# Multilevel Probing Methods for Approximating Selected Entries of the Matrix Inverse

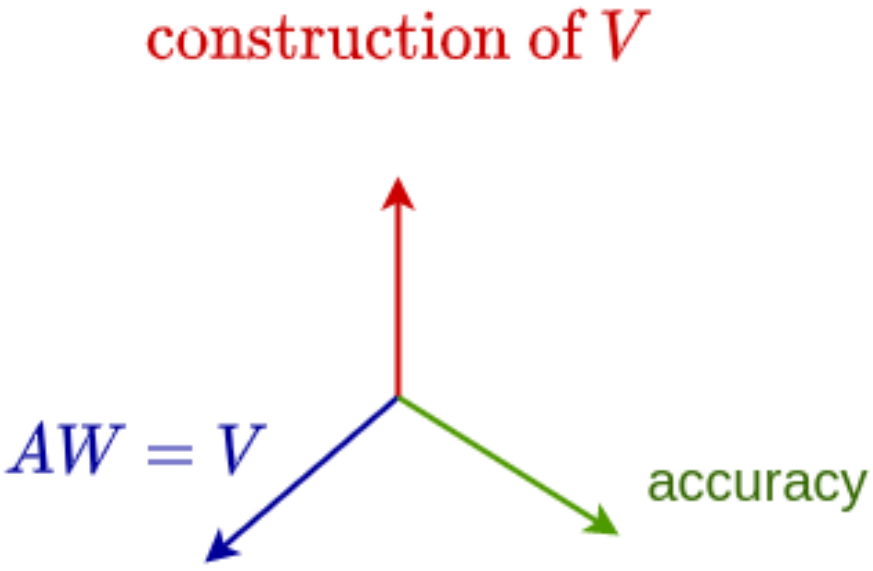
Vasileios Georgiou, Efstratios Gallopoulos

Computer Engineering Department
University of Patras

SIAM Conference on Applied Linear Algebra
May 20, 2021

# Goals

Approximate diag($A^{-1}$) via MV products with orthogonal basis $V$




1. memory and #operations for constructing $V \downarrow$ and #RHS $\uparrow$
2. approximation of the diagonal blocks and off-diagonal entries
3. solvers for MRHS & sequences of linear systems

*Numerical library for large scale computation for manycore + GPUs

# The Selected Inversion Problem

Queries (Q) on entries of $A^{-1}$

**Exact Selection**

$Q_1$: For $A \in \mathbb{C}^{n \times n}$ and subset $Z$, return $(A^{-1})_{ij}$ for all $(i,j) \in Z$

**Approximate Selection:** First define $T_Z : \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$

$$(T_Z(A))_{ij} = \begin{cases} (A)_{ij}, & \text{if } (i,j) \in Z \\ 0, & \text{otherwise} \end{cases}$$

$Q_2$: For $n \times n$ matrix $A$ and the index subset $Z$, return $(B)_{ij}$ for all $(i,j) \in Z$ such that $\|B - T_Z(A^{-1})\| \leq \epsilon$

**Special Case:** $\text{diag}(A^{-1})$

# Applications

- Quantum transport simulations




- Uncertainty quantification

$$\min \|Yz - b\|_2 \longrightarrow (C^{-1})_{ii} \text{ confidence on i-th sample}$$

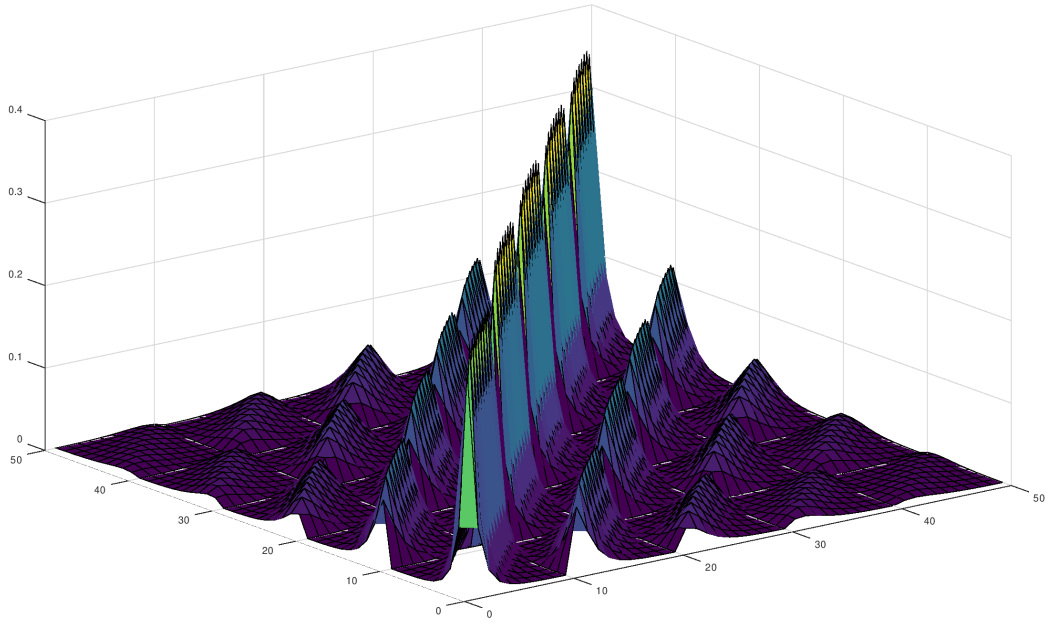
- Covariance estimation

$$f_\lambda(\Theta) = \underbrace{-\log(\det(\Theta)) + \mathrm{tr}(S\Theta)}_{\text{log-likelihood}} + \underbrace{\lambda|\Theta|_1}_{\text{regularizer}}$$

- Network analysis

$$f(A) = \alpha_0 I + \alpha_1 A + \alpha_2 A^2 + \ldots \alpha_{p-1} A^{p-1} \text{ where } A = I - \alpha G$$

# Localization




*"Sparsity does not take into account the size of the nonzeros, all being treated as equal"* *M. Benzi*

$$\|A^{-1} - T_\delta(A^{-1})\| \leq \epsilon \text{ and } \mathrm{nnz}(T_\delta(A^{-1})) \leq N_\delta$$

$T_\delta$: thresholding operator, $N_\delta$: maximum number of nonzeros

**Probing methods** utilize significant structure $S(T_\delta(A^{-1}))$

# Extracting diag($A^{-1}$) with Probing Vectors

Probing method by [Tang and Saad, 2011] motivated by [Coleman and More, 1983] proposed method for diag($A^{-1}$)

$$S(T_\delta(A^{-1})) \approx (\wedge_{i=1}^{m} S(A))$$

probing vectors $v_i$ tailored to extract diag($A^{-1}$) generated by coloring graph with adj.matrix $\wedge_{i=1}^{m} S_A$

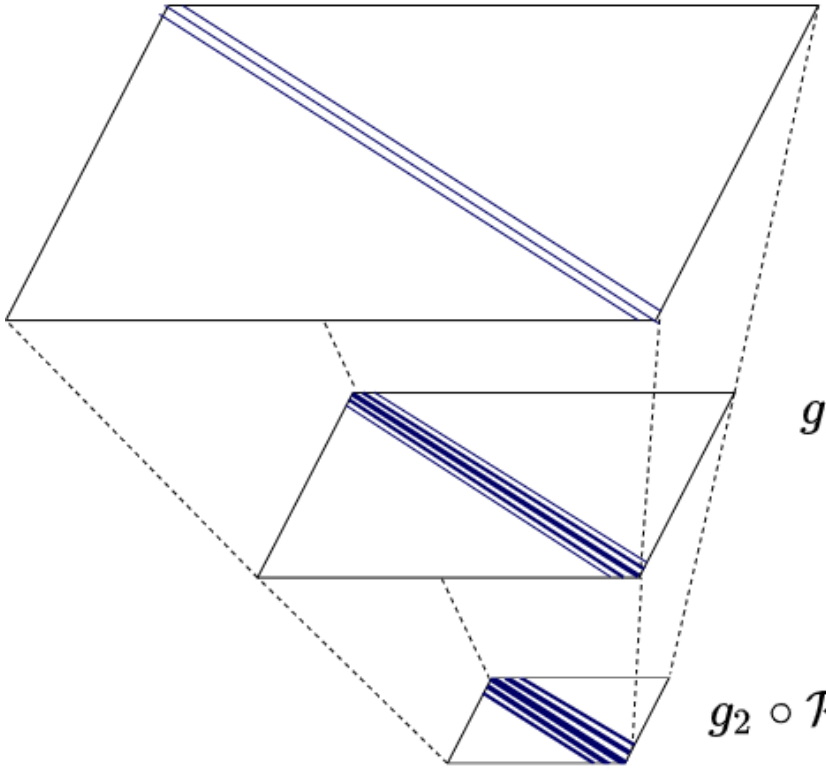
$$diag(A^{-1}) \approx \sum_{i=1}^{s} v_i \odot (A^{-1} v_i)$$

Error depends on entries not captured by $\wedge_{i=1}^{m} S(A)$

+ **Exact result for $A^{-1}$ with the same sparsity pattern as $A^k$**
- **Requires computing and storing $\wedge_{i=1}^{m} S_A$, m can be large**

Related methods [Stathopoulos, Laeuchli, orginos 2013], [Laeuchli, Stathopoulos, Orginos 2020]

# Multilevel Probing



$S(A)$

$S(A)$ : nonzero patter of $A$
$R_i$ : coarsening operator
$g_i$ : expansion operator

$g_1 \circ \mathcal{R}_1(S(A))$

$g_2 \circ \mathcal{R}_2(g_1 \circ \mathcal{R}_1(S(A)))$

$$S(T_\delta(A^{-1})) \approx \underbrace{\wedge_{i=1}^{k} S(A)}_{\text{probing for } k=\xi_1\xi_2} \approx \overbrace{P}^{\text{prolong}} \circ \Bigg( \underbrace{\wedge_{i=1}^{\xi_2}(\mathcal{R}_2 \circ (\underbrace{\wedge_{i=1}^{\xi_1}(R_1 \circ S(A))}_{\text{level 1}}))}_{\text{level 2}} \Bigg)$$

# General Algorithm

**Algorithm 1** Multilevel Probing

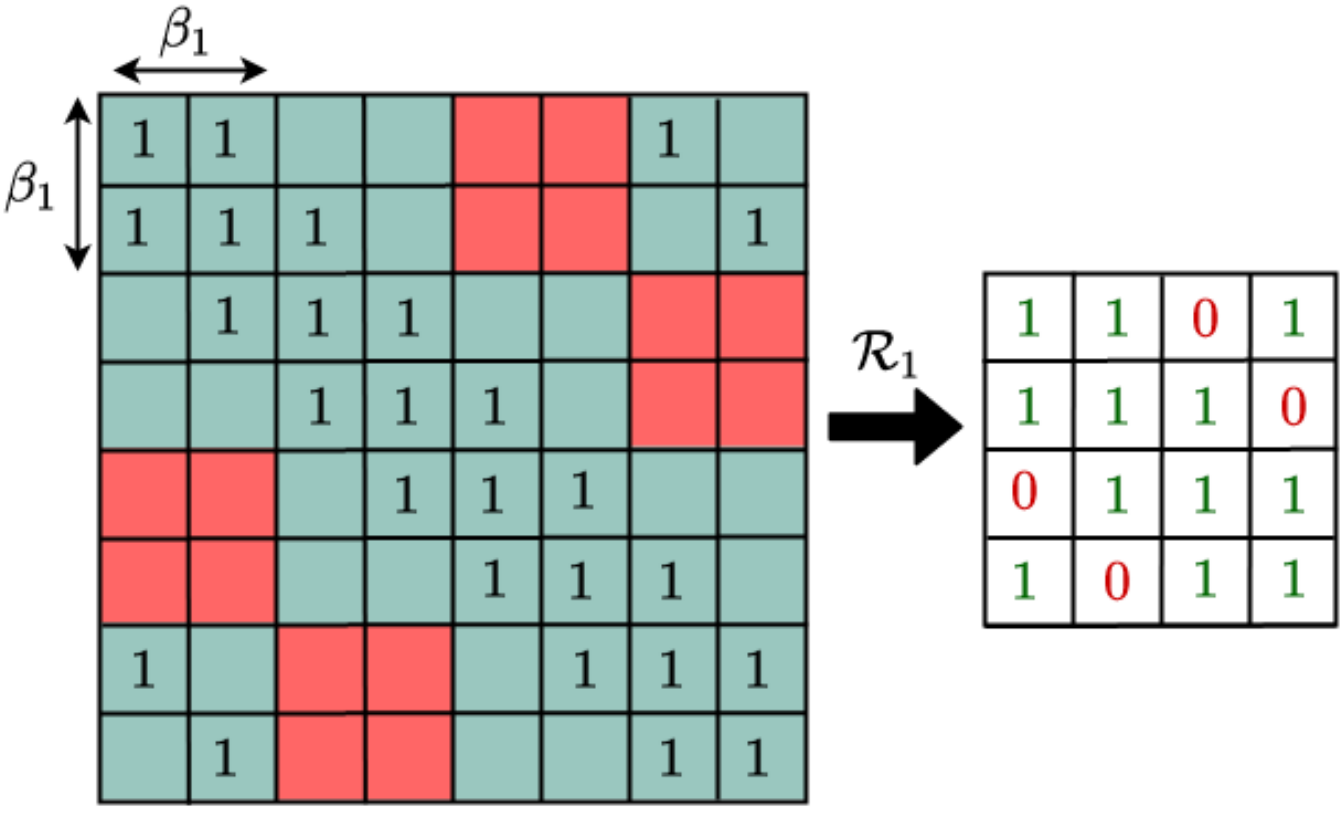
**Input:** $A \in \mathbb{C}^{n \times n}$, $\beta$, $k$, $m$

**Output:** $d$

1: Initialize $S_0 = S(A)$, $c = [\,]$
2: **for** $i = 1$ to $p$ **do**
3: compute $\hat{S}_k \leftarrow R_i \circ S_{k-1}$ // coarsening
4: compute $S_k \leftarrow \wedge_{i=1}^{\xi_i} \hat{S}_k$ // expansion
5: $c \leftarrow \text{color}(S_p)$
6: generate right-hand sides V // vector-prolongation
7: solve $AW = V$
8: compute $d = \sum_{i=1}^{s\beta} W_i \odot V_i$

pros: Preserve $\text{nnz}(S_k) \leq \text{nnz}(S_0) = \text{nnz}(S(A))$ for all $k$
cons # RHS $\uparrow \Rightarrow$ amount of time spent on solving linear systems

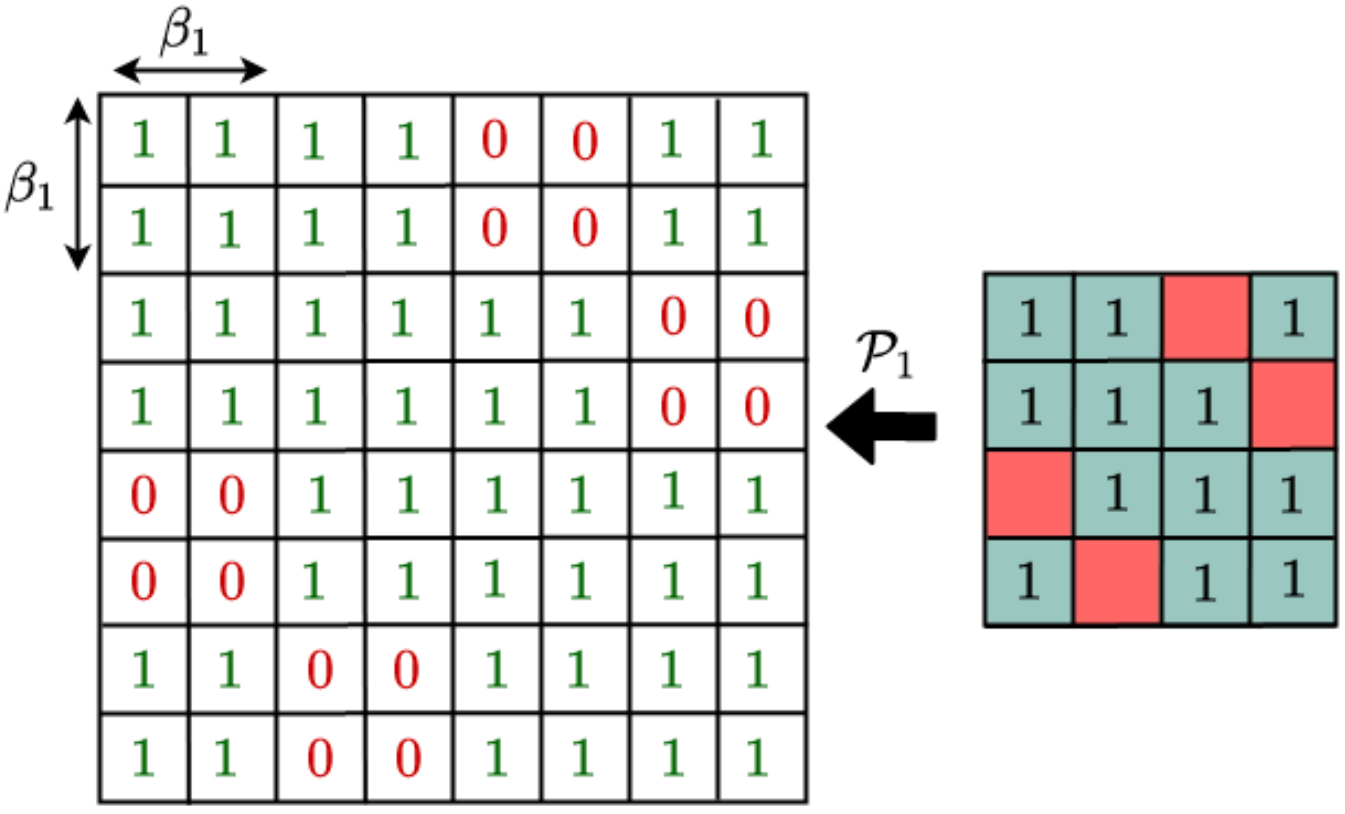
# Coarsening Operator - Blocking




**coarsening operator** $\mathcal{R} : \{0,1\}^{n \times n} \to \{0,1\}^{n/\beta_1 \times n/\beta_1}$

$$(R(A))_{ij} = \begin{cases} 1 \text{ if } \; A_{IJ} \neq 0_{\beta \times \beta} \\ 0 \text{ else} \end{cases}$$

# Prolongation Operator - Blocking




**prolongation operator** $\mathcal{P} : \{0,1\}^{n/\beta_1 \times n/\beta_1} \to \{0,1\}^{n \times n}$

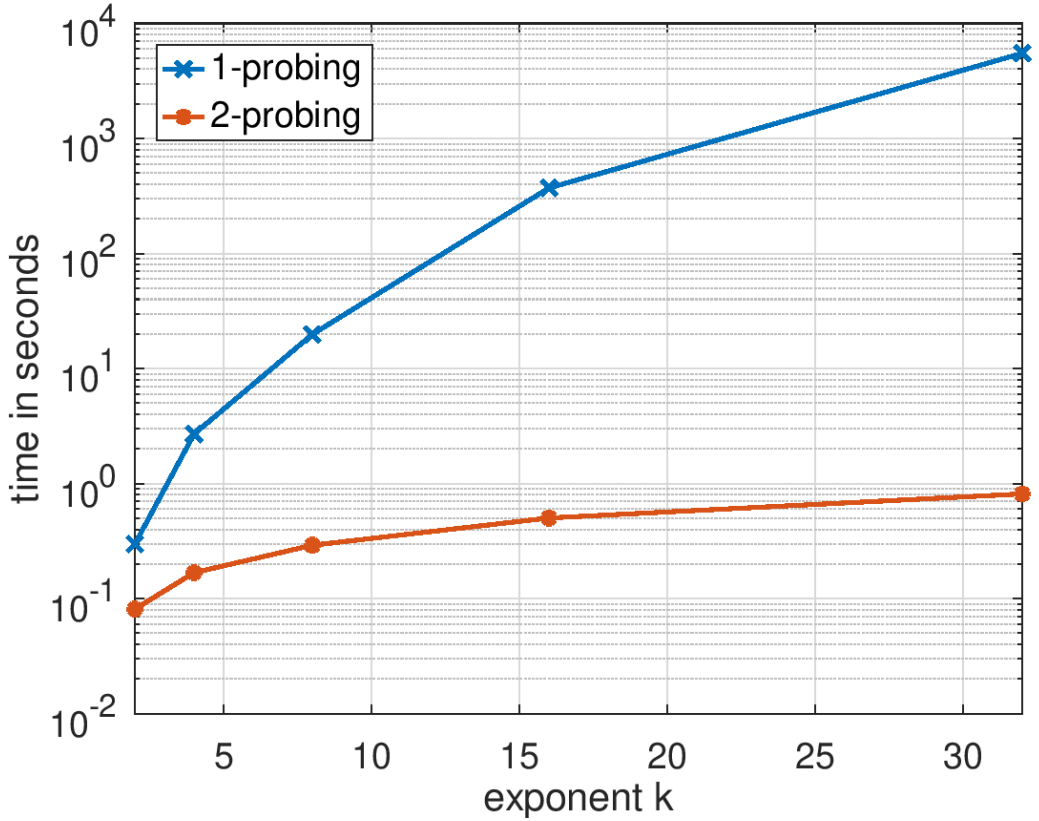
$$\mathcal{P}(A) = A \otimes ee^\top, \; (P(A))_{ij} = \begin{cases} 1 \\ 0 \end{cases}$$

# Block-probing vs Probing - probing runtimes

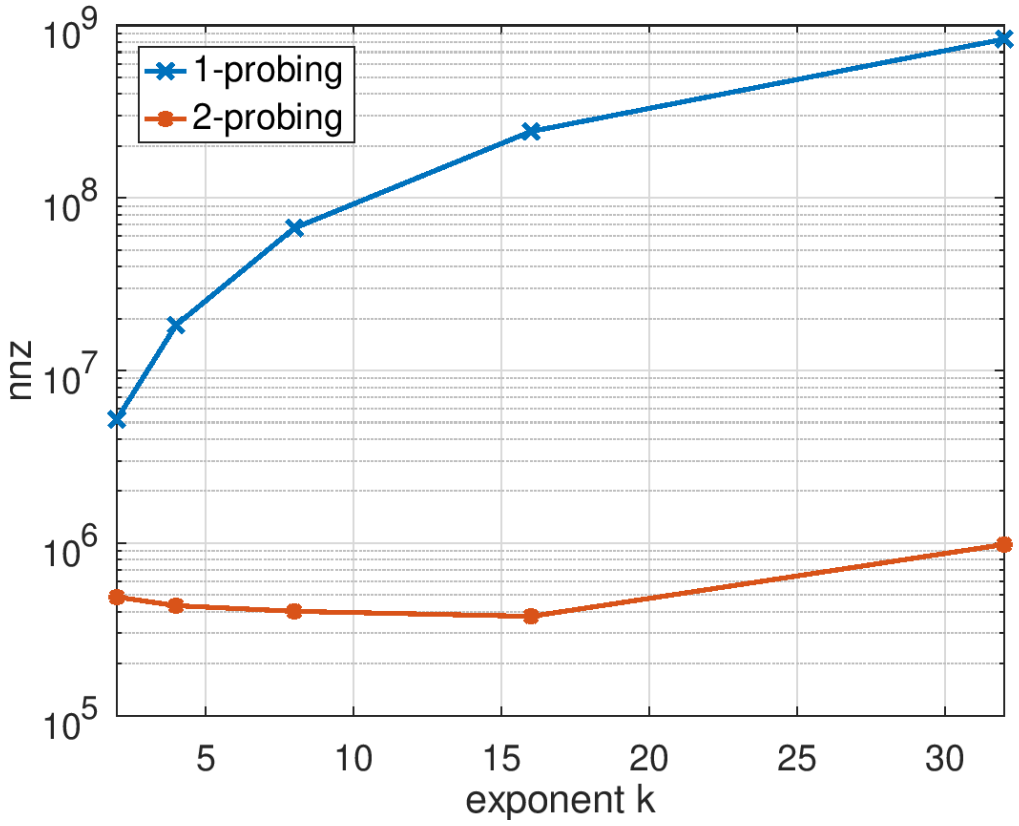
65356 $\times$ 65356 covariance matrix generated from

$$(A)_{ij} = \begin{cases} \left(1 - \frac{\text{dist(i, j)}}{\alpha}\right)^{\beta}, & \text{if dist}(i,j) \leq \alpha \\ 0, & \text{otherwise} \end{cases}$$

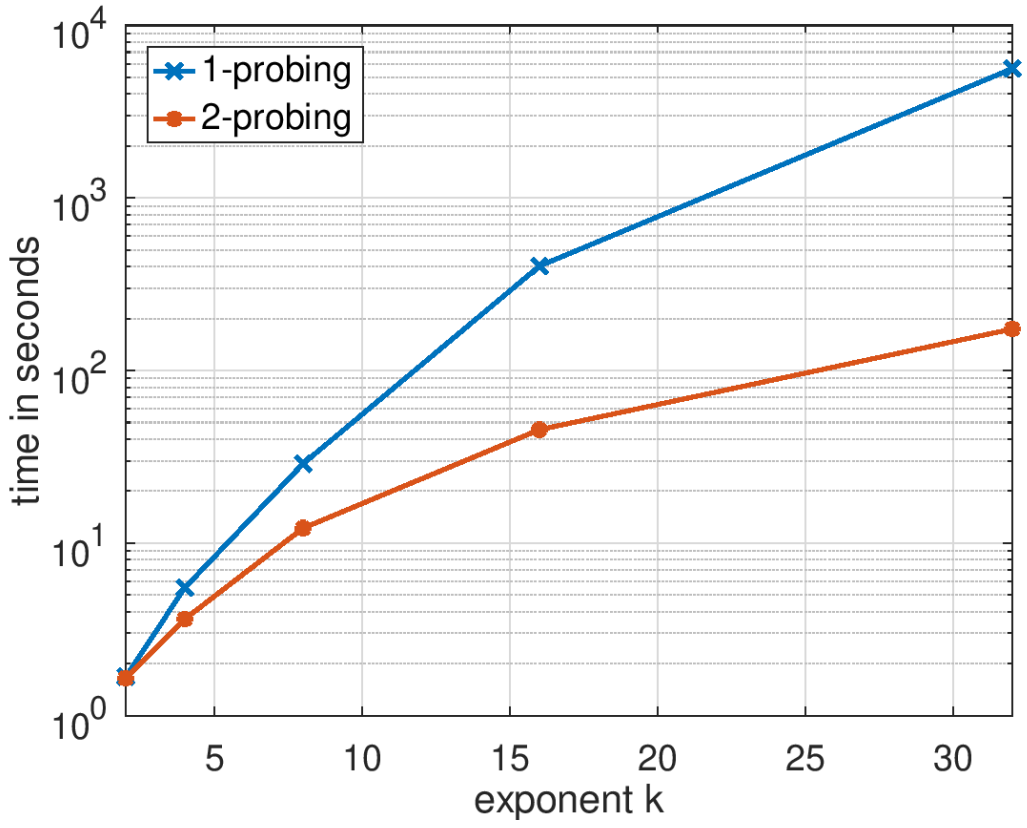
for $\alpha = 3$, $\beta = 5$. Use PCG with Jacobi preconditioning & 25 iterations



# Block-probing vs Probing - number of nonzeros



# Block-probing vs Probing - total runtimes




Matrices with localized inverse usually expected to be well conditioned. Several results suggest that decay rate $\sim \kappa(A)$

# Error

Let vectors $d, \hat{d}, d^*$ be

- $d \leftarrow \text{diag}(A^{-1}) = (A^{-1} \odot I)e$: exact values
- $\hat{d} \leftarrow (A^{-1}V)e$, as computed by numerical solver
- $\tilde{d} \leftarrow (A^{-1}V)e$, for exact solution of $AW = V$

Relative error $\mathcal{E}$ bounded as follows:

$$\mathcal{E} = \left\| (\hat{d} - d) \oslash d \right\|_\infty \leq \underbrace{\left\| (\hat{d} - \tilde{d}) \oslash d \right\|_\infty}_{\mathcal{E}_{\text{solve}}} + \underbrace{\left\| (\tilde{d} - d) \oslash d \right\|_\infty}_{\mathcal{E}_{\text{probe}}}$$

Necessary to bound both terms and consider tradeoffs

$$\mathcal{E}_{\text{probe}} = \left\| [(A^{-1} \odot VV^\top)e] \oslash [(A^{-1} \odot I)e] - e \right\|_\infty$$

# Block-probing vs Probing - Error

For symmetric $n \times n$ matrix $A$ and hierarchy $H(\beta, k, p) = \{S^{(i)}_{\beta,k}\}$ generated by the block coarsening

$$\mathcal{E}_{\text{probe}}(S^{(p)}_{k,\beta}) \leq \mathcal{E}_{\text{probe}}(\wedge_{i=1}^{k^p} S(A))$$

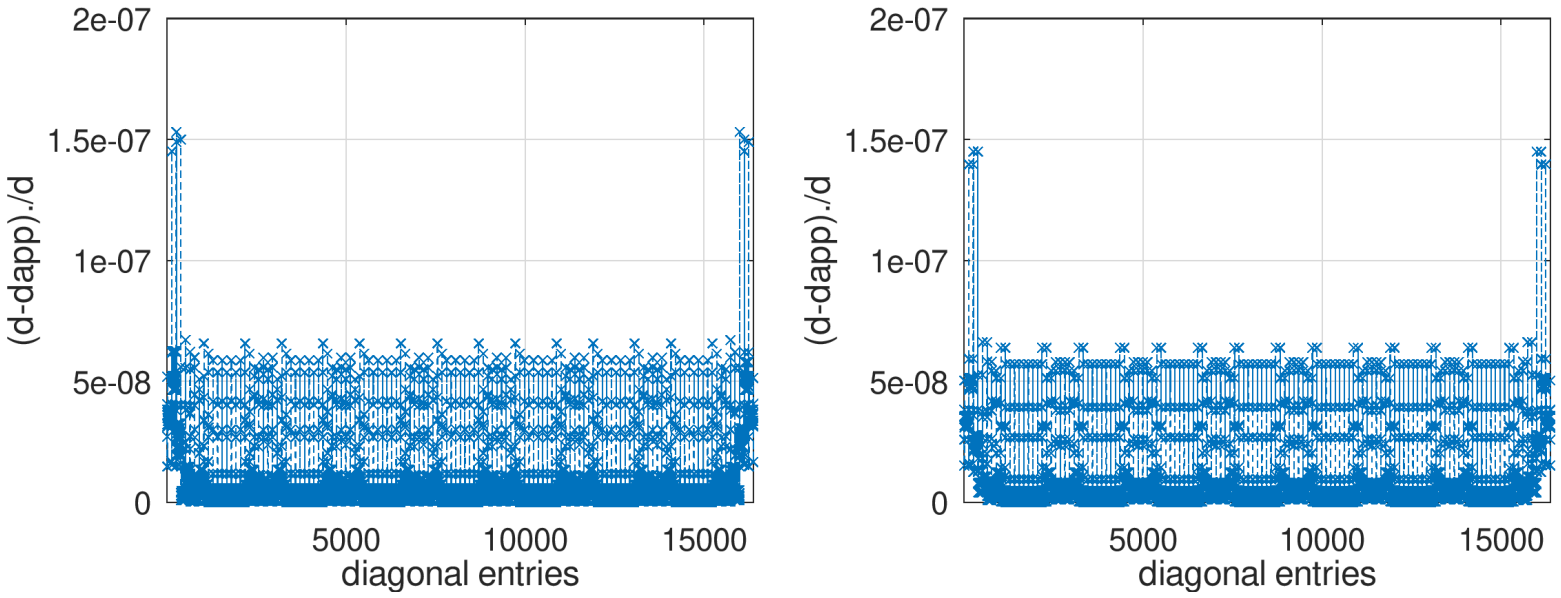



Figure: On the left: (1, 3)-probing. On the right (2, 3)-probing

## Theorem (Benzi and Razouk)

*Let $A \in \mathbb{C}^{n \times n}$ a diagonalizable matrix and $f(A)$ exhibits decay away from an nonzero pattern described by*

$$|(f(A))_{ij}| \leq \kappa(X) C \lambda^{dist(i,j)} \text{ where}$$

$0 \leq \lambda < 1$, $\kappa(X)$ *the condition number of the diagonalizer matrix $X$ and $C$ a constant.*

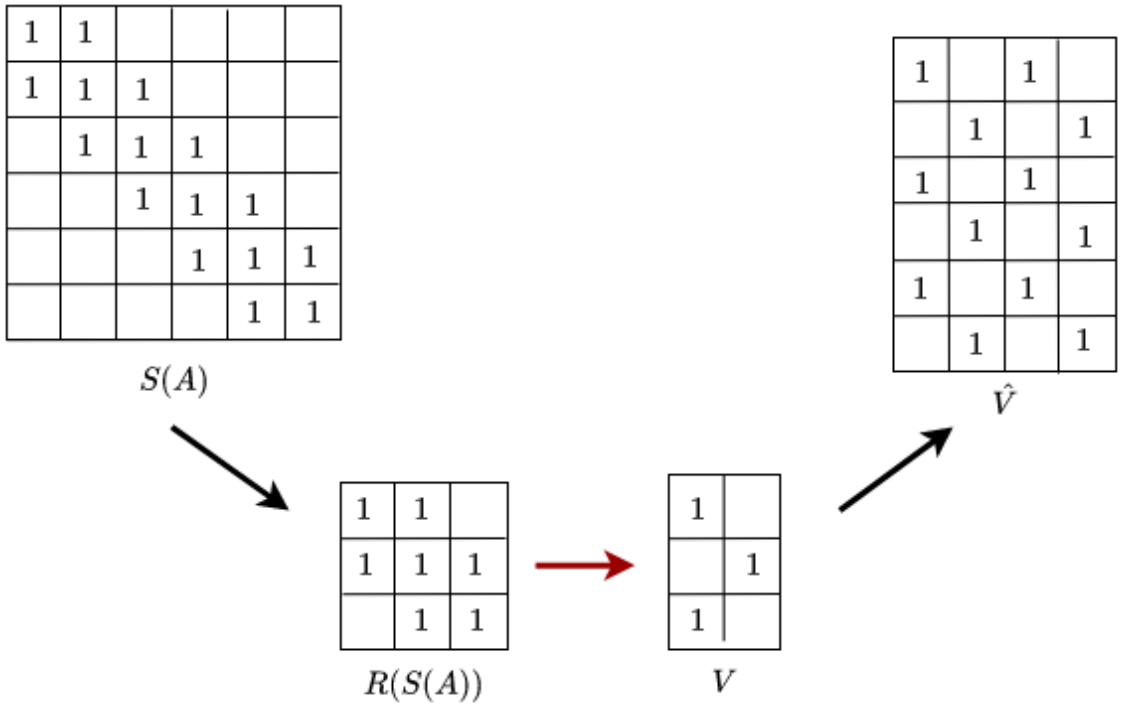
If the number of z-distant neighbors of every node of $G_A$ is given by $N_z = \alpha z$ for constant $\alpha$ and all $z = 1, \ldots, \text{diam}(G_A)$, and probing based on $\wedge_{i=1}^{p} S(A)$ , then

$$\mathcal{E}_{\text{probe}} \leq \kappa(X) C \frac{(1+p)\lambda^{p+1} - p\lambda^{p}}{|f_{\text{min}}|(1-\lambda)^2} \text{ where}$$

$f_{\text{min}}$, $C$ and $\lambda$ follow from the theorem by [Benzi and Razouk].

# Approximating the diagonal blocks of $A^{-1}$

For $A$ with sparsity structure $S(A)$




The diagonal blocks can be extracted exatly from the equation

$$d_\beta(A) = \sum_i (A^{-1} V_i) \bullet V_i \text{ where } \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \bullet \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 Y_1 \\ X_2 Y_2 \end{bmatrix}$$

# Approximating off-diagonal entries (I)

**Proposition (Coleman & More, 1983)**

*All nonzeros $A$ are contained in columns of $W = AV$. Vectors $V_{:,i}$ formed by coloring graph with adjacency matrix $\wedge_{i=1}^{2} S(A)$*

Since $\wedge_{i=1}^{k} S(A) \approx S(T_\delta(f(A)))$ and

$$\wedge_{i=1}^{2}(\wedge_{j=1}^{k} S(A)) \odot \left[ P_{\beta^{k+1}} \circ \overbrace{(\underbrace{\wedge_{i=1}^{2}(R_\beta}_{\text{extra step}} \circ \underbrace{S_{\beta,k}^{p}}_{p \text{ level}}))}^{G} \right] = \wedge_{i=1}^{2}(\wedge_{j=1}^{k} S(A))$$

$$V \leftarrow \text{color}(G)$$

**Filtering Problem:**

1) Extracting nonzero $(P_{\beta^{k+1}} \circ S_{\beta,k}^{p+1})_{ij}$ can require a lot of memory
2) Slow decay can lead to noise terms being evaluated

**Solution:** fill entries of $A^{-1}V$ based on $\wedge_{i=1}^{\xi} S(A)$ where $\xi \ll p$

# Approximating off-diagonal entries (III) - Reconstruction

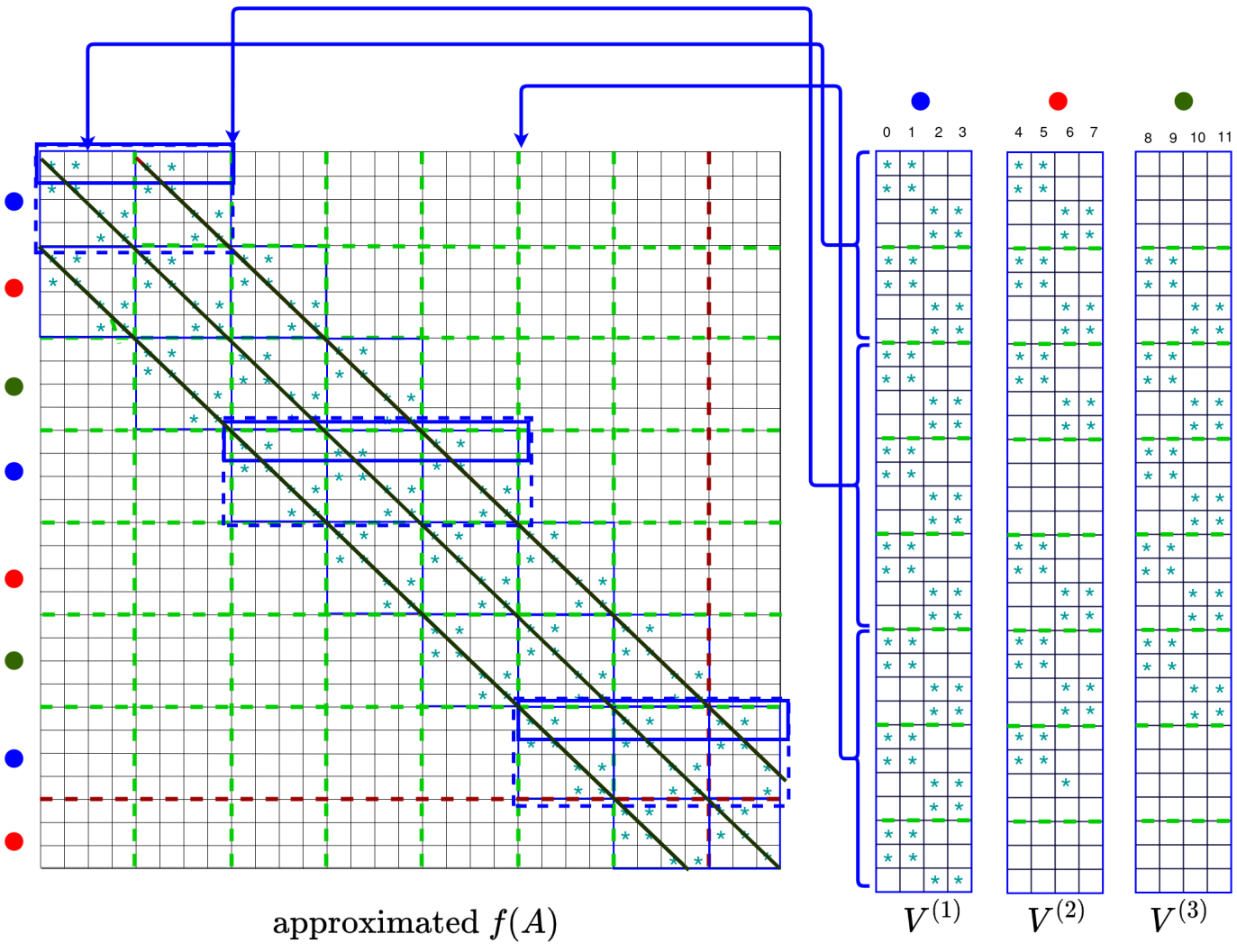



Figure: reconstructing entries of approximate inverse

# Software & numerical experiments

approximate inverse structure + solving system with MRHS

**IMPLEMENTATIONS ON HIGHLY PARALLEL HETEROGENEOUS ARCITECTURES**

- Written in C, uses MKL & CUDA libraries
- Block probing, Pattern sampling, low-memory variants
- SRHS, block, global variants of CG, QMR, GMRES for symmetric, complex-symmetric, hermitian matrices
- Approximate inverse preconditioning
- Least squares fitting methods

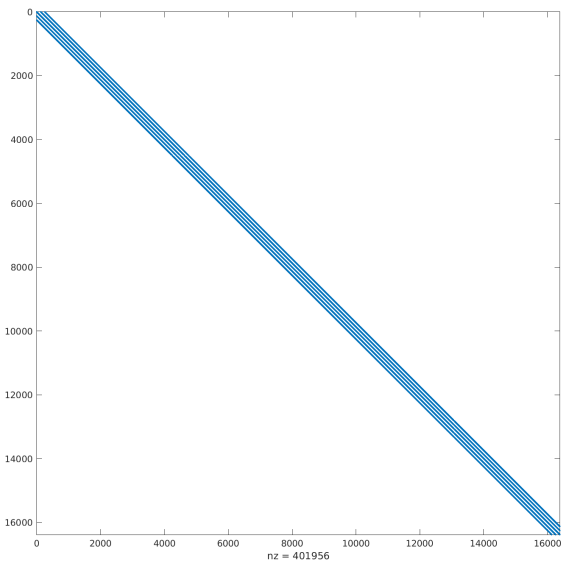
Numerical experiments (in this talk): Emphasis on demonstrating advantages of the block probing method. Not focused on solvers & scalability.

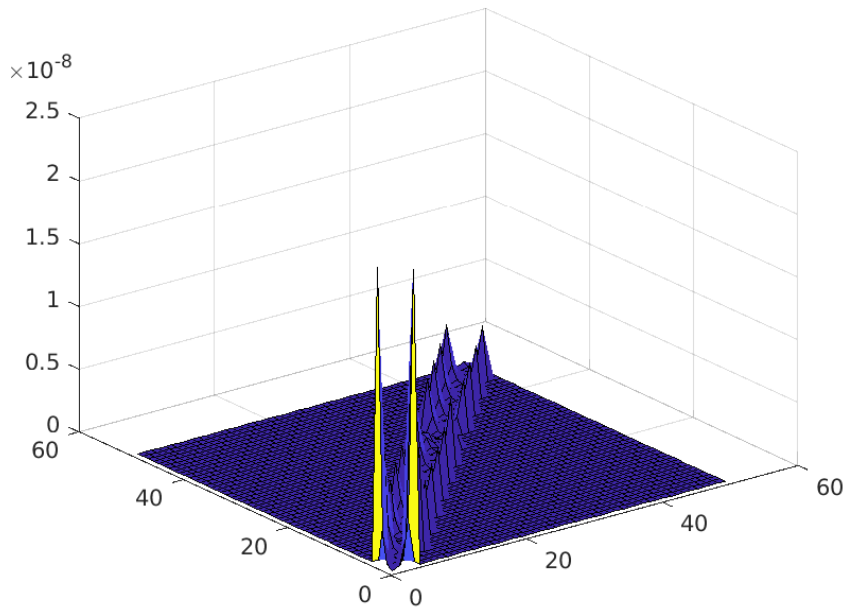
# Experiments with covariance matrices

Dataset 1: Uncertainty quantification, on 2D grid

$$(A)_{ij} = \begin{cases} \left(1 - \frac{\text{dist(i, j)}}{\alpha}\right)^{\beta}, & \text{if dist}(i,j) \leq \alpha \\ 0, & \text{otherwise} \end{cases}$$

Use 35 iterations of PCG (max) with jacobi preconditioning, (2, 3)-probing, $\mathcal{E} = 3.7 \times 10^{-7}$, computation of $4 \times 4$ blocks



# Experiments with standard laplacian operators

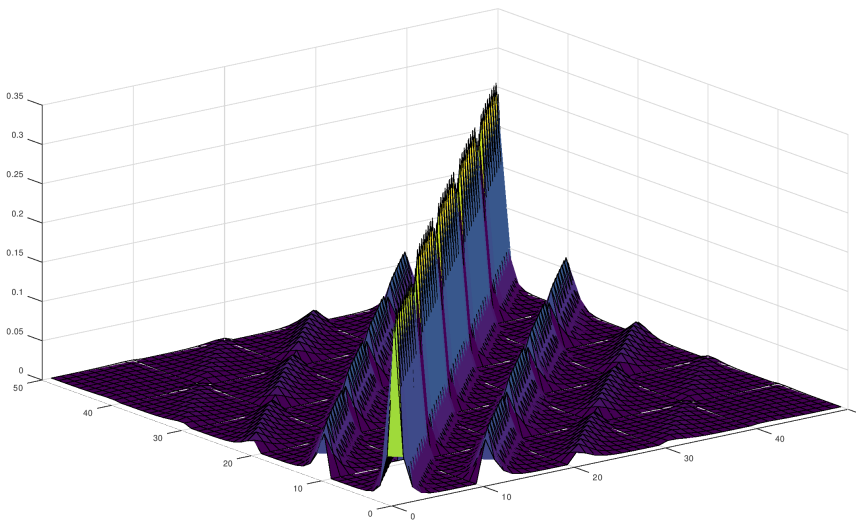
Dataset 2: Laplacian matrices with diagonal shifts:

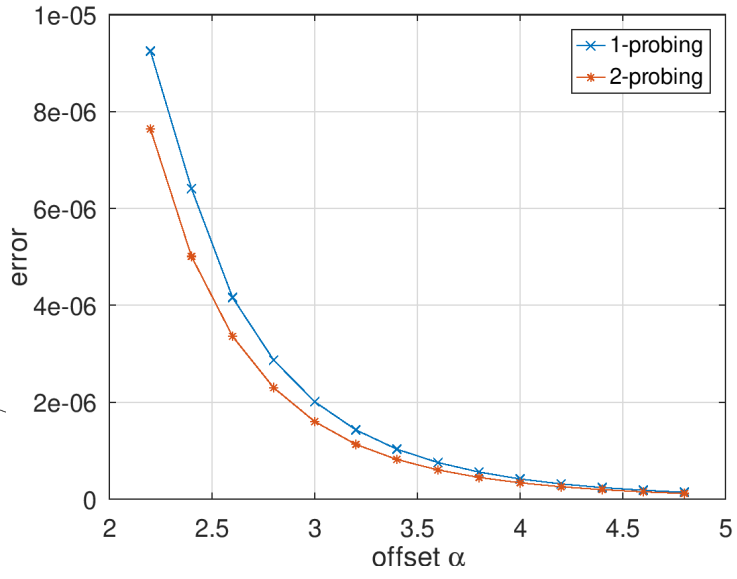
$$A = D - H + \alpha I$$

$H \in \{0,1\}^{n \times n}$: depends on the domain (2D, 3D)

$D$: diagonal matrix, depends on boundary conditions

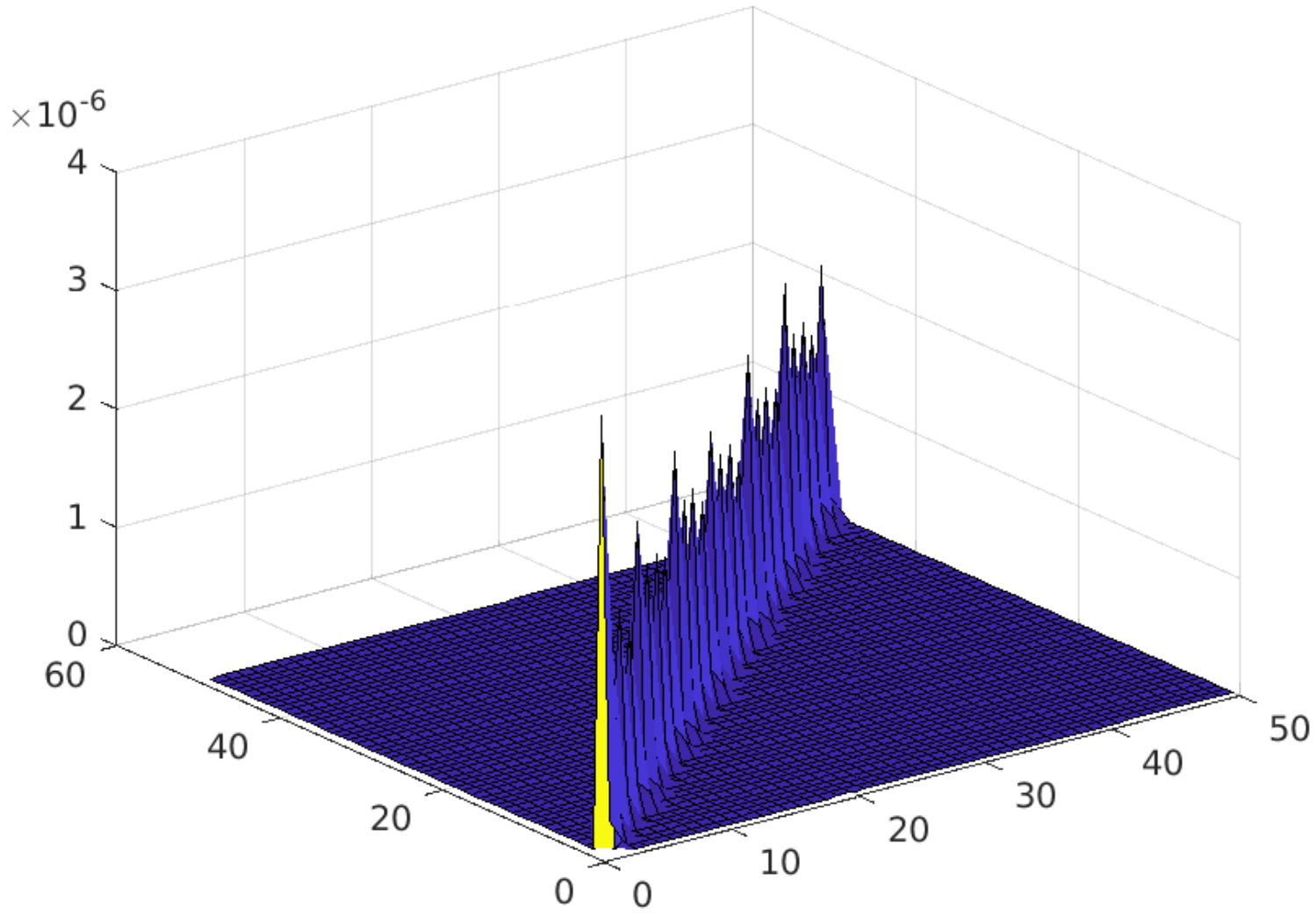
$\alpha$: shift that controls diagonal dominance





# Numerical Experiments for the diagonal blocks

(2, 2)-block probing, 30 iterations of PCG, jacobi preconditioning
$\alpha = 3.2, \mathcal{E} = 8.2 \times 10^{-6}$



# Numerical Experiments for off-diagonal entries

(2, 3)-block probing, 30 iterations of PCG, jacobi preconditioning
$\alpha = 3.6$, mask: $S(A)$, $\mathcal{E} = 3.8 \times 10^{-3}$

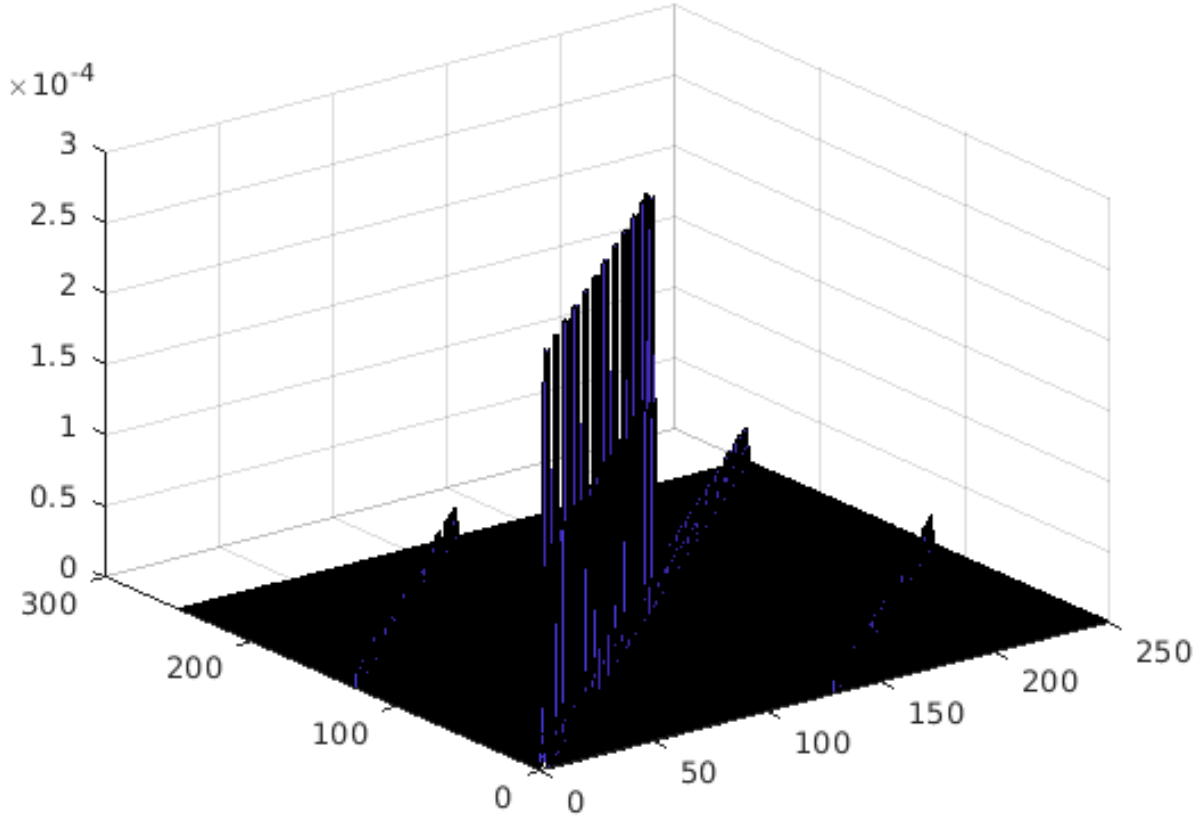



Figure: Entrywise relative error, of entries on $S(A)$ (top $250 \times 250$)

(2, 2)-block probing, 30 iterations of PCG, jacobi preconditioning
$A = H + D +$ 6*diag(rand(n, 1)), mask: $S(A)$ $\mathcal{E} = 1.6 \times 10^{-6}$

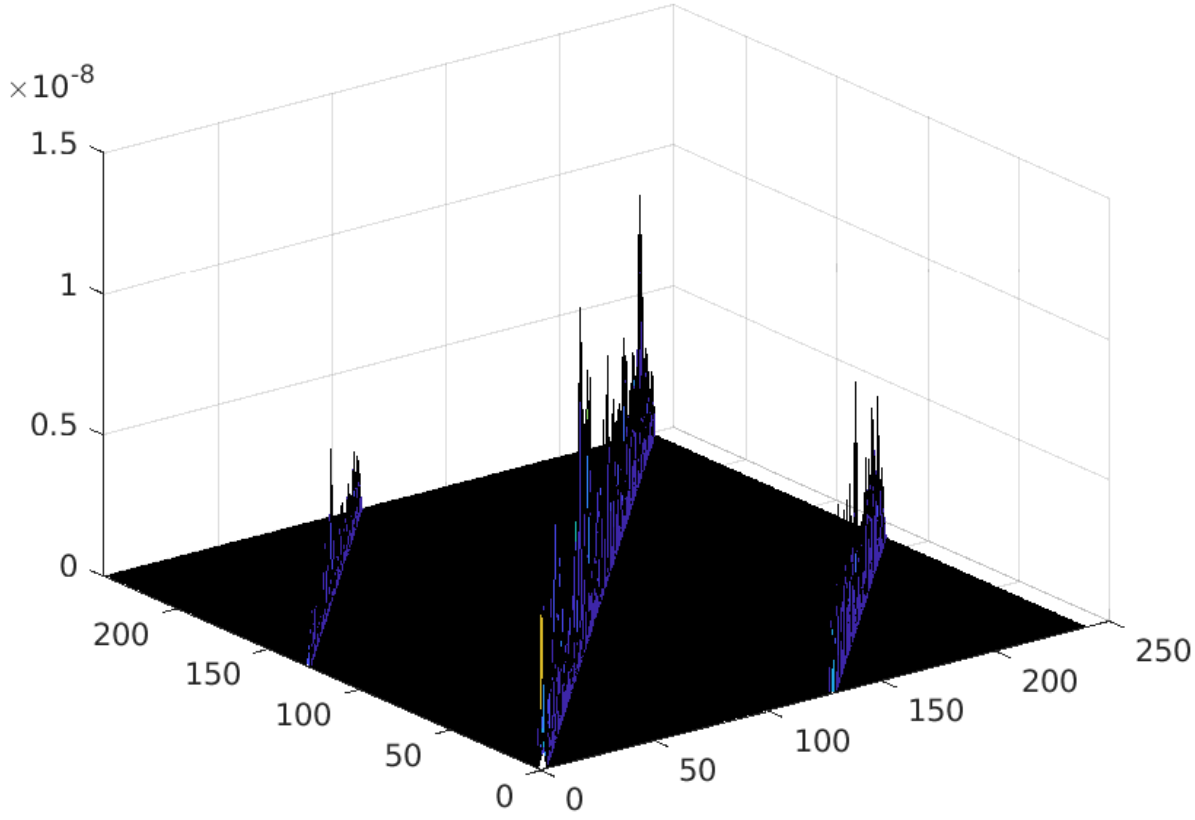
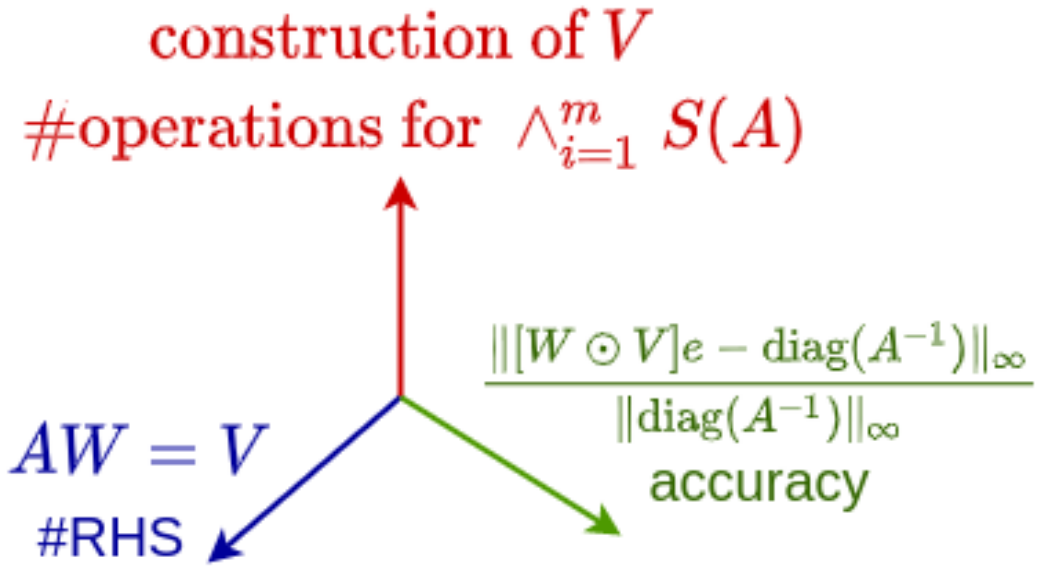



Figure: Entrywise relative error, of entries on $S(A)$ (top $250 \times 250$)

**Conclusions:** Focus on utilizing possible tradeoffs of probing




**Future work:**

- Mixed precision support (+half precision)
- Combination with SQUIC method for covariance estimation
- Multilevel solvers (?) (similar to multigrid)
- Exploit approximate low-rank of the off-diagonal blocks

**In preparation for publication:**

- "Multilevel probing for approximating entries of the matrix inverse" (available soon)
- Software library including a technical report on optimizations

Thank you

Questions?