IBM Research Georgia Tech

Dense Associative Memory through the Lens of Random Features

Benjamin Hoover **Polo** Chau **Hendrik** Strobelt **Parikshit** Ram **Dmitry** Krotov

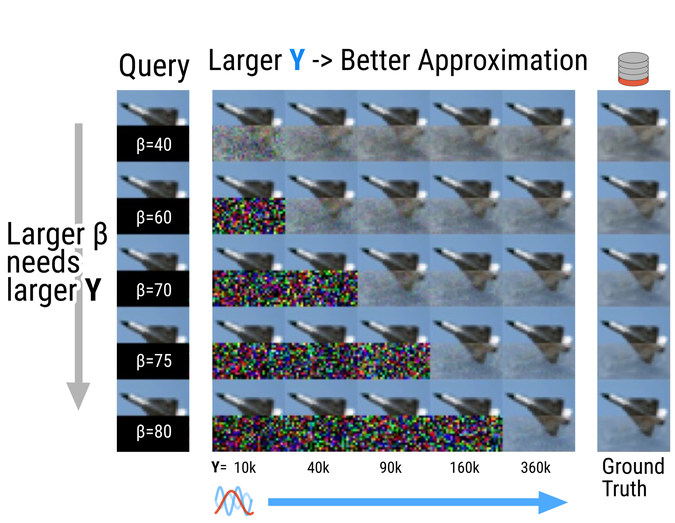
Traditional Memory representations of Dense Associative Memory (MrDAM) requires introduction of new weights every time a new memory pattern is stored in the network. New patterns must be added as rows to the memory matrix.

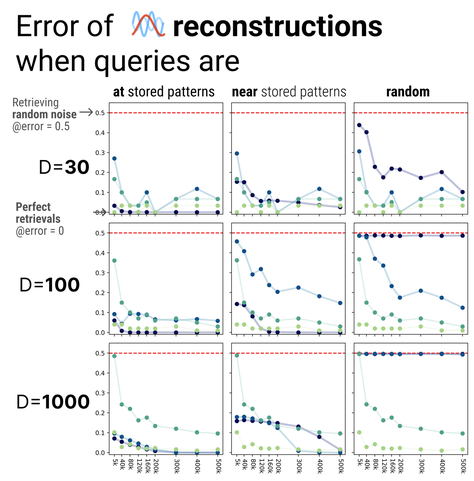
Our Method

Distributed representations of Dense Associative Memory (DrDAM) makes it possible to add new patterns without adding new weights. New patterns are entangled in a memory tensor of fixed width. DrDAM's energy approximates the energy of MrDAM using Random Features.

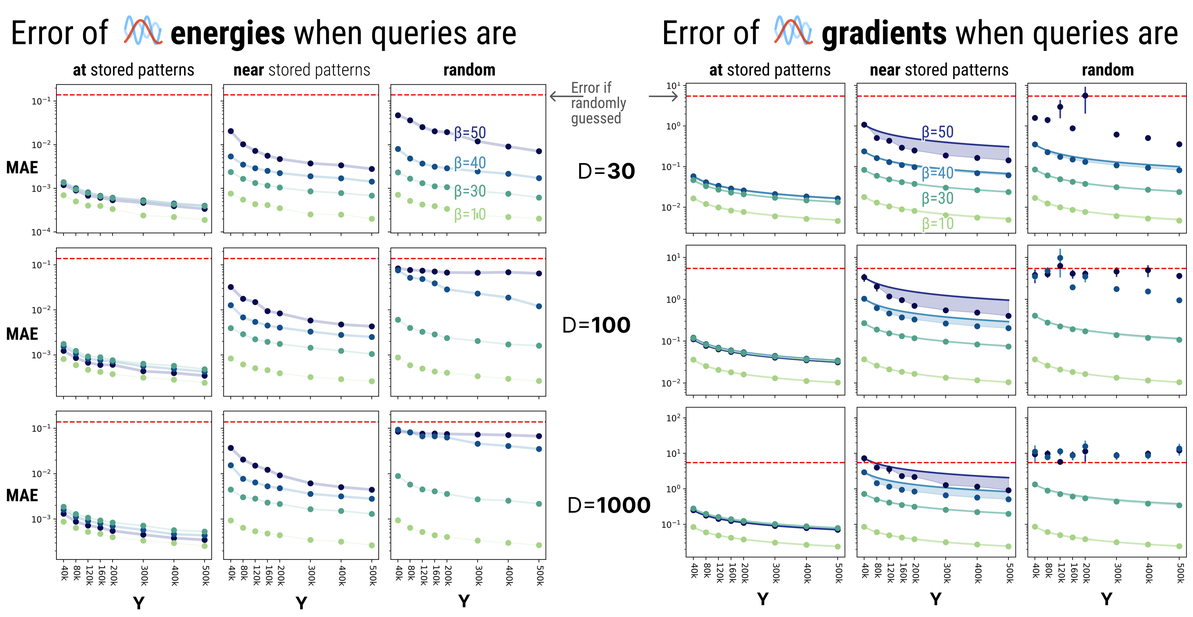
approximates the energies and gradients of

Use more random features Y for better approximations





Use more random features Y at low temp and high corruption



approximates memory retrievals of

Energy minimization retrieves same memories in both models

How is this possible?

noisy query's energy sum of similarities between query and stored patterns

$$E(\mathbf{x}) = -\frac{1}{\beta}\log\left(\sum_\mu \exp\left(-\frac{\beta}{2}||\mathbf{x}-\xi^\mu||^2\right)\right)$$

Use kernel trick to replace RBF Kernel with inner product of random feature maps φ

$$E(\mathbf{x}) = -\frac{1}{\beta}\log\left(\sum_\mu \varphi\left(\sqrt{\beta}\,\mathbf{x}\right)\cdot\varphi\left(\sqrt{\beta}\,\xi^\mu\right)\right)$$

where $\varphi(\mathbf{x}) = \frac{1}{\sqrt{Y}}\begin{bmatrix}\cos(\langle\omega^1,\mathbf{x}\rangle)\\ \sin(\langle\omega^1,\mathbf{x}\rangle)\\ \dots,\\ \cos(\langle\omega^Y,\mathbf{x}\rangle)\\ \sin(\langle\omega^Y,\mathbf{x}\rangle)\end{bmatrix}$, $\omega^\alpha \sim \mathcal{N}(0, I_D)$ are the random features

noisy query's approx. energy inner prod of query features and entangled memory features

$$\tilde{E}(\mathbf{x}) = -\frac{1}{\beta}\log\left(\varphi\left(\sqrt{\beta}\,\mathbf{x}\right)\cdot\sum_\mu \varphi\left(\sqrt{\beta}\,\xi^\mu\right)\right)$$

Memory tensor T of constant size Y

Once **memory tensor T** is computed, the original patterns ξ^μ can be discarded. New memories can be introduced with additive updates to **T**!