RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030
PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

Fixed-accuracy Randomized Interpolative Decomposition

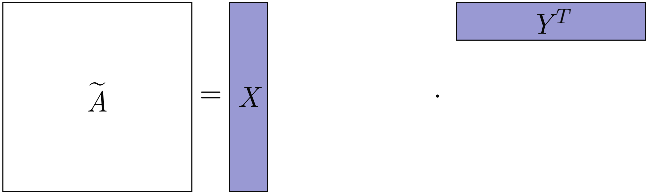
Exa-SofT General Assembly (September 2025)

K. Hoogveld[†]
Supervisors: A. Buttari[†], T. Mary[*]

[†]CNRS-IRIT (Exa-SofT), [*]CNRS-LIP6 (Exa-MA)

Fixed-rank *versus* Fixed-accuracy

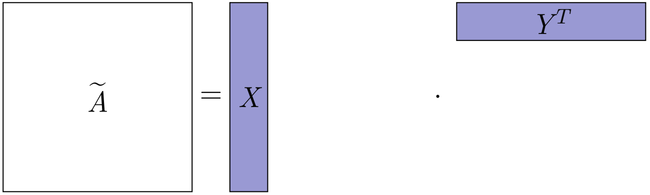
Let A be a dense matrix of size $m \times n$ with $m \geq n$. Take a target rank k. If k is small enough such that the rank-k approximation



requires less storage than the full-rank matrix A, i.e., if $k(m+n) \leq mn$, then XY^T is a low-rank approximation of the original matrix $\Rightarrow$ fixed-rank strategy.

Fixed-rank *versus* Fixed-accuracy

Let A be a dense matrix of size $m \times n$ with $m \geq n$.



In pratice such a rank k is often not known. A target accuracy ε is rather chosen and the low-rank approximation algorithm will determine the associated k_ε for which $\left\|A - \widetilde{A}\right\| \leq \varepsilon \Rightarrow$ fixed-accuracy strategy.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030

PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

cea cnrs inria

Randomized Interpolative Decomposition *(in literature)*

Randomized Interpolative Decomposition (RID)
Fixed-rank

F. Woolfe, E. Liberty, V. Rokhlin and M. Tygert. "A fast randomized algorithm for the approximation of matrices". In: Elsevier ACHA (2008).

1. Sampling Generate the $k \times n$ sketch S for fixed rank k where Ω is a $k \times m$ sampling matrix;

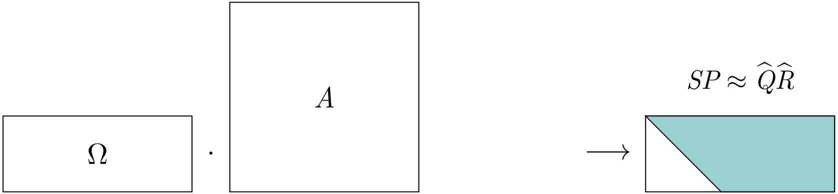
$$\boxed{\Omega} \cdot \boxed{A} \longrightarrow \boxed{S = \Omega A}$$

Randomized Interpolative Decomposition (RID)

Fixed-rank

F. Woolfe, E. Liberty, V. Rokhlin and M. Tygert. "A fast randomized algorithm for the approximation of matrices". In: Elsevier ACHA (2008).

1. Sampling Generate the $k \times n$ sketch S for fixed rank k where Ω is a $k \times m$ sampling matrix;
2. QRCP Compute a QRCP factorization of S;

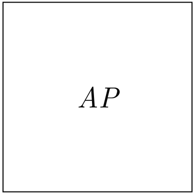


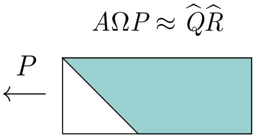
Randomized Interpolative Decomposition (RID)

Fixed-rank

F. Woolfe, E. Liberty, V. Rokhlin and M. Tygert. "A fast randomized algorithm for the approximation of matrices". In: Elsevier ACHA (2008).

1. Sampling Generate the $k \times n$ sketch S for fixed rank k where Ω is a $k \times m$ sampling matrix;
2. QRCP Compute a QRCP factorization of S;
3. $S \to A$ Using P.

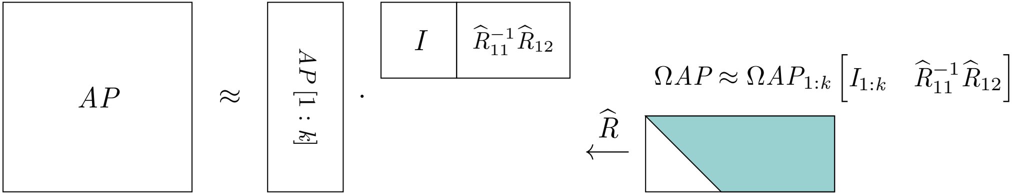




Randomized Interpolative Decomposition (RID)
Fixed-rank

F. Woolfe, E. Liberty, V. Rokhlin and M. Tygert. "A fast randomized algorithm for the approximation of matrices". In: Elsevier ACHA (2008).

1. Sampling Generate the $k \times n$ sketch S for fixed rank k where Ω is a $k \times m$ sampling matrix;
2. QRCP Compute a QRCP factorization of S;
3. $S \rightarrow A$ Using P and $\widehat{R}$.



Randomized Interpolative Decomposition (RID)

Fixed-accuracy (*in literature*)

T. Mary, I. Yamazaki, J. Kurzak, P. Luszczek, S. Tomov, and J. Dongarra. "Performance of random sampling for computing low-rank approximations of a dense matrix on GPUs". In: SC (2015).

```
repeat
   S_i = A^T Ω_i
   Q_i, R_i = qr(S_i, NoPivot)
   Q_i, R_i = qr(Q_i − Q Q^T Q_i)
   B_i = Q_i^T A^T
   A^T = A^T − Q_i B_i
   if norm(A) ≤ ε then stop
end
return S = [S_1 ... S_i]
```

- Reliable
- High complexity

P-G. Martinsson and S. Voronin. "A Randomized Blocked Algorithm for Efficiently Computing Rank-revealing Factorizations of Matrices". In: SIAM SISC (2016).

```
S_i = A^T Ω_i
repeat
   Q_i, R_i = qr(S_i, NoPivot)
   S_{i+1} = A^T Ω_{i+1}
   S_{i+1} = S_{i+1} − Q Q^T S_{i+1}
   if norm(S_{i+1}) ≤ ε then stop
end
return S = [S_1 ... S_i]
```

- Low complexity
- Unreliable

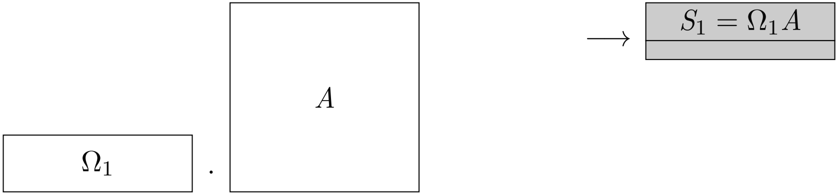
N. Halko, P-G. Martinsson and J. Tropp. "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions". In: SIAM Review (2011).

A novel approach

Randomized Interpolative Decomposition (RID)
by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. "Randomized QR with column pivoting". In: SIAM Journal on Scientific Computing (2017).

Let b be the block size and p the oversampling factor.

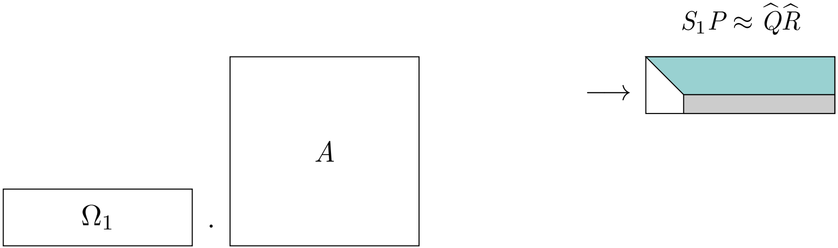


Randomized Interpolative Decomposition (RID)

by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. "Randomized QR with column pivoting". In: SIAM Journal on Scientific Computing (2017).

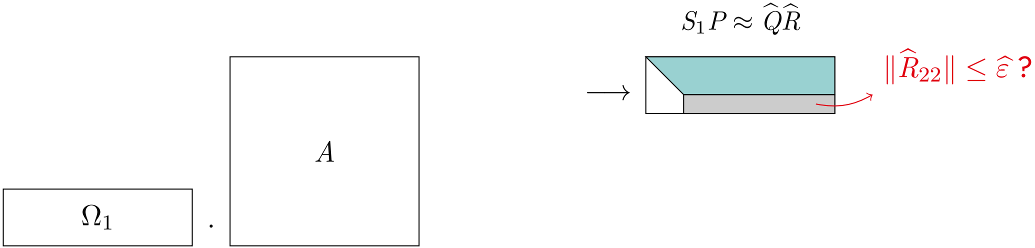
Let b be the block size and p the oversampling factor.



Randomized Interpolative Decomposition (RID)
by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. "Randomized QR with column pivoting". In: SIAM Journal on Scientific Computing (2017).

Let b be the block size and p the oversampling factor.

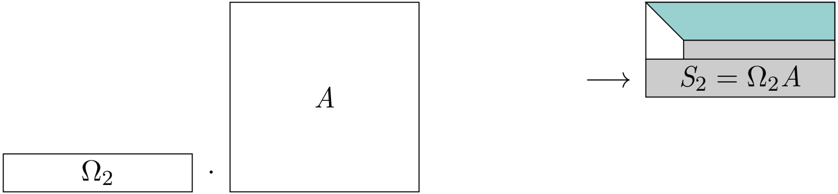


Randomized Interpolative Decomposition (RID) by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. "Randomized QR with column pivoting". In: SIAM Journal on Scientific Computing (2017).

Let b be the block size and p the oversampling factor.

$$S = \begin{bmatrix} \widehat{R} \\ S_2 \end{bmatrix}$$



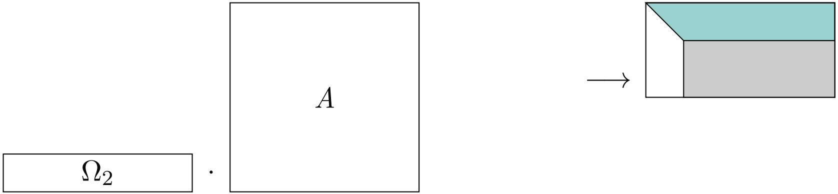
Randomized Interpolative Decomposition (RID)

by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. "Randomized QR with column pivoting". In: SIAM Journal on Scientific Computing (2017).

Let b be the block size and p the oversampling factor.

$$S = \begin{bmatrix} \widehat{R} \\ \widehat{S_2} \end{bmatrix}$$

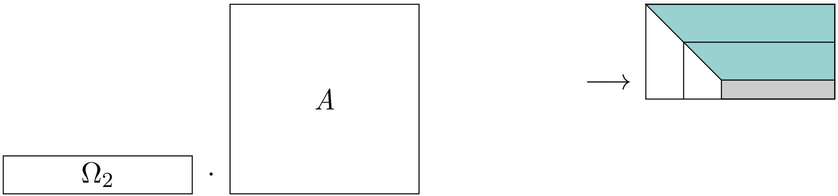


Randomized Interpolative Decomposition (RID)
by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. “Randomized QR with column pivoting”. In: SIAM Journal on Scientific Computing (2017).

Let b be the block size and p the oversampling factor.

$$SP \approx \widehat{Q}\widehat{R}$$

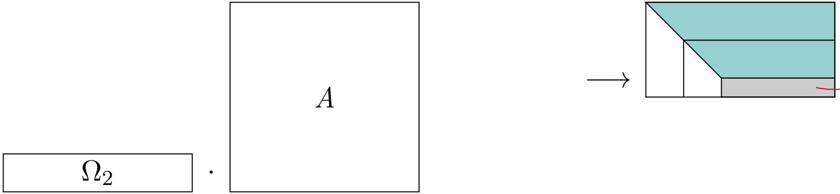


Randomized Interpolative Decomposition (RID)
by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. "Randomized QR with column pivoting". In: SIAM Journal on Scientific Computing (2017).

Let b be the block size and p the oversampling factor.

$$SP \approx \widehat{Q}\widehat{R}$$



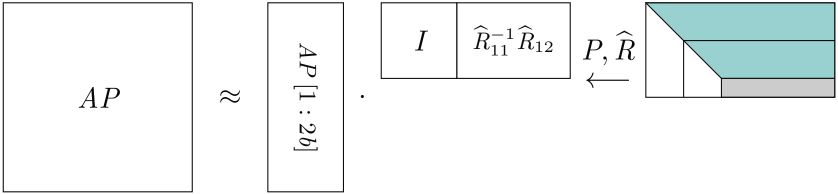
$\|\widehat{R}_{22}\| \leq \widehat{\varepsilon}$?

Randomized Interpolative Decomposition (RID)
by E. Agullo, A. Buttari, K. Hoogveld and T. Mary

J. Duersch and M. Gu. "Randomized QR with column pivoting". In: SIAM Journal on Scientific Computing (2017).

Let b be the block size and p the oversampling factor.

$$\Omega AP \approx \Omega AP_{1:2b} \begin{bmatrix} I_{1:2b} & \widehat{R}_{11}^{-1}\widehat{R}_{12} \end{bmatrix}$$

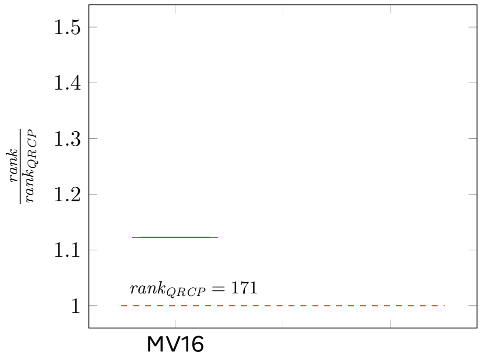


Experimental results

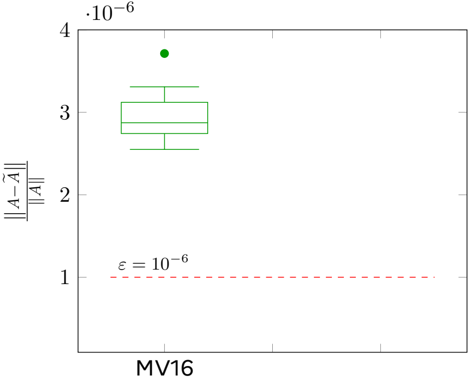
Backward error

Let us assess the numerical behavior of these fixed-accuracy RID variants on a 2048×2048 Heat test matrix using a Gaussian sampling matrix.

Relative truncation rank



Relative backward error

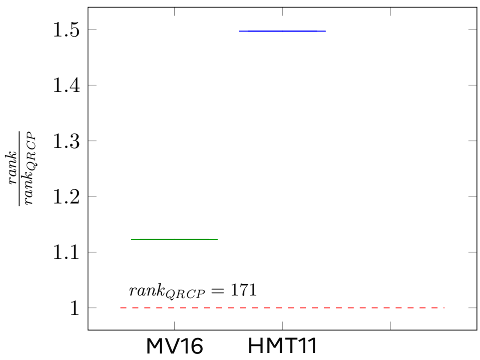


Experimental results

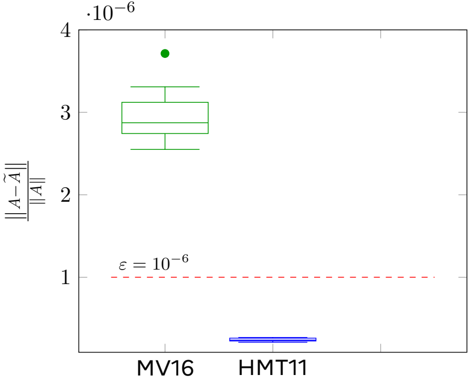
Backward error

Let us assess the numerical behavior of these fixed-accuracy RID variants on a 2048×2048 Heat test matrix using a Gaussian sampling matrix.

Relative truncation rank



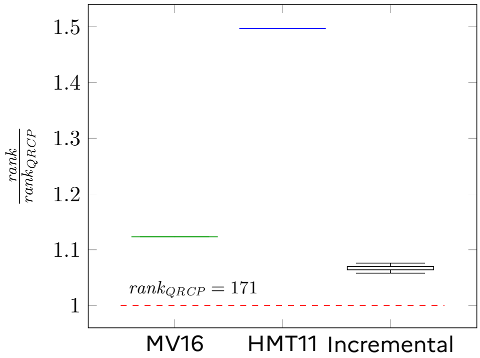
Relative backward error



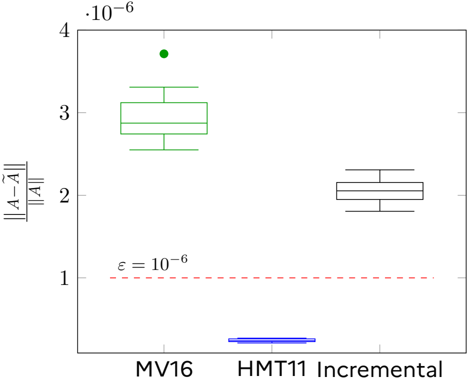
Experimental results
Backward error

Let us assess the numerical behavior of these fixed-accuracy RID variants on a 2048×2048 Heat test matrix using a Gaussian sampling matrix.

Relative truncation rank



Relative backward error



Variants

Variants

Subsampled Randomized Hadamard Transform (SRHT)

Algorithm	Gaussian sketching	
Incremental RID	$2mnk + 4nk^2 - \frac{8}{3}k^3$	
HMT11 RID	$2mnk + 6nk^2 - \frac{8}{3}k^3$	
MV16 RID	$6mnk + 6nk^2 - \frac{8}{3}k^3$	

Variants

Subsampled Randomized Hadamard Transform (SRHT)

Algorithm	Gaussian sketching	SRHT sketching
Incremental RID	$2mnk + 4nk^2 - \frac{8}{3}k^3$	$2mnk\frac{\log(b)}{b} + 4nk^2 - \frac{8}{3}k^3$
HMT11 RID	$2mnk + 6nk^2 - \frac{8}{3}k^3$	$2mnk\frac{\log(b)}{b} + 6nk^2 - \frac{8}{3}k^3$
MV16 RID	$6mnk + 6nk^2 - \frac{8}{3}k^3$	$2mnk\frac{\log(b)}{b} + 4mnk + 6nk^2 - \frac{8}{3}k^3$

Key Ideas:

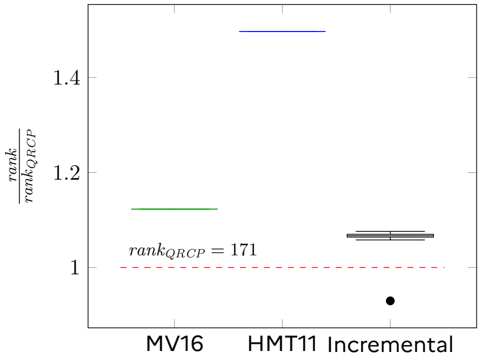
- $\Omega \rightarrow$ Sampling matrix defined as $\Omega = \sqrt{\frac{m}{l}}RHD$;
- $H \rightarrow$ Fast to apply via Fast Hadamard Transform (FHT) in $\mathcal{O}(m \log m)$ time;
- $D \rightarrow$ Random sign flipping spreads out information;
- $R \rightarrow$ Uniform row sampling reduces dimensionality.

Experimental results

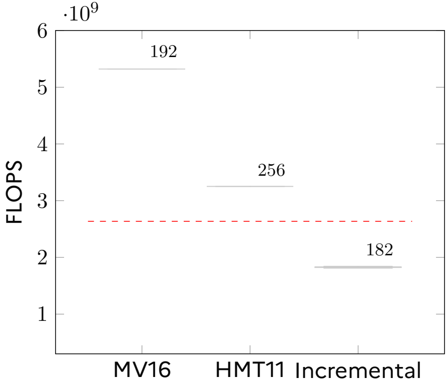
Subsampled Randomized Hadamard Transform (SRHT)

Let us assess the numerical behavior of these fixed-accuracy RID variants on a 2048×2048 Heat test matrix using a SRHT sampling matrix.

Relative truncation rank



Operational complexity

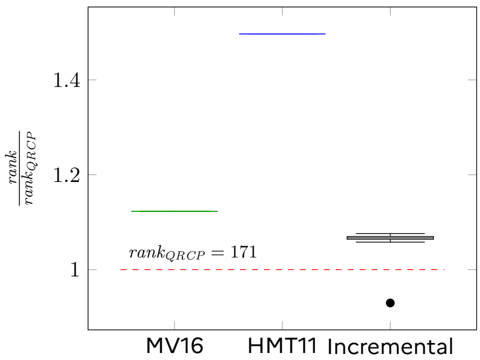


Experimental results

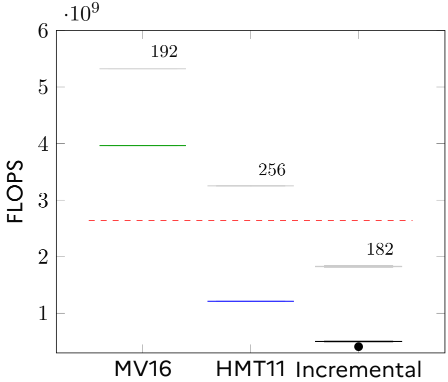
Subsampled Randomized Hadamard Transform (SRHT)

Let us assess the numerical behavior of these fixed-accuracy RID variants on a 2048×2048 Heat test matrix using a SRHT sampling matrix.

Relative truncation rank



Operational complexity

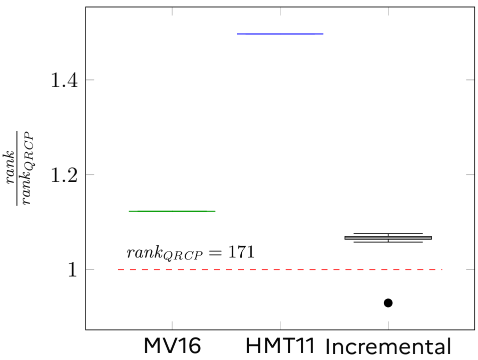


Experimental results

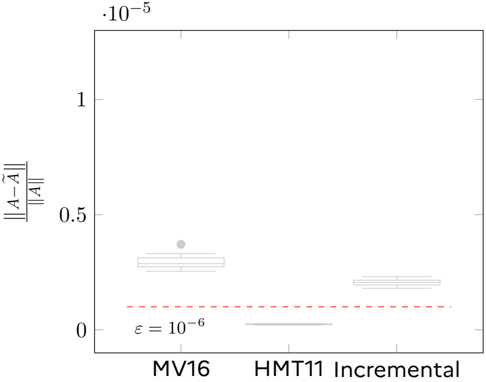
Subsampled Randomized Hadamard Transform (SRHT)

Let us assess the numerical behavior of these fixed-accuracy RID variants on a 2048×2048 Heat test matrix using a SRHT sampling matrix.

Relative truncation rank



Relative backward error

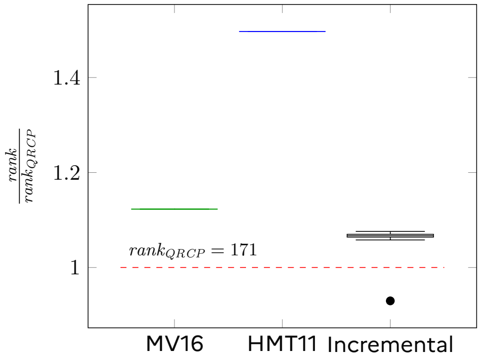


Experimental results

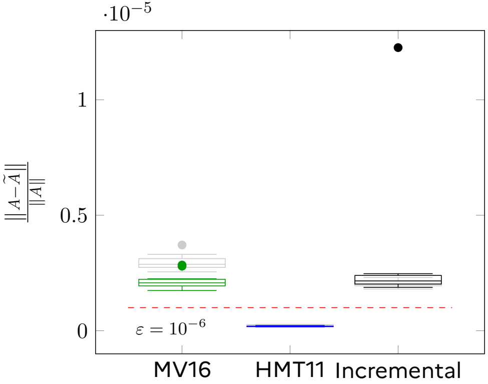
Subsampled Randomized Hadamard Transform (SRHT)

Let us assess the numerical behavior of these fixed-accuracy RID variants on a 2048×2048 Heat test matrix using a SRHT sampling matrix.

Relative truncation rank



Relative backward error



Variants

Orthogonal post-processing

Sketch based (ORID$_S$)

$$\widetilde{Q}, \widetilde{R} = \texttt{qr}(A[:,\ 1{:}k],\ \texttt{NoPivot}())$$
$$R = \widetilde{R}\ [\texttt{I}(k),\ \widehat{R}_{11}^{-1}\widehat{R}_{12}]$$

- Orthogonal basis at a low operational cost

Original matrix based (ORID$_A$)

$$\widetilde{Q}, \widetilde{R} = \texttt{qr}(A[:,\ 1{:}k],\ \texttt{NoPivot}())$$
$$R = [\ \widetilde{R},\ \widetilde{Q}^T A_{1:k,k+1:n}]$$

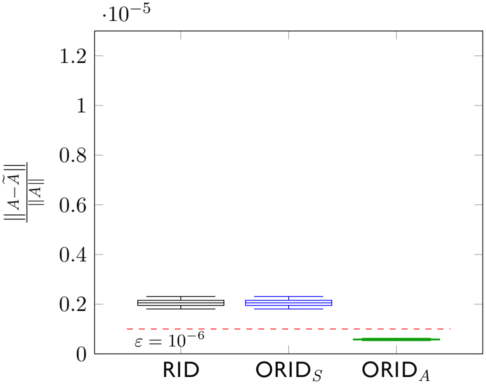
- Orthogonal basis with good approximation quality

Algorithm	Gaussian sketching
Incremental RID	$2mnk + 4nk^2 - \frac{8}{3}k^3$
Incremental ORID$_S$	$2mnk + 4nk^2 + 4mk^2 - 2k^3$
Incremental ORID$_A$	$4mnk + 5nk^2 + 2mk^2 - 3k^3$

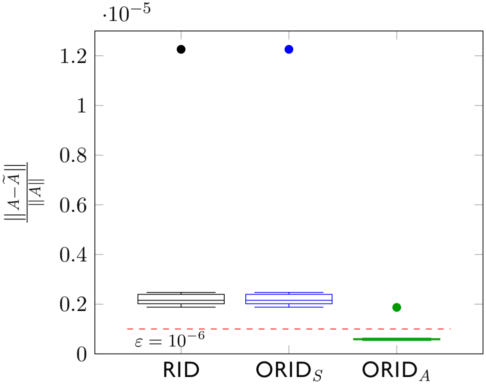
Experimental results
Backward error

Let us compare, using again the Heat test matrix, the post-processing techniques for incremental RID.

Gaussian sketching



SRHT sketching



RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030 PROGRAMME DE RECHERCHE NUMÉRIQUE POUR L'EXASCALE

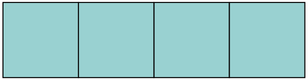
cea cnrs inria

Software

Software
Tournament pivoting

M. Beaupère and L. Grigori. "Communication avoiding low rank approximation based on QR with tournament pivoting". In: HAL (2021).

Let's take an example where the sketch is partitioned in 4 blocks of size $b \times b$.



Goal: Task-based column selection in a parallel distributed setting in $Chameleon$.

Key Ideas:

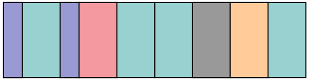
- Each task computes QRCP on a block pair;
- Elegant implementation using REDUX from $StarPU$.

Software

Tournament pivoting

M. Beaupère and L. Grigori. "Communication avoiding low rank approximation based on QR with tournament pivoting". In: HAL (2021).

Let's take an example where the sketch is partitioned in 4 blocks of size $b \times b$.



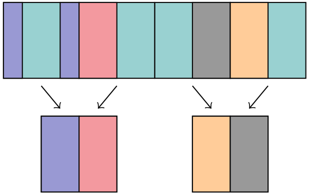
Goal: Task-based column selection in a parallel distributed setting in $Chameleon$.

Key Ideas:

- Each task computes QRCP on a block pair;
- Elegant implementation using REDUX from $StarPU$.

M. Beaupère and L. Grigori. "Communication avoiding low rank approximation based on QR with tournament pivoting". In: HAL (2021).

Let's take an example where the sketch is partitioned in 4 blocks of size $b \times b$.



Goal: Task-based column selection in a parallel distributed setting in $Chameleon$.

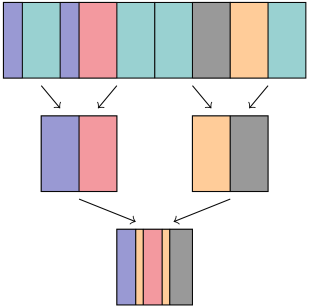
Key Ideas:

- Each task computes QRCP on a block pair;
- Elegant implementation using REDUX from $StarPU$.

Software
Tournament pivoting

M. Beaupère and L. Grigori. "Communication avoiding low rank approximation based on QR with tournament pivoting". In: HAL (2021).

Let's take an example where the sketch is partitioned in 4 blocks of size $b \times b$.



Goal: Task-based column selection in a parallel distributed setting in $\mathit{Chameleon}$.

Key Ideas:

- Each task computes QRCP on a block pair;
- Elegant implementation using REDUX from StarPU.

$\Rightarrow$ MR in StarPU on SCRATCH data in REDUX.

Column permutation

Goal: Permute columns of a distributed matrix based on the intership work of Matteo Marcos (supervised by Alycia Lisito, Pierre Ramet and Mathieu Faverge).

Key Ideas of his internship:

- Row permutation of matrices with irregular distributions in *Chameleon*;
- Communication volume reduction by use of workspaces;
- **Extension to column permutation**.

Conclusion

Conclusion and future work

Conclusion

- Extended RID to a novel fixed-accuracy variant;
- Lead a unified operational complexity analysis;
- `Julia` numerical behavior assessment.

Conclusion and future work

Conclusion

- Extended RID to a novel fixed-accuracy variant;
- Lead a unified operational complexity analysis;
- `Julia` numerical behavior assessment.

Future work

- Write task-based SRHT sketching;
- Consolidate Tournament pivoting implementation;
- Familiarize with the permutation developments by Matteo Marcos.

Thank you for your attention. Any questions ?

As part of the "France 2030" initiative, this work has benefited from a national grant managed by the French National Research Agency (Agence Nationale de la Recherche) attributed to the Exa-SofT project of the NumPEx PEPR program, under the reference ANR-22-EXNU-0003.