Expander Dictionary Learning

Michael Murray[1,2], Jared Tanner[1,2]

Contact Email: mmurray@turing.ac.uk

[1]University of Oxford, [2]Alan Turing Institute

1. Overview of Dictionary Learning

- Dictionary Learning is a feature learning technique in which it is assumed that each data point can be represented as a linear combination of a common set of vectors.
- Given a data matrix $\mathbf{Y}$, the objective is to learn a 'dictionary' $\mathbf{A}$ and latent representation $\mathbf{X}$ such that $\mathbf{Y} \approx \mathbf{AX}$.
- The original motivation was computing sparse representations. Hence $\mathbf{X}$ is typically assumed to be sparse, meaning the data lies on a union of low dimensional subspaces.
- Dictionary learning typically involves solving an optimisation problem which is not jointly convex in $\mathbf{A}$ and $\mathbf{X}$, and is NP-hard in general.
- Despite this a number of algorithms work well in practise, these algorithms alternate between the $\mathbf{A}$ and $\mathbf{X}$, iteratively updating one while holding the other constant.

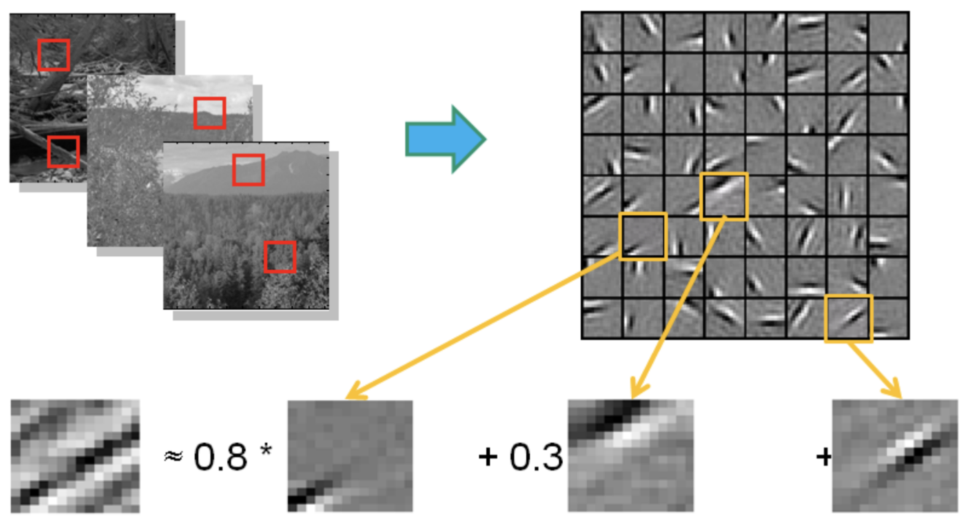
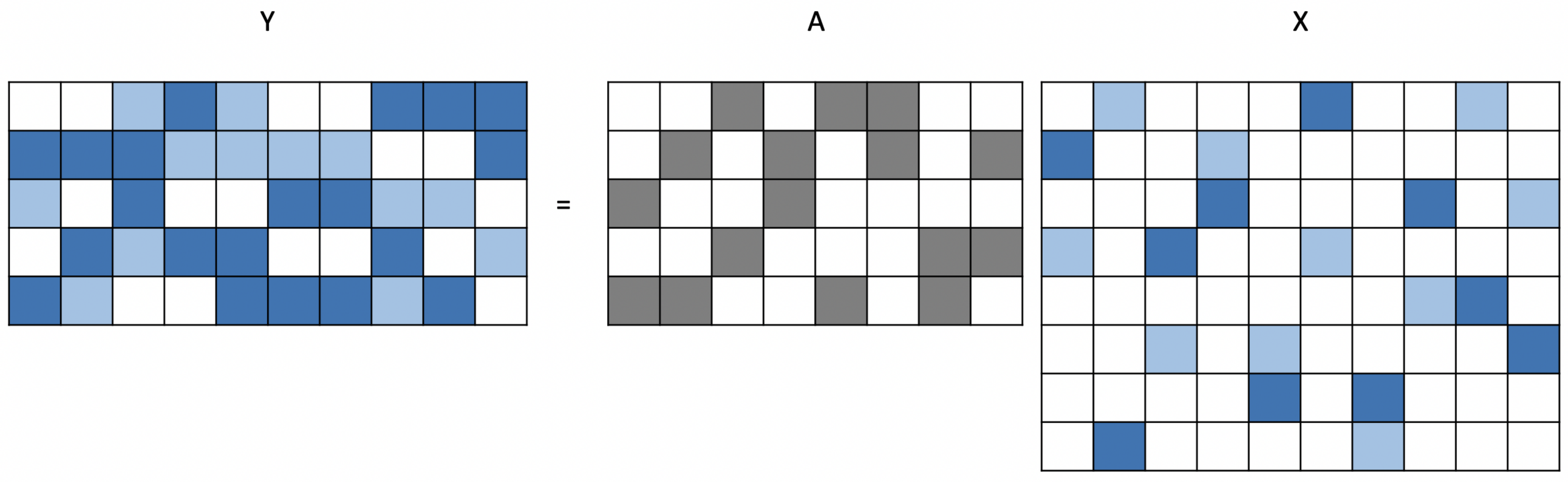


Image by Andrew Ng

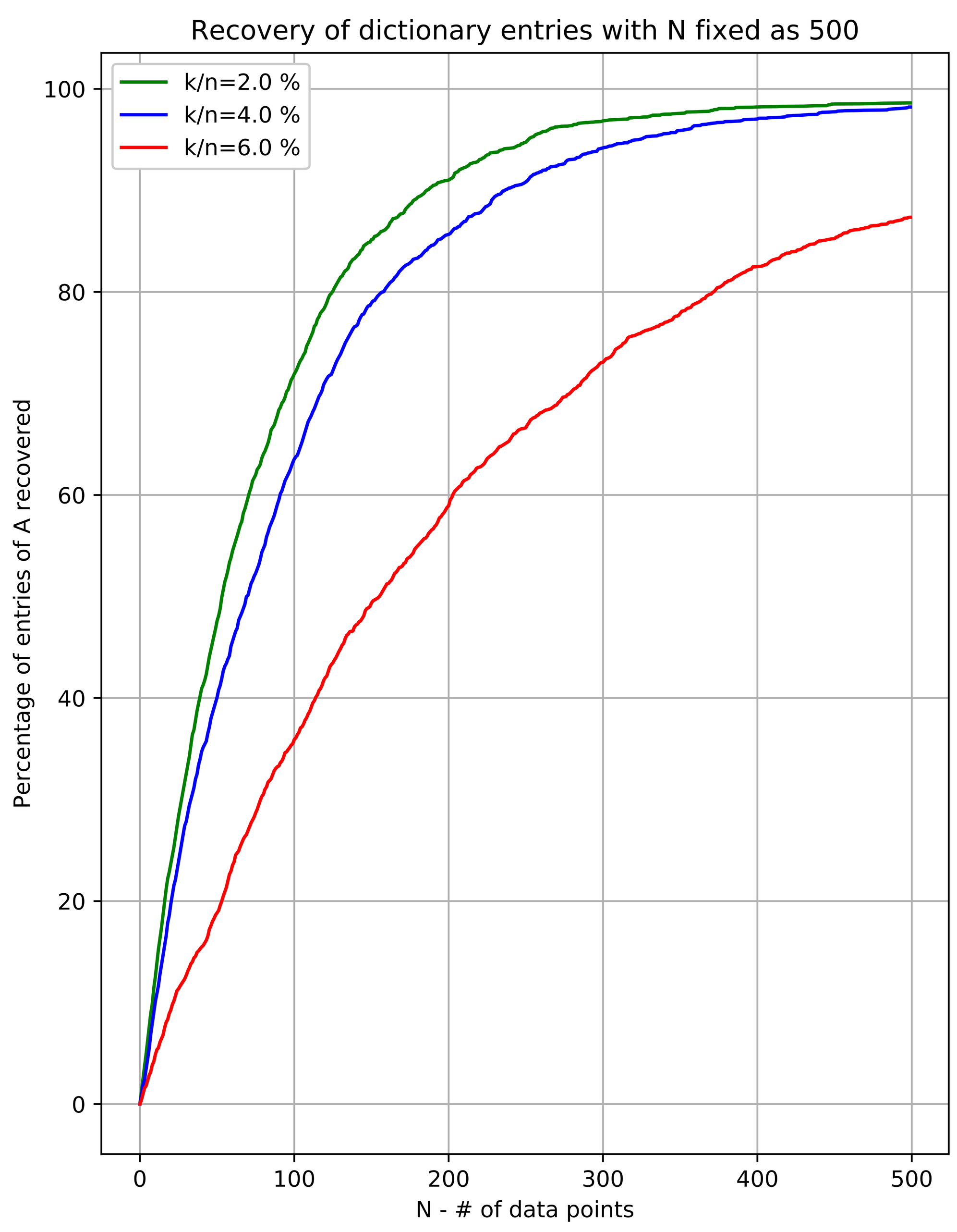
2. Problem Definition

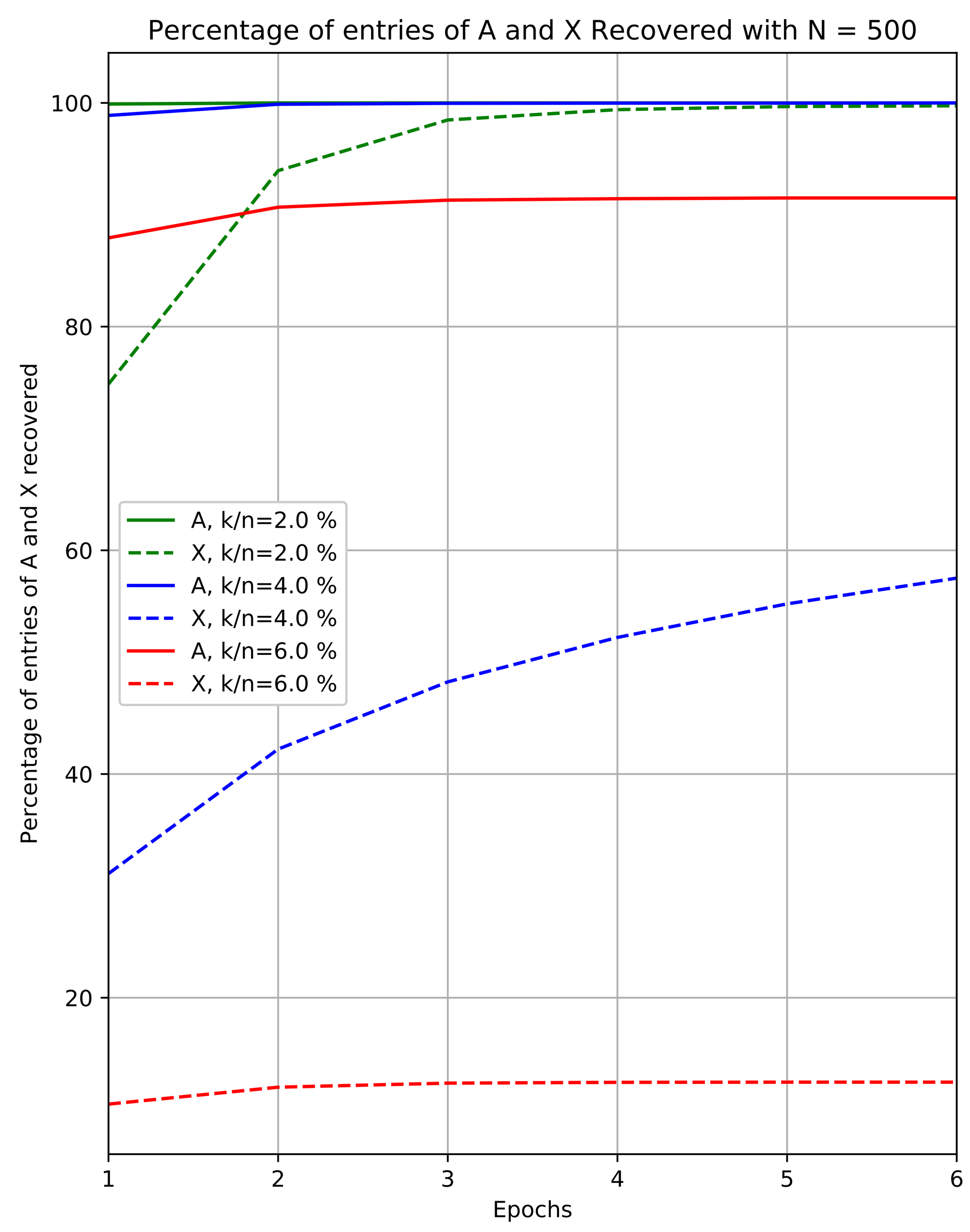
- The algorithms used in practice are heuristic and lack recovery guarantees.
- A number of researchers have tackled this problem in recent years by making additional assumptions on $\mathbf{A}$, e.g., square and full column rank [8], incoherent [3] [1], sparse and random [7] [2].
- In our work we consider a new problem, in which $\mathbf{A}$ is column d sparse and binary and $\mathbf{X}$ is real, column k sparse and dissociated. It can hence be viewed, in the context of combinatorial compressed sensing [6], as sparse recovery with an unknown sensing matrix.
- Our goal is to derive an algorithm and then analyse its typical performance by providing lower bounds on the probability of success, given that the dictionary and latent representation are samples from a particular pair of distributions.



3. Contribution I - Expander Based Dictionary Recovery (EBDR) Algorithm

Fixing m, n, d and k, we sample the dictionary from the uniform distribution over the set of all column d sparse binary matrices with dimension $m \times n$. Under an appropriate regime for the parameters m, n, d and k, the dictionary is with high probability the adjacency matrix of a k, ϵ, d expander graph [4] [5]. This insight forms the basis of our approach.





$N = 500$, $n = 1200$, $m = 960$, $d = 9$ and $\epsilon = 1/6.1$. RHS computational running time; 60s for $k/n = 2\%$, 495s for $k/n = 4\%$, 412s for $k/n = 6\%$. Ran on 2017 MacBook Pro with a 3.1 GHz Intel Core i7 processor, no parallelisation.

Algorithmic Ideas:

1. **Unique Neighbour Property:** some entries in a column of $\mathbf{Y}$ will be directly entries of $\mathbf{X}$, we call these 'singletons'.
2. **Identifying singletons:** the frequency with which a value appears in a column of $\mathbf{Y}$ is a certificate for whether it is a singleton.
3. **Extracting partial supports:** the locations of a singleton give a subset of the locations of a column of $\mathbf{A}$, we call these 'partial supports'.
4. **Clustering:** partial supports can be grouped by column of origin, so long as $\epsilon < 1/6$, simply by computing their pairwise inner products.
5. **Reconstruction:** a column of $\mathbf{A}$ can be reconstructed by calculating the union of the supports of its partial supports.

4. Contribution II - Recovery Guarantees for EBDR

For fixed m, n, d and k let A be a random matrix whose pmf is the uniform distribution over all d column sparse binary matrices. Let X be a random matrix whose pdf is uniform over the set of k column sparse real matrices with entries in some interval $\mathcal{C} \subset \mathbb{R}$, $0 \notin \mathcal{C}$. Given $Y = AX$ suppose $EBDR(Y, m, n, d, k) = (\hat{A}, \hat{X})$ and let P be the permutation matching each reconstructed column to its target index.

Theorem 1. Let Λ_0 be the event that $\{A \in \mathcal{E}^{m \times n}_{k,\epsilon,d}\} \cap \{\epsilon \leq 1/6\}$. Define

$$\delta_+ = \min\{1, 2d(n+1)e^{-2(n-1)(\frac{1}{n} - \frac{d}{m})^2}\}$$
$$\delta_- = 2d(n+1)e^{-2(n-1)(\frac{n+k-1}{n} - \frac{d}{m})^2},$$

For any $\delta_* \in (\delta_-, \delta_+)$ let $\alpha = \frac{d}{m} + \sqrt{\frac{\ln\left(\frac{2d(n+1)}{\delta_*}\right)}{2(n-1)}}$ and $L = \left(\frac{n - n\alpha - k + 2}{n - k + 1}\right)^{-(k-1)} \ln\left(\frac{2d(n+1)}{\delta_*}\right)$. So long as

$$N \geq \delta_*^{-1}(1 + 2\epsilon d) Ln(n+1)(\ln(n) + 1)$$

then there exists a random permutation $P \in \mathcal{P}$ such that

$$\mathbb{P}\left((\hat{A}P,\ P\hat{X}) = (A, X) \mid \Lambda_0\right) > 1 - \delta_*.$$

5. Appendix - Expander Graphs

(Definition of (k, ϵ, d) expander) An unbalanced, left d-regular bipartite graph $G = ([n], [m], E)$ is a (k, ϵ, d) expander iff

$$|\mathcal{N}(\mathcal{S})| > (1 - \epsilon)d|\mathcal{S}| \quad \forall \quad \mathcal{S} \in [n]^{(\leq k)}.$$

(Unique Neighbour Property) Suppose that G is an unbalanced, left d-regular bipartite graph $G = ([n], [m], E)$. Let $\mathcal{S}$ be any subset of nodes $\mathcal{S} \in [n]^{(\leq k)}$ and define $\mathcal{N}_1(\mathcal{S}) = \{i \in \mathcal{N}(\mathcal{S}) \ \ s.t. \ \ |\mathcal{N}(i) \cap \mathcal{S}| = 1\}$. If G is a (k, ϵ, d) expander graph then

$$|\mathcal{N}_1(\mathcal{S})| > (1 - 2\epsilon)d|\mathcal{S}| \ \ d\forall \ \ \mathcal{S} \in [n]^{(\leq k)}.$$

References

[1] A. Agarwal, A. Anandkumar, and P. Netrapalli. A Clustering Approach to Learn Sparsely-Used Overcomplete Dictionaries. *arXiv e-prints*, page arXiv:1309.1952, Sep 2013.

[2] S. Arora, A. Bhaskara, R. Ge, and T. Ma. Provable bounds for learning some deep representations. *CoRR*, abs/1310.6343, 2013.

[3] S. Arora, A. Bhaskara, R. Ge, and T. Ma. More algorithms for provable dictionary learning. *CoRR*, abs/1401.0579, 2014.

[4] B. Bah and J. Tanner. Vanishingly Sparse Matrices and Expander Graphs, With Application to Compressed Sensing. *ArXiv e-prints*, July 2012.

[5] B. Bah and J. Tanner. On the construction of sparse matrices from expander graphs. *CoRR*, abs/1804.09212, 2018.

[6] R. Mendoza-Smith and J. Tanner. Expander l0-decoding. *Applied and Computational Harmonic Analysis*, 2017.

[7] D. A. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. *CoRR*, abs/1206.5882, 2012.

[8] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere I: overview and the geometric picture. *CoRR*, abs/1511.03607, 2015.