Supervised Kernel Based Nonlinear Unmixing of Hyperspectral Data

Hong Xiao
School of Computer Science, Xi'an Shiyou University Shannxi, Xi'an 710065, China
xiaohong@xsyu.edu.cn

Abstract. In hyperspectral imagery problem, pixels are mixtures of spectral component associated with pure materials. Recently, nonlinear models have been taken into consideration to surmount some limitations of linear model. In this paper, the nonlinear hyperspectral image unmixing problem is formulated with kernel learning theory, with the number of kernels being controlled by the coherence rule. To be more physically interpretable, a relationship between endmembers and abundance vectors is introduced as a constraint of the optimization problem. An iterative learning algorithm derived from augmented Lagrangian method is proposed to solve the defined problem. Simulation results show the efficacy of the proposed model and algorithm.

1. Introduction
Hyperspectral imagery has gained considerable attention in the last decade. It possesses more accurate and detailed information extraction ability than other types of remotely sensed data as the images provide rich spectral information. Applications include agriculture monitoring, natural disaster analysis and ecosystem monitoring, etc.

In hyperspectral imagery, a pixel is generally mixed by a number of materials present in the scene. A widely studied approach that is used to solve a mixed pixel classification problem is the linear unmixing, which assumes that an image pixel vector is linearly mixed by material signatures. To be physically explainable, two constraints, nonnegativity and full additivity are required to impose on the linear mixing model. Extensive studies have been exploited on this unmixing model [1]. However, if the components of interest are in an intimate association, light interacts with more than one component. The mixing mechanism will be no more linear. If a linear mixing model is used on data where the mixing systems are nonlinear, the estimated result will be significantly in error. Some algorithms have been proposed with the consideration of such nonlinear effects, e.g. [2] [3] [4] [5].

As kernel-based algorithms in Hilbert space have been a topic of considerable interest in machine learning community and gained notable success to deal nonlinear problems in many applications. In this paper, author formulates the abundance estimation problem under supervised kernel learning schema, on assuming N training pixels available (in practice, the training information could be obtained by special pixels, or by performing dedicated algorithm over certain pixels, or by generating with a dedicated model). A kernel dictionary is established by calculating coherence parameter of data in feature space. To be physically interpretable in the nonlinear context, the relationships between endmembers and abundances are also taken into consideration as constraints for the model. Finally a learning algorithm based on augmented Lagrangian method is proposed for solving the defined optimization problem.
2. Kernel Based Abundance Nonlinear Estimation Problem

Suppose \( L \) is the number of spectral bands. Let \( \mathbf{r} = [r_1, r_2, \cdots, r_L]^T \) be an \( L \times 1 \) column mixed pixel that is a mixture of \( R \) endmembers with additive noise. Assume that \( \mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_R]^T \) is the \( L \times R \) with each column \( \mathbf{m}_i \) a target signature. Let \( \mathbf{a} = [\alpha_1, \alpha_2, \cdots, \alpha_R]^T \) be an \( R \times 1 \) abundance column vector associate with the pixel \( \mathbf{r} \). Consider the mixing mechanism with a model specified by

\[
\mathbf{r} = \mathbf{F}(\mathbf{a}, \mathbf{M}) + \mathbf{n}
\]  

(1)

where \( \mathbf{n} \) is an additive white noise sequence. The function \( \mathbf{F} \) denotes the mixing mechanism in general sense. With this model, author is interested in finding a function \( \Psi \) which is an inverse map of \( \mathbf{F} \) to interpolate \( \alpha = \Psi(\mathbf{r}(\alpha, \mathbf{M})) \). Considering the nonlinear unmixing, \( \mathbf{r} \) is transformed into a high-dimensional space with a nonlinear map \( \phi \), and it is expected that \( \alpha \) can be expressed by a linear combination of the components of \( \phi(\mathbf{r}) \), i.e., \( \tilde{\alpha} = \mathbf{B}^T \phi(\mathbf{r}) \), where the matrix \( \mathbf{B} \) consists the columns \( b_i \) as the weight vectors in the feature space. As these weight vectors lie in the space spanned by the training points, which implies \( \mathbf{B} \) can be represented by the combination of these points, namely, \( \mathbf{B} = \begin{bmatrix} \phi(\mathbf{r}_1)^T, \phi(\mathbf{r}_2)^T, \cdots, \phi(\mathbf{r}_N)^T \end{bmatrix} \mathbf{W} \), where \( \mathbf{W} \) is an \( N \times R \) matrix. With this expansion, the estimated abundance vector is written as

\[
\tilde{\alpha} = \mathbf{W}^T [\phi(\mathbf{r}_1)^T, \phi(\mathbf{r}_2)^T, \cdots, \phi(\mathbf{r}_N)^T]^T \phi(\mathbf{r})
\]  

(2)

By the virtue of the Riesz representation theorem, an inner product in Hilbert space such as \( \langle \phi(\mathbf{r}_i), \phi(\mathbf{r}) \rangle \) can be expressed by a kernel function \( \kappa(\mathbf{r}_i, \mathbf{r}) \). Thus it is unnecessary to know explicitly what the map \( \phi \) is. The estimated abundance vector \( \tilde{\alpha} \) is now written as the linear combination of kernel functions

\[
\tilde{\alpha} = \sum_{n=1}^{N} w(i) \kappa(\mathbf{r}, \mathbf{r}_i) = \mathbf{W}^T \mathbf{K}_r
\]  

where \( w(i) \) denotes the \( i \)-th row of \( \mathbf{W} \) and \( \mathbf{K}_r \) denotes the vector \( [\kappa(\mathbf{r}, \mathbf{r}_1), \cdots, \kappa(\mathbf{r}, \mathbf{r}_N)]^T \) which consists of \( N \) kernels. With all available training pixels, the cost function is defined as the sum of squared norm of the abundance estimate error. It is written as following

\[
J(\mathbf{W}) = \sum_{n=1}^{N} \| \mathbf{a}_n - \mathbf{W}^T \kappa_{r_n} \|^2
\]  

(3)

In this nonlinear unmixing context, in order to improve the physical explicable of the model, marginal conditions are taken into consideration, namely, for a pixel made up by a pure spectral, \( \mathbf{r} = \mathbf{m}_i \), the estimated abundance should be an all zero vector except that the element corresponding to this endmember equals to 1. This constraint should be considered with respect to every endmember. As a result, Equation (4) is obtained

\[
\mathbf{I}_R = \mathbf{W}^T [\kappa_{m_1}, \kappa_{m_2}, \cdots, \kappa_{m_R}] \quad \text{(4)}
\]

where \( \mathbf{I}_R \) denotes the \( R \times R \) identity matrix. Considering both of the cost function and this marginal constraint, the optimization problem is now cast as

\[
\min_{\mathbf{W}} J(\mathbf{W}) \quad \text{s.t.} \quad \mathbf{I}_R = \mathbf{W}^T [\kappa_{m_1}, \kappa_{m_2}, \cdots, \kappa_{m_R}]
\]

3. Supervised Kernel Learning Algorithm To The Constrained Problem

The constraint defined by Equation (4) actually consists of \( R \times R \) equality constraints. Considering these equality constraints, the Lagrangian function of this problem is written as

\[
L(\mathbf{W}, \mathbf{A}) = J(\mathbf{W}) + \| \mathbf{A} \otimes (\mathbf{W}^T \mathbf{K}_r - \mathbf{I}_R) \|
\]  

(5)

where the Lagrangian multiplier matrix \( \mathbf{A} \) is defined with the \( (i, j) \)-th element \( \lambda_{ij} \), the matrix \( \mathbf{K}_r \) is defined as \( [\kappa_{m_1}, \kappa_{m_2}, \cdots, \kappa_{m_R}] \) and \( \mathbf{l} \) is the all-one vector of dimension \( R \times 1 \). In order to solve the defined problem, the first step is to calculate the gradient of the cost function \( \nabla J(\mathbf{W}) \)

\[
\nabla_{\mathbf{W}} J(\mathbf{W}) = 2 \sum_{n=1}^{N} (\kappa_{r_n} \kappa_{r_n}^T) \mathbf{W} - 2 \sum_{n=1}^{N} \kappa_{r_n} \mathbf{a}_n^T
\]  

(6)
The stationary points of the Lagrangian function gives the optimal solution of the constrained problem. Taking derivatives of Equation (5) with respect to $W$ and $\Lambda$.

$$\frac{\partial L(W, \Lambda)}{\partial W} = \nabla_w J(W) + K_M \Lambda^T$$  \hspace{1cm} (7)

$$\frac{\partial L(W, \Lambda)}{\partial \Lambda} = W^T K_M - I_R$$  \hspace{1cm} (8)

Inspired by online kernel learning techniques, author proposes an iterative learning algorithm to alleviate these problems. As it may not be necessary to use every learning pixel in the kernel expansion Equation (2), among all of the $N$ training pixels, only $P \ll N$ is selected into the dictionary. The model thus becomes

$$\alpha = \sum_{i=1}^{P} w(n_i) \kappa(r, r_n)$$  \hspace{1cm} (9)

which is the same with Equation (2) except that only $P$ selected kernels are used here. In the kernel learning domain, one of the most used sparsification rules is the coherence criterion which has the advantage of simplicity and satisfying performance. Therefore, a similar way is employed to select training pixels into the dictionary in this paper.

3.1. Dictionary selection with coherence criterion

In this context, author decides whether to add a new training pixel into the dictionary based on the coherence value between this data and all the other members in the dictionary. It is defined as

$$\mu_r = \max_{r \in D} |\kappa(r_n, r_{w_i})|$$

It reflects the largest correlations between $r$ and elements in the dictionary $D$. Author suggests inserting $r$ into the dictionary provided that its coherence remains below a given threshold $\mu_0$, namely, $\max_{r \in D} |\kappa(r_n, r_{w_i})| \leq \mu_0$ where $\mu_0$ is a parameter in $[0, 1]$ determining the level of sparsity.

3.2. Iterative learning algorithm

One of the most effective methods to deal with equality constrained problem is the augmented Lagrangian method, alternatively referred to as multiplier methods. The augmented Lagrangian function of our problem is formulated by modifying Equation (5) as

$$L_a(W, \Lambda) = J(W) + ||[\Lambda \otimes (W^T K_M - I_R)]|| + \frac{1}{2} \nu ||W^T K_M - I_R||_F^2$$  \hspace{1cm} (10)

where $\| \cdot \|_F$ denotes the Frobenius norm of a matrix. The derivatives of the first and the second item have been calculated by Equation (7) and Equation (8). The derivative of the third item is obtained by

$$\frac{\partial}{\partial W} \frac{1}{2} \nu ||W^T K_M - I_R||_F^2 = \nu K_M (K_M^T W - I_R)$$

As the typical step of an augmented Lagrangian method, author starts with a matrix of Lagrangian multipliers $\Lambda_k$. Then $W_k$ is calculated by gradient update methods. Next $\Lambda_k$ is updated to $\Lambda_{k+1}$ by the following standard method $\Lambda_{k+1} = \Lambda_k + \nu (W_k^T K_M - I_R)$. Finally the following expressions are iteratively performed in order to get the optimal solution

$$W_{k+1} = W_k - \eta \left( \nabla_w J(W) + K_M \Lambda_k^T + \nu K_M (K_M^T W - I_R) \right)$$  \hspace{1cm} (11)

$$\Lambda_{k+1} = \Lambda_k + \nu (W_k^T K_M - I_R)$$  \hspace{1cm} (12)

Where $\nabla_w J(W)$ has been calculated by Equation (6) and $\eta$ is a small positive step size. Another parameter $\nu$ is also a small number which can also vary to accelerate the convergence.

3.3. Determining the abundance of other pixels

Once this coefficient matrix $W$ has been determined, the abundance of any given pixel $r$ by Equation (9) could be estimated. However, it should be noticed that this estimate does not ensure the non-
negativity and the sum-to-one constraint. If these constraints are strictly required to be satisfied, it is suggested to determine the abundance by

$$\hat{\alpha} = \arg\min_{\alpha \geq 0, \|\alpha\|_1 = 1} \|\alpha - W^\top \kappa_r\|^2$$

(13)

where $\|\cdot\|_1$ denotes the L1-norm of the vector. In fact, Equation (13) is a task of performing Euclidean projection onto the positive simplex. An algorithm of this projection has been described in [6].

4. Computer Simulations

4.1. Synthetic mixture of real spectra

In this section, the proposed algorithm is studied by data generated by linear/nonlinear mixing of three pure materials with spectral signatures from the USGC library. These spectra consist of 2151 contiguous bands. Two 50-by-50 hyperspectral images have been generated respectively by linear mixing model and bilinear mixing model in [2], denoted by $I_1$ and $I_2$, respectively. Among all 2500 pixels 500 are used for training. The abundance vector $\alpha_n$ are generated in the simplex defined by the positivity and sum-to-one constraints. Both images have been corrupted by an additive Gaussian noise with $\text{SNR} \approx 20$dB. The Gaussian kernel $\kappa(r_1, r_2) = \exp(-\frac{\|r_1-r_2\|^2}{2\sigma^2})$ is used. The kernel width is set to $\sigma = 2$ and the coherence threshold is set to $\mu_0 = 0.582$.

The numbers of selected kernel of two images are respectively 18 and 23 out of 500 training pixels. The abundance vectors corresponding to selected pixels of the bilinear model are shown in Figure 1. The quality of the unmixing approaches is measured by the root mean square error

| Table 1. RMSE comparison. |
|---------------------------|
| $I_1$ | $I_2$ |
| Fully constrained LS [1] | 0.0123 | 0.0452 |
| Proposed without projection (13) | 0.0301 | 0.0316 |
| Proposed with projection (13) | 0.0257 | 0.0284 |
| RBF neural networks in [4] | 0.0214 | 0.0216 |

$$\text{RMSE} = \sqrt{\frac{1}{|\text{test set}|} \sum_{n \in \text{test set}} \|\alpha_{n1} - \bar{\alpha}_{n2}\|^2 / N_{\text{test}}}$$

For both models, the results of the proposed approach with the fully constrained solution [1] are compared with those with nonlinear unmixing method of RBF neural network [4] as shown in Table 1. The method in [1] was derived based on linear mixing model, it gives the optimal solution for $I_1$ and has the lowest RMSE for $I_1$. However, it does not work well with the nonlinearly mixed image $I_2$. The RBF based neural network in [4] uses all the training pixels as Gaussian kernels, it thus gives a lowest RMSE for $I_2$ at the cost of inverting the matrix with the dimension of the number of training pixels (namely $500 \times 500$). Whereas the proposed algorithm operates matrices of much lower dimension due to the coherence criterion and no matrix inversion operation is required. The projected results lead to have lower RMSEs than that without projection.

Figure 1. Abundance vectors for $I_2$. Red ‘o’ s are abundance vectors corresponding to the selected kernels.
4.2. Simulation on real data

The studied image is the scene over Moffett Field (CA, USA), captured by the airborne visible infrared imaging spectrometer (AVIRIS). A sub-image of size $50 \times 50$ pixels was chosen to evaluate the proposed algorithm. This scene is mainly composed of water, vegetation and soil. The endmembers have been extracted by the VCA algorithm with $R = 3$. The training data were generated by the bilinear model. Figure 2 shows the estimated abundance maps. The proposed algorithm has a good ability to determine the relationship between the observed pixels and the abundance vectors.

![Abundance maps estimated by the proposed algorithm (soil, vegetation and water respectively).](image)

5. Conclusion

In this paper, a nonlinear unmixing schema of hyperspectral images has been presented, based on advances on kernel learning theory. The use of coherence sparsification criterion reduces notably the number of kernels selected in the dictionary, which releases the calculation burden without severe performance degradation. Future works may concentrate on the study of other kernel selection criteria and other learning strategies.

References

[1] D. C. Heinz and C.-I Chang, Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery, *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 529–545, 2001.

[2] W. Fan, B. Hu, J. Miller, and M. Li, Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data, *International Journal of Remote Sensing*, vol. 30, pp. 2951–2962, 2009.

[3] A. Halimi, Y. Altman, N. Dobigeon, and J.-Y. Tourneret, Nonlinear unmixing of hyperspectral images using a generalized bilinear model, *IEEE Trans. Geosci. Remote Sens.*, 2010.

[4] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks, *IEEE Geosci. and Remote Sensing Lett.*, vol. 39, no. 8, pp. 2314–2318, Aug. 2001.

[5] J. Chen, C. Richard, P. Honeine, Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model”. IEEE Trans. Signal Process., vol. 61, no. 2, pp. 480-492, Jan. 2013.

[6] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, Efficient projections onto the $l_1$-ball for learning in high dimensions, in *Proc. of the 25th international conference on Machine learning*. ACM, 2008, pp. 272–279.