MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

COMPUTATIONAL METHODS IN
SYSTEMS AND CONTROL THEORY

Mixed precision HODLR matrices

Xiaobo Liu
MPI for Dynamics of Complex Technical Systems, Germany

SIAM PP26
Zuse Institute Berlin, Germany
March 3, 2026

Joint work with Erin Carson[a] • Xinye Chen[b]

[a]Department of Numerical Mathematics, Charles University, Czech Republic.
[b]LIP6, Sorbonne University, France.

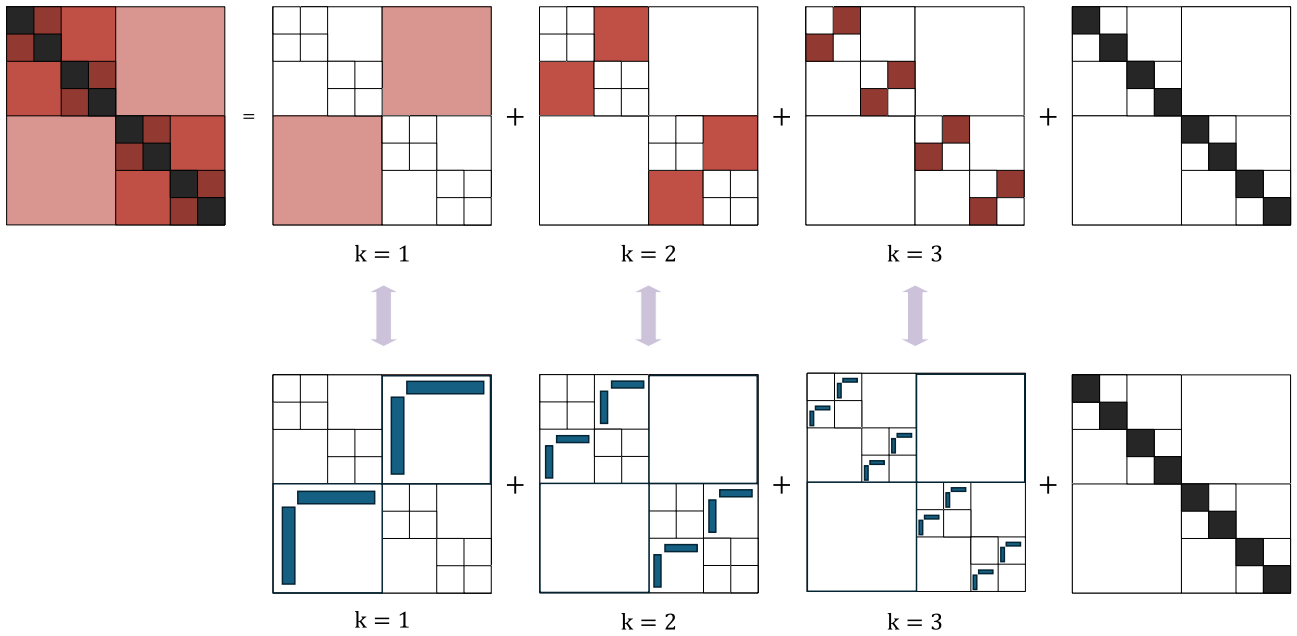
HODLR Matrix

Hierarchical Off-Diagonal Low-Rank (HODLR) matrices: a class of hierarchical matrices, used as data-sparse approximations of *non-sparse* matrices.

- **Key property**:
  - Hierarchical partitioning of off-diagonal blocks reveals rank deficiency at levels $1, 2, \cdots, \ell$ (given tree depth ℓ)
  - Off-diagonal blocks stored as low-rank (rank $\leq p$) matrix products UV^T
- **Reduced storage and arithmetic cost**:
  - $\mathcal{O}(pn \log n)$ storage cost: $O(np)$ for each level $\times$ $O(\log n)$ levels
  - Typically $\mathcal{O}(p^\alpha n \log^\beta n)$, $\alpha, \beta \in \{1, 2\}$ arithmetic operation cost, e.g., matrix–vector product, matrix multiplication, inversion, and factorizations
- **Applications**: kernel methods, integral equations, covariance matrices, etc.

HODLR Matrix: Visualization and Construction

Convert a matrix into HODLR format:



1. Partition $H^{(k)}$ into $\begin{bmatrix} H_{11}^{(k+1)} & H_{12}^{(k+1)} \\ H_{21}^{(k+1)} & H_{22}^{(k+1)} \end{bmatrix}$
2. Compute low-rank approximations $H_{12}^{(k+1)} \approx U_1^{(k+1)} (V_2^{(k+1)})^T$ and $H_{21}^{(k+1)} \approx U_2^{(k+1)} (V_1^{(k+1)})^T$ until bottom level/min block size reached

(p, ε)-HODLR Matrix

To quantify the error incurred in the low-rank factorization of the off-diagonal blocks, we introduced the definition.

Definition ((p, ε)-HODLR matrix)

$\widetilde{H}$ is defined to be a (p, ε)-HODLR matrix to a p-HODLR matrix H, if every off-diagonal block at any level k satisfies

$$\|\widetilde{H}_{ij}^{(k)} - H_{ij}^{(k)}\|_F \leq \varepsilon \|H_{ij}^{(k)}\|_F, \quad 0 \leq \varepsilon < 1.$$

- Rank-constrained $\rightsquigarrow$ (practically) tolerance-constrained HODLR format.

Theorem (Approximation error in diagonal blocks)

Let $\widetilde{H}$ be a (p, ε)-HODLR matrix associated with H. Then for the HODLR matrices $\widetilde{H}_{ii}^{(k)}$, $i = 1 : 2^k$, at level $k \in \{0, \ldots, \ell\}$, it holds that

$$\|\widetilde{H}_{ii}^{(k)} - H_{ii}^{(k)}\|_F \leq \varepsilon \|H_{ii}^{(k)}\|_F, \quad i = 1 : 2^k.$$

(p, ε)-HODLR Matrix

To quantify the error incurred in the low-rank factorization of the off-diagonal blocks, we introduced the definition.

Definition ((p, ε)-HODLR matrix)

$\widetilde{H}$ is defined to be a (p, ε)-HODLR matrix to a p-HODLR matrix H, if every off-diagonal block at any level k satisfies

$$\|\widetilde{H}_{ij}^{(k)} - H_{ij}^{(k)}\|_F \leq \varepsilon \|H_{ij}^{(k)}\|_F, \quad 0 \leq \varepsilon < 1.$$

- Rank-constrained $\rightsquigarrow$ (practically) tolerance-constrained HODLR format.

Theorem (Approximation error in diagonal blocks)

Let $\widetilde{H}$ be a (p, ε)-HODLR matrix associated with H. Then for the HODLR matrices $\widetilde{H}_{ii}^{(k)}$, $i = 1\colon 2^k$, at level $k \in \{0, \ldots, \ell\}$, it holds that

$$\|\widetilde{H}_{ii}^{(k)} - H_{ii}^{(k)}\|_F \leq \varepsilon \|H_{ii}^{(k)}\|_F, \quad i = 1\colon 2^k.$$

Floating-point Arithmetics

Computers can represent numbers as

$$x = \pm m \cdot 2^e$$

where

- m is a t-digit number in $[0, 2)$
- e is an integer between $e_{\min}$ and $e_{\max}$ (in IEEE 754 $e_{\min} = 1 - e_{\max}$)

Table: Parameters of five floating point formats.

Type	Signif. bits (t)	Exp. bits	Range	$u = 2^{-t}$
fp8-q52	3	5	$10^{\pm 5}$	1.3×10^{-1}
bfloat16	8	8	$10^{\pm 38}$	3.9×10^{-3}
binary16	11	5	$10^{\pm 5}$	4.9×10^{-4}
binary32	24	8	$10^{\pm 38}$	6.0×10^{-8}
binary64	53	11	$10^{\pm 308}$	1.1×10^{-16}

- Lower precision $\rightsquigarrow$ faster flops, less comm., lower energy consumption.

Mixed-precision HODLR Representation

Idea: Reduced precisions for storing off-diagonal blocks in different levels

1: $k \leftarrow 0, H \leftarrow H^{(0)}$
2: **while** $k < \ell$ **do**
3: **for** $i := 1 : 2^k$ **do**
4: $\widehat{H}^{(k+1)}_{2i-1,2i} \leftarrow \widehat{U}^{(k+1)}_{2i-1}(\widehat{V}^{(k+1)}_{2i})^T$ [Compute in u, and store in u_{k+1}]
5: $\widehat{H}^{(k+1)}_{2i,2i-1} \leftarrow \widehat{U}^{(k+1)}_{2i}(\widehat{V}^{(k+1)}_{2i-1})^T$ [Compute in u, and store in u_{k+1}]
6: **end for**
7: $k \leftarrow k + 1$
8: **end while**
9: **for** $i := 1 : 2^\ell$ **do**
10: $\widehat{H}^{(\ell)}_{i,i} \leftarrow H^{(\ell)}_{i,i}$ [Store $\widehat{H}^{(\ell)}_{i,i}$ in u]
11: **end for**

Define $\xi_k := \max_{|i-j|=1} \|\widetilde{H}^{(k)}_{ij}\|_F / \|\widetilde{H}\|_F, 1 \le k \le \ell$. If $u_k \le \varepsilon/(2^{k/2}\xi_k)$, then $\|H - \widehat{H}\|_F \lesssim (2\sqrt{2\ell}+1)\varepsilon\|H\|_F$.

- Mixed-precision representation error comparable to $\|H - \widetilde{H}\|_F \le \varepsilon\|H\|_F$.

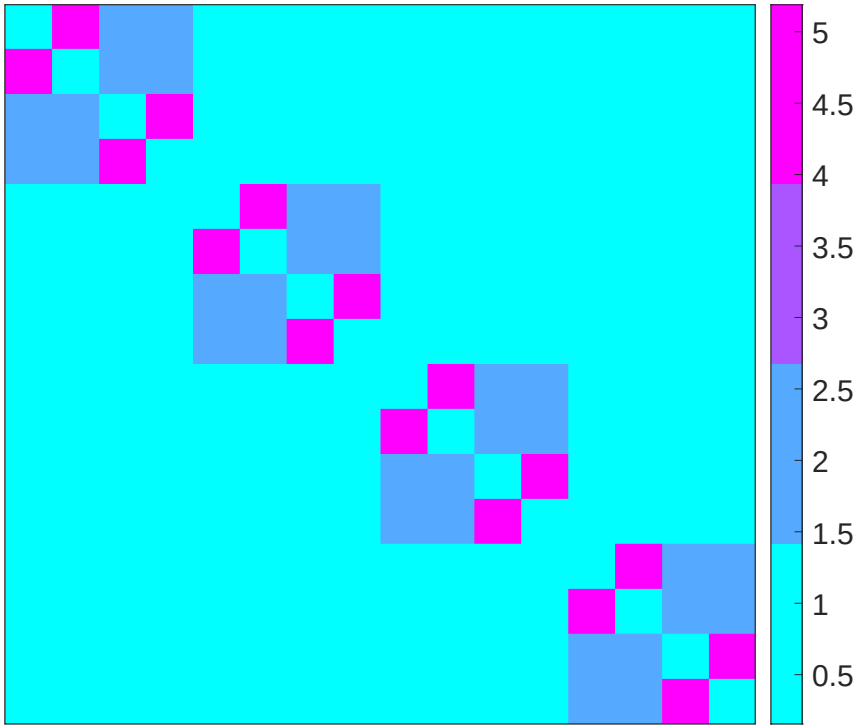
Datasets from SuiteSparse

Table: Summary of test matrices from the SuiteSparse collection.

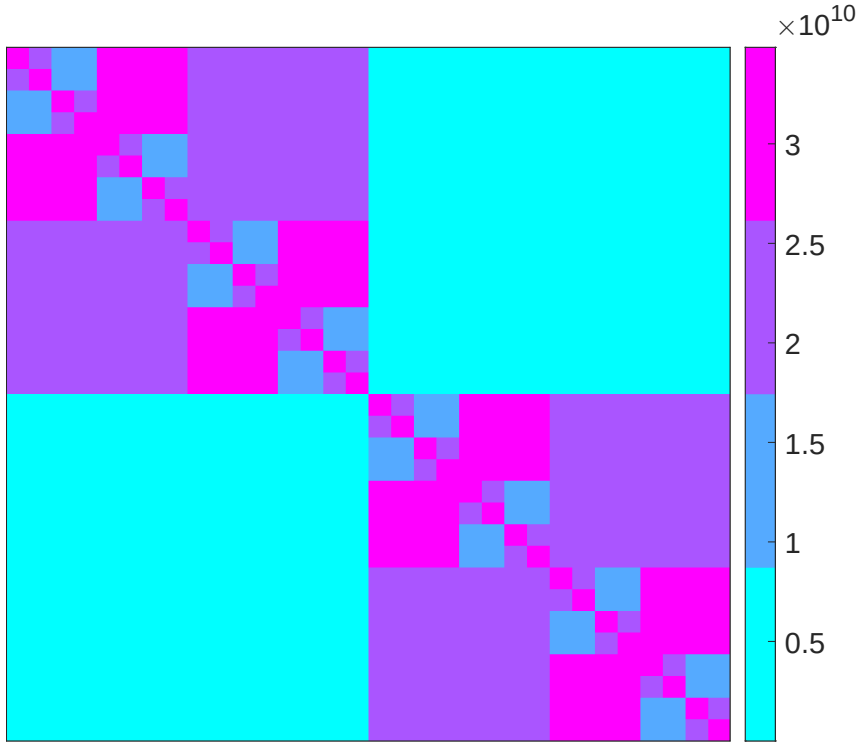
Dataset	Size	Nonzeros	Description
saylr3	1,000	3,750	Computational fluid dynamics problem
LeGresley_2508	2,508	16,727	Power network problem
ex37	3,565	67,591	Computational fluid dynamics problem
1138_bus	1,138	4,054	Power network problem
bcsstk08	1,074	12,960	Structural problem
cavity18	4,562	138,040	Computational fluid dynamics problem

Schur complement in use: correspond to root separator of hierarchical partitionings, whose approximation crucial in sparse linear solvers

Mixed-precision HODLR Representation



(a) saylr3



(b) LeGresley_2508

Figure: Chosen precisions: left: {bf16, fp16, fp16, fp32, fp32, fp32}; right: {q52, fp32, fp32, fp32, fp32, fp32}.

- Depth $\ell = 6$, $\varepsilon = 10^{-4}$
- Set of available precisions {q52, bf16, fp16, fp32, fp64}

Mixed-precision HODLR: Reconstruction Error

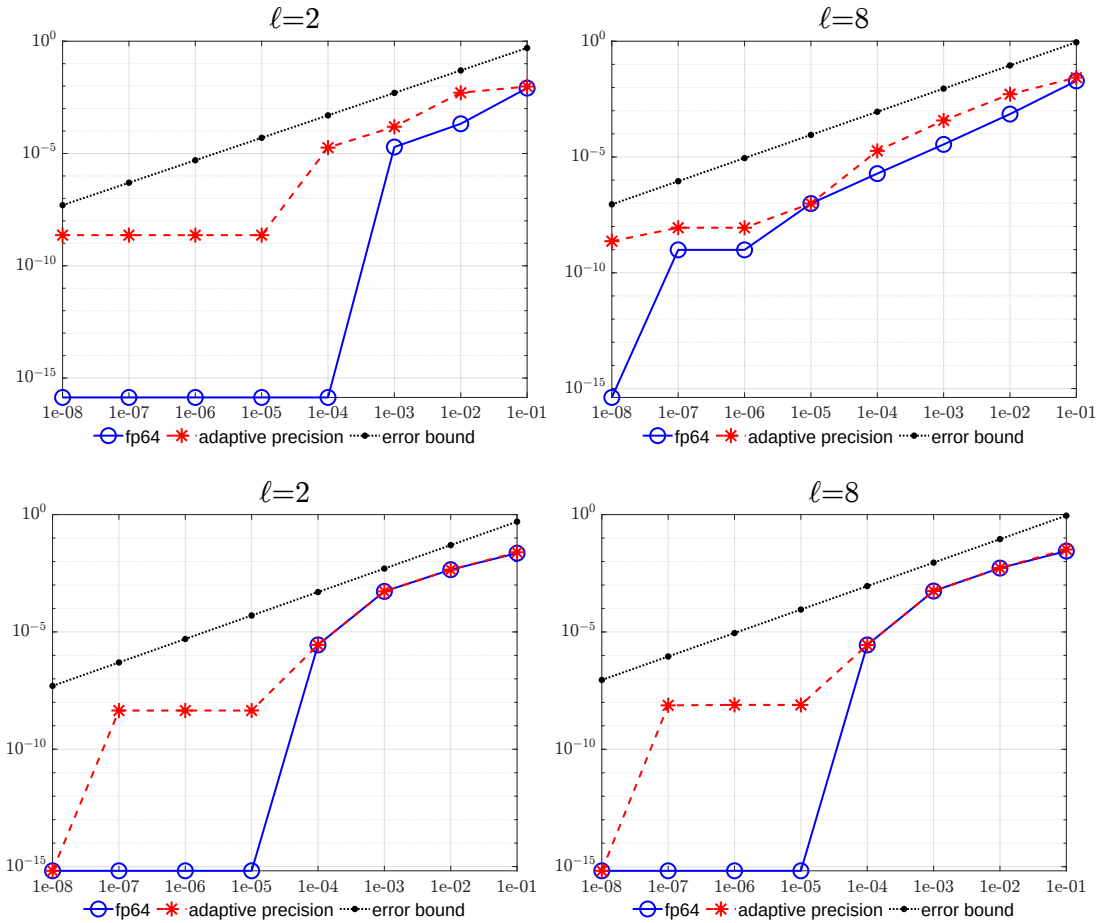
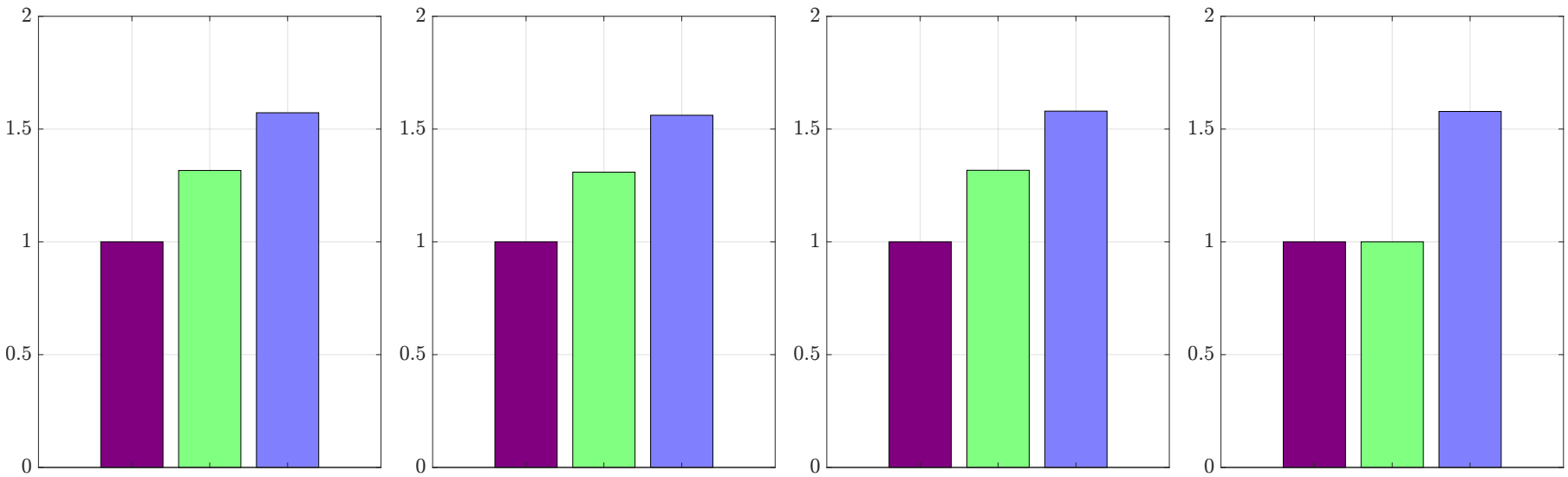


Figure: Top: ex37; bottom: 1138_bus. x-axis: ε; y-axis: $\|H - \widehat{H}\|_F / \|H\|_F$.

- Finite precision/low-rank approximation error dominates as ε varies
- fp64 not needed for satisfying the error bound

Mixed-precision HODLR : Storage Reduction

Setting: $\ell = 8$. Storage cost computed as total bits used for storing the low-rank factors in each level.



(a) bcsstk08 (b) cavity18 (c) ex37 (d) LeGresley_2508

Figure: Storage savings of mixed-precision HODLR matrices relative to uniform (double) precision HODLR matrices. Purple: $\varepsilon = 10^{-7}$, green: $\varepsilon = 10^{-4}$, blue: $\varepsilon = 10^{-1}$.

- Storage savings increase as ε increases

Matrix–vector Product: Algorithm

- Matrix–vector products by level, using the low-rank approximations

1: $b \leftarrow \mathbf{0} \in \mathbb{R}^n$
2: **for** $k = 1 : \ell$ **do**
3: Partition b into 2^k blocks, i.e., $(b_i^{(k)})_{i=1}^{2^k}$
4: Partition x into 2^k blocks, i.e., $(x_i^{(k)})_{i=1}^{2^k}$
5: **for** $i = 1 : 2^{k-1}$ **do**
6: $b_{2i-1}^{(k)} \leftarrow b_{2i-1}^{(k)} + U_{2i-1}^{(k)} (V_{2i}^{(k)})^T x_{2i}^{(k)}$ [Compute in precision u]
7: $b_{2i}^{(k)} \leftarrow b_{2i}^{(k)} + U_{2i}^{(k)} (V_{2i-1}^{(k)})^T x_{2i-1}^{(k)}$ [Compute in precision u]
8: **end for**
9: **end for**
10: **for** $i = 1 : 2^{\ell}$ **do**
11: $b_i^{(\ell)} \leftarrow b_i^{(\ell)} + H_{i,i}^{(\ell)} x_i^{(\ell)}$ [Compute in precision u]
12: **end for**

Thought: Approximation errors in the off-diagonal blocks, reduced precision?

Matrix–vector Product: New Results

Idea: Balance the approximation error in $\widehat{H}$ (mixed-precision HODLR representation) and the finite-precision computation error in $b \leftarrow \widehat{H}x$.

Lemma (Working precision for matrix–vector products)

Let $A \approx A_p = X_p \Sigma_p Y_p^T = \sum_{i=1}^{p} \sigma_i x_i y_i^T$ be the best rank-p approximation. Then the finite precision computation error in $\widehat{b} = \mathrm{fl}(A_p x) \leq$ the low-rank approximation error, if the working precision has unit roundoff $u \leq \sigma_{p+1}/(pn\sigma_1)$.

- Intuition: more inexact the low-rank representation, lower the precision

Theorem (Working precision and backward error)

If $b = \widehat{H}x$ is computed (in the usual fashion of HODLR matrix–vector product) in a working precision $u \leq \varepsilon/n$, then the computed $\widehat{b}$ satisfies

$$\widehat{b} = \mathrm{fl}(\widehat{H}x) = (H + \Delta H)x, \quad \|\Delta H\|_F \leq \mathcal{O}(2^{\ell/2})\varepsilon \|H\|_F.$$

- If $u \leq \varepsilon/n$, overall error dominated by the matrix representation error

Matrix–vector Product: New Results

Idea: Balance the approximation error in $\widehat{H}$ (mixed-precision HODLR representation) and the finite-precision computation error in $b \leftarrow \widehat{H}x$.

Lemma (Working precision for matrix–vector products)

Let $A \approx A_p = X_p \Sigma_p Y_p^T = \sum_{i=1}^{p} \sigma_i x_i y_i^T$ be the best rank-p approximation. Then the finite precision computation error in $\widehat{b} = \mathrm{fl}(A_p x) \leq$ the low-rank approximation error, if the working precision has unit roundoff $u \leq \sigma_{p+1}/(pn\sigma_1)$.

- Intuition: more inexact the low-rank representation, lower the precision

Theorem (Working precision and backward error)

If $b = \widehat{H}x$ is computed (in the usual fashion of HODLR matrix–vector product) in a working precision $u \leq \varepsilon/n$, then the computed $\widehat{b}$ satisfies

$$\widehat{b} = \mathrm{fl}(\widehat{H}x) = (H + \Delta H)x, \quad \|\Delta H\|_F \leq \mathcal{O}(2^{\ell/2})\varepsilon \|H\|_F.$$

- If $u \leq \varepsilon/n$, overall error dominated by the matrix representation error

Matrix-vector Product: Test Matrices

Kernel matrix K:

$$\text{(i)} \quad K_{ij} = \begin{cases} \frac{1}{x-y}, & \text{if} \quad x \neq y; \\ 1, & \text{otherwise.} \end{cases}$$

$$\text{(ii)} \quad K_{ij} = \begin{cases} \log \|x_i - x_j\|_2, & \text{if} \quad x \neq y; \\ 0, & \text{otherwise.} \end{cases}$$

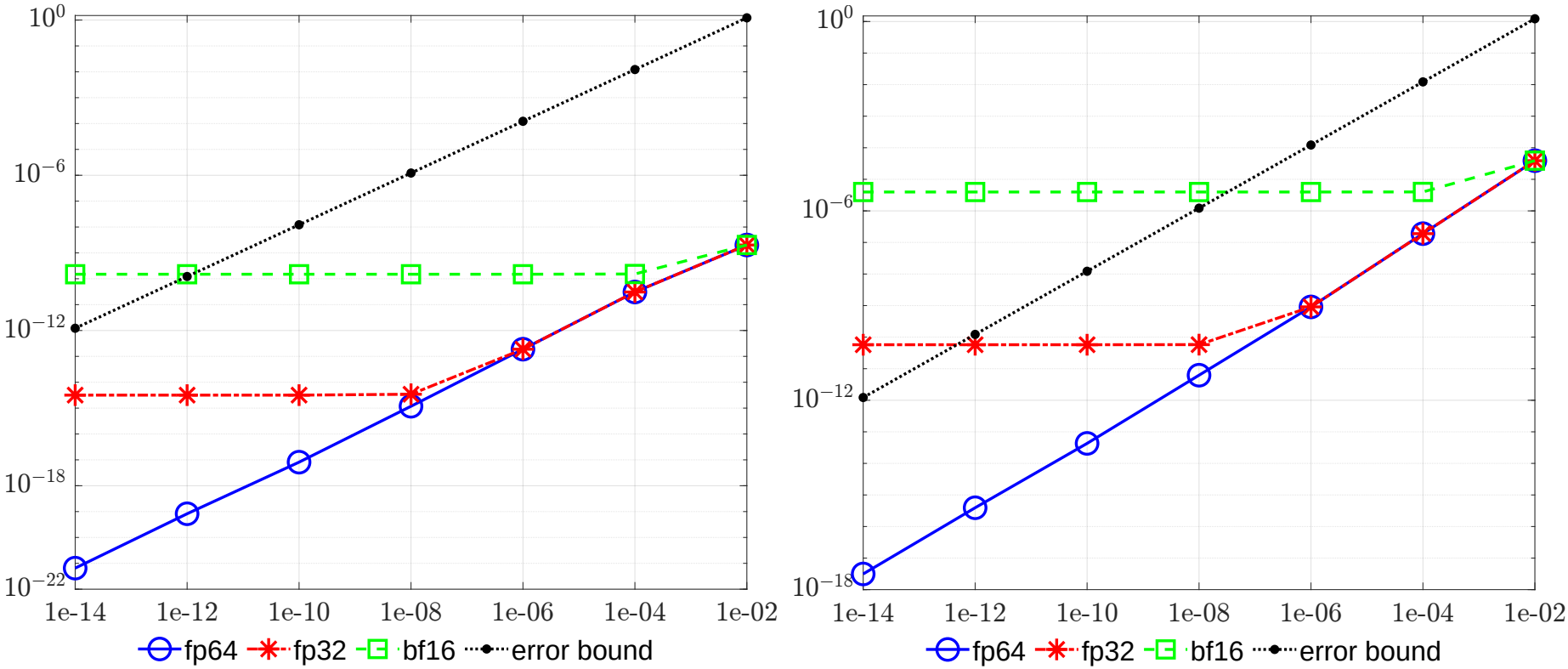
Evaluated at 1D and 2D point sets:

- Set $\mathbf{s_1}$: A set of uniform grid points in $[0, 1]$.
- Set $\mathbf{s_2}$: A set of uniform grid points in $[-1, 1] \times [-1, 1]$.

1. **mat-1** generated by kernel (i) at $\mathbf{s}_1$; $n = 2000$.
2. **mat-2** generated by kernel (ii) at $\mathbf{s}_2$; $n = 2000$.

Matrix–vector Product: Experiments

Setting: $\ell = 8$, x generated from $\mathcal{U}(1, 1)$, in different working precisions



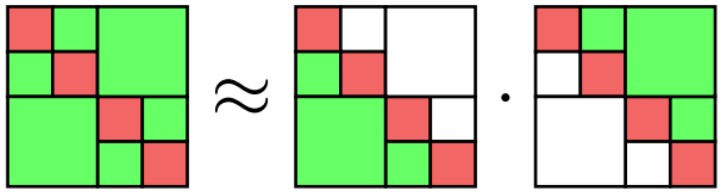
(a) mat-1 (b) mat-2

Figure: x-axis: ε; y-axis: *averaged* $\|b - \widehat{H}x\|_F / \|K\|_F \|x\|_F$ from 10 runs.

- Finite precision/low-rank approximation error dominates as ε varies

LU Factorization: Algorithm

HODLR LU factorization: $H \approx LU$:



Key steps in the recursive algorithm:

1: Partition $H^{(k)}$ into $\begin{bmatrix} H_{11}^{(k+1)} & H_{12}^{(k+1)} \\ H_{21}^{(k+1)} & H_{22}^{(k+1)} \end{bmatrix}$

2: $L_{11}, U_{11} \leftarrow \texttt{HODLR\_LU}(H_{11}^{(k+1)}, k+1)$

3: $U_{12} \leftarrow$ Solve triangular system $L_{11}U_{12} = H_{12}^{(k+1)}$

4: $L_{21} \leftarrow$ Solve triangular system $L_{21}U_{11} = H_{21}^{(k+1)}$

5: $H_{22}^{\varepsilon} \leftarrow H_{22}^{(k+1)} - L_{21}U_{12}$ (possible rank-p truncation to (p, ε)-HODLR)

6: $L_{22}, U_{22} \leftarrow \texttt{HODLR\_LU}(H_{22}^{\varepsilon}, k+1)$

7: $L \leftarrow \begin{bmatrix} L_{11} & \\ L_{21} & L_{22} \end{bmatrix}, U \leftarrow \begin{bmatrix} U_{11} & U_{12} \\ & U_{22} \end{bmatrix}$

- At the bottom level $k = \ell$, dense LU factorization of $H_{ii}^{(\ell)}$ is computed.

LU Factorization: Error Bound

Theorem (Working precision and backward error)

Let $\widehat{H}$ be the mixed-precision ℓ-level HODLR representation. If the LU decomposition of $\widehat{H}$ is computed in a working precision $u \lesssim \varepsilon/n$, then the LU factorization of $\widehat{H}$ satisfies

$$\widehat{L}\widehat{U} = H + \Delta H, \quad \|\Delta H\|_F \lesssim \mathcal{O}(2^\ell)\varepsilon\|H\|_F + \mathcal{O}(2^\ell)\varepsilon\|\widehat{L}\|_F\|\widehat{U}\|_F.$$

- For a relatively large ε (a coarse approximation), we can compute the LU factorization in a low precision without affecting the backward error

LU factorization: Experiments

Example matrix: ex37, in different working precisions

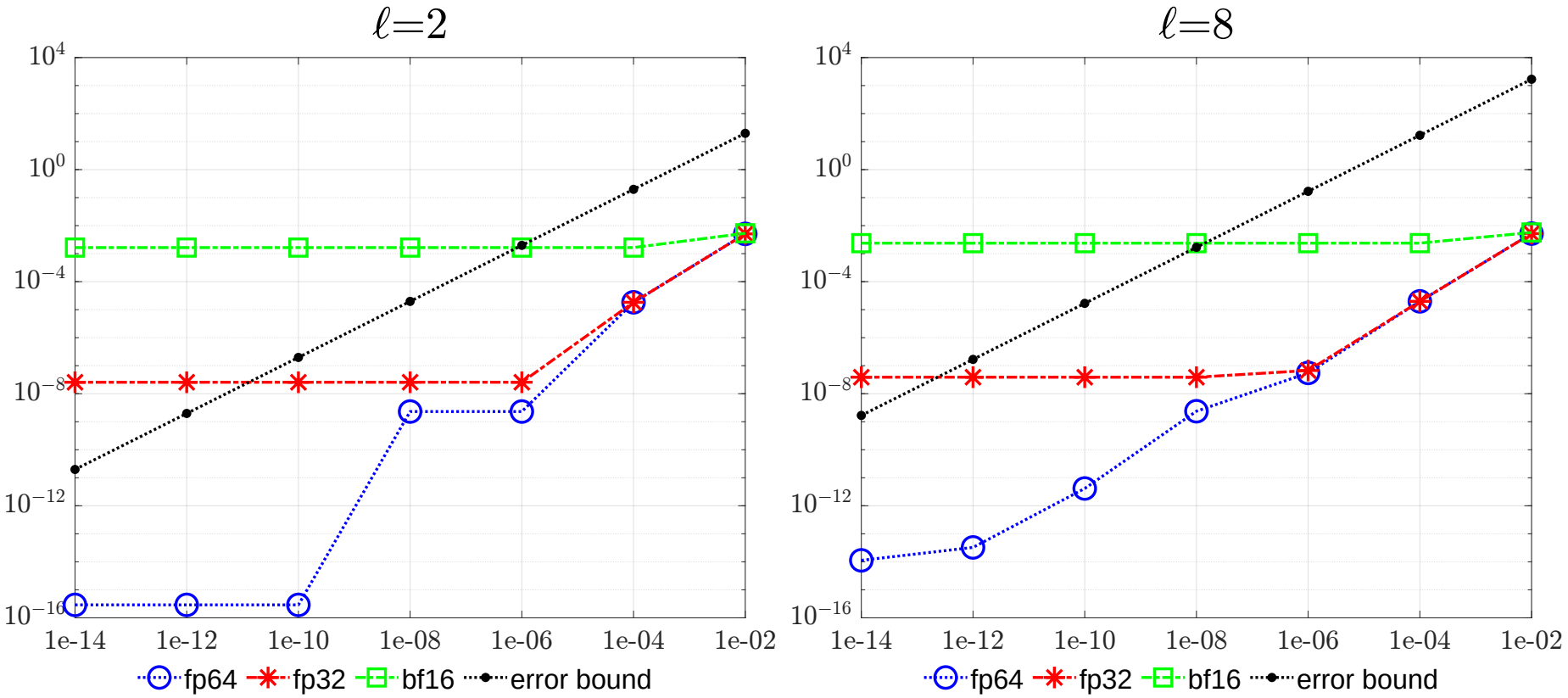


Figure: x-axis: ε; y-axis: relative backward error $\frac{\|\widehat{L}\widehat{U}-A\|_F}{\|A\|_F}$.

- Smaller ε requires higher precision for balancing the overall error
- Low-rank approximation error dominates after around the level $\varepsilon \approx u$

Conclusions

Mixed-precision approach for storing and operating with HODLR matrices:

- Global representation error of mixed-precision HODLR matrices
- Backward error of HODLR matrix–vector products
- Backward error of LU factorization of HODLR matrices

$\rightsquigarrow$ Larger approximation parameter, lower the precision can be used

- Error analyses remain applicable for uniform-precision HODLR matrices

▶ E. Carson, X. Chen, X. Liu. Mixed precision HODLR matrices. SIAM J. Sci. Comput., 47(3):A1408-A1435, 2025.

▶ Software: `https://github.com/chenxinye/mhodlr`

Next?

- Error analysis for fused multiply-add (for the Schur complement in LU)
- Mixed precision in more general hierarchical formats

Conclusions

Mixed-precision approach for storing and operating with HODLR matrices:

- Global representation error of mixed-precision HODLR matrices
- Backward error of HODLR matrix–vector products
- Backward error of LU factorization of HODLR matrices

$\rightsquigarrow$ Larger approximation parameter, lower the precision can be used

- Error analyses remain applicable for uniform-precision HODLR matrices

▶ E. Carson, X. Chen, X. Liu. Mixed precision HODLR matrices. SIAM J. Sci. Comput., 47(3):A1408-A1435, 2025.

▶ Software: `https://github.com/chenxinye/mhodlr`

Next?

- Error analysis for fused multiply-add (for the Schur complement in LU)
- Mixed precision in more general hierarchical formats

Conclusions

Mixed-precision approach for storing and operating with HODLR matrices:

- Global representation error of mixed-precision HODLR matrices
- Backward error of HODLR matrix–vector products
- Backward error of LU factorization of HODLR matrices

$\rightsquigarrow$ Larger approximation parameter, lower the precision can be used

- Error analyses remain applicable for uniform-precision HODLR matrices

▶ E. Carson, X. Chen, X. Liu. Mixed precision HODLR matrices. SIAM J. Sci. Comput., 47(3):A1408-A1435, 2025.

▶ Software: `https://github.com/chenxinye/mhodlr`

Next?

- Error analysis for fused multiply-add (for the Schur complement in LU)
- Mixed precision in more general hierarchical formats