Accelerating seismic processing with Tile-Low Rank matrix approximations

Yuxi Hong, Hatem Ltaief, David Keyes, Matteo Ravasi*
King Abdullah University of Science and Technology

Over the last decade, seismic processing has experienced a shift from adjoint-based computations followed by adaptive subtraction, to inversion-based algorithms. Examples of the first kind are the two workhorses in the seismic industry for multiple attenuation, namely Surface-Related Multiple Elimination and Internal Multiple Elimination. Methods that fall in the second category are Estimation of Primaries by Sparse Inversion (van Groenestijn and Verschuur, 2009), Up/Down Multi-Dimensional Deconvolution (Wapenaar et al., 2011) and Marchenko-based redatuming (Wapenaar et al., 2014). Whilst attractive from a theoretical point of view, inversion-based methods greatly increase the I/O (and/or memory access) and computational burden of these applications. Their deployment to large-scale 3D seismic datasets is therefore still in its infancy.

A common feature of such methods is the requirement of applying a numerical integral operator of the convolution type, also generally referred to as Multi-Dimensional Convolution (MDC):

$$\mathbf{y} = \mathbf{R}\mathbf{x}: \quad y(t, \mathbf{x}_B, \mathbf{x}_A) = \mathcal{F}^{-1}\left(\int_{\delta\mathbb{D}} R(\omega, \mathbf{x}_B, \mathbf{x}_R)\mathcal{F}\left(x(t, \mathbf{x}_R, \mathbf{x}_A)\right) d\mathbf{x}_R\right). \tag{1}$$

where t and ω are the time and the angular frequency, respectively, $\mathbf{x}_A$, $\mathbf{x}_B$ and $\mathbf{x}_R$ represent spatial locations with the latter spanning the integration domain $\delta\mathbb{D}$, $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the forward and inverse Fourier transforms. Here, $R(\omega, \mathbf{x}_B, \mathbf{x}_R)$ represents the kernel of the integral operator that in this context is represented by the entire reflection seismic data or a pre-processed version of it. Once the spatial integral is discretized, the kernel becomes a stack of matrices for each frequency ω and the integral can be interpreted as a batched matrix-vector multiplication (MVM).

In this work we propose to leverage *data sparsity* in frequency domain seismic data by means of so-called tile low-rank (TLR) matrix compression (Amestoy et al., 2015; Akbudak et al., 2018). Whilst dense MVM requires the matrix elements to be stored in row- or column-major data layout format, TLR-MVM splits the dense matrix into tiles and compresses each tile using an algebraic method of choice (e.g., rank-revealing QR, randomized SVD, etc.). More specifically, given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ divided into tiles of size $n_b \times n_b$ and a certain threshold ϵ, each tile is decomposed into $\mathbf{U}_k^\epsilon$ and $\mathbf{V}_k^\epsilon$ eigenvectors and its k most significant singular values $\mathbf{\Sigma}_k^\epsilon$ such that $||\mathbf{A} - \mathbf{U}_k^\epsilon \mathbf{\Sigma}_k^\epsilon \mathbf{V}_k^{T\epsilon}||_F \leq \epsilon||\mathbf{A}||_F$. Finally, since we are here interested to perform a number of TLR-MVM for all of the available seismic frequency slices and given that low frequencies have generally a higher compression ratio compared to high frequencies, additional load balancing strategies must been devised. More specifically, a *Merge-Phase* strategy is designed to achieve intra-node load balancing by evenly distributing the computation on each processing unit across the three different phases on the TLR-MVM computation. On the other hand, to leverage performance on distributed-memory systems a *Zig-Zag* strategy is used to allocate frequency matrices across computational nodes, such that all nodes process some of the low and some of the high frequencies (Hong et al., 2021).

Using the benchmark synthetic dataset in Ravasi and Vasconcelos (2021), the proposed strategy is applied to the Marchenko redatuming problem. Figure 1a compares the redatumed wavefields from the dense reflection seismic data and two TLR compressed versions of it. Looking at the reconstruction errors after solving the Marchenko equations with 10 iterations of conjugate gradient, we conclude that minimal degradation in the quality of wavefields is obtained whilst drastically reducing the size of the kernel and the computational cost of the modelling operator. More specifically, given an acquisition geometry with 9801 sources and receivers, the dense kernel for a set of 150 frequencies has size of 115 GB. Its corresponding TLR compressed matrix has size of 31 GB for $\epsilon = 0.001$ (~ 3 compression ratio) and 19 GB for $\epsilon = 0.005$ (~ 6 compression ratio). Figure 1b displays the time-to-solution for batched dense-MVM and TLR-MVM for 2 different hardware architectures, namely Intel Cascade Lake and NEC SX-Aurora TSUBASA. In both cases we observe that TLR-MVM is faster than dense-MVM and that the two additional load balancing strategies can further speedup computations in both systems. Our implementation obtains a 16X and 3X performance speedup against Intel and NEC optimized dense MVMs, respectively.

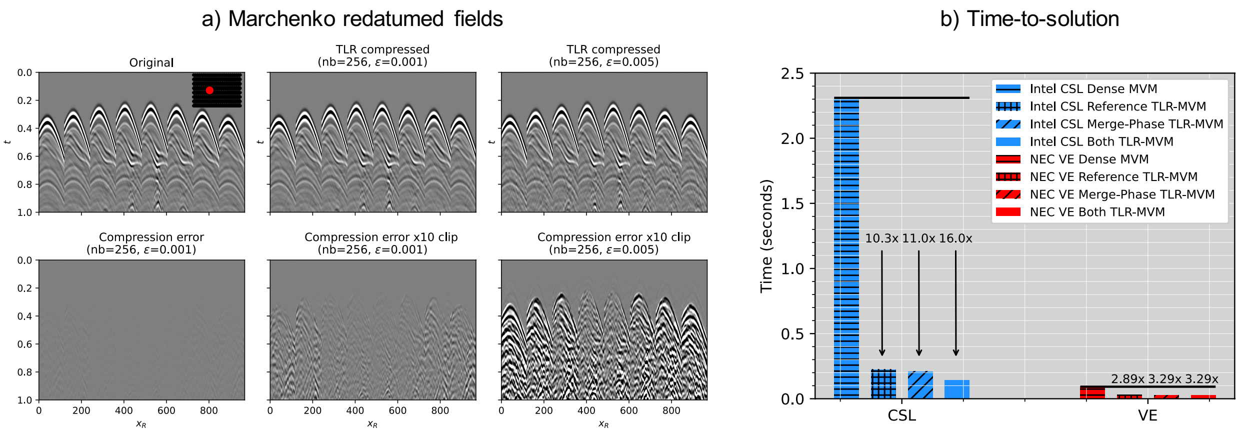

Figure 1: a) Reconstructed subsurface Green's function using the original dense reflection response as kernel, and two TLR compressed versions with different accuracy $\epsilon = 0.001$ and $\epsilon = 0.005$, respectively. b) Time-to-solution for Intel Cascade Lake (blue) and NEC SX-Aurora (red).

References

Akbudak, K., H. Ltaief, A. Mikhalev, A. Charara, A. Esposito, and D. Keyes, 2018, Exploiting Data Sparsity For Large-scale Matrix Computations: European Conference on Parallel Processing, Springer, 721–734.

Amestoy, P., C. Ashcraft, O. Boiteau, A. Buttari, J.-Y. L'Excellent, , and C. Weisbecker, 2015, Improving Multifrontal Methods by Means of Block Low-Rank Representations: SIAM Journal on Scientific Computing, **37**, no. 3. (doi: 10.1137/120903476).

Hong, Y., H. Ltaief, M. Ravasi, L. Gatineau, and D. E. Keyes, 2021, Accelerating Seismic Redatuming Using Tile Low-Rank Approximations on NEC SX-Aurora TSUBASA: Supercomputing Frontiers and Innovation, **8**, no. 2. (doi: 10.14529/jsfi210201).

Ravasi, M., and I. Vasconcelos, 2021, An open-source framework for the implementation of large-scale integral operators with flexible, modern high-performance computing solutions: Enabling 3d marchenko imaging by least-squares inversion: Geophysics, **86**, no. 5, WC177–WC194. (doi: 10.1190/geo2020-0796.1).

van Groenestijn, G. J., and D. J. Verschuur, 2009, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, no. 3, 1MJ–Z54. (doi: 10.1190/1.3111115).

Wapenaar, K., J. Thorbecke, J. van der Neut, F. Broggini, E. Slob, and R. Snieder, 2014, Marchenko imaging: Geophysics, **79**, no. 3, WA39–WA57. (doi: 10.1190/geo2013-0302.1).

Wapenaar, K., J. van der Neut, E. Ruigrok, D. Draganov, J. Hunziker, E. Slob, J. Thorbecke, and R. Snieder, 2011, Seismic interferometry by crosscorrelation and by multidimensional deconvolution: A systematic comparison: Geophysical Journal International, **185**, 1335–1364. (doi: 10.1111/j.1365-246X.2011.05007.x).