MLMC++ as a variance reduction method

M. Khalil, A. Frommer

Wuppertal University
Lattice conference, Aug 8 – 13 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No' 765048

Content

Problem Statement

Consider the problem of computing

$$\mathrm{tr}(f(A)) := \sum_{i=1}^{n} [f(A)]_{ii} \tag{1}$$

- In our case: $f(A) = A^{-1}$
- for $A \in \mathbb{C}^{n \times n}$ large, sparse matrix.
- Compute (1) directly $\rightarrow$ not possible (storage, cost).
- Hutchinson's method $\rightarrow$ estimate $\mathrm{tr}(A^{-1})$ stochastically.

Hutchinson approach: (Hutchinson, 1989)

- Assume: a vector $x \in \mathbb{C}^n$ with i.id distribution as:

$$x_i \in \{-1, 1\} \text{ with equal probabililty } \tfrac{1}{2}, \tag{2}$$
$$x_i \text{ is } N(0,1) \text{ normally distributed.} \tag{3}$$

- the unbiased trace estimator of A^{-1} is given by:

$$\text{tr}(A^{-1}) \approx \frac{1}{s}\sum_{i=1}^{s} x_i^* A^{-1} x_i. \tag{4}$$

where $\tau = x^* A^{-1} x$ is the mean value,

- the variance:

$$\mathbb{V}[x^* A^{-1} x] = \frac{1}{2}\|\text{offdiag}(A^{-1})\|_F^2. \tag{5}$$

MC trace algorithm

Input: A, x, ϵ, m
for $s = 1$ to m **do**
  $\tau_i \leftarrow x^* A^{-1} x$
  if $\mathrm{Var}(\tau)/s \leq \epsilon$ **then**
    stop
  end if
end for
Output: mean of τ

Properties:

- simple, requires a solver for A^{-1}.
- convergence rate of MC is slow $\rightarrow$ as $O(1/\sqrt{s})$.
- variance $\rightarrow$ very large when a_{ij} large.

Variance Reduction: via, Deflation, .., Hutch++, MLMC

Content

Approximate/Inexact deflation

Projections $\rightarrow$ does not depend on eigenmodes

- Projection vector $\leftarrow V = WU^*$
- $W, U \in \mathbb{C}^{n \times d}$
- $\mathrm{tr}(A) = \mathrm{tr}(A(I - V)) + \mathrm{tr}(AV)$
- use the cyclic trace property :
  $\mathrm{tr}(AV) = \mathrm{tr}(AWU^*) = \mathrm{tr}(U^*AW), \;\; U^*AW \in \mathbb{C}^{d \times d}$

Our case: $B = A^{-1}$

- stochastic column vectors $\rightarrow W = U = BV$, $V \in \mathbb{C}^{n \times d}$
  [Hutch++]

Hutch ++ method

The idea of the Hutch++ based on the projection technique..

- given, $A \in C^{n\times n}$ is a PSD , ϵ relative accuracy , d nr. deflation. vects.
- Rademacher matrix: $S = 2 \times \text{randi}(2, n, d) - 3$
  $S \in C^{n\times d}$ with entries $\{-1, 1\}$
- solve the system $Y = A^{-1}S$, and $Y \in C^{n\times d}$
- Project eigenvectors: $[V, \ ] = qr(Y, 0)$
- the projection: $W = VV'$
- $\text{tr}(A^{-1}) = \underbrace{\text{tr}(A^{-1}W)}_{\text{directly}} + \underbrace{\text{tr}(A^{-1}(I - W))}_{\text{stochastically}}$

Hutch++ algorithm

Compute Stochastic part: $\mathrm{tr}(A_s^{-1}) = \mathrm{tr}(A^{-1}(I - W))$

```
Input: A, S, d, m, x
for s = 1 → m do
    τ ← x* W A^{-1} W x
    if Var(τ)/s ≤ ϵ then
        stop
    end if
end for
Output: mean of τ
```

Imporvments:

- easy and simple, reduce the accuracy dependence from $O(n^2) \to O(n)$.

Content

MLMC: main idea (Giles, 2015)

Assume a random function q_0 splits as

$$q_0 = \sum_{\ell=1}^{L} q_l, \quad \ell \text{ nr of level differnce}$$

$$\mathbb{E}[q_0] = \sum_{\ell=1}^{L-1} \underbrace{\mathbb{E}[q_\ell - q_{\ell+1}]}_{=w_\ell} + \underbrace{q_L}_{=w_L}$$

where $w_\ell^{(i)}$ independent samples on each level.
In case: $w = x^* A^{-1} x$, the unbiased estimator for $\mathrm{tr}(A^{-1})$ given by

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

The variance: $\sum_{\ell=1}^{L} \frac{1}{N_\ell} \mathbb{V}[w_\ell]$.

MLMC: setup phase

- The goal: Solver & Hierarchy of the linear system
- define Prolongation $P_l \rightarrow l = 0, 1, ..., L$
- define Restriction $R_l \rightarrow l = 0, 1, ..., L$
- usually $\rightarrow R = P^*$
- define coarse matrix $\rightarrow B_{l+1} = R_l B_l P_l$

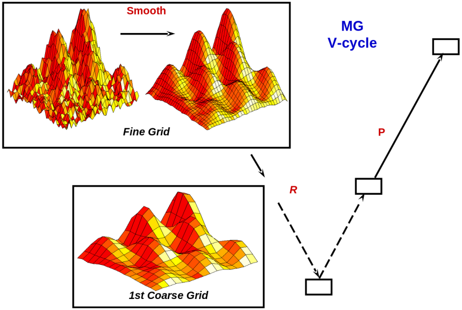


Figure: MG V-cycle: setup phase

2-grid mlmc

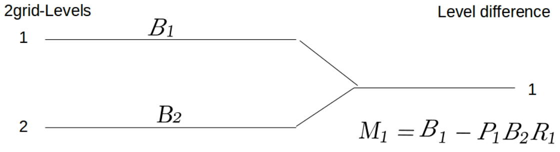


Figure: 2-Grid example

- Accumulated prolongation and restriction: $\hat{P}_\ell = P_0 \cdots P_\ell, \quad \hat{R}_\ell = R_\ell \cdots R_0.$
- Multilevel decomposition:

$$\underbrace{x^* B x}_{q_0} = \sum_{\ell=0}^{L-1} \underbrace{x^* \left(\hat{P}_\ell B_\ell \hat{R}_\ell - \hat{P}_{\ell+1} B_{\ell+1} \hat{R}_{\ell+1} \right) x}_{w_\ell} + \underbrace{x^* (\hat{P}_L B_L \hat{R}_L) x}_{w_L}$$

Target accuracy ρ $\leftarrow$ measured standard deviation

Uniform accuracy:

- Idea: distribute the target accuracy in equal.
  Achieve: $\rho_\ell = \rho/\sqrt{L-1}$ for all ℓ,

Optimized accuracy:

- Idea: minimize the work if we know the cost C_ℓ ... and the variance V_ℓ
- for each sample: update C_ℓ and measured V_ℓ.
- define the new optimal target tolerance ρ_ℓ

Content

- **Idea:** MLMC & Hutch++ together..
- The level difference matrix given as:

$$M_l = \hat{P}_l B_l \hat{R}_l - \hat{P}_{l+1} B_{l+1} \hat{R}_{l+1} \quad , \quad M_l \in C^{n_l \times n_l} \tag{6}$$

- generate random matrix $S_\ell \in C^{n_l \times d_l}$, $l = 1, .., L-1$
- applying the projection vectors: $Q_l = qr[M_l^{-1} S_l, 0]$,
- split the matrix level difference:

$$M_l = (M_l)_d + (M_l - (M_l)_d) \tag{7}$$

- trace of multilevel decomposition of M_l:

$$\text{tr}(M_l^{-1}) = \text{tr}(\underbrace{Q_l^* M_l Q_l}_{\text{directly}}) + \text{tr}(M_l \underbrace{(I - Q_l^* Q_l)}_{\text{stochastically}}) \tag{8}$$

- **challenge:** find an optimal way to determine d_l to reduce the variance enough

Content

Schwinger model: Table

Schwinger model						
N		$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	L
128	n_ℓ	$2 \cdot 128^2$	$4 \cdot 32^2$	$8 \cdot 8^2$	$8 \cdot 2^2$	4
	$\mathrm{nnz}(S_\ell^N)$	$2.94e5$	$1.64e5$	$2.46e4$	1024	
m	-0.1320	-0.1325	-0.1329	-0.1332	-0.1333	
n_{defl}	384	384	512	512	512	

Table: Parameters and quantities for Schwinger example

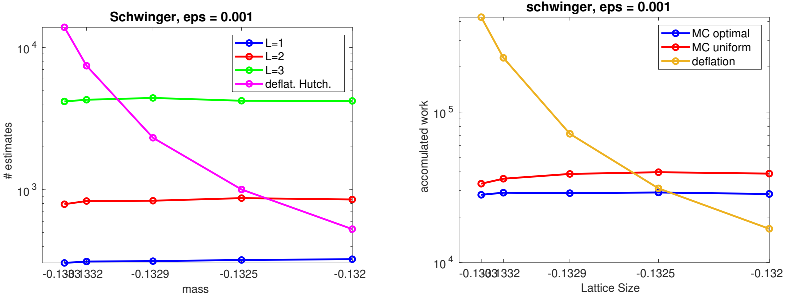
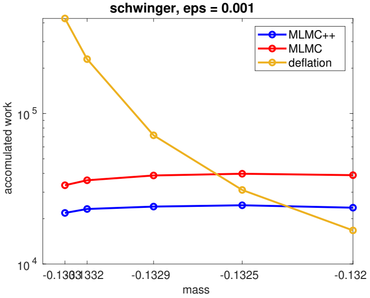


Figure: Comparison of uniform and optimal mlmc for schwinger matrix: no of samples on each level difference (left) and total work for different (right) m.

mlmcPP work: Figures



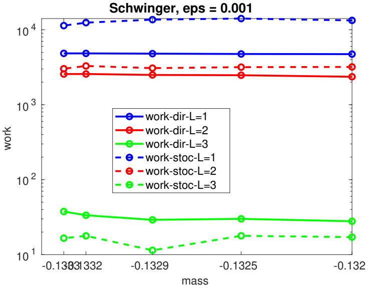


Figure: On each level diff: (1) explore accomulated work of defl.Hutch, mlmc and mlmc++ for schwinger matrix (left), (2) comparison between setup and stochastic work of mlmc++ (right).

schwinger variance: Figures

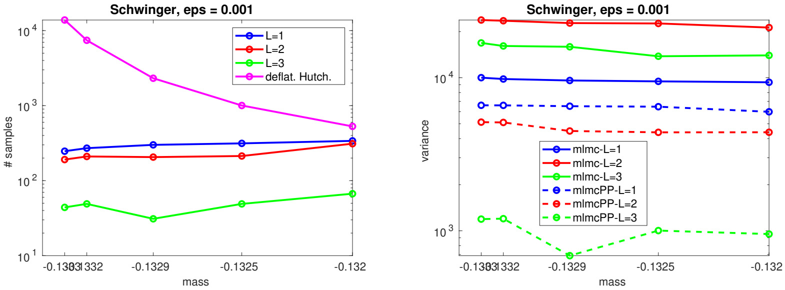
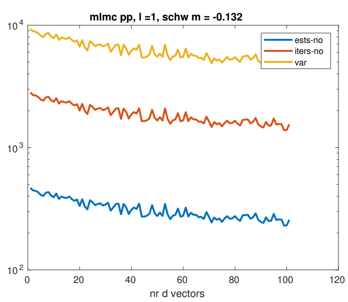


Figure: show samples of mlmc++ and deflation Hutch. for schwinger matrix (left). variance comparison between mlmc and mlmc++ on each level difference.

mlmc++ & deflation mlmc: compromise I



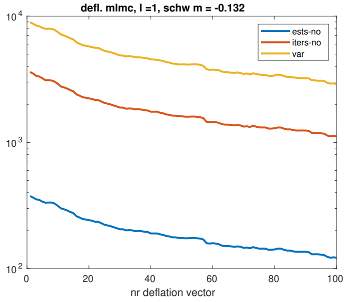


Figure: compromise of mlmc++ and def. mlmc for schwinger matrix at first level difference: samples, MG V-cycles, and the variance.

mlmc++ & deflation mlmc: compromise I

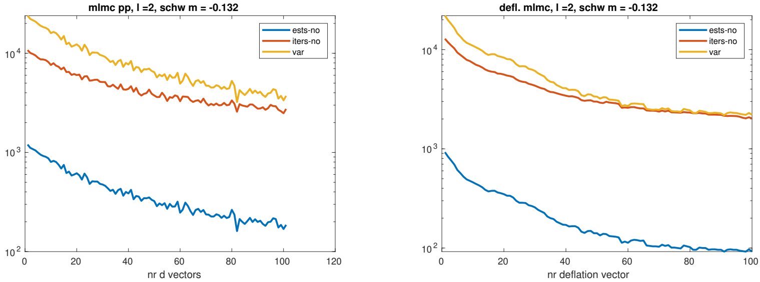


Figure: compromise of mlmc++ and def. mlmc for schwinger matrix at second level difference: samples, MG V-cycles, and the variance.

mlmc++ & deflation mlmc: compromise I

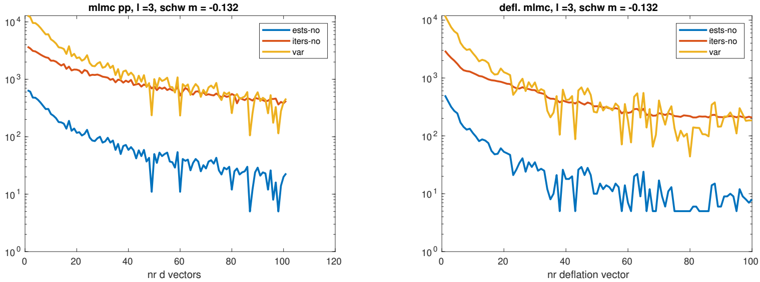


Figure: compromise of mlmc++ and def. mlmc for schwinger matrix at third level difference: samples, MG V-cycles, and the variance.

Content

Summery

- We exploreed another approach $\rightarrow$ reduction of the variance of $\text{tr}(A^{-1})$
- higher precision can be obtained at much less work.
- the optimal way to project eigenmodes is not yet reproduced

Outlook:

- turn into $4D$ problems of QCD.

Thank You for your attention!