On the Performance of Random-Projection-based Dimensionality Reduction for Endmember Extraction

Qian Du, James E. Fowler
Department of Electrical and Computer Engineering
Mississippi State University

ABSTRACT

Random projection (RP) is a computationally efficient and data-independent method for dimensionality reduction (DR). The transformation matrix of RP generally includes Gaussian distributed random vectors, which are mutually orthogonal unit vectors. Theoretical results indicate that it can well preserve distances among data points as well as the structure of data cloud [1-3]. Recently, RP is of great interest because the mathematical theory of popular compressed sensing shows that sparsely representable signals can be recovered exactly from RP [4].

The traditional DR methods, such as principal component analysis (PCA) and maximum noise fraction (MNF) transform, are data-dependent and involve burden-some computations. For a hyperspectral image with $M \times N$ pixels and L bands, PCA needs L^2MN for covariance matrix calculation and $O(L^3)$ for eigen-decomposition; MNF needs the computations for noise covariance matrix estimation and its eigen-decomposition, in addition to those in the original PCA. If RP is adopted, tremendous savings can be achieved. If the original data is reduced to K-dimensional, the computation time involved in the random matrix generation is only $O(K^2L)$ including the expensive Gram-Schmidt process; it can be further reduced to $O(KL)$ if using uniformly distributed random variables without orthogonalization. Note that $K << L << MN$.

Endmember extraction is an important step in spectral mixture analysis for hyperspectral imagery. Usually, DR is a preprocessing step for endmember extraction, which can not only save computational time but also improve the performance. The performance improvement comes from the automated removal of spectrally trivial variation in pixel signatures [5]. In general, MNF is preferred since it can compact the major data information in terms of signal-to-noise ratio, a better criterion than variance used in PCA. So, in this paper, we will evaluate the

performance of RP-based DR for endmember extraction and compare it with the MNF-based DR. Widely used algorithms, such as N-FINDR [6] and VCA [7], will be examined. When a spectral library is known, extracted endmembers can be simply compared with true endmember signatures using spectral angle. If no ground truth about endmember signatures is available, the volume of the simplex formed by extracted endmembers is calculated, and the set resulting in a larger volume is considered as a better one since it includes more distinctive endmembers; or, the pairwise spectral angles are computed, and the set providing the larger average spectral angle is claimed to be the better one.

Some preliminary result was shown in Table I, which is about AVIRS Lunar Lake data of size 200×200 with 158 bands (after removing bad bands). In this case, we know the 6 endmember classes but not their precise spectral signatures. The data dimensionality was reduced to 6 and 12 using MNF and RP. The N-FINDR algorithm was applied, whose initials were randomly chosen. 50 runs were executed, and for each run the initials for N-FINDR were different and the transformation matrix for RP was also changed. When the data dimensionality was reduced to 6, 7 out of 50 times the 6 true endmember classes were found for both MNF and RP-transformed data. The volumes of the simplex constructed by the six extracted endmembers (in the original space) were computed. As show in Table I, the average volume (and the minimum volume) when using RP was smaller than using MNF, indicating poor quality of extracted endmembers; however, the maximum volume from RP was larger than that from MNF. When the data dimensionality was reduced to 12, all the 6 true endmember classes were found for both MNF and RP in each trial; the simplex constructed by these 6 endmembers was evaluated. As listed in Table I, the average volume (and the minimum volume) when using RP was smaller than using MNF but the maximum volume from RP was greater.

TABLE I

SIMPLEX VOLUME CONSTRUCTED BY THE SIX EXTRACTED ENDMEMBERS

	Reduced Dimensionality					
	6			12		
	MAX	MIN	AVERAGE	MAX	MIN	AVERAGE
MNF	2.0574×10^{12}	1.8053×10^{12}	2.0302×10^{12}	2.6376×10^{12}	1.6486×10^{12}	2.1281×10^{12}
RP	2.5039×10^{12}	0.3674×10^{12}	1.2758×10^{12}	2.6742×10^{12}	1.0263×10^{12}	1.7672×10^{12}

With extensive evaluations using several datasets, this paper will draw the following conclusions.

1. RP-based DR results in comparable endmembers, particularly when the reduced dimensionality is higher than the number of endmembers.
2. If multiple runs are allowed, RP-based DR may provide better performance than conventional PCA and MNF-based DR.
3. A simplest RP matrix from uniformly distributed random variables without Gram-Schmidt orthogonalization provides similar performance, but needs the lowest computations in forming the random matrix.

REFERENCES

[1] S. Dasgupta, "Experiments with random projections," In *Proceedings of the Sixteen Conference on Uncertainty in Artificial Intelligence*, pp. 143-151, 2000.

[2] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *Proceedings of ACM Conference on Knowledge Discovery and Data Mining*, pp. 245-250, 2001.

[3] D. Achlioptas, "Database-friendly random projections," In *Proceedings of ACM Symposium on the Principles of Database Systems*, pp. 271-281, 2001.

[4] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[5] Q. Du, N. Raksuntorn, N. H. Younan, and R. L. King, "Endmember extraction for hyperspectral imagery," *Applied Optics*, vol. 47, no. 28, pp. F77-F84, Oct. 2008.

[6] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE 3753*, pp. 266-275, 1999.

[7] J. M. P. Nascimento and J. M. Bioucas Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898-910, Apr. 2005.