Multilevel Approach to Stochastic Estimation of the Trace

M. Khalil & (A. Frommer & G. RAMIREZ)

UNIFE & WUP & CYI

November 22, 2021

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

STIMULATE
European Joint Doctorates

Outline:

- Problem Statement
- Hutchinson approach
- Deflation approach
- MLMC approach
- Numerical Results
- Conclustions

Content

Problem Statement

Hutchinson approach

Deflation approach

MLMC approach

Numerical Results

Conclusions

Problem Statement

Consider the problem of computing

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

- 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).
- standard approach to estimate $\text{tr}(A^{-1})$ stochastically is Hutchinson's method.

Content

Problem Statement

Hutchinson approach

Deflation approach

MLMC approach

Numerical Results

Conclusions

Hutchinson approach I

- Assume: a vector $x \in \mathbb{C}^n$ with probability:

  1. x_i i.i.d random variables, normally distributed
  2. $x_i \in \{-1, 1\}$ with equal probability $\frac{1}{2}$

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

$$\text{tr}(A^{-1}) \approx \frac{1}{s} \sum_{i=1}^{s} x_i^* A^{-1} x_i. \tag{2}$$

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

- the variance:

$$\mathbb{V}[x^* A^{-1} x] = \frac{1}{2} \|\text{offdiag}(A^{-1} + (A^{-1})^*)\|_F^2. \tag{3}$$

- the accuracy $\sim \frac{1}{\sqrt{s}} \times$ RMSD

Hutchinson approach II

Properties:

- simple, requires a solver for A^{-1}.
- **variance** $\rightarrow$ very large when a_{ij} large.
- accuracy $\sim \frac{1}{\sqrt{s}} \rightarrow s$ too much, slow convergence.
- cost of MC simulation $\rightarrow$ very expensive.

Improvements:

Reduce variance through:
- Deflation (known) approach,
- MLMC (new) approach.

Content

Problem Statement

Hutchinson approach

Deflation approach

MLMC approach

Numerical Results

Conclusions

deflation I

- suppose $A \in \mathsf{C}^{n\times n}$ a non-singular matrix

$$A = U\Sigma V^* \tag{4}$$

with $\rightarrow \quad U, \Sigma, V \in \mathsf{C}^{n\times n}, U^*U = V^*V = I$

- Then For the off-diagonal of A^{-1} we have

$$\|\text{offdiag}(A^{-1})\|_F^2 = \sum_{i=1}^{n} \sigma_i^{-2} - \sum_{i=1}^{n} |(A^{-1})_{ii}|^2. \tag{5}$$

- **problem:** small eigenvalues of $A \rightarrow$ increasing the variance for $\mathsf{tr}(A^{-1})$.
- **solution:** multiplying A^{-1} in a proper projection P as:

$$A = \mathbb{P}A + (I - \mathbb{P})A$$

deflation II

Assume an orthogonal projector $\mathbb{P} \in \mathbb{C}^{s \times s}$ as:

$$P = V_s V_s^* \quad \text{where} \quad V_s = [v_1|...|v_s]$$

The trace inverse is given by:

$$\text{tr}(A^{-1}) = \text{tr}(A^{-1}(I - P)) + \text{tr}(A^{-1}P)$$

1) $\text{tr}(A^{-1}(I - \mathbb{P})) \rightarrow$ deflated smallest eigenpairs (stochasically).

2) $\text{tr}(A^{-1}P) = \text{tr}((V_s^* A V_s)^{-1}) \quad \in \mathbb{C}^{s \times s} \rightarrow$ directly.

- **Properties:**
- variance is decreased
- requires singular eigenpairs $\rightarrow$ still quite costly

Content

Problem Statement

Hutchinson approach

Deflation approach

MLMC approach

Numerical Results

Conclusions

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 $\text{tr}(A^{-1})$ given by

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

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

MLMC: for $\mathrm{tr}(A^{-1})$ **setup phase**

using **MG Hierarchy**, for number of levels L we have:

- sequence of matrices

$$A_\ell \in \mathbb{C}^{n_\ell \times n_\ell}, \ell = 1, \dots, L, \quad \underbrace{A = A_1}_{\text{original matrix}} \in \mathbb{C}^{n_1 \times n_1}$$

- use prolongation and restriction operators $\rightarrow$ transfer data between the levels:

$$P_\ell \in \mathbb{C}^{n_\ell \times n_{\ell+1}}, \quad R_\ell \in \mathbb{C}^{n_{\ell+1} \times n_\ell},\ \ell = 1, \dots, L-1.$$

- use Petrov-Galerkin approach $\rightarrow$ construct coarse system matrices A_ℓ

$$A_{\ell+1} = R_\ell A_\ell P_\ell,\ \ell = 1, \dots, L-1.$$

- Using the accumulated prolongation and restriction operators

$$\hat{P}_\ell = P_1 \cdots P_{\ell-1} \in \mathbb{C}^{n \times n_\ell}, \quad \hat{R}_\ell = R_{\ell-1} \cdots R_1 \in \mathbb{C}^{n_\ell \times n}, \quad \ell = 1, \dots, L,$$

- where we put $\hat{R}_1 = \hat{P}_1 = I \in \mathbb{C}^{n \times n}$ by convention, $\rightarrow \hat{P}_\ell A_\ell^{-1} \hat{R}_\ell$ as the approximation to A^{-1} at level ℓ.

MLMC: trace estimator I

then the trace of the multilevel difference as:

$$\text{tr}(A^{-1}) = \sum_{\ell=1}^{L-1} \text{tr} \underbrace{\left(\hat{P}_\ell A_\ell^{-1} \hat{R}_\ell - \hat{P}_{\ell+1} A_{\ell+1}^{-1} \hat{R}_{\ell+1}\right)}_{\text{differences}} + \text{tr} \underbrace{(\hat{P}_L A_L^{-1} \hat{R}_L)}_{\text{coarsest}}. \qquad (6)$$

Note: compute coarsest term directly as

$$\text{tr}(\hat{P}_L A_L^{-1} \hat{R}_L) = \text{tr}(A_L^{-1} \hat{R}_L \hat{P}_L).$$

In some cases $\hat{R}_\ell \hat{P}_\ell = I \in \mathbb{C}^{n_\ell \times n_\ell}$, Then

$$\text{tr}(\hat{P}_\ell A_\ell^{-1} \hat{R}_\ell) \quad = \quad \text{tr}(A_\ell^{-1} \hat{R}_\ell \hat{P}_\ell) = \text{tr}(A_\ell^{-1}),$$

This means that instead of the multilevel decomposition (6) we can use

$$\text{tr}(A) = \sum_{\ell=1}^{L-1} \left(\text{tr}(A_\ell^{-1}) - \text{tr}(P_\ell A_{\ell+1}^{-1} R_\ell)\right) + \text{tr}(A_L^{-1}),$$

mlmc: trace estimator II

The unbiased mlmc trace estimator is then

$$\text{tr}(A^{-1}) \approx \sum_{\ell=1}^{L-1}\sum_{i=1}^{N_\ell}\Big((x^{i,\ell})^*\hat{P}_\ell A_\ell^{-1}\hat{R}_\ell x^{i,\ell} - (x^{i,\ell})^*\hat{P}_{\ell+1}A_{\ell+1}^{-1}\hat{R}_{\ell+1}x^{i,\ell}\Big)$$
$$+\sum_{i=1}^{N_L}(x^{n,L})^*\hat{P}_L A_L^{-1}\hat{R}_L x^{i,L},$$

where the vectors $x^{i,\ell} \in \mathsf{C}^n$ are s.i samples of $x \in \mathbb{C}^n$

Accuracy:

- we aim at a relative accuracy of $\epsilon = 10^{-3}$.
- perform 5 stochastic estimates, compute the value τ (difference between mean and RMSD)
- The requirement is to have
$$\sum_{\ell=1}^{L-1}\rho_\ell^2 = (\epsilon\tau)^2,$$
- stopping criteria given as: $\rho_\ell = \epsilon\tau/\sqrt{L-1}$ for all ℓ,

Content

Problem Statement

Hutchinson approach

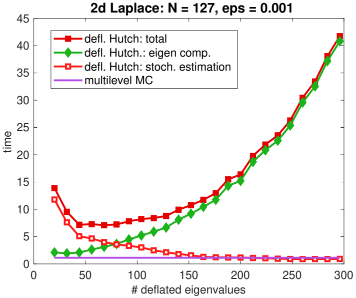
Deflation approach

MLMC approach

Numerical Results

Conclusions

2d Laplace: rough scan



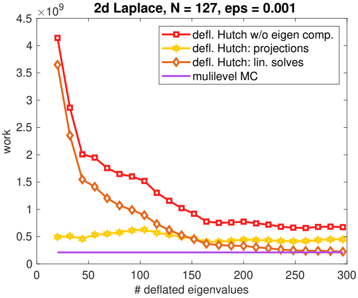
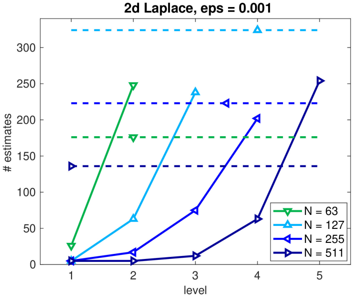


Figure: 2d Laplace: Comparison of time and cost for MLMC and defl. Hutchinson, varying n_{defl}.

2d Laplace									
N		$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	n_{defl}	L
63	n_ℓ	63^2	31^2	15^2				92	3
	$\text{nnz}(L_\ell^N)$	1.96e4	8.28e3	1.85e3					
127	n_ℓ	127^2	63^2	31^2	15^2			44	4
	$\text{nnz}(L_\ell^N)$	8.01e4	3.50e4	8.28e3	1.85e3				
255	n_ℓ	255^2	127^2	63^2	31^2	15^2		64	5
	$\text{nnz}(L_\ell^N)$	3.24e5	1.44e5	3.50e4	8.28e3	1.85e3			
511	n_ℓ	511^2	255^2	127^2	63^2	31^2	15^2	76	6
	$\text{nnz}(L_\ell^N)$	1.30e6	5.82e5	1.44e5	3.50e4	8.28e3	1.85e3		

Table Parameters and quantities for 2d Laplace.

2d Laplace: Figures



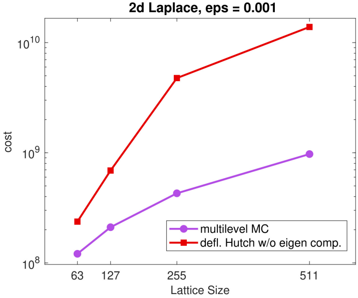


Figure: Comparison of mlmc and defl. Hutch. for 2d Laplace matrices: no of stoch. estimates on each level difference and total work for different N.

Gauge Laplace: Table

Gauge Laplace							
N		$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	n_{defl}	L
64	n_ℓ	4096	1354	134		60	3
	$\text{nnz}(G_\ell^N)$	20480	24900	3172			
128	n_ℓ	16384	5440	554		60	3
	$\text{nnz}(G_\ell^N)$	81920	99448	11300			
256	n_ℓ	65536	21802	2348	196	20	4
	$\text{nnz}(G_\ell^N)$	327680	394628	49416	6352		
512	n_ℓ	262144	87296	9562	924	20	4
	$\text{nnz}(G_\ell^N)$	327680	394628	49416	6352		

Table: Parameters and quantities for Gauge Laplace Example

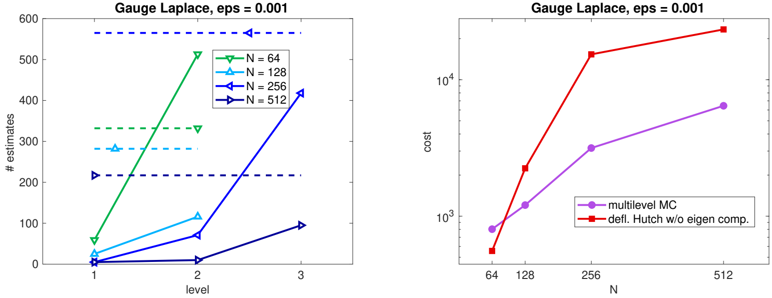


Figure: Comparison of mlmc and defl. Hutch. for the gauge Laplace matrices: no of stoch. estimates on each level difference and total work for different N.

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

Schwinger Model: Figures

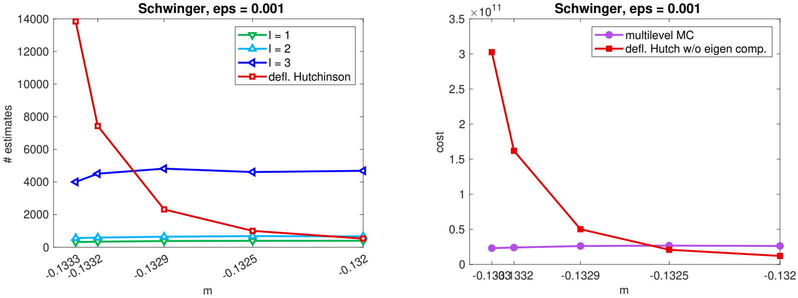


Figure: mlmc and defl. Hutch. for the Schwinger matrix: no of stoch. estimates on each level difference and total work for different masses m.

Content

Problem Statement

Hutchinson approach

Deflation approach

MLMC approach

Numerical Results

Conclusions

Summery

- A novel approach developed $\rightarrow$ substantial reduction of the variance of $\text{tr}(A^{-1})$
- higher precision can be obtained at much less work.
- We illustrate this by using three different classes of matrices.

Outlook:

- turn into $4D$ problems of QCD.
- Apply the mlmc approach on Hutch $++$

Thank You for your attention!