Asymptotic Complexity of Low-rank Sparse Direct Solvers with Sparse Right-hand Sides

Theo Mary, CNRS
Joint work with P. Amestoy, A. Buttari, J.-Y. L'Excellent, and G. Moreau

SIAM CSC, 2020

Systems of linear equations:

$Ax = b$, where A is sparse. In direct methods, 3 phases:

- analysis: nested dissection;
- factorization: $A \to LU$;
- solve: $Ly = b$ and $Ux = y$.

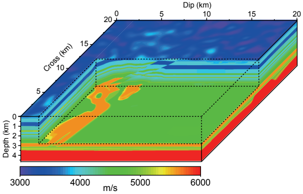
	2D ($N \times N$)	3D ($N \times N \times N$)
$\mathcal{C}_{\text{fac}}$	$\Theta(N^3)$	$\Theta(N^6)$
$\mathcal{C}_{\text{sol}}$ (per RHS)	$\Theta(N^2 \log N)$	$\Theta(N^4)$

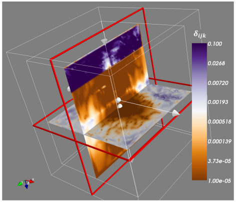
Complexities on regular 2D/3D problems (N is the grid size)

Factorization is usually the most expensive part, however...

Applications with many RHS

Several important applications possess **many RHS**, e.g.,
exploration geophysics: FWI (Helmholtz), CSEM (Maxwell)





In 3D domains, $\Theta(N^2)$ sources

$$\Rightarrow \mathcal{C}_{\text{sol}} = \Theta(N^4) \times \Theta(N^2) = \Theta(N^6) \equiv \mathcal{C}_{\text{fac}}$$

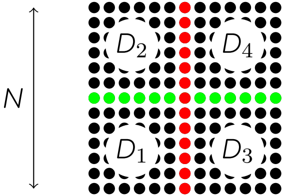
Practical $AX = B$ example, matrix S21 (CSEM application):

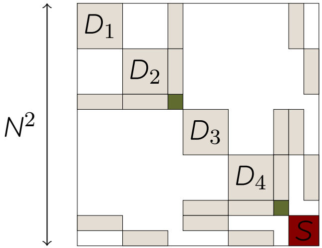
n	n_{rhs}	$\text{nnz}(A)/n$	$\text{nnz}(B)/n_{\text{rhs}}$	$\mathcal{T}_{\text{fac}}$	$\mathcal{T}_{\text{sol}}$
20.6 M	12340	13	9.5	10825	15029

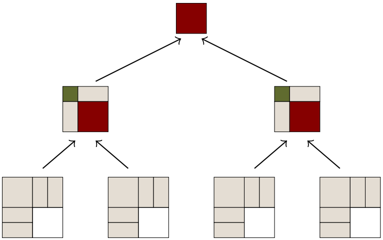
Run on EOS computer (90 MPI)

To tackle large scale problems,
crucial to exploit **all types** of sparsity of the problem

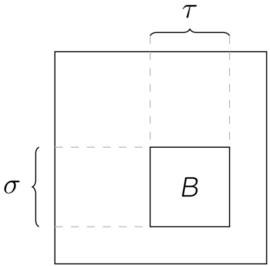
Exploiting the sparsity of A: nested dissection

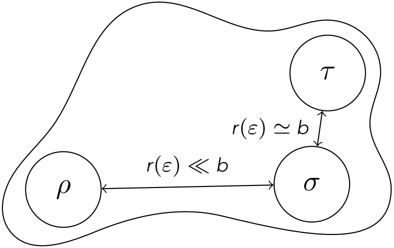






- Factorization ($A = LU$): bottom up traversal, dense factorization at each node
- Forward solve ($Ly = b$): bottom up traversal, dense solve
- Backward solve ($Ux = y$): top down traversal, dense solve





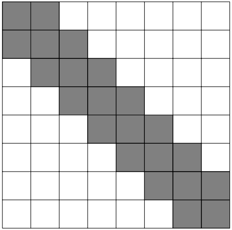
A block B represents the interaction between two subdomains.
Far away subdomains $\Rightarrow$ block of low numerical rank:

$$\begin{array}{cccc} B & \approx & X & Y^T \\ b \times b & & b \times r(\varepsilon) & r(\varepsilon) \times b \end{array}$$

with $r(\varepsilon) \ll b$ such that $\|B - XY^T\| \leq \varepsilon$

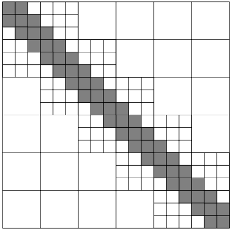
Low-rank matrix formats

Block low-rank (BLR)



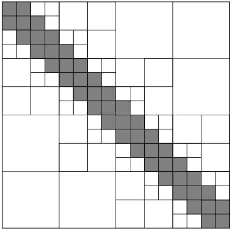
$\Theta(m^3) \to \Theta(m^2)$ flops
$\Theta(m^2) \to \Theta(m^{3/2})$ space

Multilevel BLR (MBLR)



$\Theta(m^3) \to \Theta(m^{\frac{\ell+3}{\ell+1}})$ flops
$\Theta(m^2) \to \Theta(m^{\frac{\ell+2}{\ell+1}})$ space

Hierarchical ($\mathcal{H}$, HSS, ...)



$\Theta(m^3) \to \widetilde{\Theta}(m)$ flops
$\Theta(m^2) \to \widetilde{\Theta}(m)$ space

For all formats, the gain in flops $\mathcal{G}_{\text{flops}}$ is asymptotically greater than the gain in space $\mathcal{G}_{\text{space}}$ (to be precise $\mathcal{G}_{\text{flops}} = \mathcal{G}_{\text{space}}^2$)

Complexity of sparse low-rank direct solvers

Sparsity and data sparsity can be combined by using low-rank formats to approximate the dense separators

2D regular problem

	$\mathcal{C}_{\text{fac}}$	$\mathcal{C}_{\text{sol}}$ (per RHS)
FR	$\Theta(N^3)$	$\Theta(N^2 \log N)$
BLR	$\Theta(N^2 \log N)$	$\Theta(N^2)$
$\mathcal{H}$, MBLR	$\Theta(N^2)$	$\Theta(N^2)$

Complexity of sparse low-rank direct solvers

Sparsity and data sparsity can be combined by using low-rank formats to approximate the dense separators

3D regular problem

	$\mathcal{C}_{\text{fac}}$	$\mathcal{C}_{\text{sol}}$ (per RHS)
FR	$\Theta(N^6)$	$\Theta(N^4)$
BLR	$\Theta(N^4)$	$\Theta(N^3 \log N)$
MBLR ($\ell = 2$)	$\Theta(N^{10/3})$	$\Theta(N^3)$
MBLR ($\ell = 3$)	$\Theta(N^3 \log N)$	$\Theta(N^3)$
$\mathcal{H}$, MBLR ($\ell > 3$)	$\Theta(N^3)$	$\Theta(N^3)$

Complexity of sparse low-rank direct solvers

Sparsity and data sparsity can be combined by using low-rank formats to approximate the dense separators

3D regular problem

	$\mathcal{C}_{\text{fac}}$	$\mathcal{C}_{\text{sol}}$ (per RHS)
FR	$\Theta(N^6)$	$\Theta(N^4)$
BLR	$\Theta(N^4)$	$\Theta(N^3 \log N)$
MBLR ($\ell = 2$)	$\Theta(N^{10/3})$	$\Theta(N^3)$
MBLR ($\ell = 3$)	$\Theta(N^3 \log N)$	$\Theta(N^3)$
$\mathcal{H}$, MBLR ($\ell > 3$)	$\Theta(N^3)$	$\Theta(N^3)$

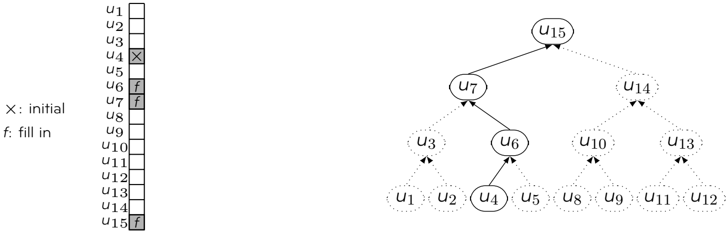
Back to the CSEM example (S21 matrix):

	$\mathcal{T}_{\text{fac}}$	$\mathcal{T}_{\text{sol}}$
FR	10825	15029
BLR	1568	7046

$\Rightarrow$ **With data sparsity and many RHS, the solve phase becomes asymptotically dominant!**

Gilbert and Liu

- **Forward solve**: nodes associated with RHS zeros can be pruned
  ⇒ only need to traverse branches from RHS nonzeros to root
- If X is also sparse (only part of the solution needed), then same thing applies to backward solve



Complexity reduction achieved by sparsity and data sparsity of A well known BUT reduction from sparsity of B never analyzed!

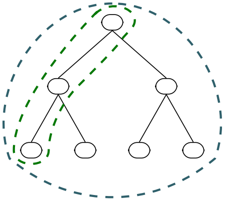
⇒ **What asymptotic gain can we obtain by exploiting sparse RHS?**

RHS sparsity model

Consider a RHS with nnz nonzeros:

- If nnz $= \Theta(1)$, then the forward solve amounts to traverse $\Theta(1)$ branches $\Rightarrow$ its complexity is that of the critical path

  The gain from exploiting RHS sparsity then is



$$\mathcal{G}_{\text{spRHS}}(N) = \frac{\mathcal{C}_{\text{fwd}}(N)}{\mathcal{C}^{c}_{\text{fwd}}(N)}$$

  where
  - $\mathcal{C}_{\text{fwd}}(N)$ is the complexity of the forward solve
  - $\mathcal{C}^{c}_{\text{fwd}}(N)$ is the complexity of its critical path

- Applications where nnz $= \Theta(1)$ are actually very common (cf. our CSEM example, FWI, ...)

$\Rightarrow$ Let us first assume nnz $= \Theta(1)$ (will generalize later)

Complexity analysis

Consider a separator tree with L levels and n_ℓ separators of order m_ℓ at level ℓ. Then

$$\mathcal{C}_{\mathrm{fwd}}(N) = \sum_{\ell}^{L} n_\ell \times \mathcal{C}_{\mathrm{dense}}(m_\ell)$$

$$\mathcal{C}^{c}_{\mathrm{fwd}}(N) = \sum_{\ell}^{L} \cancel{n_\ell} \times \mathcal{C}_{\mathrm{dense}}(m_\ell)$$

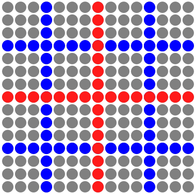
where $\mathcal{C}_{\mathrm{dense}}(m_\ell) = \Theta(m_\ell^{\alpha})$ is the complexity of the dense solve.

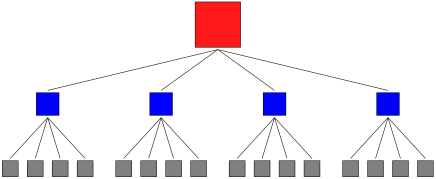
Consider a separator tree with L levels and n_ℓ separators of order m_ℓ at level ℓ. Then

$$\mathcal{C}_{\text{fwd}}(N) = \sum_{\ell}^{L} n_\ell \times \mathcal{C}_{\text{dense}}(m_\ell)$$

$$\mathcal{C}^{c}_{\text{fwd}}(N) = \sum_{\ell}^{L} \cancel{n_\ell} \times \mathcal{C}_{\text{dense}}(m_\ell)$$

where $\mathcal{C}_{\text{dense}}(m_\ell) = \Theta(m_\ell^\alpha)$ is the complexity of the dense solve.





In the 2D case (with cross separators), $L = \log_2 N$, $n\ell = 4^\ell$, and $m_\ell = \Theta(N/2^\ell)$

Complexity analysis, cont'd

$$\mathcal{C}_{\mathsf{fwd}}(N) = \sum_{\ell=0}^{\log_2 N} \Theta(4^\ell \times (N/2^\ell)^\alpha) = \Theta(N^\alpha \sum_{\ell=0}^{\log_2 N} 2^{(2-\alpha)\ell})$$

$$= \begin{cases} \Theta(N^2 \log N) & \text{if } \alpha = 2 \\ \Theta(N^\alpha) \times \frac{2^{(2-\alpha)\log_2 N} - 1}{2^{2-\alpha} - 1} = \Theta(N^2) & \text{if } \alpha < 2 \end{cases}$$

Complexity analysis, cont'd

$$\mathcal{C}_{\mathsf{fwd}}(N) = \sum_{\ell=0}^{\log_2 N} \Theta(4^\ell \times (N/2^\ell)^\alpha) = \Theta(N^\alpha \sum_{\ell=0}^{\log_2 N} 2^{(2-\alpha)\ell})$$

$$= \begin{cases} \Theta(N^2 \log N) & \text{if } \alpha = 2 \\ \Theta(N^\alpha) \times \frac{2^{(2-\alpha)\log_2 N} - 1}{2^{2-\alpha} - 1} = \Theta(N^2) & \text{if } \alpha < 2 \end{cases}$$

$$\mathcal{C}^c_{\mathsf{fwd}}(N) = \sum_{\ell=0}^{\log_2 N} \Theta(\cancel{4^\ell} \times (N/2^\ell)^\alpha) = \Theta(N^\alpha \sum_{\ell=0}^{\log_2 N} 2^{-\alpha\ell})$$

$$= \Theta(N^\alpha) \times \underbrace{\frac{1 - 2^{-\alpha \log_2 N}}{1 - 2^{-\alpha}}}_{\Theta(1)} = \Theta(N^\alpha)$$

Complexity analysis, cont'd

$$\mathcal{C}_{\mathsf{fwd}}(N) = \sum_{\ell=0}^{\log_2 N} \Theta(4^\ell \times (N/2^\ell)^\alpha) = \Theta(N^\alpha \sum_{\ell=0}^{\log_2 N} 2^{(2-\alpha)\ell})$$

$$= \begin{cases} \Theta(N^2 \log N) & \text{if } \alpha = 2 \\ \Theta(N^\alpha) \times \frac{2^{(2-\alpha)\log_2 N} - 1}{2^{2-\alpha} - 1} = \Theta(N^2) & \text{if } \alpha < 2 \end{cases}$$

$$\mathcal{C}^c_{\mathsf{fwd}}(N) = \sum_{\ell=0}^{\log_2 N} \Theta(\cancel{4^\ell} \times (N/2^\ell)^\alpha) = \Theta(N^\alpha \sum_{\ell=0}^{\log_2 N} 2^{-\alpha\ell})$$

$$= \Theta(N^\alpha) \times \underbrace{\frac{1 - 2^{-\alpha \log_2 N}}{1 - 2^{-\alpha}}}_{\Theta(1)} = \Theta(N^\alpha)$$

Conclusion: $\mathcal{G}^{2D}_{\mathsf{spRHS}}(N) = \begin{cases} \Theta(\log N) & \text{if } \alpha = 2 \quad \text{(FR)} \\ \Theta(N^{2-\alpha}) & \text{if } \alpha < 2 \quad \text{(LR)} \end{cases}$

Same applies for 3D problems.

	$\mathcal{G}^{2D}_{\text{spRHS}}(N)$	$\mathcal{G}^{3D}_{\text{spRHS}}(N)$
FR ($\alpha = 2$)	$\Theta(\log N)$	$\Theta(1)$
BLR ($\alpha = 1.5$)	$\Theta(N^{\frac{1}{2}})$	$\Theta(\log N)$
MBLR ($\alpha = \frac{\ell+2}{\ell+1}$)	$\Theta(N^{\frac{\ell}{\ell+1}})$	$\Theta(N^{\frac{\ell-1}{\ell+1}})$
Hierarchical ($\alpha = 1$)	$\Theta(N)$	$\Theta(N)$

Asymptotic value of $\mathcal{G}_{\text{spRHS}}(N)$ increases as α decreases

$\Leftrightarrow$

Gain from exploiting RHS sparsity increases with data sparsity

$$\frac{\mathcal{C}_{LR}}{\mathcal{C}_{LR+spRHS}} \gg \frac{\mathcal{C}_{FR}}{\mathcal{C}_{FR+spRHS}}$$

Same applies for 3D problems.

	$\mathcal{G}^{2D}_{\text{spRHS}}(N)$	$\mathcal{G}^{3D}_{\text{spRHS}}(N)$
FR ($\alpha = 2$)	$\Theta(\log N)$	$\Theta(1)$
BLR ($\alpha = 1.5$)	$\Theta(N^{\frac{1}{2}})$	$\Theta(\log N)$
MBLR ($\alpha = \frac{\ell+2}{\ell+1}$)	$\Theta(N^{\frac{\ell}{\ell+1}})$	$\Theta(N^{\frac{\ell-1}{\ell+1}})$
Hierarchical ($\alpha = 1$)	$\Theta(N)$	$\Theta(N)$

Asymptotic value of $\mathcal{G}_{\text{spRHS}}(N)$ increases as α decreases

$\Leftrightarrow$

Gain from exploiting RHS sparsity increases with data sparsity

$$\frac{\mathcal{C}_{LR}}{\mathcal{C}_{LR+spRHS}} \gg \frac{\mathcal{C}_{FR}}{\mathcal{C}_{FR+spRHS}} \quad \Leftrightarrow \quad \frac{\mathcal{C}_{FR+spRHS}}{\mathcal{C}_{LR+spRHS}} \gg \frac{\mathcal{C}_{FR}}{\mathcal{C}_{LR}}$$

Complexity bounds and interpretation

Same applies for 3D problems.

	$\mathcal{G}^{2D}_{\text{spRHS}}(N)$	$\mathcal{G}^{3D}_{\text{spRHS}}(N)$
FR ($\alpha = 2$)	$\Theta(\log N)$	$\Theta(1)$
BLR ($\alpha = 1.5$)	$\Theta(N^{\frac{1}{2}})$	$\Theta(\log N)$
MBLR ($\alpha = \frac{\ell+2}{\ell+1}$)	$\Theta(N^{\frac{\ell}{\ell+1}})$	$\Theta(N^{\frac{\ell-1}{\ell+1}})$
Hierarchical ($\alpha = 1$)	$\Theta(N)$	$\Theta(N)$

Asymptotic value of $\mathcal{G}_{\text{spRHS}}(N)$ increases as α decreases

$\Leftrightarrow$

Gain from exploiting RHS sparsity increases with data sparsity

$\Leftrightarrow$

Gain from exploiting data sparsity increases with RHS sparsity

$$\frac{\mathcal{C}_{LR}}{\mathcal{C}_{LR+spRHS}} \gg \frac{\mathcal{C}_{FR}}{\mathcal{C}_{FR+spRHS}} \quad \Leftrightarrow \quad \frac{\mathcal{C}_{FR+spRHS}}{\mathcal{C}_{LR+spRHS}} \gg \frac{\mathcal{C}_{FR}}{\mathcal{C}_{LR}}$$

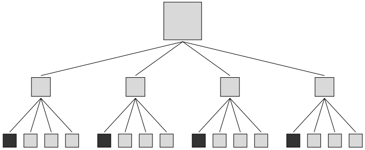
Generalization to RHS with nnz(B) not in $\Theta(1)$

If nnz(B) is not in $\Theta(1)$, the complexity of the forward solve depends on the distribution of the nonzeros in the tree:

Generalization to RHS with nnz(B) not in $\Theta(1)$

If nnz(B) is not in $\Theta(1)$, the complexity of the forward solve depends on the distribution of the nonzeros in the tree:

- Spread nonzeros: maximizes total flops for a fixed nnz. For $\alpha < 2$, $\mathcal{G}_{\text{spRHS}}(N, \alpha, \text{nnz}) = \Theta\left(\frac{\mathcal{G}_{\text{spRHS}}(N,\alpha,1)}{\text{nnz}^{\alpha/2-1}}\right) \Rightarrow$ gain decreases when nnz increases, but nonconstant $\mathcal{G}_{\text{spRHS}}$ is maintained as long as $\text{nnz} = o(N^2)$ (2D) or $\text{nnz} = o(N^3)$

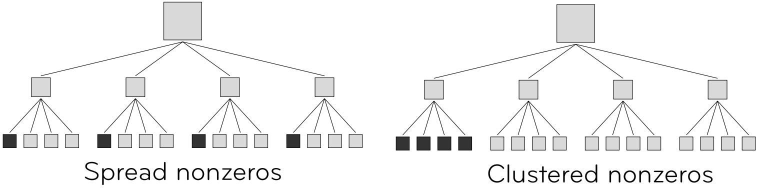


Spread nonzeros

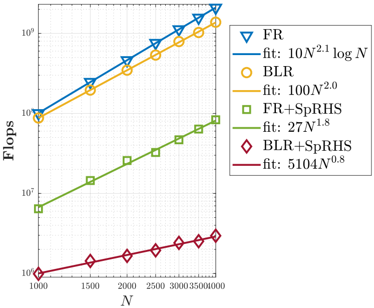
Generalization to RHS with $\text{nnz}(B)$ not in $\Theta(1)$

If $\text{nnz}(B)$ is not in $\Theta(1)$, the complexity of the forward solve depends on the distribution of the nonzeros in the tree:

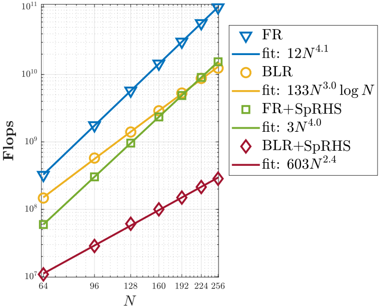
- Spread nonzeros: maximizes total flops for a fixed nnz. For $\alpha < 2$, $\mathcal{G}_{\text{spRHS}}(N, \alpha, \text{nnz}) = \Theta\left(\frac{\mathcal{G}_{\text{spRHS}}(N,\alpha,1)}{\text{nnz}^{\alpha/2-1}}\right) \Rightarrow$ gain decreases when nnz increases, but nonconstant $\mathcal{G}_{\text{spRHS}}$ is maintained as long as $\text{nnz} = o(N^2)$ (2D) or $\text{nnz} = o(N^3)$
- Clustered nonzeros: more favourable case, and more realistic of typical applications where sources are localized in the domain. $\mathcal{G}_{\text{spRHS}}(N, \alpha, \text{nnz}) = \Theta(\mathcal{G}_{\text{spRHS}}(N, \alpha, 1))$ is maintained as long as $\text{nnz} = O(N^{\alpha})$ (2D) or $\text{nnz} = O(N^{2\alpha})$ (3D)



Poisson equation, one RHS with one nonzero (placed on a leaf)



2D Poisson problem



3D Poisson problem

	$\mathcal{G}^{2D}_{\text{spRHS}}$ FR	BLR	$\mathcal{G}^{3D}_{\text{spRHS}}$ FR	BLR
Theoretical	$\Theta(\log N)$	$\Theta(N^{0.5})$	$\Theta(1)$	$\Theta(\log N)$
Experimental	$\Theta(N^{0.3}\log N)$	$\Theta(N^{1.2})$	$\Theta(N^{0.1})$	$\Theta(N^{0.6}\log N)$

Results on real-life problems (CSEM application)

Flop results ($\times 10^{12}$)

	Small problem (2.9M)			Large problem (17.4M)		
	FR	BLR	$\mathcal{G}_{\text{BLR}}$	FR	BLR	$\mathcal{G}_{\text{BLR}}$
Dense RHS	70	35	2.0	777	405	1.9
Sparse RHS	8	3	2.6	73	4	18.7
$\mathcal{G}_{\text{spRHS}}$	8.7	11.1		10.7	103.8	

$\Rightarrow$ $\mathcal{G}_{\text{spRHS}}$ is higher with BLR and $\mathcal{G}_{\text{BLR}}$ is higher with sparse RHS

(and trend becomes more and more visible as problem gets larger)

Results on real-life problems (CSEM application)

Flop results ($\times 10^{12}$)

	Small problem (2.9M)			Large problem (17.4M)		
	FR	BLR	$\mathcal{G}_{BLR}$	FR	BLR	$\mathcal{G}_{BLR}$
Dense RHS	70	35	2.0	777	405	1.9
Sparse RHS	8	3	2.6	73	4	18.7
$\mathcal{G}_{spRHS}$	8.7	11.1		10.7	103.8	

Time results (EOS computer, 90 MPI)

	Small problem (2.9M)			Large problem (17.4M)		
	FR	BLR	$\mathcal{G}_{BLR}$	FR	BLR	$\mathcal{G}_{BLR}$
Dense RHS	377	273	1.4	5449	3097	1.8
Sparse RHS	105	76	1.4	845	386	2.2
$\mathcal{G}_{spRHS}$	3.6	3.6		6.4	8.0	

$\Rightarrow$ $\mathcal{G}_{spRHS}$ **is higher with BLR and** $\mathcal{G}_{BLR}$ **is higher with sparse RHS**

(and trend becomes more and more visible as problem gets larger)

Time results follow same trend as flops, but less pronounced

Results on real-life problems (CSEM application)

Flop results ($\times 10^{12}$)

	Small problem (2.9M) FR	Small problem (2.9M) BLR	Small problem (2.9M) $\mathcal{G}_{BLR}$	Large problem (17.4M) FR	Large problem (17.4M) BLR	Large problem (17.4M) $\mathcal{G}_{BLR}$
Dense RHS	70	35	2.0	777	405	1.9
Sparse RHS	8	3	2.6	73	4	18.7
$\mathcal{G}_{spRHS}$	8.7	11.1	22.5	10.7	103.8	199.2

Time results (EOS computer, 90 MPI)

	Small problem (2.9M) FR	Small problem (2.9M) BLR	Small problem (2.9M) $\mathcal{G}_{BLR}$	Large problem (17.4M) FR	Large problem (17.4M) BLR	Large problem (17.4M) $\mathcal{G}_{BLR}$
Dense RHS	377	273	1.4	5449	3097	1.8
Sparse RHS	105	76	1.4	845	386	2.2
$\mathcal{G}_{spRHS}$	3.6	3.6	5.0	6.4	8.0	14.1

$\Rightarrow$ $\mathcal{G}_{spRHS}$ is higher with BLR and $\mathcal{G}_{BLR}$ is higher with sparse RHS

(and trend becomes more and more visible as problem gets larger)

Time results follow same trend as flops, but less pronounced

Cumulated gain from RHS sparsity and BLR very significant!

Conclusion

Exploit three types of sparsity to accelerate $AX = B$:

- Sparsity of matrix A (e.g., nested dissection)
- Data sparsity of separators (low-rank matrix formats: BLR, $\mathcal{H}$, ...)
- Sparsity of right-hand side B (pruned tree)

$\Rightarrow$ Take-home message: **Gain from exploiting RHS sparsity increases with data sparsity, and vice versa**

- Conclusions apply to forward solve only, except...

Conclusion

Exploit three types of sparsity to accelerate $AX = B$:

- Sparsity of matrix A (e.g., nested dissection)
- Data sparsity of separators (low-rank matrix formats: BLR, $\mathcal{H}$, ...)
- Sparsity of right-hand side B (pruned tree)

$\Rightarrow$ Take-home message: **Gain from exploiting RHS sparsity increases with data sparsity, and vice versa**

- Conclusions apply to forward solve only, except...
- Fourth type of sparsity: sparsity of solution X
  If only part of X is of interest, our complexity bounds also apply to backward solve and hence to the overall solve phase!
- Examples: augmented systems, Schur complement approaches, inversion of selected entries, ...

References

- P. R. Amestoy, J.-Y. L'Excellent, and G. Moreau, **On Exploiting Sparsity of Multiple Right-Hand Sides in Sparse Direct Solvers**, *SIAM J. Sci. Comput.*, 41(1), A269–A291 (2019).
- P. R. Amestoy, A. Buttari, J.-Y. L'Excellent, and T. Mary, **On the Complexity of the Block Low-Rank Multifrontal Factorization**, *SIAM J. Sci. Comput.*, 39(4), A1710–A1740 (2017).
- P. R. Amestoy, A. Buttari, J.-Y. L'Excellent, and T. Mary, **Bridging the Gap between Flat and Hierarchical Low-rank Matrix Formats: the Multilevel Block Low-Rank Format**, *SIAM J. Sci. Comput.*, 41(3), A1414–A1442 (2019).
- P. R. Amestoy, S. de la Kéthulle de Ryhove, J.-Y. L'Excellent, G. Moreau, and D. V. Shantsev, **Efficient use of sparsity by direct solvers applied to 3D controlled-source EM problems**, *Comput. Geosci.*, 23, 1237–1258 (2019).