geoconvention
Calgary • Canada • May 11-13 2026

CountSketch MSSA for Seismic Reconstruction

Akash Nair and Mauricio D. Sacchi

Signal Analysis and Imaging Group (SAIG), Department of Physics, University of Alberta

Summary

Seismic data reconstruction aims to mitigate the impact of irregular and sparse spatial sampling prior to imaging and inversion. Among rank-reduction approaches, Multichannel Singular Spectrum Analysis (MSSA), also known as the Cadzow filter, has proven particularly effective due to its ability to exploit the intrinsic low-rank structure of seismic data in the frequency–space domain.

A significant limitation of classical MSSA implementations is their reliance on repeated low-rank projections of large block-Hankel matrices, which can lead to prohibitive memory usage and computational cost. Fast MSSA (FSSA) alleviates this burden through FFT-based implicit Hankel operators and randomized QR decompositions, achieving log-linear complexity. However, for modern large-scale and time-lapse surveys, even $\mathcal{O}(N \log N)$ scaling can become a bottleneck.

We introduce a memory-efficient MSSA framework based on the CountSketch transform that avoids explicit matrix construction and achieves strictly linear computational complexity. The proposed approach preserves the signal subspace with high probability while enabling scalable reconstruction of large post-stack and pre-stack datasets. Field-data examples from Aquistore and Penobscot demonstrate reconstruction quality comparable to FSSA, with an order-of-magnitude reduction in computational cost.

Introduction

Irregular spatial sampling is increasingly common in modern seismic acquisition due to operational, environmental, and economic constraints. Such sampling patterns degrade data quality and negatively impact subsequent imaging and inversion. As a result, seismic data reconstruction has become a standard preprocessing step in seismic processing workflows.

Existing reconstruction methods include prediction filtering, Fourier-based interpolation, local transform methods, rank-reduction techniques, and, more recently, machine-learning-based approaches (Stanton and Sacchi, 2013). Among these, rank-reduction methods based on Multichannel Singular Spectrum Analysis (MSSA) have gained widespread adoption due to their robustness and strong theoretical foundation (Trickett, 2008; Sacchi, 2009; Trickett and Burroughs, 2009; Trickett et al., 2010).

MSSA is typically applied in the frequency–space (FX or FXY) domain, where monochromatic seismic data exhibit low-rank structure when embedded into block-Hankel matrices. The computational burden of MSSA stems from repeated singular value decompositions (SVDs) of large matrices within an iterative framework. Randomized numerical linear algebra has been successfully used to reduce this cost, including randomized SVD (Oropeza and Sacchi, 2011) and Fast MSSA (FSSA) (Cheng et al., 2019), which combines randomized QR decomposition with FFT-based Hankel operators (Chiron et al., 2014; Cheng and Sacchi, 2016).

Despite these advances, the rapid growth in data volume—particularly for wide-azimuth, multi-component, and time-lapse surveys—motivates the development of reconstruction methods with strictly linear computational and memory complexity. In this work, we propose an implicit MSSA framework based

on CountSketch operators (Charikar et al., 2004) that preserves the essential signal subspace while avoiding the explicit construction of a Hankel matrix.

MSSA in the Frequency–Space Domain

We briefly review MSSA in the FXY domain to establish notation and to highlight the origin of its computational cost.

Consider a seismic data cube transformed into the frequency domain. For a fixed angular frequency ω, the monochromatic slice is denoted by

$$\mathbf{D}(\omega) \in \mathbb{C}^{N_x \times N_y}, \tag{1}$$

where N_x and N_y denote the spatial dimensions. For seismic data dominated by locally planar or gently curved events, the monochromatic slice exhibits strong spatial coherence. When embedded into a Hankel structure, this coherence manifests as a low-rank matrix (Oropeza and Sacchi, 2011).

In the presence of noise or missing observations—particularly when sampling is irregular or incoherent—the corresponding Hankel matrix becomes full rank. Rank reduction can then be used to approximate this matrix by the Hankel matrix of an ideal, noise-free and fully sampled dataset. This simple but powerful idea lies at the core of rank-reduction methods for seismic denoising and reconstruction.

MSSA proceeds by embedding $\mathbf{D}(\omega)$ into a block-Hankel matrix

$$\mathbf{A}(\omega) \in \mathbb{C}^{N \times M}, \tag{2}$$

For notational simplicity, we drop the explicit dependence on ω in what follows, with the understanding that the procedure is applied independently to all frequencies of interest. Standard frequency-domain symmetries are enforced so that the inverse Fourier transform yields real-valued time-domain data.

Under the assumption of locally stationary dips within each window, the Hankel matrix $\mathbf{A}$ is approximately low-rank and admits the truncated singular value decomposition

$$\mathbf{A} \approx \mathbf{A}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H, \tag{3}$$

where $k \ll \min(N, M)$. Rank reduction is achieved by retaining only the k dominant singular values and vectors.

The filtered Hankel matrix must then be mapped back to the spatial domain. This is achieved through *anti-diagonal averaging*, which can be expressed explicitly as a linear operator $\mathscr{A}$ that undoes the action of the embedding operator $\mathscr{H}$. Denoting the Hankelization by

$$\mathbf{A} = \mathscr{H}(\mathbf{D}),$$

the reconstruction step is given by

$$\widehat{\mathbf{D}} = \mathscr{A}(\mathbf{A}_k), \tag{4}$$

In essence, $\mathscr{A}$ redistributes the elements of $\mathbf{A}_k$ back onto the spatial grid by summing along anti-diagonals and normalizing by the number of contributions at each spatial location.

For denoising, this procedure is typically applied once. For reconstruction with missing data, the rank-reduction and anti-diagonal averaging steps are embedded within an iterative algorithm—commonly of POCS type—where data consistency is enforced at each iteration (Oropeza and Sacchi, 2011).

The dominant computational and memory cost arises from forming and factorizing $\mathbf{A}$ for every frequency slice and at every iteration. In practice, additional windowing is often required to limit the number of local dips and ensure the validity of the low-rank assumption. Consequently, the combined effect of looping over frequencies, iterations, and spatial windows leads to repeated large-scale SVD computations. Without careful algorithmic design, classical MSSA implementations quickly become impractical, motivating the use of randomized and implicit formulations.

Sketching and Subspace Preservation

Sketching methods from randomized numerical linear algebra provide an efficient mechanism to reduce the dimensionality of large matrices while approximately preserving their dominant column (or row) spaces (Woodruff, 2014). The key idea is to replace expensive factorizations of a large matrix by analogous computations on a much smaller surrogate.

Let

$$\mathbf{A} \in \mathbb{C}^{N\times M},$$

and let

$$\mathbf{S} \in \mathbb{R}^{M\times s}, \qquad k \le s \ll M,$$

be a random sketching matrix, where k is the target numerical rank. Typical choices for $\mathbf{S}$ include Gaussian, Rademacher, or CountSketch matrices. The sketch dimension s must be sufficiently large (often $s \approx k+p$, with a small oversampling parameter p) to ensure stable subspace capture.

The sketched matrix is defined as

$$\mathbf{Y} = \mathbf{A}\mathbf{S} \in \mathbb{C}^{N\times s}. \tag{5}$$

By construction, the column space of $\mathbf{Y}$ approximates the column space of $\mathbf{A}$, and equivalently, the row space of $\mathbf{Y}^H$ approximates the row space of $\mathbf{A}^H$.

Instead of computing the singular value decomposition (SVD) of $\mathbf{A}$ directly, we compute the thin SVD of the sketch,

$$\mathbf{Y} = \mathbf{U}_Y \mathbf{\Sigma}_Y \mathbf{V}_Y^H, \tag{6}$$

and retain only the leading k left singular vectors,

$$\mathbf{U}_k \in \mathbb{C}^{N\times k}.$$

These vectors span an approximate dominant subspace of $\mathbf{A}$ after lifting back to the original space.

A rank-k approximation of $\mathbf{A}$ is then obtained by projecting $\mathbf{A}$ onto this subspace,

$$\widehat{\mathbf{A}}_k = \mathbf{P}_k \mathbf{A}, \qquad \mathbf{P}_k = \mathbf{U}_k \mathbf{U}_k^H, \tag{7}$$

where $\mathbf{P}_k$ is an orthogonal projector (in the Euclidean sense) onto the sketched dominant subspace. Importantly, the singular values $\mathbf{\Sigma}_Y$ and right singular vectors $\mathbf{V}_Y$ of the sketch are not used; only the subspace information carried by $\mathbf{U}_k$ is retained.

This procedure does not guarantee recovery of the exact truncated SVD of $\mathbf{A}$. However, in applications such as denoising and data reconstruction, this heuristic is typically sufficient. In particular, for Hankel or block-Hankel matrices arising in MSSA, the singular-value spectrum is often characterized by a small number of dominant singular values associated with a coherent signal, followed by a rapidly decaying tail associated with noise. In this regime, preserving the dominant subspace

is more important than accurately estimating the full singular spectrum, and sketch-based subspace projection provides an effective and computationally efficient alternative to the full SVD.

CountSketch MSSA

Standard MSSA is applied to a monochromatic slice of a seismic data cube and consists of three main steps:

1. *Hankelization*: the monochromatic slice is reshaped into a block-Hankel matrix, mapping linear or gently curved events in the data to a low-rank matrix structure. This step is conceptually simple but incurs substantial memory overhead due to the explicit formation of the Hankel matrix.
2. *Rank reduction*: a truncated singular value decomposition (SVD) is applied to the block-Hankel matrix to separate coherent signal components (low-rank) from incoherent noise (high-rank). This step is computationally expensive, with a cost that scales cubically with matrix dimensions.
3. *Anti-diagonal averaging*: the filtered block-Hankel matrix is mapped back to the physical data domain by averaging along its anti-diagonals.

To mitigate the computational and memory costs of these operations, Cheng et al. (2019) proposed fast MSSA implementations that exploit convolutional structure and FFTs, reducing the overall complexity to $\mathcal{O}(N \log N)$. However, as seismic reconstruction problems extend to large-scale and time-lapse surveys—where one must handle four spatial and two temporal dimensions—the data volumes increase dramatically. In this regime, even FFT-based MSSA becomes computationally demanding.

Motivated by this challenge, we propose an implicit MSSA solver based on the CountSketch transform, designed to scale linearly with data size. However, first, we describe the particular $\mathbf{S}$ used by CountSketch. In principle, the sketching matrix $\mathbf{S}$ can be chosen from a variety of random ensembles, such as dense Gaussian or Rademacher matrices, which are known to provide strong subspace embedding guarantees. However, these choices are impractical for large-scale MSSA applications due to their high memory footprint and computational cost.

Here, we instead adopt the CountSketch (CS) transform, which yields a highly sparse sketching matrix and enables efficient implicit implementations with linear computational complexity.

CountSketch matrix. Let $\mathbf{S} \in \mathbb{R}^{M \times s}$ denote a CountSketch matrix. Each row of $\mathbf{S}$ contains exactly one nonzero entry. Specifically, two random functions are drawn:

$$h : \{1, \ldots, M\} \to \{1, \ldots, s\}, \qquad \xi : \{1, \ldots, M\} \to \{-1, +1\},$$

where $h(i)$ is a uniformly distributed hash function and $\xi(i)$ is an independent Rademacher random variable. The entries of $\mathbf{S}$ are defined as

$$S_{i,j} = \begin{cases} \xi(i), & j = h(i), \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

As a result, $\mathbf{S}$ is extremely sparse, with exactly one nonzero per row and a total of M nonzero entries.

The sketch $\mathbf{Y} = \mathbf{AS}$ is formed by hashing and aggregating signed columns of $\mathbf{A}$ according to h, providing an efficient approximation to the column space of $\mathbf{A}$ while avoiding explicit construction of either $\mathbf{S}$ or the Hankel matrix.

Now that we have defined the particular form of our Sketching operator, we provide the final algorithm proposed in this paper:

- *Implicit Hankelization*: rather than explicitly forming the block-Hankel matrix $\mathbf{A}$, we compute its sketch

$$\mathbf{Y} = \mathbf{AS},$$

  directly from the monochromatic data slice. This is achieved through matrix–vector products that emulate the action of $\mathbf{A}$ without materializing it. The sketching operator implicitly accesses and aggregates the required data samples, effectively performing Hankelization and sketching in a single pass.

- *Rank reduction via CountSketch*: the SVD is applied to the sketched matrix $\mathbf{Y}$ to estimate a low-dimensional subspace $\mathbf{U}_k$ associated with the dominant signal components. Importantly, we avoid forming the projected matrix $\mathbf{P}_k\mathbf{A}$ explicitly, thereby eliminating the need to store or manipulate large intermediate matrices. This step is critical for achieving both memory efficiency and computational scalability.

- *Implicit anti-diagonal averaging*: the filtered result is mapped back to the original data geometry by implicitly undoing the Hankelization step, followed by appropriate normalization. As in the previous steps, this operation is performed without ever forming the reduced-rank Hankel matrix $\mathbf{A}_k$, relying instead on the adjoint action of the implicit Hankel operator.

Together, these steps define a fully implicit MSSA algorithm in which neither the block-Hankel matrix nor its reduced-rank approximation is explicitly constructed, yielding a method that is both memory-efficient and well-suited for large-scale seismic reconstruction problems.

Finally, the implicit CountSketch–MSSA filter $\mathscr{F}$ is embedded within an iterative algorithm reminiscent of Projection Onto Convex Sets (POCS) (Abma and Kabir, 2006) to enforce data consistency. We emphasize *reminiscent* because rank reduction is a non-convex operation; therefore, although the algorithm follows the structural pattern of POCS, it does not constitute a true POCS method in the strict mathematical sense.

When the observed data are considered free of noise, we enforce strict data consistency, and the observed traces are preserved exactly, and the filter $\mathscr{F}$ only updates the missing traces:

$$\mathbf{d}^{(i+1)} = a\mathbf{d}_{obs} + (\mathbf{I} - a\mathscr{S})\mathscr{F}(\mathbf{d}^{(i)}) \tag{9}$$

where $\mathscr{S}$ is the binary sampling operator ($\mathscr{S}_{ii} = 1$ if sampled, 0 otherwise), $\mathbf{I}$ is the identity matrix and i is the iteration index. This formulation aligns with the method adopted by Oropeza and Sacchi (2011). The relaxation parameter $a \in (0,1)$ controls the strength of reinsertion of the observed data during the iteration. In practice, a is often chosen heuristically to balance data fidelity and convergence stability, for example by monitoring the residual norm and the overall reconstruction quality across iterations.

Results

We tested the performance of the CountSketch MSSA on real post-stack and pre-stack datasets and compared its computational speed with FSSA Cheng et al. (2019). First, it was tested on post-stack data from Aquistore, a CCS facility in Estevan, Canada (Roach et al., 2017). The data was randomly decimated by 50%, and a sketch with twice the rank ($k = 11$) was used. The window size was set to 5×5. The reconstruction result was excellent, with a reconstruction error of less than 1% given by:

$$\text{Reconstruction Error (\%)} = \frac{\|\mathbf{d}_{recon} - \mathbf{d}_{true}\|_2}{\|\mathbf{d}_{true}\|_2} \times 100\% \tag{10}$$

where $\mathbf{d}_{recon}$ is the reconstructed data and $\mathbf{d}_{true}$ is the ground truth data.

The computational times are detailed in Table 1. For the largest tested dimension (469×481), CountSketch MSSA required approximately 11.87 seconds for the total job (50 slices, 10 iterations), whereas

FSSA required 101.85 seconds, demonstrating an order-of-magnitude speedup. The scaling behaviour is visually represented in Figure 1, which illustrates that the computational cost of FSSA scales super-linearly as N increases, while CountSketch MSSA maintains a near-linear profile.

Table 1: Computational time comparison between the proposed CountSketch MSSA and FSSA across varying spatial dimensions. Results represent the total runtime for 50 monochromatic frequency slices over 10 POCS iterations using the Aquistore dataset.

Data Size	CountSketch MSSA (s)	FSSA (s)
20×20	0.14	0.90
40×40	0.21	1.07
80×80	0.44	2.97
160×160	1.31	13.37
320×320	5.33	48.95
469×481	11.87	101.85

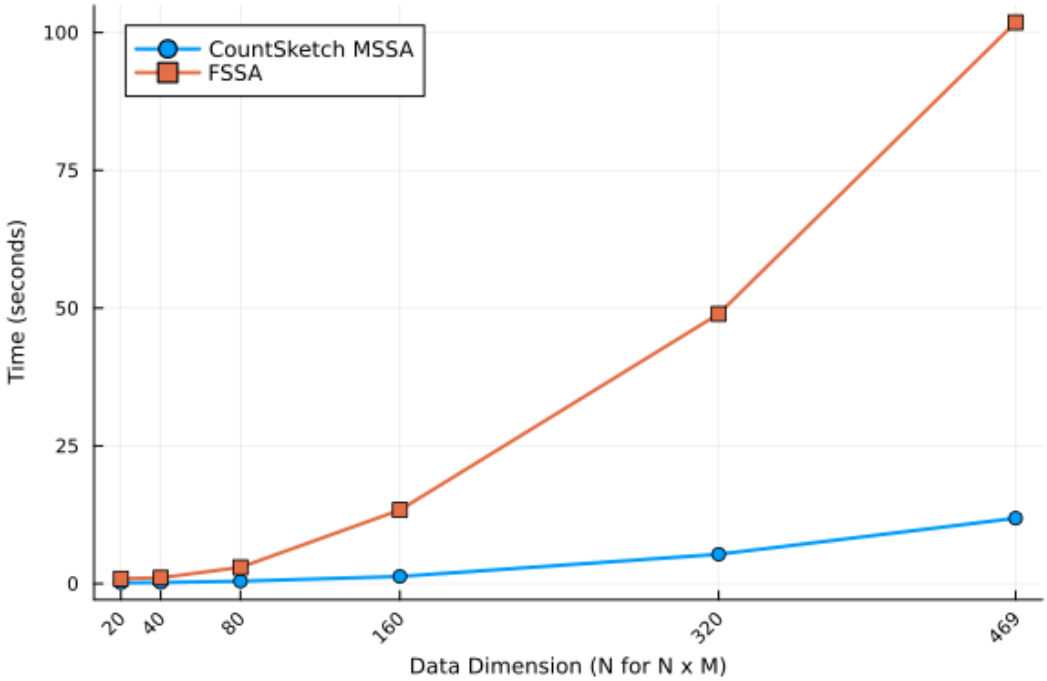


Figure 1: Computational time comparison between the proposed CountSketch MSSA ($\mathcal{O}(NL)$ complexity) and the FSSA algorithm ($\mathcal{O}(N \log N)$ complexity).

Figure 2 displays the reconstruction results, the difference section, and the variance (uncertainty) map of the same slice from the 3D volume at iterations 10, 30, and 50 across 10 different stochastic realizations of the sketch. The low variance indicates that the reconstruction is stable across different realizations.

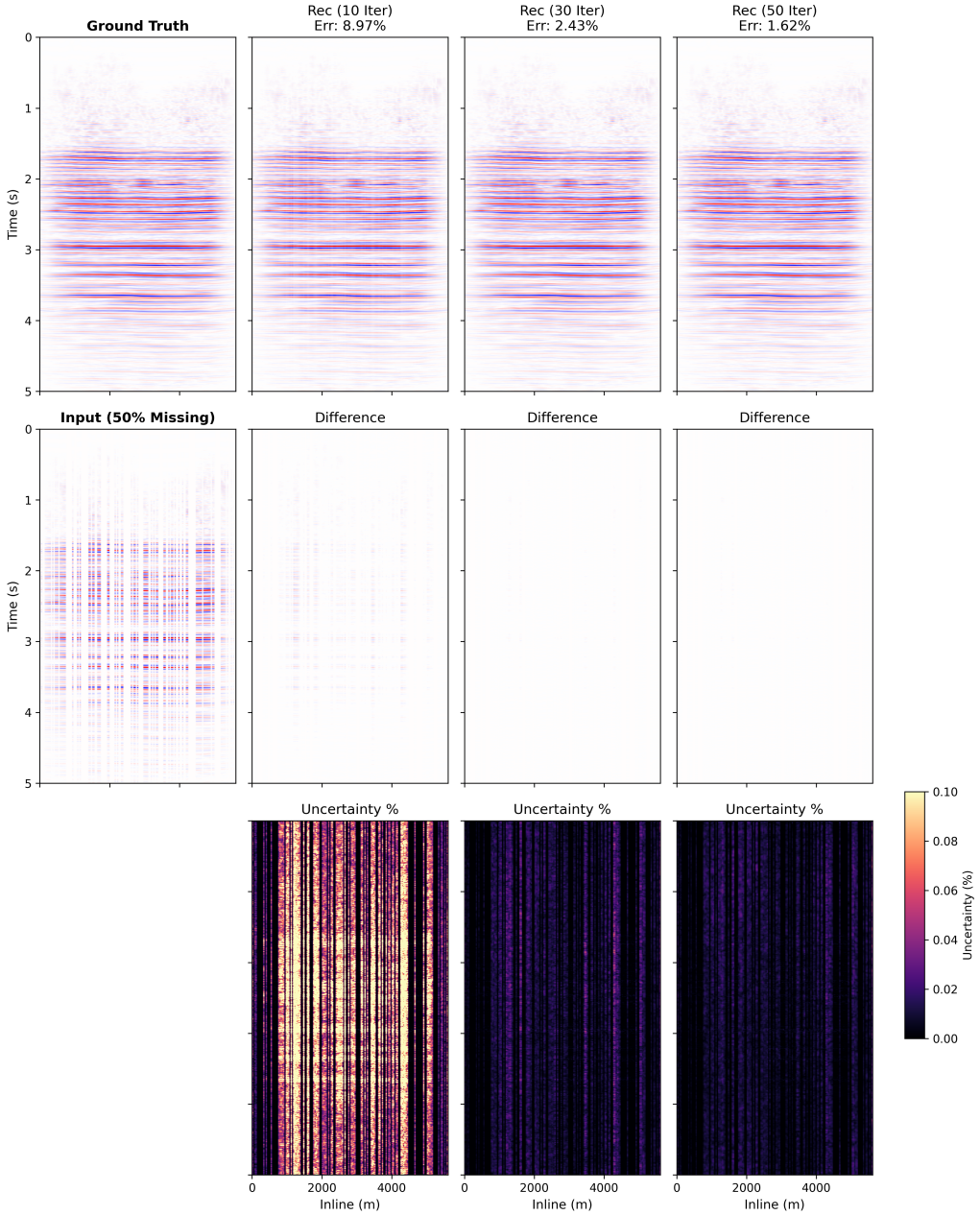


Figure 2: Result of CountSketch MSSA on post-stack Aquistore dataset at different numbers of iterations with their uncertainty maps.

We further validated the method on a pre-stack CMP gather from Project Penobscot in offshore Nova Scotia, Canada. To evaluate the performance on this field data, we simulated a missing trace scenario by decimating 50% of the traces using a random pattern. The proposed Sketched MSSA algorithm, implemented within an ensemble POCS framework ($N = 10$), successfully reconstructs the missing data, seamlessly restoring the continuity of the hyperbolic reflection events and preserving the signal's amplitude character, as shown in Figure 3. We performed 10 random realizations of the sketch, and the plot shows the mean of all those realizations. The associated uncertainty map confirmed low variance, again indicating a stable reconstruction.

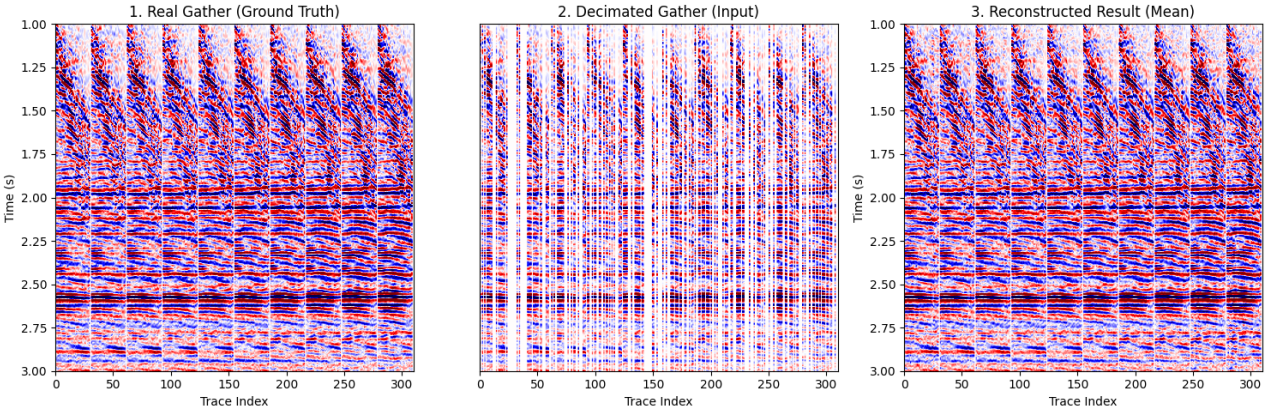


Figure 3: Result of CountSketch MSSA for pre-stack CMP gathers from Penobscot Project.

Conclusions

We propose a CountSketch MSSA framework for seismic reconstruction, which is a linear-time, memory-efficient alternative to Fast MSSA. The method preserves MSSA's ability to recover coherent seismic events while enabling scalable reconstruction of large time-lapse post-stack and pre-stack datasets. It exploits the Hankel structure with the sketch properties to implicitly reconstruct monochromatic data slices. Uncertainty quantification through ensemble variance showed stable convergence across different realizations of the sketch. Field-data examples confirm high-fidelity reconstruction with errors around 1%, establishing the proposed approach as a robust and trustworthy workflow for large-scale seismic processing. Future work will explore simultaneous denoising and despiking within this framework, as well as joint reconstruction of time-lapse datasets.

Acknowledgements

We want to thank the Signal Analysis and Imaging Group sponsors and Future Energy Systems at the University of Alberta for supporting the stimulating research environment that enabled the preparation of this work. We are also thankful to the Petroleum Technology Research Centre and the Geological Survey of Canada for making the Aquistore data available to us. We also thank dGB Earth Sciences for making the Penobscot data available as an OpendTect project via their TerraNubis portal terranubis.com.

References

Abma, R., and N. Kabir, 2006, 3d interpolation of irregular data with a POCS algorithm: Geophysics, **71**, E91–E97.

Charikar, M., K. Chen, and M. Farach-Colton, 2004, Finding frequent items in data streams: Theoretical Computer Science, **312**, 3–15.

Cheng, J., M. Sacchi, and J. Gao, 2019, Computational efficient multidimensional singular spectrum analysis for prestack seismic data reconstruction: Geophysics, **84**, V111–V119.

Cheng, J., and M. D. Sacchi, 2016, Fast dual-domain reduced-rank algorithm for 3D deblending via randomized QR decomposition: Geophysics, **81**, no. 1, V89–V101.

Chiron, L., M. A. van Agthoven, B. Kieffer, C. Rolando, and M.-A. Delsuc, 2014, Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry: Proceedings of the National Academy of Sciences, **111**, no. 4, 1385–1390.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, V25–V32.
Roach, L. A. N., D. J. White, B. Roberts, and D. Angus, 2017, Initial 4d seismic results after co2 injection start-up at the aquistore storage site: Geophysics, **82**, B95–B107.
Sacchi, M., 2009, *FX* Singular Spectrum Analysis: CSPG CSEG CWLS Convention, 392–395.
Stanton, A., and M. D. Sacchi, 2013, All roads lead to rome: predictability, sparsity, rank and pre-stack seismic data reconstruction: CSEG Recorder, **38**, 32–37.
Trickett, S., 2008, *F-XY* Cadzow noise suppression: SEG Technical Program Expanded Abstracts 2008, **27**, no. 1, 2586–2590.
Trickett, S., and L. Burroughs, 2009, Prestack rank-reducing noise suppression: Theory: SEG Technical Program Expanded Abstracts 2009, **28**, no. 1, 3332–3336.
Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: SEG Technical Program Expanded Abstracts 2010, **29**, no. 1, 3829–3833.
Woodruff, D. P., 2014, Sketching as a tool for numerical linear algebra: Foundations and Trends® in Theoretical Computer Science, **10**, 1–157.