# MLMC++ as a variance reduction method 

M. Khalil, A. Frommer

Wuppertal University<br>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

# Monte Carlo Trace Estimation 

## Variance Redcution Methods

MLMC approach
mlmc ++

## Conclusions

## Numerical Results

## Problem Statement

Consider the problem of computing

$$
\begin{equation*}
\operatorname{tr}(f(A)):=\sum_{i=1}^{n}[f(A)]_{i i} \tag{1}
\end{equation*}
$$

- 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 $\operatorname{tr}\left(A^{-1}\right)$ stochastically.


## Hutchinson approach: (Hutchinson, 1989)

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

$$
\begin{align*}
& x_{i} \in\{-1,1\} \text { with equal probabililty } \frac{1}{2},  \tag{2}\\
& x_{i} \text { is } N(0,1) \text { normally distributed. } \tag{3}
\end{align*}
$$

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

$$
\begin{equation*}
\operatorname{tr}\left(A^{-1}\right) \approx \frac{1}{s} \sum_{i=1}^{s} x_{i}^{*} A^{-1} x_{i} \tag{4}
\end{equation*}
$$

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

- the variance:

$$
\begin{equation*}
\mathbb{V}\left[x^{*} A^{-1} x\right]=\frac{1}{2}\left\|\operatorname{offdiag}\left(A^{-1}\right)\right\|_{F}^{2} \tag{5}
\end{equation*}
$$

## MC trace algorithm

1: Input: $A, x, \epsilon, m$
2: for $s=1$ to $m$ do
3: $\quad \tau_{i} \leftarrow x^{*} A^{-1} x$
4: if $\operatorname{Var}(\tau) / s \leq \epsilon$ then
5: stop
6: end if
7: end for
8: Output: mean of $\tau$

## 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 $a_{i j}$ large.

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

## Content

## MLMC approach

MLMC approach

## $\mathrm{mlmc}++$

,

Numerical Results
Numerical Results
Numerical Results



Variance Redcution Methods<br>Variance Redcution Methods

$\qquad$
$\square$

-
$\qquad$都
.....

Conclusions<br>Conclusions ? <br><br>

$\qquad$
$\qquad$ (
$\qquad$
$\qquad$
$\qquad$
Monte Carlo Trace Estimation
Monte Carlo Trace Estimation
Monte Carlo Trace Estimation . .
$\qquad$
$\square$

## Approximate/Inexact deflation

Projections $\rightarrow$ does not depend on eigenmodes

- Projection vector $\leftarrow V=W U^{*}$
- $W, U \in \mathbb{C}^{n \times d}$
- $\operatorname{tr}(A)=\operatorname{tr}(A(I-V))+\operatorname{tr}(A V)$
- use the cyclic trace property : $\operatorname{tr}(A V)=\operatorname{tr}\left(A W U^{*}\right)=\operatorname{tr}\left(U^{*} A W\right), \quad U^{*} A W \in \mathbb{C}^{d \times d}$

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

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


## Hutch ++ method

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

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


## Hutch++ algorithm

Compute Stochastic part: $\operatorname{tr}\left(A_{s}^{-1}\right)=\operatorname{tr}\left(A^{-1}(I-W)\right)$
1: Input: $A, S, d, m, x$
2: for $s=1 \rightarrow m$ do
3: $\quad \tau \leftarrow x^{*} W A^{-1} W x$
4: if $\operatorname{Var}(\tau) / s \leq \epsilon$ then
5: stop
6: end if
7: end for
8: Output: mean of $\tau$

Imporvments:

- easy and simple, reduce the accuracy dependence from $O\left(n^{2}\right) \rightarrow O(n)$.


## 

Content
Monte Carl
Variance R
MLMC app
mamet+
Numerical
Conclusion <br> \section*{\section*{MLMC approach <br> \section*{\section*{MLMC approach <br> <br>  <br> <br>  <br> <br>  <br> <br>  <br> <br>  <br> <br> $\qquad$ <br> <br> Monte Carlo} <br> \title{

} <br> $\qquad$ <br>  <br>  <br>  <br> \section*{} Monte} Monte}

Monte Carlo Trace Estimation
Variance Redcution Methods
MLMC approach Monte Carlo Trace Estimation
Variance Redcution Methods
MLMC approach
$m \min ++$
Content
Monte Carl
Variance R
MLMC app
mime+ +
Numerical
Conclusions
Content
Monte Carlo
Variance Rec
MLMC appr
mime ++
Numerical R
Conclusions

Numerical Results

Content
Monte Carlo
Variance Rec
MLMC appr
mlmc++
Numerical R
Conclusions
Content
Monte Carl
Variance R
MLMC app
mime ++
Numerical
Conclusions
Content
Monte Carl
Variance R
MLMC app
mime ++
Numerical
Conclusions
Content
Monte Carlo Trace Estimation
Variance Redcution Methods
MLMC approach
mlmc ++
Numerical Results
Conclusions
Content
Monte Carl
Variance R
MLMC app
mime ++
Numerical
Conclusions
Content
Monte Carla
Variance Re
MLMC app
mlmc++
Numerical
Conclusions
Carlo Tracerastina Monte Carlo Trace Estimation





```
*
``` Carlo Trace Estimation
\(\qquad\)
路


\(\qquad\)
\(\qquad\)
\(\qquad\)
 \(\qquad\)

        
         Carlo Trace Estimation
\(\square\)

\(\qquad\) \(\square\)
 \(+\) \(+\) ,
\(\qquad\)


 (
\(\qquad\)
\(\qquad\)
\(\qquad\)
\(\square\)
\(\square\)

\begin{abstract}
\(\qquad\)
\(\square\)
T
\end{abstract}
\(\square\)
. \(\square\)
\(\square\)
,
\(\qquad\)
\(\square\) (.) f
Content
Monte Carl
Variance R
MLMC app
mime ++
Numerical
Conclusions
Content
Monte Carl
Variance R
MLMC app
mime ++
Numerical
Conclusions

\begin{abstract}

\end{abstract}


\begin{abstract}

\end{abstract}

\section*{MLMC: main idea (Giles, 2015)}

Assume a random function \(q_{0}\) splits as
\[
\begin{gathered}
q_{0}=\sum_{\ell=1}^{L} q_{l}, \quad \ell \mathrm{nr} \text { of level differnce } \\
\mathbb{E}\left[q_{0}\right]=\sum_{\ell=1}^{L-1} \underbrace{\mathbb{E}\left[q_{\ell}-q_{\ell+1}\right]}_{=w_{\ell}}+\underbrace{q_{L}}_{=w_{L}}
\end{gathered}
\]
where \(w_{\ell}^{(i)}\) independent samples on each level.
In case: \(w=x^{*} A^{-1} x\), the unbiased estimator for \(\operatorname{tr}\left(A^{-1}\right)\) given by
\[
\frac{1}{N} \sum_{i=1}^{N} x^{(i)} A^{-1} x^{(i)} \approx \operatorname{tr}\left(A^{-1}\right)
\]

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

\section*{MLMC: setup phase}
- The goal: Solver \& Hierarchy of the linear system
- define Prolongation \(P_{l} \rightarrow l=0,1, \ldots, L\)
- define Restriction \(R_{l} \rightarrow l=0,1, \ldots, L\)
- usually \(\rightarrow R=P^{*}\)
- define coarse matrix \(\rightarrow B_{l+1}=R_{l} B_{l} P_{l}\)


Figure: MG V-cycle: setup phase

\section*{MLMC-stochastic phase}

\section*{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:
\[
\underbrace{x^{*} B x}_{q_{0}}=\sum_{\ell=0}^{L-1} \underbrace{x^{*}\left(\hat{P}_{\ell} B_{\ell} \hat{R}_{\ell}-\hat{P}_{\ell+1} B_{\ell+1} \hat{R}_{\ell+1}\right)}_{w_{\ell}} x+\underbrace{x^{*}\left(\hat{P}_{L} B_{L} \hat{R}_{L}\right) x}_{w_{L}}
\]

\section*{accuracy type}

Target accuracy \(\rho \leftarrow\) measured standard deviation

\section*{Uniform accuracy:}
- Idea: distribute the target accuracy in equal. Achieve: \(\rho_{\ell}=\rho / \sqrt{L-1}\) for all \(\ell\),

\section*{Optimized accuracy:}
- Idea: minimize the work if we know the cost \(C_{\ell} \ldots\) and the variance \(V_{\ell}\)
- for each sample: update \(C_{\ell}\) and measured \(V_{\ell}\).
- define the new optimal target tolerance \(\rho_{\ell}\)

\section*{Content \\ Content}
\(\square\)

Monte Carlo Trace Estimation
Variance Redcution Methods





mete Carlo Trace Estimation



\author{
Monte Carlo Trace Estimation
Variance Redcution Methods \\ \(\longrightarrow\) \\ \(\qquad\)
} Monte Carlo Trace
Variance Redcution
MLMC approach
mIme++ Monte Carlo Trace
Variance Redcution
MLMC approach
mIme++
Numerical Results

\author{
Conclusions
}

\author{
\(\qquad\)
}

Numerical Results

2
 Monte Carlo Trace
Variance Redcution
Maim approach
Mime+ Monte Carlo Trace Es
Variance Redcution
MLMC approach
mime+ +路



\[
2
\]
\(\qquad\)







\(\qquad\)
\(\qquad\) \(\downarrow\)

 \(\square\)
```

,

```
 (- \(\square\)
\(\qquad\)
\(\qquad\)
\(\qquad\)
\(\qquad\)









\section*{\(\mathrm{mlmc}++\)}
- Idea: MLMC \& Hutch++ together..
- The level difference matrix given as:
\[
\begin{equation*}
M_{l}=\hat{P}_{l} B_{l} \hat{R}_{l}-\hat{P}_{l+1} B_{l+1} \hat{R}_{l+1} \quad, \quad M_{l} \in C^{n_{l} \times n_{l}} \tag{6}
\end{equation*}
\]
- generate random matrix \(S_{\ell} \in C^{n_{l} \times d_{l}}, l=1, . ., L-1\)
- applying the projection vectors: \(Q_{l}=q r\left[M_{l}^{-1} S_{l}, 0\right]\),
- split the matrix level difference:
\[
\begin{equation*}
M_{l}=\left(M_{l}\right)_{d}+\left(M_{l}-\left(M_{l}\right)_{d}\right) \tag{7}
\end{equation*}
\]
- trace of multilevel decomposition of \(M_{l}\) :
\[
\begin{equation*}
\operatorname{tr}\left(M_{l}^{-1}\right)=\operatorname{tr}(\underbrace{Q_{l}^{*} M_{l} Q_{l}}_{\text {directly }})+\operatorname{tr}(M_{l} \underbrace{\left(I-Q_{l}^{*} Q_{l}\right)}_{\text {stochastically }}) \tag{8}
\end{equation*}
\]
- challenge: find an optimal way to determine \(d_{l}\) to reduce the variance enough

\section*{Content}

\author{
Variance Redcution Methods \\ MLMC approach \\ mime ++ \\ ML MC \\ MC \\  \\ \(\qquad\) \\  \\ \title{

} \\  \\  \\ 都 \\ mime ++
}

Numerical Results
Conclusions

\title{
Monte Carlo Trace Estimation
}
\(\square\)

\author{

}


\author{

}

五

        
    (

\section*{Schwinger model: Table}
\begin{tabular}{|r|c|cccc|c|}
\hline \multicolumn{7}{|c|}{ Schwinger model } \\
\hline\(N\) & & \(\ell=1\) & \(\ell=2\) & \(\ell=3\) & \(\ell=4\) & \(L\) \\
\hline 128 & \(n_{\ell}\) & \(2 \cdot 128^{2}\) & \(4 \cdot 32^{2}\) & \(8 \cdot 8^{2}\) & \(8 \cdot 2^{2}\) & 4 \\
& \(\mathrm{nnz}\left(S_{\ell}^{N}\right)\) & \(2.94 e 5\) & \(1.64 e 5\) & \(2.46 e 4\) & 1024 & \\
\hline \hline\(m\) & -0.1320 & -0.1325 & -0.1329 & -0.1332 & -0.1333 & \\
\(n_{\text {defl }}\) & 384 & 384 & 512 & 512 & 512 & \\
\hline
\end{tabular}

Table: Parameters and quantities for Schwinger example

\section*{schwinger accuracy: Figures}



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

\section*{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).

\section*{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.

\section*{\(\mathrm{mlmc}++\&\) deflation \(\mathrm{mlmc}:\) compromise I}


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

\section*{\(\mathrm{mlmc}++\) \& 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.

\section*{\(\mathrm{mlmc}++\&\) deflation \(\mathrm{mlmc}:\) compromise I}



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

\section*{Content}

\author{
Monte Carlo Trace Estimation \\ \section*{Variance Redcution Methods} \\ MLMC approach \\ \section*{\(\mathrm{mlmc}++\) \\ \\ } \\ Numerical Results \\ Conclusions
}
- We exploreed another approach \(\rightarrow\) reduction of the variance of \(\operatorname{tr}\left(A^{-1}\right)\)
- higher precision can be obtained at much less work.
- the optimal way to project eigenmodes is not yet reproduced

\section*{Outlook:}
- turn into \(4 D\) problems of QCD .

\section*{Thank You for your attention!}```

