Efficient construction of an HSS preconditioner for symmetric positive definite $\mathcal{H}^2$ matrices

Xin Xing, Hua Huang, Edmond Chow

University of California, Berkeley

October 23, 2021

Outline

Problem

Dense and symmetric positive definite (SPD) linear system

$$Ax = b$$

where A is an $\mathcal{H}^2$ matrix or can be applied by FMM

Approaches:

- iterative solve using fast matvec by $\mathcal{H}^2$ or FMM
- direct solve using special rank-structured approx. with high-accuracy
- iterative solve using low-accuracy direct solver as preconditioner

HSS as a preconditioner

HSS format (Hierarchically SemiSeparable):

- Has fast decomposition and inversion algorithm.
- High accuracy: $\mathcal{O}(N)$, $\mathcal{O}(N^{\frac{3}{2}})$, $\mathcal{O}(N^2)$ construction costs for 1/2/3-dimensional problems
- Low accuracy: (e.g., fix compression rank) may not be SPD

Existing preconditioners

- HSS Cholesky factorization (Xia, 2010; Li, 2012)
$$A \approx RR^T$$
where R is an HSS matrix.

- Structured incomplete factorization (Xia, 2018; Xia, 2021)
$$A \approx LL^T$$
where L is a ULV-type factor as $L = Q_s L_s Q_{s-1} L_{s-1} \ldots Q_1 L_1$.

- SPDHSS approximation (Xing, 2018)
$$A \approx A_{\text{spdhss}} = LL^T$$
where A_{spdhss} is an SPD HSS matrix.

$\mathcal{H}^2$-based construction of SPDHSS

Target: An HSS preconditioner

- SPD
- efficient to construct

$\mathcal{H}^2$-based construction of SPDHSS

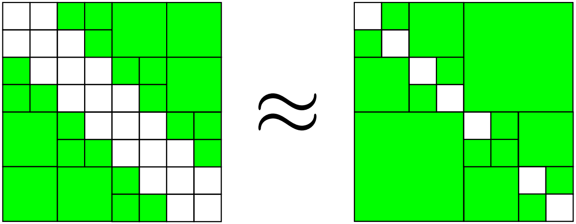
Target: An HSS preconditioner

- SPD
- efficient to construct

Idea:

$$A \approx A_{\text{spdhss}} = LL^T$$

- HSS + scaling-and-compressing technique
  $\longrightarrow \mathcal{O}(N^2 r)$ SPDHSS construction
- Exploiting low-rank blocks in $\mathcal{H}^2$ representation of A
  $\longrightarrow \mathcal{O}(rN \log N)$ construction of A_{spdhss}

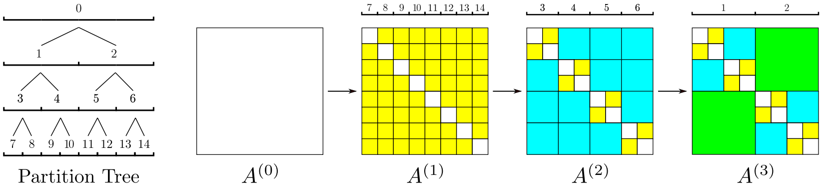


$A_{\mathcal{H}^2}$ $\approx$ A_{SPDHSS}

Outline

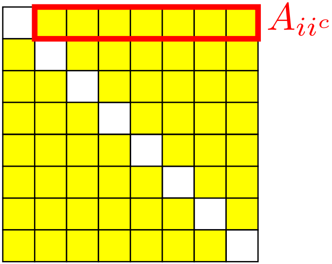
HSS representation

- Matrix hierarchical partitioning
- Compression of off-diagonal blocks at different levels



Off-diagonal block compression

Uniform basis property:



Compression of off-diag block row

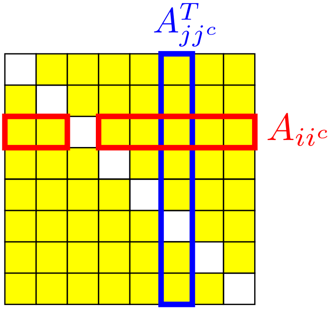
$$A_{ii^c} \approx U_i E_{ii^c}$$

Induced compression:

$$A_{ij} \approx U_i B_{ij} U_j^T$$

Off-diagonal block compression

Uniform basis property:



Compression of off-diag block row

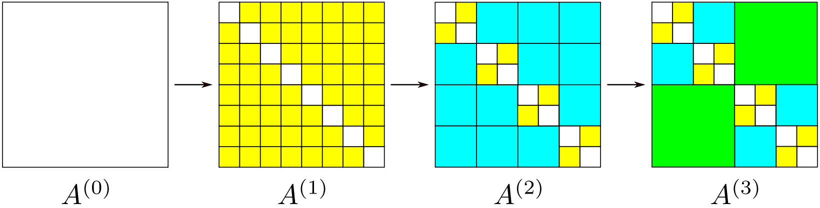
$$A_{ii^c} \approx U_i E_{ii^c}$$

Induced compression:

$$A_{ij} \approx U_i B_{ij} U_j^T$$

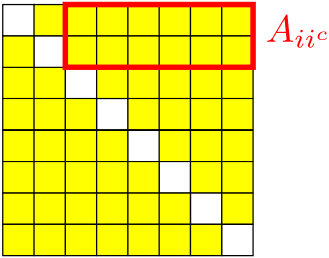
Off-diagonal block compression

Nested basis property:



Off-diagonal block compression

Nested basis property:



Nested compression of off-diag block row

$$A_{ii^c} \approx \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} \begin{bmatrix} E_{1i^c} \\ E_{2i^c} \end{bmatrix}$$
$$\approx \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} R_i E_{ii^c} = U_i E_{ii^c}$$

Nested basis:

$$U_i = \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} R_i$$

HSS construction

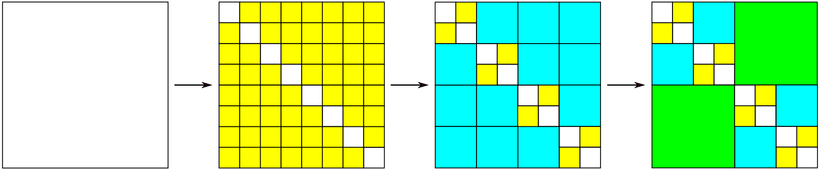
HSS construction :

- Nested compression of off-diagonal block row

$$A_{ii^c} \approx U_i E_{ii^c}$$

- Induced compression of off-diag blocks

$$A_{ij} \approx U_i B_{ij} U_j^T$$

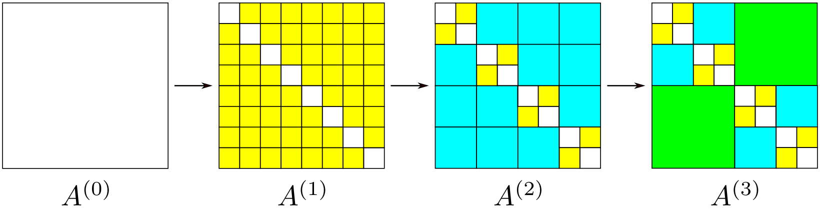


$A^{(0)}$ $A^{(1)}$ $A^{(2)}$ $A^{(3)}$

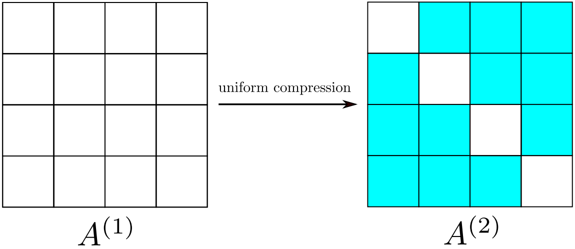
Outline

Scaling and compressing technique

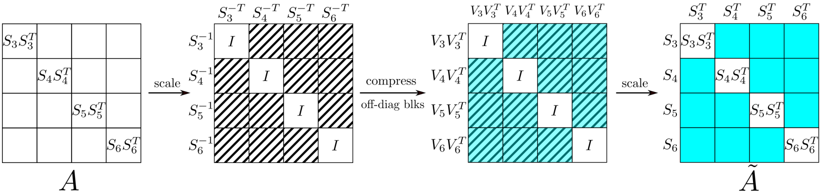
Level-by-level construction of HSS:



Uniform compression at one level:



Scaling and compressing technique



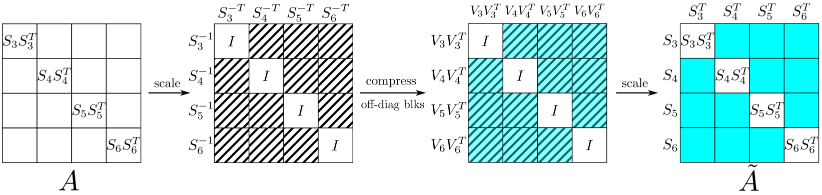
$$A_{ij} \xrightarrow{\text{scale}} S_i^{-1} A_{ij} S_j^{-T} \xrightarrow{\text{uniform compress}} V_i V_i^T \left(S_i^{-1} A_{ij} S_j^{-T} \right) V_j V_j^T$$

$$\xrightarrow{\text{scale back}} S_i V_i V_i^T S_i^{-1} A_{ij} S_j^{-T} V_j V_j^T S_j^T$$

Resulting block compression:

$$A_{ij} \approx (S_i V_i) \left(V_i^T S_i^{-1} A_{ij} S_j^{-T} V_j \right) (S_j V_j)^T = U_i B_{ij} U_j^T$$

Scaling and compressing technique



Positive definiteness:

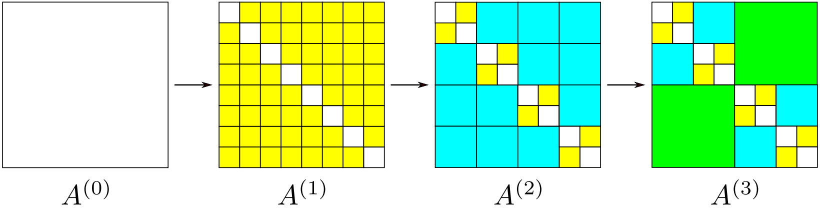
$$\tilde{A} = \mathrm{diag}(A_{ii}) + \mathrm{diag}(S_iV_iV_i^TS_i^{-1}) \underbrace{(A - \mathrm{diag}(A_{ii}))}_{\text{off-diagonal blocks}} \mathrm{diag}(S_iV_iV_i^TS_i^{-1})^T$$

$$= \mathrm{diag}(S_iV_iV_i^TS_i^{-1})\ A\ \mathrm{diag}(S_iV_iV_i^TS_i^{-1})^T + \mathrm{diag}(S_i(I - V_iV_i^T)S_i^T)$$

Result: $\tilde{A}$ is SPD when A is SPD for any orthonormal matrix V_i.

SPDHSS

SPDHSS = scaling-and-compressing + nested compression



Nested compression of off-diag block row:

$$A_{ii^c} \approx S_i V_i V_i^T S_i^{-1} A_{ii^c} = U_i E_{ii^c}, \quad U_i = S_i V_i$$

Compression of off-diag blocks:

$$A_{ij} \approx U_i B_{ij} U_j^T, \quad B_{ij} = V_i^T S_i^{-1} A_{ij} S_j^{-T} V_j$$

Remark: uniform and nested basis properties still hold

SPDHSS: scaling step

Symmetric decomposition of diagonal block:

$$A_{ii}^{(k-1)} = \begin{bmatrix} A_{i_1 i_1}^{(k-1)} & \dots & A_{i_1 i_m}^{(k-1)} \\ \vdots & \ddots & \vdots \\ A_{i_m i_1}^{(k-1)} & \dots & A_{i_m i_m}^{(k-1)} \end{bmatrix} = \begin{bmatrix} S_{i_1} S_{i_1}^T & \dots & U_{i_1} B_{i_1,i_m} U_{i_m}^T \\ \vdots & \ddots & \vdots \\ U_{i_m} B_{i_m,i_1} U_{i_1}^T & \dots & S_{i_m} S_{i_m}^T \end{bmatrix}$$

$$= \begin{bmatrix} S_{i_1} & & \\ & \ddots & \\ & & S_{i_m} \end{bmatrix} \left(I + \mathbf{V}_i \mathbf{B}_{ii} \mathbf{V}_i^T \right) \begin{bmatrix} S_{i_1} & & \\ & \ddots & \\ & & S_{i_m} \end{bmatrix}^T$$

with

$$\mathbf{B}_{ii} = \begin{bmatrix} 0 & B_{i_1,i_2} & \dots & B_{i_1,i_m} \\ B_{i_2,i_1} & 0 & \dots & B_{i_2,i_m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{i_m,i_1} & B_{i_m,i_2} & \dots & 0 \end{bmatrix}, \quad \mathbf{V}_i = \begin{bmatrix} V_{i_1} & & \\ & \ddots & \\ & & V_{i_m} \end{bmatrix}.$$

SPDHSS: scaling step

Symmetric decomposition of diagonal block:

$$I + \mathbf{V}_i \mathbf{B}_{ii} \mathbf{V}_i^T = \bar{S}_i \bar{S}_i^T$$
$$\text{with} \quad \bar{S}_i = I + \mathbf{V}_i((I + \mathbf{B}_{ii})^{1/2} - I)\mathbf{V}_i^T$$

$$\Longrightarrow A_{ii}^{(k-1)} = S_i S_i^T, \quad S_i = \begin{bmatrix} S_{i_1} & & \\ & \ddots & \\ & & S_{i_m} \end{bmatrix} \bar{S}_i.$$

SPDHSS: scaling step

Symmetric decomposition of diagonal block:

$$
\begin{aligned}
&I + \mathbf{V}_i \mathbf{B}_{ii} \mathbf{V}_i^T = \bar{S}_i \bar{S}_i^T \\
&\text{with} \quad \bar{S}_i = I + \mathbf{V}_i((I + \mathbf{B}_{ii})^{1/2} - I)\mathbf{V}_i^T
\end{aligned}
$$

$$
\Longrightarrow A_{ii}^{(k-1)} = S_i S_i^T, \quad S_i = \begin{bmatrix} S_{i_1} & & \\ & \ddots & \\ & & S_{i_m} \end{bmatrix} \bar{S}_i .
$$

Scaled off-diagonal blocks $S_i^{-1} A_{ij}^{(k-1)} S_j^{-T}$

$$
\begin{aligned}
&S_i^{-1} \begin{bmatrix} U_{i_1} & & \\ & \ddots & \\ & & U_{i_m} \end{bmatrix} \begin{bmatrix} B_{i_1 j_1} & \dots & B_{i_1 j_m} \\ \vdots & \ddots & \vdots \\ B_{i_m j_1} & \dots & B_{i_m j_m} \end{bmatrix} \begin{bmatrix} U_{j_1} & & \\ & \ddots & \\ & & U_{j_m} \end{bmatrix}^T S_j^{-T} \\
=&\mathbf{V}_i (I + \mathbf{B}_{ii})^{-1/2} \mathbf{B}_{ij} (I + \mathbf{B}_{jj})^{-1/2} \mathbf{V}_j^T
\end{aligned}
$$

SPDHSS: compression step

Scaled off-diagonal block

$$S_i^{-1} A_{ij}^{(k-1)} S_j^{-T} = \mathbf{V}_i (I + \mathbf{B}_{ii})^{-1/2} \mathbf{B}_{ij} (I + \mathbf{B}_{jj})^{-1/2} \mathbf{V}_j^T$$

Note: V_i captures the column space of scaled block row

$$V_i = \begin{bmatrix} V_{i_1} & & \\ & \ddots & \\ & & V_{i_m} \end{bmatrix} \bar{V}_i \qquad (1)$$

with $\bar{V}_i$ capturing the column space of

$$\left[(I + \mathbf{B}_{ii})^{-1/2} \mathbf{B}_{ij} (I + \mathbf{B}_{jj})^{-1/2} \right]_{\text{for all } j \neq i}$$

SPDHSS: full algorithm

Input: HSS rank r, an SPD matrix A
Output: an SPD HSS approximation with $\{A_{ii}\}, \{B_{ij}\}, \{U_i\}, \{R_i\}$

At the leaf level

compute $A_{ii} = S_i S_i^T$ directly

compute scaled off-diagonal block $C_{ij}^{(0)} = S_i^{-1} A_{ij} S_j^{-T}$

compute $V_i \in \mathbb{R}^{|I_i| \times r}$ minimizing $\|C_{ii^c}^{(0)} - V_i V_i^T C_{ii^c}^{(0)}\|_F$

compute $B_{ij} = V_i^T C_{ij}^{(0)} V_j$

set $U_i = S_i V_i$

for $k = 2, 3, \ldots, L-1$ **do**

compute $(I + \mathbf{B}_{ii})^{\pm 1/2}$ via the eigen-decomposition of $\mathbf{B}_{ii}$

compute $E_{ij} = (I + \mathbf{B}_{ii})^{-1/2} \mathbf{B}_{ij} (I + \mathbf{B}_{jj})^{-1/2}$

assemble E_{ii^c} and compute $\bar{V}_i$ minimizing $\|E_{ii^c} - \bar{V}_i \bar{V}_i^T E_{ii^c}\|_F$

compute $B_{ij} = \bar{V}_i^T (I + \mathbf{B}_{ii})^{-1/2} \mathbf{B}_{ij} (I + \mathbf{B}_{jj})^{-1/2} \bar{V}_j$

set $R_i = (I + \mathbf{B}_{ii})^{1/2} \bar{V}_i$, $\forall i \in \text{lvl}(k)$

end for

$\mathcal{H}^2$-based SPDHSS

Bottleneck:
Compute $V_i/\bar{V}_i$ for scaled off-diagonal block row
$\Longrightarrow \mathcal{O}(N^2)$ computational and storage costs

Leaf level: $V_i = \arg\min_V \|S_i^{-1}A_{ii^c} - VV^TS_i^{-1}A_{ii^c}\|$

Nonleaf level: $\bar{V}_i = \arg\min_V \|\Psi_i A_{ii^c} - VV^T\Psi_i A_{ii^c}\|$

Key Idea: randomized algorithm with $S_i^{-1}(A_{ii^c}\Omega_{i^c})$ and $\Psi_i(A_{ii^c}\Omega_{i^c})$ accelerated by $\mathcal{H}^2$ representation of A
$\Longrightarrow \mathcal{O}(N\log Nr)$ computational and storage costs

Outline

Numerical tests

Test problem:

$$A = \sigma I + K(X, X)$$

with non-oscillatory kernel function $K(x, y)$ and point set X

- random sampling on a sphere in 3D
- random sampling in a ball in 3D

H2Pack

H2Pack (https://github.com/scalable-matrix/H2Pack)

- $\mathcal{O}(N)$ $\mathcal{H}^2$ matrix construction using the proxy point method
- $\mathcal{O}(N)$ $\mathcal{H}^2$ matrix-vector multiplication.
- HSS matrix construction, matrix-vector multiplication, and ULV decomposition.

Notes:

- inputs: $K(x, y)$, X, and an error threshold
- any non-oscillatory translationally invariant $K(x, y)$

H2Pack

Relative error of $\mathcal{H}^2$-matvec in H2Pack. Each test has 1×10^6 points.

relerr threshold		1.00E-2	1.00E-4	1.00E-8	1.00E-12
3D Laplace	sphere	8.42E-4	3.68E-6	6.35E-10	9.20E-13
	ball	8.21E-4	4.13E-6	8.05E-10	5.30E-13
3D Gaussian	sphere	3.38E-3	1.89E-5	4.25E-9	2.38E-13
	ball	3.57E-3	1.53E-5	1.13E-9	3.85E-12
3D Stokes	sphere	1.26E-3	7.06E-6	3.73E-10	2.71E-12
	ball	1.42E-3	7.72E-6	2.61E-9	2.76E-12
3D Matérn-$\frac{3}{2}$	sphere	3.22E-3	3.55E-5	2.05E-9	1.51E-13
	ball	1.55E-3	3.21E-5	2.34E-9	1.87E-13

H2Pack

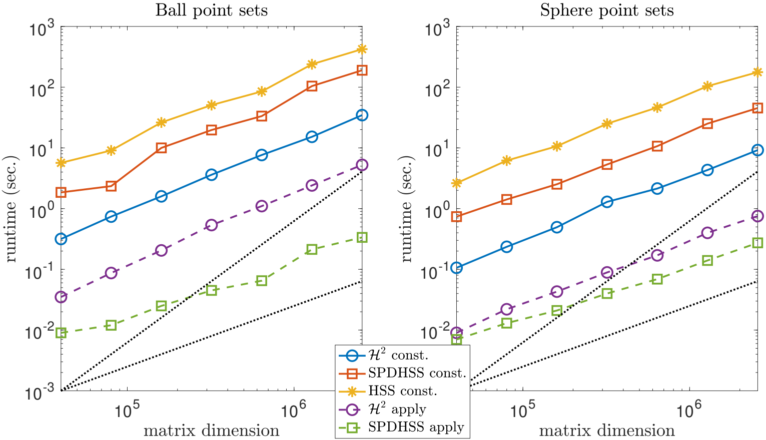
- 3D Laplace kernel
- Intel Skylake node (two Intel Xeon Platinum 8160 processors, 48 cores)

Runtime (in sec.) of three packages with relative accuracy around 10^{-5}

	H2Pack		PVFMM		FMM3D	
#pts $\times 10^5$	build	matvec	build	matvec	build	matvec
sphere 4	0.125	0.018	0.134	0.056	0.183	0.329
sphere 8	0.223	0.037	0.328	0.132	0.441	0.626
sphere 16	0.444	0.075	0.686	0.230	1.025	1.259
ball 4	0.173	0.035	0.113	0.075	0.170	0.192
ball 8	0.249	0.090	0.221	0.267	0.443	1.266
ball 16	0.784	0.149	0.691	0.219	0.955	1.261

Computational cost of SPDHSS

Matérn-$\frac{3}{2}$ kernel, HSS rank $r = 100$



Preconditioning performance

Setting:

- Matérn-$\frac{3}{2}$ kernel, $K(x,y) = \left(1 + \sqrt{3}\, l\, |x-y|\right) \exp\left(-\sqrt{3}\, l\, |x-y|\right)$
- $N = 3.2 \times 10^5$ points in a ball
- PCG with $\tau = 10^{-6}$

parameter l	0.0010	0.0025	0.005	0.010	0.025	0.05	0.10	0.25	0.5
Unprecond.	41	119	297	687	1896	-	-	1684	634
BJ	459	1202	2504	-	-	-	2765	707	172
FSAI $k=200$	2659	2623	2045	1518	960	569	266	63	18
FSAI $k=400$	1734	1531	1111	831	511	266	108	28	10
SPDHSS $r=100$	2	3	5	14	62	159	287	245	117
SPDHSS $r=200$	1	2	4	5	20	58	131	171	92
HSS $r=100$	2	3	6	/	/	/	/	/	/
HSS $r=200$	2	2	3	6	/	/	/	/	/

Preconditioning performance

Setting:

- Inverse multiquadric (IMQ) kernel, $K(x,y) = 1/\sqrt{1 + l\,|x-y|^2}$.
- $N = 3.2 \times 10^5$ points in a ball
- PCG with $\tau = 10^{-6}$

parameter l	0.001	0.005	0.01	0.1	0.5	1	5	10	50	100
Unprecond.	1239	2656	-	2812	1958	1576	915	724	394	284
BJ	-	-	-	1605	529	322	195	151	97	73
FSAI $k = 200$	2839	1092	619	121	45	32	24	24	23	23
FSAI $k = 400$	1598	535	266	51	24	21	18	18	18	17
SPDHSS $r = 100$	53	249	328	291	112	74	38	28	13	10
SPDHSS $r = 200$	12	71	129	195	84	60	33	23	10	8
HSS $r = 100$	/	/	/	/	/	/	/	/	/	/
HSS $r = 200$	/	/	/	/	/	/	/	/	/	/

Takeaways

- Scaling-and-compressing technique for preserving SPD.
- Use of existing/cheap-to-get rank structures for accelerating complicated constructions.