (D_0^{-1} - P_0 D_1^{-1} P_0^H)\right) + \operatorname{tr}(P_0^H \widetilde{P}^H P_0 D_1^{-1})\end{aligned}$$

Let us introduce: $B_1 := P_0^H \widetilde{P}^H P_0$.

MGMLMC for a lattice with a displacement

I have recently implemented a minimal version of DD-αAMG in MATLAB, which is available here:
`https://github.com/Gustavroot/MGMLMC_for_mat_with_displ`

In that MATLAB code, also orthogonal deflation with low-accuracy vectors has been implemented for both Hutchinson and MGMLMC. This code allows one to compute variances and compare both methods.

BERGISCHE UNIVERSITÄT WUPPERTAL

MGMLMC for a lattice with a displacement

Some results (4-D problem, LQCD, 16^4 lattice $\rightarrow 786,432$ matrix dim., 5×12 matrix displacement $\widetilde{P}$ – $\beta = 3.55$, $\kappa = 0.137$, $m_\pi \approx 420$ MeV, $N_f = 3$, p.b.c.):

BERGISCHE UNIVERSITÄT WUPPERTAL

MGMLMC for a lattice with a displacement

Some results (4-D problem, LQCD, 16^4 lattice $\rightarrow$ $786,432$ matrix dim., 5×12 matrix displacement $\widetilde{P}$ – $\beta = 3.55$, $\kappa = 0.137$, $m_\pi \approx 420$ MeV, $N_f = 3$, p.b.c.):

- sampled variances (at $\ell = 0$):
  - $\mathrm{var}(\Pi D_0^{-1} \widetilde{P}^H)$:

n_{defl}	var ($\times 10^3$)
0	145.5
32	131.8
128	120.5
512	115.3

  - $\mathrm{var}\left(\Pi(D_0^{-1} - P_0 D_1^{-1} P_0^H)\widetilde{P}^H\right)$:

n_{defl}	var ($\times 10^3$)
0	110.7
32	109.8
128	105.5

BERGISCHE UNIVERSITÄT WUPPERTAL

MGMLMC for a lattice with a displacement

Some results (4-D problem, LQCD, 16^4 lattice $\rightarrow$ $786,432$ matrix dim., 5×12 matrix displacement $\widetilde{P}$ – $\beta = 3.55$, $\kappa = 0.137$, $m_\pi \approx 420$ MeV, $N_f = 3$, p.b.c.):

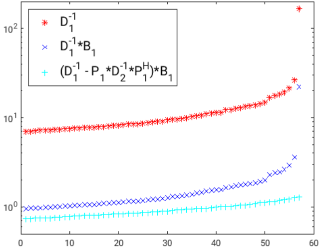
- sampled variances (at $\ell = 0$):
  - $\mathrm{var}(\Pi D_0^{-1} \widetilde{P}^H)$:

n_{defl}	var ($\times 10^3$)
0	145.5
32	131.8
128	120.5
512	115.3

  - $\mathrm{var}\left(\Pi (D_0^{-1} - P_0 D_1^{-1} P_0^H) \widetilde{P}^H \right)$:

n_{defl}	var ($\times 10^3$)
0	110.7
32	109.8
128	105.5

- singular values (at $\ell = 1$, ascending order):



BERGISCHE UNIVERSITÄT WUPPERTAL

MGMLMC for a lattice with a displacement

- more details of these results for a lattice with a displacement will be made available in a proceedings at the end of this week
- large-scale tests (C, MPI, OpenMP, large lattices) are under preparation and will be released soon in a paper, as part of the work of Jose Jimenez-Merchan (PhD student in our group)

BERGISCHE
UNIVERSITÄT
WUPPERTAL

This work is part of the Research Unit 5269, funded by the DFG under FOR 5269.

Thank you!

Questions?