SUBMODULAR VOLUME SIMPLEX ANALYSIS: A GREEDY ALGORITHM FOR HYPERSPECTRAL UNMIXNG

Seong H. Lee, Sildomar T. Monteiro and Steven J. Scheding

School of Aerospace, Mechanical and Mechatronic Engineering
The University of Sydney, NSW 2006, Australia
s.lee@acfr.usyd.edu.au

ABSTRACT

The hyperspectral unmixing problem can be formulated as a combinatorial optimization which selects the spectral vectors that maximize the volume of a simplex, with the assumptions that the dataset contain pure pixels and the mixture is linear. Submodularity presents an intuitive diminishing returns property which arises naturally in discrete and combinatorial optimization problems. Submodular functions enable the application of fast, greedy algorithms which possess near optimal approximation guarantees. This paper proposes a submodular greedy-based approach for solving the spectral unmixing problem by modifying the objective function to become a non-decreasing submodular function. Theoretical and experimental results are presented to demonstrate the feasibility of the method.

Index Terms— Submodular optimization, greedy algorithm, hyperspectral unmixing, endmember extraction

1. INTRODUCTION

Hyperspectral remote sensing is widely applied to identify materials present in a scene. It is used for several applications such as mineral classification, land survey, environmental monitoring, military surveillance, etc. Due to the typically low spatial resolution of the spectrometer, multiple substances may be present in a single pixel, which generates mixed spectral signatures. The objective of unmixing is to identify the materials present (*endmembers*), the number of materials and their respective quantities (*abundance*). There are several methods for models based on different assumptions for endmember extraction. This paper considers a linear mixture model, which assumes the observed dataset is a linear combination of spectra of materials, and a pure pixel assumption, which supposes the presence of at least one “pure pixel" per endmember in the dataset. There exists several pure pixel based algorithms, such as N-FINDR [1], the vertex component analysis (VCA) [2] and the simplex growing algorithm (SGA) [3]. Among those, N-FINDR and SGA are in the class of volume-based algorithm. They are based on Winter’s maximum volume simplex criterion which states that the spectral vectors which form the maximum volume simplex would be the endmembers or at least would be near to the actual endmembers [1]. A competing approach is Craig’s minimum volume criterion which finds a minimum volume simplex that contains all or most of the vectors of dataset. With the pure pixel assumption, the algorithm based on Craig’s and Winter’s criterion performs equally, but Craig’s is relatively complicated by using the maximization of determinant of inverse matrix with constraints. The most of methods above require a dimensionality reduction which means a loss of information and the calculation of an approximated volume.

Submodular functions have received increased interest due to their diminishing marginal returns property [4] which states including an element in a smaller set makes a bigger difference than including it in a bigger set. Those functions have properties analogous for discrete optimization to the concave functions in continuous optimization. Submodular optimization have been applied successfully to several areas such as sensor placement, dictionary learning and signal processing [5], etc. This paper proposes a submodular volume simplex analysis (SVSA) method using a fast greedy based approach which iteratively obtains the vertices of simplex. This gives near optimal worst case approximation guarantees for the submodular maximization. The greedy algorithm is the best possible polynomial-time approach for solving a submodular function maximization with a near-optimal approximation guarantee. The proposed method avoids the requirement of a dimensionality reduction.

The main contribution of this paper is that it proposes a new non-decreasing submodular objective function for the maximum volume simplex analysis allowing to take advantage of the nice properties provided by submodularity.

2. BACKGROUND

2.1. Submodular function optimization

This brief introduction of submodular theory is based on [4]. It defines a submodular function, then it is presented an equiv-

This work has been supported by the Australian Centre for Field Robotics and the Rio Tinto Centre for Mine Automation.

alent definition which explicit a property of the decreasing marginal value which is a fundamental property for combinatorial optimization and has an important role for the submodular maximization under cardinality constraint. Lastly it is presented a greedy algorithm and its advantage.

Let γ be a finite set, 2^γ is a set of all subset of γ, a real valued function $F : 2^\gamma \rightarrow \mathbb{R}$ is submodular if

$$F(S \cap T) + F(S \cup T) \leq F(S) + F(T), \forall S, T \subset \gamma. \quad (1)$$

and it is monotone when: $F(S) \leq F(T), \forall S \subset T \subset \gamma$. Equivalently, F is submodular if it satisfies the following statement: for any $S \subseteq T \subseteq \gamma$ and $\forall j \in \gamma \setminus T$

$$\rho_j(S) \geq \rho_j(T) \quad (2)$$

where $\rho_j(S) = F(S \cup \{j\}) - F(S)$ is a marginal function. This is a property of the decreasing marginal value. If F is monotone, $\rho_j(S) \geq \rho_j(T) \geq 0$ for $S \subseteq T \subseteq E$ and $\forall j \in \gamma$. This means that if F is a non-decreasing submodular function, the marginal value is always positive.

Define a submodular maximization: Given a certain number k, the solution set D^* is given by:

$$D^* = \underset{D \subseteq \gamma | D| \leq k}{\arg\max} F(D) \quad (3)$$

where F is a non-decreasing submodular function. It possible to rewrite $F(D^*)$ in the following form:

$$F(D^0) + \sum_{p=1}^{k} \rho_{j_p}(D^{p-1}). \quad (4)$$

The greedy algorithm [4] selects the ${j_p}^*$ that maximizes $\rho_{j_p}(D^{p-1})$ in p-th iteration. A property (2) guarantees the maximum gain in each iteration. This makes a greedy approach efficient and provides a good approximation value. Non-decreasing property is also important factor, because it guarantees a positive increment in each iteration. According to [4], the greedy algorithm has a worst case approximation guarantee of $(1 - 1/e)$ of optimum value for monotonic submodular function, this means: $F(D^k) \geq (1 - 1/e) \max_{|D| \leq k} F(D)$, where e is the Napier's number. It was proved in [6] that the $(1 - 1/e)$ approximation is the best possible for any polynomial-time approach, unless P=NP.

2.2. Linear spectral unmixing

In this section, a linear mixing model of the hyperspectral dataset and Winter's maximum volume criterion are presented.

Let the hyperspectral dataset be with N distinct endmembers, M spectral bands and n pixels, then the dataset has the following form: $X^p = EA^p$, where $p \in \gamma = \{1, 2, ..., n\}$ is an index of pixels, $X^p = \{x_1^p, ...x_M^p\}^T \in \mathbb{R}^M$ is M spectral bands corresponding the pixel p, $E = \{e_1, ...e_N\} \in \mathbb{R}^{M \times N}$ is a matrix containing spectral signatures of i-th column vector e_i is i-th endmember spectra, $A^p = \{a_1^p, ..., a_N^p\} \in \{\mathbb{R}^N\}$ is an abundance vector in pixel p.

The linear mixing models present some constraints related to the abundance. First the abundance has a nonnegative value. This means $a_i^P \geq 0$ for all $i \in \{1, 2, ..., N\}$ and $p \in \{1, 2, ...n\}$. This is valid since there is no negative absorption of spectra. Second, the sum of the fractional abundance per pixel is one, $\sum_{i=1}^{N} a_i^p = 1$. This is also true since the fractional abundance is a percentage of the endmembers in each pixel. Third, the number of the substances is smaller than the number of spectral bands and the number of pixels. $\min\{n, M\} \geq N$ and E is of full rank. This is to assume that there are enough dimension of vectors space to select a simplex of $N - 1$ dimension. This is unlikely to be violated, because typically the number of substances present in the dataset is smaller than the number of bands and the number of pixels.

The objective of unmixing is to find the endmembers E and the abundance vector $A^1, ..., A^n$ from the dataset $X^1, ..., X^n$. With the pure pixel assumption, there are X^{p_i} such as $X^{p_i} = e_i$ for all $i \in E = \{1, 2, ..., N\}$. This means: $X^p = \sum_{i=1}^{N} a_i^p X^{p_i}$. As the sum of abundance is one, the entire set of vectors in the dataset X^p are inside the simplex of which the vertices are X^{p_i}s. Winter's maximum volume criterion is originated from this observation. The optimization function aims to find $D = \{d_1, d_2, ..., d_s\}$ such as

$$D^* = \underset{D \subseteq \gamma | D| \leq N}{\arg\max} Vol(D) \quad (5)$$

where the simplex is given by:

$$Vol(D) = Vol(X^{d_1}, ..., X^{d_s}) = \frac{|det \begin{vmatrix} 1 & ... & 1 \\ X^{d_1} & ... & X^{d_s} \end{vmatrix}|}{s!} \quad (6)$$

The term $1/s!$ of (6) does not affect the solution of the optimization problem (5). Therefore it is eliminated. This problem is similar to the maximization of submodular function under cardinality constraints which determines the subset D of cardinality N that maximizes a submodular objective set function F (3).

However, the volume (6) is not a non-decreasing submodular function.

3. SUBMODULAR VOLUME SIMPLEX ANALYSIS

In this section, the $Vol(D)$ of (5) is replaced by a new non-decreasing submodular function which possesses a same optimal solution of (5). First recall the *Gram matrix* to calculate the exact volume of simplex which is a nonnegative definite symmetric matrix. Second, it is presented a notation and an inequality of principal minors of nonnegative definite matrix. The *principal minors* is a submatrix which is defined by the

index set. The union and intersection of index set well defines the submatrix. The submodular function is obtained by composing the logarithm function with both sides of this inequality. Last condition to satisfy is a non-decreasing property. The logarithm of a determinant of a symmetric matrix is not necessarily increasing when includes a additional low and column. The necessary condition to a non-decreasing principal minor is that the smallest eigenvalue is more than one. Therefore by rescaling the Gram matrix, a monotone submodular function is obtained. These whole modifications do not change the optimal solution. Finally it is possible to apply the greedy algorithm for a non-decreasing submodular function, then it is obtained a submodular volume simplex analysis with all good properties.

Note that the matrix

$$V = \begin{pmatrix} 1 & ... & 1 \\ X^{d_1} & ... & X^{d_s} \end{pmatrix}, V \in \mathbb{R}^{(M+1)\times s} \tag{7}$$

is not a square matrix. To resolve this problem, most of the algorithms such as N-FINDR and SGA in (6) use dimensionality reduction methods to make (7) a square matrix. This results an approximated volume and also there is loss of information.

In this paper, the Gram matrix is used to calculate the volume. If there is $s-1$ vectors $\{X^{d_2}-X^{d_1}, ..., X^{d_s}-X^{d_1}\}$, the volume of parallelepiped formed by them is given by $Vol(D) =$

$$\sqrt{\|X^{d_2}-X^{d_1} \wedge ... \wedge X^{d_s}-X^{d_1}\|} = \sqrt{det(V^T V)} \tag{8}$$

The Gram matrix is nonnegative definite symmetric matrix $Y =$

$$\begin{pmatrix} 1+X^{d_1}X^{d_1} & 1+X^{d_1}X^{d_2} & ... & 1+X^{d_1}X^{d_s} \\ 1+X^{d_2}X^{d_1} & 1+X^{d_2}X^{d_2} & ... & 1+X^{d_2}X^{d_s} \\ ... & ... & ... & ... \\ 1+X^{d_s}X^{d_1} & 1+X^{d_s}X^{d_2} & ... & 1+X^{d_s}X^{d_s} \end{pmatrix} \tag{9}$$

Note that i-th row and column of (9) is the only place that is affected by i-th column of the matrix V. For example, the second row and column of (9) is the only place that is affected by X^{d_2} in (7).

A notion of principal minor is presented. Let a matrix H, a nonnegative definite symmetric matrix, n a positive integer, $\gamma = \{1, 2, ..., n\}$, for $I \subset \gamma$, denote by $H[I]$ the principal submatrix of H that deletes the rows and columns in the set $\gamma \setminus I$ from matrix H. This means that for $I = \{e, f\}$, $H[I] = \begin{pmatrix} H_{e,e} & H_{e,f} \\ H_{f,e} & H_{f,f} \end{pmatrix}$ Therefore with this notation, it is possible to manipulate the dimension of the matrix with the set of indexes. As a convention the $det(H[\emptyset]) = 0$.

Recall an inequality [7] which relates the determinant of matrix (9) and the submodularity. For a nonnegative definite symmetric matrix, the following inequality holds:

$$detH[I \cap J] detH[I \cup J] \leq detH[I] detH[J], \tag{10}$$

Algorithm 1 Submodular volume simplex analysis

Input: Index set γ, data set X, number of endmembers N
Ouput: D is the set of index of the endmembers.
Initialize $D = \emptyset$, $V = \gamma$, $j = 0$. Create the Gram matrix Y from the dataset: $Y = X^T X$
repeat
  Find the index with the maximum marginal value. $i^* = \arg\max_{i \in V} F'(D, i)$ (14) and include i^* into the solution set
  $D = D \cup \{i^*\}$
  $V = V \setminus \{i^*\}$, $j = j + 1$
until $F'(D, i^*) < 0$ or $j = N$

where $I, J \in \gamma$. To apply in the maximum volume analysis, $H = Y$. The logarithm is strictly increasing function in $(0, +\infty)$, composing the logarithm with both side of (10),

$$\begin{aligned} &\log(detY[I \cap J]) + log(detY[I \cup J]) \\ &\quad \leq log(detY[I]) + log(detY[J]). \end{aligned} \tag{11}$$

Equation (11) satisfies the definition of submodular function. Therefore the $log(detY[I])$ is a submodular function. The only condition that is not satisfied is that the $log(det(Y[I]))$ is not necessarily monotonic. Friedland's work [8] gives a necessary and sufficient condition to Y has increasing principal minors. If and only if Y is positive definite symmetric matrix, and all diagonal entries of Y^{-1} are less or equal to 1. This means that the smallest eigenvalue of Y must be more than one and principal minor of Y is invertible.

The arbitrary Y does not satisfy this condition, but rescaling a matrix by multiplying $1/\lambda_k$ where λ_k is the smallest non-zero eigenvalue, it is possible to get a matrix with all the eingenvalues bigger than one. Except the matrix with eigenvalue zero, but if matrix has eigenvalue zero, determinant is zero and $log(0) = -\infty$. This means that marginal function has a negative value. So the column that makes eigenvalue zero will be eliminated by the greedy algorithm.

The optimization function (5) is replaced by a following submodular optimization: Let λ_k be the minimum nonsingular eigenvalue of $Y[\gamma]$, the objective function

$$D^* = \arg\max_{D \subseteq \gamma | D| \leq n} log(det(\frac{1}{\lambda_k}(Y[D]))) \tag{12}$$

is a (partially) non-decreasing submodular function (because there are columns that makes the determinant zero) and the optimal solution of (12) and (5) is the same.

Finally applying the greedy algorithm, it results the SVSA described in algorithm 1. SVSA chooses in its first iteration a column with the maximum module. It is, in fact, a vertex of simplex. Also observe that

$$F([D \cup \{i\}]) - F([D])) = log \frac{\frac{1}{\lambda_k} det(Y[D \cup \{i\}])}{det((Y[D]))}. \tag{13}$$

The $det(Y[D])$ is a fixed number defined in the previous iteration and the eigenvalue λ_k does not influence the optimiza-

tion. So it is sufficient to find the maximum argument of

$$F'(D,i) = det(Y[D \cup \{i\}]). \quad (14)$$

SVSA shares a similar characteristic with VCA and SGA. They obtain the vertex iteratively by searching the entire space of endmembers. The difference is that VCA uses an orthogonal subspace projection while SGA and SVSA sequentially search the vertex that creates the maximum volume, but SGA requires dimensionality reduction in each iteration. SVSA does not need it and also it calculates the exact value of the volume.

Compared to N-FINDR ($\mathcal{O}(nN^{\nu+1})$(best case)), SVSA has lower computational complexity ($\mathcal{O}(n\sum_{l=1}^{N} l^{\nu})$), where l^{ν} is a computational complexity to calculate a determinant of $l \times l$ matrix. This happens due to N-FINDR calculating the whole N endmembers once and replacing the spectral vectors of n pixels. To perform this exhaustive process, it needs a combination $\binom{N}{n}$ and it needs to calculate determinant of a matrix $N \times N$ in each replacement of the vector. Compare to this, SVSA is greedy, N iterations calculating n determinant of dimension $s \times s$ where s increases iteratively until $s = N$. SVSA is relatively fast compared to N-FINDR.

4. EXPERIMENTS

Four simulated datasets based on [10] are used in the experiment. SVMSA, VCA and N-FINDR are tested for comparison of performance. The datasets are composed of 5 endmembers, 244 spectral bands, 5000 pixels and with a signal to noise ratio (SNR) of 30 dB.

First it is tested for the dataset with the pure pixel and abundances uniformly distributed. The second case, a dataset without pure pixels and the abundances uniformly distributed is used. The third case, a no pure-pixel dataset with the abundances uniformly distributed, but limited to 0.8 is tested. The fourth case, a no pure-pixel dataset with Dirichilet distributed abundances with concetration parameter 10 is tested. Table 1 presents a summary of the results of the experiments.

All the VCA, N-FINDR, SVSA have an excellent performance in pure pixel case. They select the exact solution set. The methods, which are pure-pixel based, do not perform as well as in the first case. For the no pure-pixel cases, the vertices selected by algorithms are far from the true endmembers. As VCA has a different selection criterion, it obtains a different result. N-FINDR and SVSA share the same criterion, and therefore tend to have the similar mean of angle error. Compared to the N-FINDR, SVSA has a slightly smaller standard deviation. SVSA achieved the smallest mean square errors.

5. CONCLUSION

SVSA proved experimentally competitive to the conventional methods N-FINDR and VCA. The SVSA does not require dimensionality reduction as most of the other algorithms. SVSA calculates the exact volume using the Gram matrix and selects the best vector of the subspace in each iteration, which is guaranteed by the submodular property of decreasing marginal value. This selection process is fast since it uses a greedy algorithm which provides the best approximation possible for a polynomial-time approach.

Table 1: Comparison of 1) Pure pixel, 2) No pure pixel, 3) Abundances truncated to 0.8 and 4) Highly mixed set.

	VCA		N-FINDR		SVSA	
1	0.2(0.2)	0	0.2(0.2)	0	0.2(0.2)	0
2	2.5(3.0)	0.08	2.6(2.9)	0.07	2.9(2.8)	0.07
3	6.4(4.3)	0.15	4.8(3.2)	0.13	4.9(3.2)	0.12
4	12.3(6.0)	0.36	12.8(6.8)	0.36	12.6(6.6)	0.35

Mean of the angle error(Standard deviation) | Mean squared error

6. REFERENCES

[1] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *in Proc. SPIE Image Spectromety V*, vol. vol. 3753, pp. 266–277, 1999.

[2] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE TGRS*, vol. 43, no. 4, pp. 898–910, 2005.

[3] C.-I Chang, C,-C. Wu, W. Liu, and Y. C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm,"*IEEE TGRS*, vol. 44, no. 10-1, pp. 2804–2819, 2006.

[4] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher,"An analysis of approximations for maximizing submodular set functions–1," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

[5] V. Cevher and A. Krause, "Greedy dictionary selection for sparse representation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 99, no. 5, pp. 979–988, 2011.

[6] U. Feige, "A threshold of ln n for approximating set cover," *Journal of the ACM*, vol. 45, no. 4, pp. 634–652, 1998.

[7] F. R. Gantmaher and M. G. Krein, *Oszillationsmatrizen, Oszillationskerne und kleine schwingungen mechanischer Systeme (Oscillation matrices and kernels and small oscillations of mechanical systems)*, Berlin, Akademie-Verlag, 1960.

[8] S. Friedland, "Nonnegative definite hermitian matrices with increasing principal minors,"arXiv/1301.4665 2013.

[9] J.M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Whispers*, pp. 1–4 2009.

[10] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE JSTARS*, vol. 5, no. 2, pp. 354 –379, 2012.