FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

Exploiting Dynamic Sparsity in Einsum

November 7, 2025

NeurIPS 2025

Christoph Staudt, Mark Blacher, Tim Hoffmann, Kaspar Kasche, Olaf Beyersdorff, Joachim Giesen

Faculty of Mathematics and Computer Science

Overview

Einsum and Tensor Networks

Exponential Separation

Experiments

Einsum and Tensor Networks

Einsum and Tensor Networks

Einsum basics

Operation	Einsum expression
inner vector product	$\mathrm{einsum}\left(i, i \to;\ u, v\right)$
elementwise vector product	$\mathrm{einsum}\left(i, i \to i;\ u, v\right)$
matrix-matrix-vector product	$\mathrm{einsum}\left(ij, jk, k \to i;\ A, B, v\right)$

In practice large expressions are executed in pairwise steps:

$$\mathrm{einsum}\left(ij, j \to i;\ A, \mathrm{einsum}\left(jk, k \to j;\ B, v\right)\right)$$

Einsum and Tensor Networks

Tensor hypernetworks as graphical representation

$$\mathrm{einsum}\left(cb, ba, cf, be, ad, fe, ed, fi, eh, dg, ih, hg \rightarrow;\ A, B, C, D, E, F, G, H, I, J, K, L\right)$$
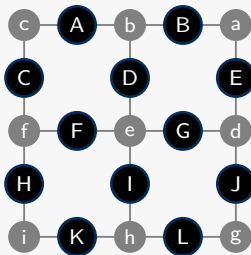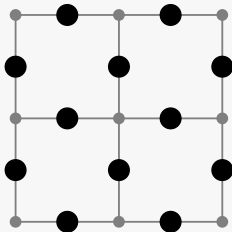
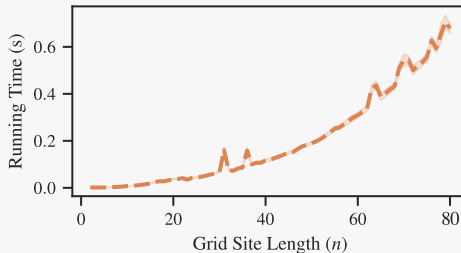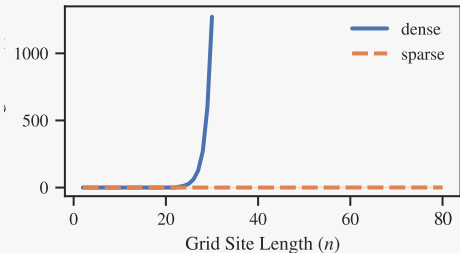# Exponential Separation

Exponential Separation

Grid instances

- Each tensor (black node) is a diagonal matrix
- We prove: For such a square grid with side length n, dense tensors require at least 2^n flops, while sparse tensors only need $O(n^2)$ flops.

Exponential Separation

Grid instances

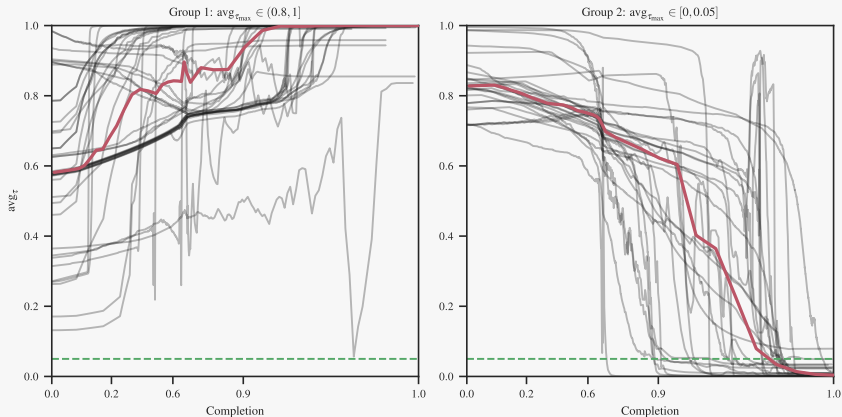Synthetic einsum expressions based on the grid formulas used in the proof.
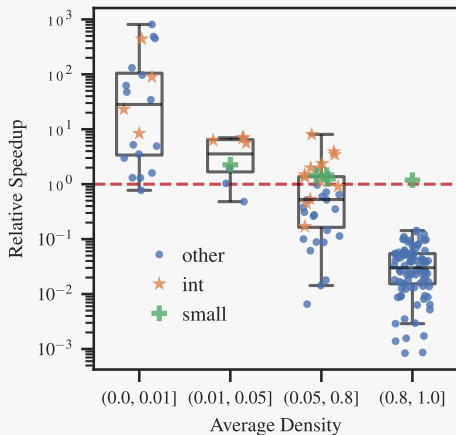
Experiments

Experiments

Dynamic sparsity in the einsum benchmark

Real world instances from the einsum benchmark: `https://benchmark.einsum.org/`.

Experiments
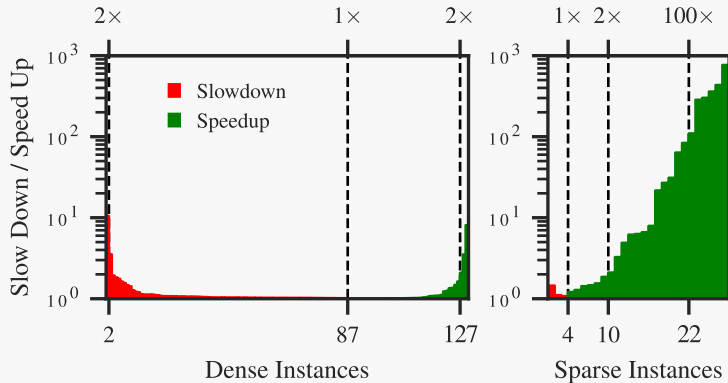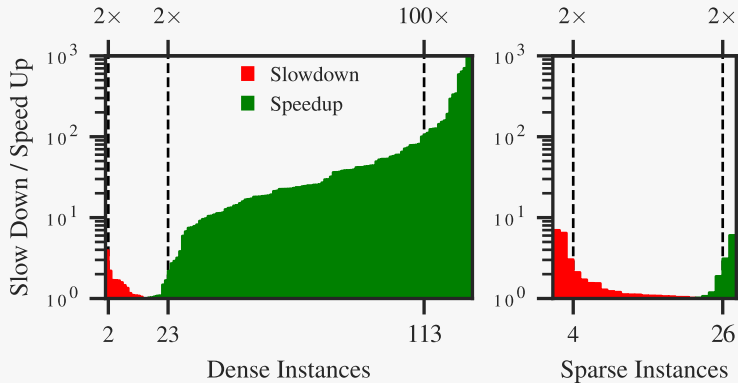
Sparse Speedup

Hybrid Algorithm

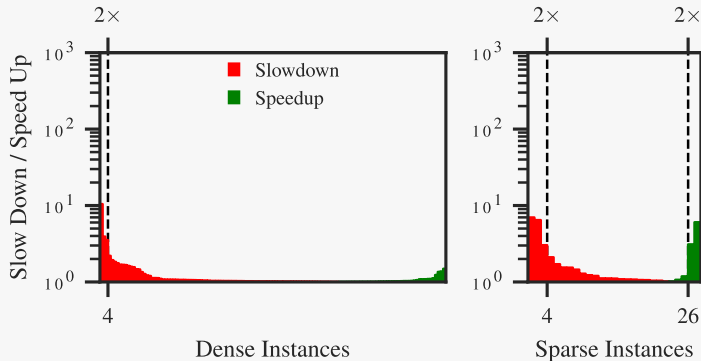# Hybrid Algorithm
Hybrid algorithm speedup

Hybrid Algorithm
Hybrid algorithm speedup

Hybrid Algorithm
Hybrid algorithm speedup

Conclusions

1. Large einsum expressions exhibit dynamic sparsity in practice.
2. Significant speed ups are possible by exploiting dynamic sparsity.