DETERMINING HYPERSPECTRAL DATA-INTRINSIC DIMENSIONALITY VIA A MODIFIED GRAM-SCHMIDT PROCESS

Oleg Kuybeda, Asaf Kagan and Yaakov Lumer

Signal and Image Processing Laboratory
Department of Electrical Engineering
Technion - IIT, Haifa, 32000, Israel
Email: koleg@tx.technion.ac.il

ABSTRACT

The overdetermined nature of hyperspectral data consitutes a serious obstacle in many applicative fields. A vital step in dimensionality reduction is determining the intrinsic number of dimensions the signal resides in.

This work proposes a Modified Gram-Schmidt (MGS) process which iteratively finds the most distant pixels within the data in terms of an orthogonal complement norm (OCN) to a subspace spanned by the extreme pixels found in previous iterations.We analyze the distribution of extreme OCN using Exreme Values Theory (EVT) and derive a termination condition for the MGS process. The dimensionality is determined by the number of found extreme pixels, which provide an estimation for the signal subspace.

Keywords: hyperspectral, endmember, rank, gram-schmidt, min-max, dimensionality, extreme value, orthogonal subspace

1. INTRODUCTION

A variety of approaches have been proposed in [1],[2],[3] for automatic determination of the pure materials spectra (endmembers) in the hyperspecral scene. Each pixel in the hyperspectral image can be represented as a linear combination of the spectra of pure materials. Apparently, in an ideal case, all hyperspectral pixels reside in a linear subspace spanned by the pure materials spectra. Determining the dimensions of this subspace is an important task for subsequent processing.

Since different endmembers are reasonably assumed to be linearly independent, we may employ a Gram-Schmidt process to find each endmember orthogonal contribution to the signal subspace. By choosing pixels that contribute maximally (have the maximal OCN) to the subspace, we likely obtain the purest pixels in the scene.

Common tools for estimating the intrinsic number of dimensions are principal component analysis (PCA) and its noise adjusted version, the minimum noise fraction (MNF) algorithms. These algorithms employ second order statistics that evaluate the spatial signal distribution in terms of energy. As a result, rare materials spectra in the scene that appear in few pixels only and don't contribute sufficiently to the energy, are disregarded. In many applications this is not affordable. An advantage of the proposed algorithm is that it does not ignore such "rare" pixels and, nevertheless, estimates the rank of the signal subspace.

2. THE LINEAR MIXTURE MODEL

Given that the number of spectral bands is b and $\mathbf{x}_i \in \mathbb{R}^b$ are the observed pixels, the linear mixture model is defined as following

$$\mathbf{x}_i = \mathbf{s}_i + \mathbf{n}_i \tag{1}$$

where $\mathbf{s}_i = \mathbf{A}\boldsymbol{\beta}_i$ is the signal part of the observations, which is a nonnegative linear combination of r linearly independent endmembers $\mathbf{A} = [\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_r]$ and $\boldsymbol{\beta}_i$ is corresponding vector of their proportions in an observation pixel $\mathbf{x}_i$. The $\mathbf{n}_i$ are assumed to be spatially and spectrally white, zero-mean Gaussian noise, with $E\{\mathbf{n}\mathbf{n}'\} = \sigma^2\mathbf{I}^b$.

The index $i = 1, \dots, T$ denotes the spatial position of a pixel under cosideration out of total T data pixels. It will be omitted in the sequel when mentioning all data pixels.

3. MODIFIED GRAM-SCHMIDT

Our objective is to estimate $\mathbf{A}$ or, more precisely, the range of $\mathbf{A}$ by finding the purest pixels and the rank of $\mathbf{A}$. The proposed way to do that is to use iterative MGS.

The iterations begin with $\hat{\mathbf{A}}_1 = \hat{\mathbf{a}}_1 = \mathbf{x}_m$, where $||\mathbf{x}_m|| = \max_{i=1\dots T} ||\mathbf{x}_i||$. At each step k, every $\mathbf{x}$ is decomposed as

$$\mathbf{x} = \mathbf{x}^p + \mathbf{x}^c \tag{2}$$

where $\mathbf{x}^p \in \mathcal{R}(\hat{\mathbf{A}}_{k-1})$ and $\mathbf{x}^c \in \mathcal{N}(\hat{\mathbf{A}}_{k-1})$. The $\mathcal{R}$ and $\mathcal{N}$ denote the range and null spaces of $\hat{\mathbf{A}}_{k-1}$ respectively. Thus $\mathbf{x}^c$ is an orthogonal complement of $\mathbf{x}$ onto a subspace spanned by the columns of $\hat{\mathbf{A}}_{k-1}$. The iteration from $k-1$ to k is performed as following

$$\begin{aligned} m &= \underset{i}{\text{argmax}} ||\mathbf{x}_i^c|| \\ \hat{\mathbf{a}}_k &= \mathbf{x}_m \\ \hat{\mathbf{A}}_k &= \left[\hat{\mathbf{A}}_{k-1} | \hat{\mathbf{a}}_k\right]. \end{aligned} \tag{3}$$

At each step, $\mathbf{s}$ and $\mathbf{n}$ can be uniquely decomposed as

$$\mathbf{s} = \mathbf{s}^p + \mathbf{s}^c \tag{4}$$

$$\mathbf{n} = \mathbf{n}^p + \mathbf{n}^c \tag{5}$$

where $\mathbf{s}^p, \mathbf{n}^p \in \mathcal{R}(\hat{\mathbf{A}}_{k-1})$ and $\mathbf{s}^c, \mathbf{n}^c \in \mathcal{N}(\hat{\mathbf{A}}_{k-1})$. Due to uniqueness of a direct sum $\mathbb{R}^b = \mathcal{R} \oplus \mathcal{N}$, the following holds:

$$\mathbf{x}^c = \mathbf{s}^c + \mathbf{n}^c. \tag{6}$$

Obviously, we would like to terminate the process, when the next $\hat{\mathbf{a}}_k$ does not provide a reliable contribution to the signal subspace, in other words, $\hat{\mathbf{a}}_k$ is mainly managed by the noise. In the next sections we analyze the behavior of noise component in $\hat{\mathbf{a}}_k$ and derive the termination condition.

4. THE NORM DISTRIBUTION

Let $\mathbf{C}_l$ be a matrix, whose rows constitue an orthogonal basis to $\mathcal{N}(\hat{\mathbf{A}}_k)$ with

$$\text{rank } \mathbf{C}_l = l = b - k \tag{7}$$

Then

$$||\mathbf{n}^c||^2 = ||\mathbf{C}_l\, \mathbf{n}^c||^2 = ||\mathbf{C}_l \mathbf{n}||^2 \tag{8}$$

Lets denote $\boldsymbol{\lambda} \doteq \mathbf{C}_l \mathbf{n}$ to be a random vector with l components.

$$E\{\boldsymbol{\lambda}\boldsymbol{\lambda}'\} = E\{\mathbf{C}_l \mathbf{n}\mathbf{n}'\mathbf{C}_l'\} = \sigma^2 \mathbf{I}^l \tag{9}$$

It is a white Gaussian noise vector with l components. Moreover,

$$\max_i ||\boldsymbol{\lambda}_i|| \equiv \max_i ||\mathbf{n}_i^c|| \tag{10}$$

Consider a random variable $||\boldsymbol{\lambda}||^2$. It has a Chi-squared distribution of order l denoted by $\chi^2(l, \sigma^2)$ with the following pdf:

$$f(u) = \frac{1}{2^{l/2}\Gamma(l/2)\sigma^2} \left(\frac{u}{\sigma^2}\right)^{(l/2)-1} e^{-u/2\sigma^2} \tag{11}$$

For sufficiently large l as we have at hand ($l >> 1$), $\chi^2(l, \sigma^2)$ can be approximated by $N(l\sigma^2, 2l\sigma^4)$ as the limiting distribution of a sum of l i.i.d squared components of $\boldsymbol{\lambda}$. Thus, for qualitative evaluations, we can use

$$||\boldsymbol{\lambda}||^2 \sim N(l\sigma^2, 2l\sigma^4). \tag{12}$$

If the noise n is not spectrally white, then $E\{\boldsymbol{\lambda}\boldsymbol{\lambda}'\}$ will not nesserarily be given by (9). Suppose, that the noise covariance-matrix eigenvalues are decreasingly ordered as ($\lambda_1^2 \geq \lambda_2^2 \geq \ldots \geq \lambda_l^2$). Then $||\boldsymbol{\lambda}||^2$ can be written as a sum of l independent $\chi^2(1, \lambda_j^2)$ variables with means and variances equal to λ_j^2 and $2\lambda_j^4$, respectively, for $j = 1 \ldots l$. Therefore, the approximation given in (12) can be rewritten as follows:

$$\begin{aligned} ||\boldsymbol{\lambda}||^2 &\sim N(l_m\sigma_e^2, 2l_v\sigma_e^4), \qquad \sigma_e^2 = \lambda_1^2 \\ l_m &= \sum_{j=1}^{l} \frac{\lambda_j^2}{\lambda_1^2}, \qquad l_v = \sum_{j=1}^{l} \frac{\lambda_j^4}{\lambda_1^4} \end{aligned} \tag{13}$$

5. EVT AND THE MAXIMUM NORM

The modified Gram-Schmidt uses maximum norm of the residuals $\{||\mathbf{x}_i^c||\}_{i=1}^T$ in order to obtain the next column of $\hat{A}$. Let's consider the distribution of $M_T \doteq \max_{i=1\ldots T} ||\boldsymbol{\lambda}_i||^2$.

5.1. Gaussian approximation analysis

Theorem 1. *If $\{\xi_n\}$ is an i.i.d. (standard) normal sequence of random variables, then the asymptotic distribution of*

$$M_n = \max\{\xi_1 \ldots \xi_n\}$$

satisfies

$$P(a_n(M_n - b_n) \leq x) \underset{n\to\infty}{\longrightarrow} \exp(-e^{-x}) \tag{14}$$

where

$$\begin{aligned} a_n &= (2\log n)^{1/2} \\ b_n &= (2\log n)^{1/2} - \\ &\quad \frac{1}{2}(2\log n)^{-1/2}(\log\log n + \log 4\pi). \end{aligned}$$

The proof is found in [4].

In general, the distribution $G(x) \doteq \exp(-e^{-x})$, also known as the Gumbel distribution, is a limiting distribution of a maximum of n random variables with exponential distribution tails. Its mean and std in the given standard form are $\eta = 0.5772$ and $\gamma = 1.6450$ respectively. Although, Theorem 1 provides us with normalizing coefficients a_n, b_n for the Gaussian variables, the coefficients for $\chi^2(l, \sigma^2)$ variables (as $||\boldsymbol{\lambda}_i||^2$ are) don't have a known analytical form. We proceed with our quantitative analysis adopting the Gaussian approximation (13) in order to obtain

$$P(M_T \leq x) \approx G\left(a_T \left[\frac{x - \sigma_e^2 l_m}{\sigma_e^2\sqrt{2l_v}} - b_T\right]\right) \tag{15}$$

with mean and std as following

$$\mu_T = \sigma_e^2 \left(\frac{\sqrt{2l_v}}{a_T}\eta + b_T\sqrt{2l_v} + l_m\right) \tag{16}$$

$$\sigma_T = \frac{\sigma_e^2\sqrt{2l_v}}{a_T}\gamma \tag{17}$$

While this approximation doesn't provide us with accurate mean and std of M_T, it is instructive to look at the ratio

$$\frac{\mu_T}{\sigma_T} \propto 2\log(T) + \log(T)^{1/2}\frac{l_m}{\sqrt{l_v}} \tag{18}$$

$$1 \leq \frac{l_m}{\sqrt{l_v}} \leq \sqrt{l} \tag{19}$$

As can be seen, the ratio doesn't depend on $\hat{\sigma}_e^2$, it is log-dependent on T, and in terms of l_m, l_v, it accepts the maximum when ($\lambda_1^2 = \lambda_2^2 = \ldots = \lambda_l^2$) (i.e. the noise is white and $l_m/\sqrt{l_v} = \sqrt{l}$). Thus the ratio μ_T/σ_T tends to infinity as $T \to \infty$ or $l \to \infty$. For typical parameters $l = 100$, $T = 10^4$, and white noise, $\mu_T/\sigma_T \approx 30$. The dominant factor in obtaining such a high ratio is the high dimenionality l.

To conclude the above derivations, we say that for sufficiently large T and when the noise is approximately white, M_T can be substituted by μ_T almost surely, i.e. with relatively small deviation.

5.2. The Chi-squared normalization

As we have seen above, the extreme value theory proposes μ_T as a reliable estimation of the most probable extreme value of a norm of a multivariate noise. It was shown in (18), that the reliability in terms of the ratio μ_T/σ_T attains its maximum when the noise is white. Theorem 1, however, doesn't provide a way for exact calculation of μ_t and σ_T for $\chi^2(l, \sigma^2)$ variables. The next theorem proposes a recipe to do it for almost any distribution of interest.

Theorem 2. *If $\{\xi_n\}$ is an i.i.d. with absolutely continuous distribution $F(x)$ and density $f(x)$, then letting*

(i) $h(x) = f(x)/(1 - F(x))$
(ii) $b_n = F^{-1}(1 - \frac{1}{n})$
(iii) $a_n = h(b_n)$
(vi) $\omega = lim_{x \to x^*} h'(x)$*, where* x^* *is the upper end-point of* F*,*

then for $M_n = \max\{\xi_1 \ldots \xi_n\}$

$$P(a_n(M_n - b_n) \leq x) \underset{n \to \infty}{\longrightarrow} \qquad (20)$$
$$\begin{cases} \exp(-e^{-x}), & if\ \omega = \infty \\ \exp\{-[1 + \frac{x}{\omega}]^{\omega}\}, & if\ \omega < \infty \end{cases}$$

The proof is found in [5].

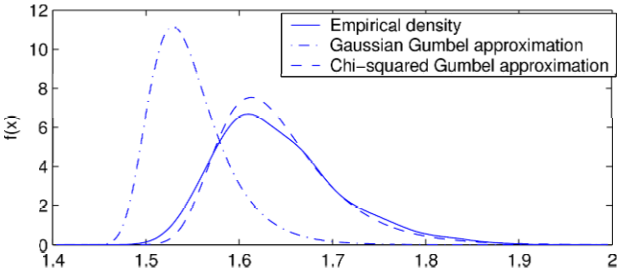

Fig. 1: Gumbel limit of the $||\boldsymbol{\lambda}||^2$ extreme value versus its distribution approximation. The Gaussian approximation (dot-dashed line) provides only a qualitative evaluation of the ratio μ_T/σ_T, whereas the exact Chi-squares (dashed line) precisely matches the Monte-Carlo simulated (solid line) density.

The determination of M_T distribution is summarized in Fig. 1. We have performed a Monte-Carlo simulation using white Gaussian (normal) multivariate noise. The solid line corresponds to the density of the simulated maximium of T noise realizations. The dot-dashed line corresponds to Gumbel density limit using Gaussian normalization. The dashed line, which is obtained using Chi-squared normalization fits well the simulated one.

6. RANK DETERMINATION

Now we can define the termination condition for the Modified Gram-Schmidt process. Our concern is to ensure that at each iteration, the signal part in $\hat{a}_k$, $||\mathbf{s}_m^c||$ dominates over the most probable value for maximum noise realization

$$M_T = \max ||\mathbf{n}^c|| \overset{a.s}{=} \sqrt{\mu_T},$$

where *a.s* denotes *almost surely*.

Heuristically, we want to ensure, that $\hat{a}_k$ contributes a "new" dimension almost surely due to the signal. We realize that with the following termination condition

$$||\mathbf{x}_m^c|| \leq \alpha\sqrt{\mu_T}. \qquad (21)$$

where $1 \leq \alpha \leq 2$. Using triangle inequality, we obtain

$$\begin{aligned} ||\mathbf{s}_m^c|| &\leq ||\mathbf{x}_m^c|| + ||\mathbf{n}_m^c|| \overset{a.s}{\leq} (\alpha + 1)\sqrt{\mu_T} \\ SNR &\doteq \frac{||\mathbf{s}_m^c||}{M_T} \overset{a.s}{\leq} (1 + \alpha) \end{aligned} \qquad (22)$$

A natural choice is seemingly $\alpha = 1$, which entails $SNR \leq 2$.

This, however, isn't the tightest bound to the SNR from above, and the choice $\alpha = 1$ seems to be not practically feasible. If the noise and the signal are statistically independent, then $\mathbf{C}_x = \mathbf{C}_s + \mathbf{C}_n$. Thus, using (13) and (16), we obtain

$$\mu_x \approx \mu_s + \mu_n, \qquad (23)$$

where

$$\begin{aligned} \mu_x &\doteq E\{\max ||\mathbf{x}^c||^2\} & (24) \\ \mu_s &\doteq E\{\max ||\mathbf{s}^c||^2\} & (25) \\ \mu_n &\doteq E\{\max ||\mathbf{n}^c||^2\}. & (26) \end{aligned}$$

Therefore

$$\max ||\mathbf{x}^c||^2 \overset{a.s}{=} \max ||\mathbf{s}^c||^2 + \max ||\mathbf{n}^c||^2 \qquad (27)$$

This result is a very coarse approximation, since the signal is not Gaussian and its dimensionality is much lower than l. Nevertheless, as soon as at each iteration we find a noisy estimation of the signal range in terms of span of columns of $\hat{A}_k$, the signal residual $\mathbf{s}^c$ cannot be eliminated. Therefore, we anticipate a low probability of finding $\mathbf{x}_m^c$ that satisfies (21) with $\alpha = 1$. By choosing α slightly above 1 and terminating the process when (21) is satisfied, we obtain

$$SNR \overset{a.s}{\leq} \alpha - 1 \qquad (28)$$

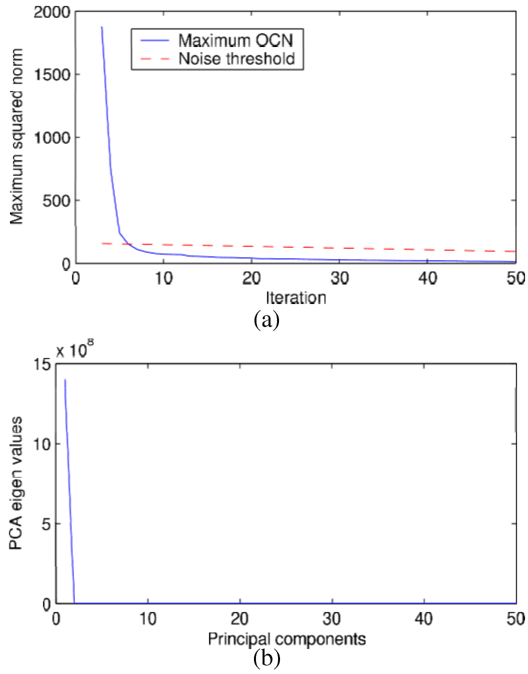

Fig. 2: Dimensionality determination. (a) the maximum squared norm of orthogonal residual thresholded by the noise-based threshold reveals dimensionality of 5, (b) the PCA eigenvalues.

7. RESULTS

We present results of the Modified Gram-Schmidt process application onto a hyperspectral image of a geological scene with 95

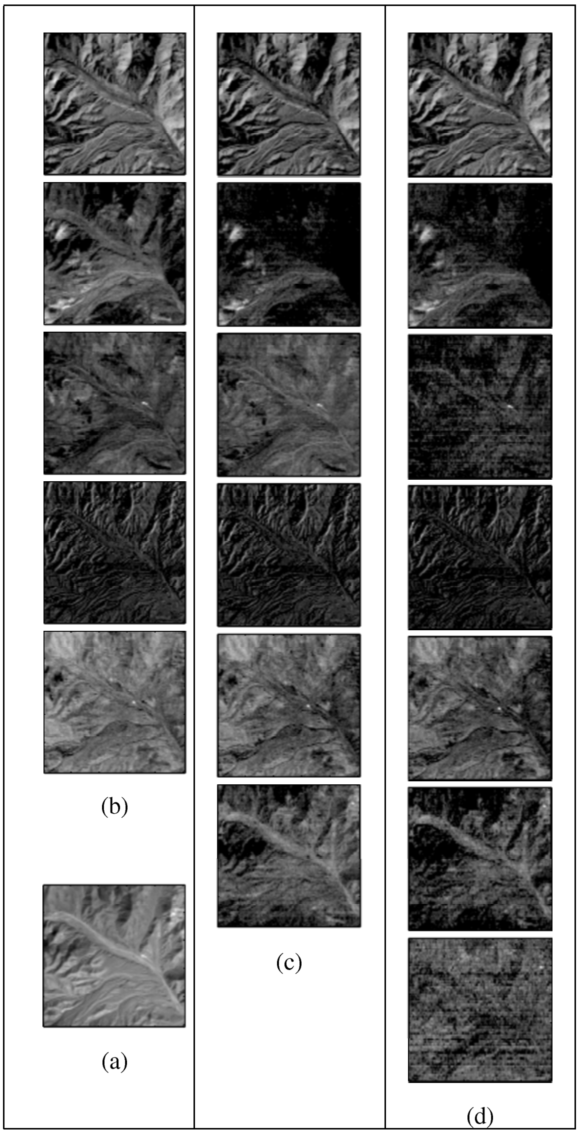

Fig. 3: **Least squares unmixing results.** (a) is a typical 10-th band, (b),(c) and (d) are the abundance maps for 5,6 and 7 endmembers.

spectral bands. The noise was assumed to be spectrally white, its variance was estimated band-wise by

$$\hat{\sigma} = \frac{\mathrm{MAD}(HH)}{0.6745}, \tag{29}$$

where the MAD signifies the *median absolute deviation* operator and HH is the finest scale wavelet (symmlet8) decomposition of the corresponding band image (see [6]). Thus the noise whitening reduces to band-wise scaling in order to obtain an equal noise variance in each spectral band.

A typical 10-th band image is shown in Fig. 3 (a). The abundance maps for 5,6 and 7 endmembers are shown in (b), (c) and (d) respectively. The thresholdng, shown in Fig. 2 (a), reveals dimensionality of 5 (for $\alpha = 1.1$). This is visually supported by Fig. 3 (b),(c) and (d). A noisy splitting of the second endmember abundance map in (c) corresponds to unmixing of 6 endmembers. Unmixing of 7 endmembers produces a competely noisy abundances of the seventh endmember (d). These observations testify that 6th and 7th endmembers have a very low SNR in their OCN onto the subspace spanned by the first 5 endmembers.

8. CONCLUSIONS

The presented analysis provides us with the termination condition for the Modified Gram-Schmidt algorithm. As a result we obtain an estimation of the signal subspace dimensionality up to the noise-related limitations. In contrast to PCA, this technique proposes a well-defined signal-noise separating threshold, which allows more reliable dimensionality estimation even with presence of rare endmembers in the data.

9. ACKNOWLEDGEMENTS

The authors wish to thank Dr. A. Zaks of the Asher Space Research Institute for providing us with the data used in our experiments. Special thanks to Prof. David Malah and Dr. Meir Barzohar for very constructive discussions and comments on the presented topic.

10. REFERENCES

[1] Michael E. Winter and Edwin M. Winter, "Comparison of approaches for determinig end-members in hyperspectral data," *Aerospace Conference Proceedings, IEEE*, vol. 3, pp. 305–313, March 2000.

[2] Shao-Shan Chiang , Chein-I Chang and Irvin W.Ginsberg "Unsupervised Hyperspectral Image Analysis Using Independent Component Analysis," *IEEE International Geoscience and Remote Sensing Symp., Hawaii, USA*, vol. 3, pp. 24-28, July 2000.

[3] Maurice D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Transactions On Geosciense And Remote Sensing*, vol. 32, no. 3, May 1994.

[4] M.Leadbetter, *Extremes and related properties of random sequences and processes*, Springler Series in Statistics, 1982.

[5] Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics., 2001.

[6] Alin Achim, Panagiotis Tsakalides and Anastasios Bezerianos. "SAR Image Denoising via Bayesian Wavelet Shrinkage Based on Heavy-Tailed Modeling". *IEEE Trans. on geoscience and remote sensing*, vol. 41, no. 8, pp. 1773-1783, August 2003.

[7] J. Gruninger, R.L. Sundberg, M.J. Fox, R.Y. Levine, W. Mundkowsky, M.S. Salisbury, and A.H. Ratcliff. Automated optimal channel selection for spectral imaging sensors. *SPIE Proceeding, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery*, VII:4381, 2001.