K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation

Summary written by Mahban Gholijafari
February 17, 2025

Summary: This paper proposes a novel algorithm for adapting dictionaries to achieve sparse signal representation named K-SVD. This algorithm generalizes K-means clustering process and can work with any pursuit method.

Related work: There are multiple applications that can benefit from the sparsity and/or overcompleteness concepts including sparsity of the wavelet coefficients of natural images leading to the success of JPEG2000 coding standard [13], denoising with wavelet methods and shift-invariant variations [6, 3, 14, 15], dynamic range compression in images [9], seperation of texture and cartoon content in images [16, 17], inpainting [7], and etc. Exact determination of sparsest representations is an NP-hard problem [4] therefore many efficient pursuit algorithms to approximate solutions have been proposed like matching pursuit (MP) [12], orthogonal matching pursuit (OMP) [1], basis pursuit (BP) [2], the focal underdetermined system solver (FOCUSS) [10], and maximum a posteriori (MAP) estimation [11]. These approximated versions has shown stability in recovery of x [5]. The motive for this work is based on the mentioned relation between sparse representation and clustering i.e. vector quantization [8].

Approach: This paper first describes K-means algorithm for vector quantization. The dictionary of VQ codewords is typically trained using the K-means algorithm. They denote the codebook matrix with codewords as columns by $C = [c_1, c_2, ..., c_K]$. Each signal is its cosest coeword under ℓ^2-norm distance when C is given. Therefore, $y_i = Cx_i$ where $x_i = e_j$ is a vector from the trivial basis, with all zero entries except a one in the jth position. The representation MSE per y_i is defined as $e_i^2 = ||y_i - Cx_i||_2^2$ and the overall MSE is $E = \sum_{i=1}^{K} e_i^2 = ||Y - CX||_F^2$. The VQ training problem is to find a codebook C to minimize error E. To find an optimal codebook for VQ, K-means algorithm is used and each iteration of it has two stages, 1. sparse coding that evaluates X and 2. updating the codebook. For sparse coding, we need to find the best possible dictionary for the sparse representation of the example set Y. The process updates one column at a time and fixes all columns in D except one, d_k, and finds a new column d_k and new values for its coefficients that best reduce the MSE. The sparse coding step is necessary to avoid falling into a local minimum trap. Our objective function is: $\min_{D,X}\{||Y - DX||_F^2\}$ subject to $\forall_i, ||x_i||_0 \leq T_0$ and with the assumption that both X and D are fixed, the penalty term can be rewritten as: $||Y - DX||_F^2 = ||E_k - d_k x_T^k||_F^2$. This way, the multiplication DX has been decomposed to the sum of K rank-1 matrices, where K-1 terms are assumed fixed and one (the kth) remains in question. The matrix E_k is the error for all the N examples when the kth atom is removed. To solve this problem, a group of indices pointing to examples $\{y_i\}$ is defined and the row vector x_T^k shrinks by discarding the zero entries and will do the same for the error matrix and then we can do the minimization via SVD. K-SVD algorithm always uses the most updated coefficient as they merge from preceding SVD steps. When the nonzero entries are small enough realtive to n, the OMP, FOCUSS, and BP approximating methods are known to perform well.

Datasets, Experiments and Results: This paper first has tried their algorithm on synthetic signals, 1500 data signals of dimension 20 were produced and white Gaussian noise with varying SNR was added to the resulting signals. Measured distance for experiments is via $1 - |d_i^T \tilde{d}_i|$ where d_i is a generating dictionary atom and $\tilde{d}_i$ is its corresponding element in the recovered dictionary. The MAP-based algorithm improved with more executed iteration and gets closer to K-SVD detection rates. Several experiments on natural image data have also been done with 11000 examples of block patches of size 8×8 pixels taken from a database of face images. A compression comparison was conducted between overcomplete Haar dictionary and DCT dictionary.

Strengths: The dictionary found by the K-SVD algorithm performs well for both synthetic and real images in applications such as filling in missing pixels and compression and outperforms other dictionaries such as nondecimated Haar and overcomplete or unitary DCT.

Weaknesses: Just like K-means, K-SVD is susceptible to local minimum traps. The convergence of the proposed algorithm depends on the success of pursuit algorithms to robustly approximate the solution and convergence is not always guaranteed. K-SVD has scalability problem when turning to work with larger image patches.

Reflections: The authors claim that this kind of dictionary can replace popular representation methods both in image enhancement and in compression. Future work can be in exploring the connection between the chosen pursuit methods in the K-SVD and the method used later in the application, studying the effect of adding weights to the atoms and handling the scalability problem of K-SVD with larger image patches.

References

[1] S. Chen, S. A. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of control*, 50(5):1873–1896, 1989.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[3] R. R. Coifman and D. L. Donoho. *Translation-invariant de-noising*. Springer, 1995.

[4] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive approximation*, 13:57–98, 1997.

[5] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on information theory*, 52(1):6–18, 2005.

[6] D. L. Donoho, I. M. Johnstone, et al. Ideal denoising in an orthonormal basis chosen from a library of bases. *Comptes rendus de l'Académie des sciences. Série I, Mathématique*, 319(12):1317–1322, 1994.

[7] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and computational harmonic analysis*, 19(3):340–358, 2005.

[8] K. Engan, S. O. Aase, and J. H. Husøy. Multi-frame compression: Theory and design. *Signal Processing*, 80(10):2121–2140, 2000.

[9] R. Gastaud and J. Starck. Dynamic range compression: A new method based on wavelet transform. In *Astron. Data Anal. Software Systems Conf*, 2003.

[10] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm. *IEEE Transactions on signal processing*, 45(3):600–616, 1997.

[11] M. S. Lewicki and B. A. Olshausen. Probabilistic framework for the adaptation and comparison of image codes. *JOSA A*, 16(7):1587–1601, 1999.

[12] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12):3397–3415, 1993.

[13] M. W. Marcellin, M. J. Gormish, A. Bilgin, and M. P. Boliek. An overview of jpeg-2000. In *Proceedings DCC 2000. Data compression conference*, pages 523–541. IEEE, 2000.

[14] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multiscale transforms. *IEEE transactions on Information Theory*, 38(2):587–607, 1992.

[15] J.-L. Starck, E. J. Candès, and D. L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on image processing*, 11(6):670–684, 2002.

[16] J.-L. Starck, M. Elad, and D. L. Donoho. Image decomposition: Separation of texture from piecewise smooth content. In *Wavelets: Applications in Signal and Image Processing X*, volume 5207, pages 571–582. SPIE, 2003.

[17] J.-L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE transactions on image processing*, 14(10):1570–1582, 2005.