A Priori MaxEnt H(**S**) independent class analysis (ica) vs. *A Posteriori MaxEnt* H(**V**) ICA

Harold Szu
Digital Media RF Lab, Tompkins Rm308, 725 23rd St., Dept. ECE, George Wash. U., Wash. DC 20052

ABSTRACT

Two mirror symmetric versions of the *maximum entropy (MaxEnt)* methodology are introduced and compared: (1) A *posteriori MaxEnt* Independent Component Analysis (ICA) H(**V**) was proposed by Bell, Sejnowski, Amari, Oja (BSAO) (early by Jutten & Herault, Comon and Cardoso (JHCC) in France). It is ambitious to factorize the unknown joint-probability density function (j-pdf) using the post processing algorithm involving the sigmoid-threshold neurons' output $\mathbf{V}(x,y)= \sigma([W]\mathbf{X}(x,y))$ of all image locations (x,y) in order to apply the pixel-ensemble averaged synaptic weight matrix [W] learning, $\partial[W]/\partial t=<\partial H(\mathbf{V})/\partial[W]>$. *The pixel ensemble average* may be necessary to factorize the unknown joint-pdf from multi-channel data vector **X**(x,y). (2) *A priori MaxEnt* H(**S**) for independent class analysis (ica) is a compliment first step to the ambitious joint-pdf factorization based on a-*posteriori MaxEnt* H(**V**) ICA. Since ica is less ambitious to ICA in finding independent classes alone without their underlying pdf, we can derive from Gibb's statistics mechanics of independent classes of irradiation sources $\mathbf{S}_j$ by *a priori MaxEnt* H(**S**), which would be a flat equal class distribution if each were not constrained by the measurements by means of Lagrangian multipliers of force vector $\boldsymbol{\lambda}_i$ and energy scalar (λ_o -1) for each pixel:

$$H(S_j) = -\sum_{j=1}^{N} S_j \ln(S_j) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i (A_{ij} S_j - X_i^{(known)}) - (\lambda_0 - 1)(\sum_{j=1}^{N} S_j - 1)$$

Physically speaking, the long distance propagation by the speed of the light insures the line of sight local validity of a linear and instantaneous within-pixel mixture model of remote optical sensing. However, only the *a priori MaxEnt* H(**S**) ica can handle a large hyperspectral image data basis by the divide-and-conquer strategy without the <u>*pixel ensemble average*</u> that limits computationally the *posteriori MaxEnt* ICA algorithm. This strategy is possible because the radiation sources vector $\mathbf{S}(x_o,y_o)$ per pixel contribute locally to the spectral image data per pixel $\mathbf{X}(x_o,y_o)= [A]\mathbf{S}(x_o,y_o)$, i.e. the irradiation collected within the foot print of the individual pixel (x_o,y_o) will not mix with other neighborhood irradiation sources that will only contribute to their corresponding neighborhood pixels. Although this is basically true for any optical imaging, the post processing BSAO algorithm attempts to de-mix all pixel data [W]**X** in a batch mode that make it not scalable to large image data basis. Being *a priori MaxEnt* H(**S**) pre-processing, we can divide and conquer image size by applying *ica* pixel by pixel in real time. Furthermore, we have derived *ab initio* from it the usual sigmoid threshold logic $\mathbf{S} = \sigma(\boldsymbol{\lambda}[A])$ and the Hebbian learning ruled $\Delta A_{ij} = \boldsymbol{\lambda}_i \mathbf{S}_j$. We can thus conjecture any linear communication theorem that a linear matrix transform (e.g. associative memory [A] recall) between data **X**=[A]**S** and its independent classes under the constraint of comprehensive decomposition $\Sigma_j S_j = 1$ must lead naturally to the coupling of sigmoid transfer logic and Hebbian learning. Thus the author has coined the Lagrangian Constraint Neural Network (LCNN) since 1997. Nonlinear ICA generalization by LCNN and two conjugate gradient ascents of two mirror symmetric MaxEnt's are indicated.

KEYWORDS: Smart Vision, Unsupervised Learning, Remote Sensing, Hyperspectral, Lagrangian Constraints Neural Net, A Priori Maximum Entropy, Independent Classification Algorithm

1. INTRODUCTION

An important milestone of artificial neural network (ANN) modeling of smart vision image processing is the discovery of *"a posteriori" MaxEnt* unsupervised learning algorithm that validates Barrow's redundancy reduction principle reproducing the sparse edges discovered early in cat eyes by Nobel Laureates Hubel and Wiesel. It taught us that a minimum redundancy is achieved at a maximum independence, of which the true one happens at the maximum entropy (like the ocean sands being decayed independently from the mountain rocks near the end of maximum entropy heat death). ANN can solve the tough mathematical problem called Independent Component Analysis (ICA), i.e. to factorize the (unknown) joint probability density (of source rocks) in order to de-mix the unknown mixtures (of color sands, for example). While the mathematical models of single eye for wavelets feature extraction and the (supervised) principle component analysis (PCA) of the covariance matrix, the binocular vision is for the unsupervised independent component analysis (ICA) analysis for blind source de-mixing. ANN has been successfully applied to the advanced brain imaging, fn-MRI and PET, which help eliminate any artifact and consequently enhance and surge the brain imaging experiments. The drawback of Bell-Sejnowski-Amari-Oja (BSAO) if any is <u>*the mandated pixel ensemble average*</u> that prevents itself from dealing with a large image data basis that may have 200 M byte consisting of 200 channels images of 100x100 pixels in the so-called hyperspectral. The remote sensing of weak optical signal is usually linear without delay and requires the Lagrangian Constraint new approach to be unsupervised because no ground truth is known except the maximum independent class entropy, Max H($\mathbf{S}_j$). On the contrary, the brain imaging has nonlinear physiological constraints and their coherent dependence of various causal delays is of interests after the customary removal of independent

artifacts. We review the a-priori maximum entropy Lagrangian Constraint neural net model that solves pixel by pixel in real time the details of hyperspectral imaging applications.

On board of a geo-synchronous satellite such as Landsat of NASA, each image pixel gathers seven spectral images from the visible to the far infrared, and can be considered as a blind mixture of radiation spectrums of all material present in the large ground footprint about 900m^2.

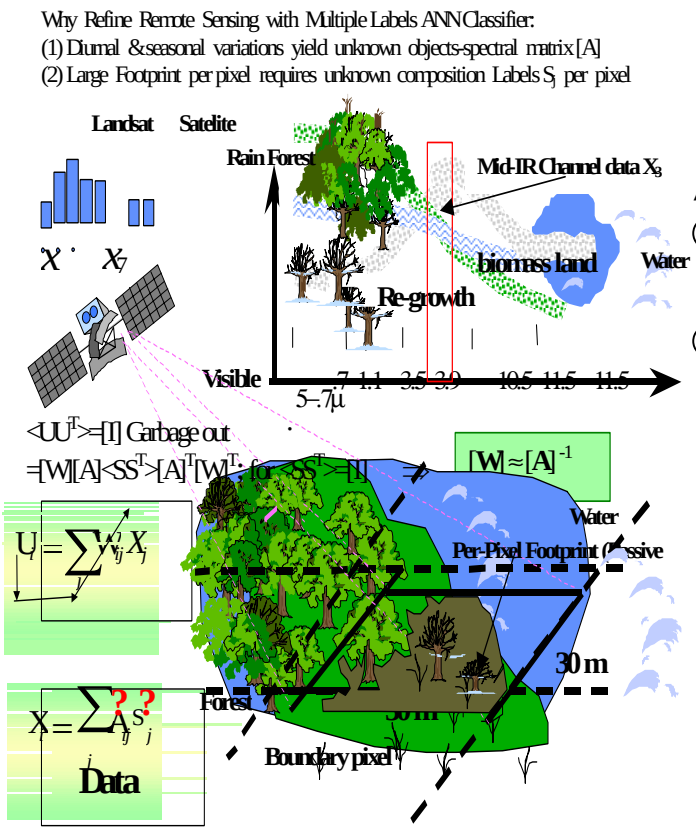
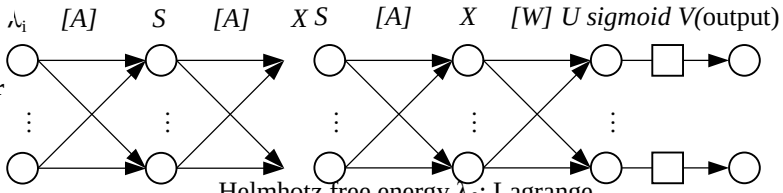


Fig. 1 Schematic diagram indicates how Bell-Sejnowski-Amari-Oja (BSAO) ICA artificial neural network (ANN) weight matrix [*W*] learns by the ensemble of gradient ascent algorithm until all neuron outputs V=σ(U) are being squeezed to become de-correlated noise $<UU^T>$ = $[W][A]<SS^T>[A]^T$ $[W]^T$ =[I] implies for independent sources $<SS^T>$=[I] the discovery of knowledge representation [W] =$[A]^{-1}$ of the external world. Thus, ANN helps unmix ground sources for subpixel composition S_j. BSAO ICA approach is *a-posteriori MaxEnt* H(V) all-pixel post-processing ANN, while LCNN ica approach is *a-priori MaxEnt* H(S) per-pixel pre-processing ANN suited for large data bases. Both implement unsupervised learning due to unknown [A].

2. LAGARANGIAN CONSTRAINT NN

Since both the objects S_j and the reflectance matrix A_{ij} in Eq(1) are unknown, it may be called a blind source separation (BSS) or independent class analysis (ica). We postulate the Lagrangain constraint maximum entropy with Lagrangian multipliers λ_i to incorporate the measurement. We shall refer to the double recursions for the double unknowns methodology as the Lagrangian Constraint Neural Network (LCNN). LCNN and ICA (by Bell and Sejnowski, 1995) Amari (1996) and Oja (1997) are different in the following senses. (1) LCNN introduced one input/hidden layer of neurons λ to form a neural network in Fig. 2(a) and BSAO is an output layer neural network in Fig. 2(b). (2) LCNN maximize the *a priori* entropy H(***S***) as pre-processing ica and BSAO maximize the *a posterior* entropy H(***V***) as post-processing ICA. (3) LCNN ica has two learning functions in conjugate dual spaces **λ** and **S**, instead of one in BSAO ICA.



$\frac{\partial \lambda_o}{\partial \lambda_k} = -X_k$ Helmhotz free energy λ_0; Lagrange Forces λ_i;the displacement data X_i derived ab-initial-ly from Basic Eq(2) Associative memory-like update of [A]

$\frac{\partial W}{\partial t} = \frac{\partial H}{\partial W}$

$\frac{\partial X_i}{\partial t} = \frac{\partial H}{\partial S_i}$

(a) LCNN Pre-processing of source entropy

(b) BSAO ICA Post-processing of output entropy

Fig. 2: Network structures of Pre-processing Lagrangian Constraint NN versus Post-processing BSAO ICA ANN

In LCNN, we define a constraint entropy function of ground radiation sources *S* to be maximized with the Lagrangian constraint multipliers as follows:

$$MaxH(S_j) = -\sum_{j=1}^{N} S_j \ln(S_j) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i (A_{ij}S_j - X_i^{(known)}) - (\lambda_0 - 1)(\sum_{j=1}^{N} S_j - 1) \quad (1)$$

so that the unknown number N of independent classes is nevertheless normalized in the probability sense.

$$\sum_{j=1}^{N} S_j = 1 \quad (2)$$

$$\sum_{j=1}^{N} A_{ij} S_j = X_i^{(known)} \quad (3)$$

While the number of photons must be real and non-negative, classes of photons associated with different ground irradiation objects may be independent in the instantaneous time scale of remote sensing measurement and shall therefore satisfy the Shannon entropy formula of the combinatorial formula of non-interacting billiard balls of different RGB colors---Log{(R+G+B)!/R!G!B!} being approximated in the Sterling factorial product. The normalized condition Eq.(2) replaces the traditional winner-take-all classification scheme for the macro-scale remote sensing per pixel. Our model is new in the sense

that the Hopfield-like quadratic energy is unknown reflectance Matrix $[A_{ij}]$ (between the irradiating S_j source neurons and constraint λ_i force neurons in the Hilbert inner product space), which is not self-dual and can learn their equilibrium without a teacher as follows.

(i) We could derive the Hebbian bilinear learning from the first principle of a priori maximum source entropy formalism, but we could equivalently use the standard Lagrangian virtual displacement of measurement data δX departing the local Maximum Entropy H(S) with the constraint λ_i force neurons.

(ii) $\frac{dX_j}{dt} = \frac{\partial H}{\partial S_j}$ of Eq.(5) to update [A]

$$\frac{\partial A_{ij}}{\partial t} = \frac{-\partial H}{\partial A_{ij}} = \lambda_i S_j \qquad (4)$$

which gives a local Hebbian learning between the Lagrangian multipliers force $\boldsymbol{\lambda}$ neurons and the source neurons $\boldsymbol{S}$.

(iii) Scalar λ_0 is the Helmhotz free energy computed at the local maximum of entropy as follows

$$\frac{\partial H}{\partial S_j} = \ln(S_j) + \sum_{i=1}^{N} \lambda_i A_{ij} + \lambda_0 \qquad (5)$$

Setting Eq(5) to zero, the real non-negative formalism of unknown source neurons is obtained and parameterized in term of unknown reflectance matrix [A] and Lagrangian force neuron $\boldsymbol{\lambda}$ to be determined;

$$S_j = \exp(-\sum_{i=1}^{N} A_{ij}\lambda_i - \lambda_0) = \exp(-\sum_{i=1}^{N} A_{ij}\lambda_i)/\exp(\lambda_0) \quad (6)$$

Substituting Eq.(6) into the normalization Eq. (2) we have derived the Helmhotz free energy as the logarithmic of the Partition function Eq(7) in Statistical Mechanics

$$\sum_{j=1}^{N} S_j = \sum_{j=1}^{N} \exp(-\sum_{i=1}^{N} A_{ij}\lambda_i - \lambda_0) = 1$$

$$\exp(\lambda_0) = \sum_{j=1}^{N} \exp(-\sum_{i=1}^{N} A_{ij}\lambda_i) \quad (7)$$

(iv) Without assumption, we have derived rigorously the sigmoid threshold logic function at the fixed point of local maximum entropy Eq(6) from the coupling between $\boldsymbol{\lambda}$ force neurons from data to independent source neurons $\mathbf{S}$ for general linear optimum communication:

$$S_j = 1/[1 + \sum_{j\neq i}^{N} \exp(-\sum_{i=1}^{N} A_{ij}\lambda_i)] = \sigma([A]\boldsymbol{\lambda}) \quad (8)$$

(v) We can expand the virtual displacement of measurement data X_j in terms of Lagrangian multipliers, force neuron $\boldsymbol{\lambda}$, in a Taylor series at the first order:

$$\Delta X_j = \sum_{k=1}^{N} \frac{\partial X_j}{\partial \lambda_k} \Delta\lambda_k \qquad (9)$$

Evaluate explicitly the partial differentiation using Eq(3) for X and Eq(6) for $\boldsymbol{S}$ in the straightforward manner, we obtain the standard perturbation relationship between the force and the displacement:

$$\Delta X_j = \sum_{k=1}^{N} (X_j^{(known)} X_k^{(known)} - \sum_{i=1}^{N} A_{ji} S_i A_{ik}) \Delta\lambda_k \quad (10)$$

where, as it should be, the Helmhotz free energy λ_o is identified to be related to the displacement vector X:

$$\frac{\partial \lambda_0}{\partial \lambda_k} = -X_k^{(known)}$$

We shall now carry out double recursion for two unknowns, namely S_j and $A_{ij,}$ as follows:

I. Initialization:

(i) Give trial Lagrangian multipliers $\lambda_i^{(o)}$,

(ii) Either substitute the in-situ laboratory measurement $\Delta A_{ij}^{(o)} = \lambda_j S_j \Delta t$ Eq(4) Hebbian-like, or approximate $A_{ij}^{(o)} \approx X_i^{(known)} \sigma(X_j^{(known)})$ +50% random noise to break the rank-one singularity, where σ is the sigmoid function of neuron transfer logic functions.

II. The first recursion:

(iii) compute λ_o from Eq. (7),

(iv) solve Eq. (6) $S_j^{(o)}$

(v) compute $X_i^{(o)}$ from Eq. (3) and thus

$$\Delta X_j^{(o)} = X_j^{(known)} - X_j^{(o)}$$

(vi) compute $\Delta\lambda_k^{(o)}$ from inverting Eq. (10),

$$\lambda_k^{(1)} = \lambda_k^{(o)} + \Delta\lambda_k^{(o)}$$

III. The Second Recursion

(viii) compute new estimate of the improved reflectance matrix $A_{ij}^{(1)}$ from Eq.(4) is possible because the inner product with known $\lambda_i^{(1)}$. Update [$A^{(1)}$] by the improved virtual displacement $\frac{dX_j}{dt} = \frac{\partial H}{\partial S_j}$ of Eq.(5) for reaching the maximum source entropy the right hand side $\Delta X_j^{(0)} = \Delta t(\ln(S_j^{(0)}) + \sum_{i=1}^{N} A_{ij}^{(1)}\lambda_i^{(1)} + \lambda_0^{(0)})$, setting Δt =1, and transpose all to the left hand side except [A]$\boldsymbol{\lambda}$ at right-hand side, and solve for [A] by the rank-one associative memory outer product formula.

IV. Go Back to Step II.

The detail coding of the algorithm published in 1997 is enclosed for reader convenience in the Matlab source code in Appendix A. Other than the questionable rank-one associative memory approximation of the exact eigen-vector decomposition, the Matlab algorithm seems to be fast and stable to implement the Lagrangian

Constraint Neural Network (LCNN) model and find S_j and A_{ij} for a known $X_j^{(known)}$ per pixel.

Furthermore, we have derived *ab initial* from it the usual sigmoid threshold logic **S** = $\sigma(\boldsymbol{\lambda}[A])$ and the Hebbian learning ruled $\Delta A_{ij} = \boldsymbol{\lambda}_i \mathbf{S}_j$. We can thus conjecture a general theorem of communication by independent classes that the linear matrix transform (associative memory [A] recall) between data **X**=[A]**S** and independent classes under the constraint of comprehensive but unknown class decomposition $\Sigma_j S_j = 1$ must lead naturally by *a priori MaxEnt* H(**S**) the biological sigmoid transfer logic and Hebbian learning.

4. EXPERIMENTS COMPARISON

We wish to compare LCNN ica and BSAO ICA solutions. LCNN approach can be adopted for searching anomalies, as revealed as singularity of jump of single source among neighborhood multiple sources. Note that the top bright spots show perhaps Poisson sun glint density in the Mediterranean Sea (top) where the strip pattern sensor noise are due to not-yet-cool down TM Sensors as NASA concurred afterwards. Tel-Avis City is located at low left. At the lower right the desert is covered with some cloud patterns, sand dome patterns and perhaps mixed with some interesting non-periodic man-made objects in the sand. We note that the Matlab can code the post-processing BSAO pixel-ensemble average algorithm to compute the relatively small data of Landsat imageries directly so that we can compare them herewith, but it could not solve the 200 hyperspectral images of million pixels**.** Only its mirror symmetric sibilant method can solve the hyperspectral images pixel-by-pixel on flight in real time. This might be important message for a large system brain imaging experiments.

5 HYPERSPECTRAL LCNN RESULTS

We are now confident and ready to apply LCNN to hyperspectral image remote sensing. Hyperspectral has 200 channel image data collected by NASA using U2 airplane, but the spectral available to us is ranging from 0.5 micron to short/middle Infrared (2 to 3 micron) only. They are divided into 200 channels. Black color means no energy present at the particular channel. Figure 6 shows the unsupervised and comprehensive decomposition results of LCNN; Figure (a) shows playa and also in figure (b), but there is a very strong interference in (b), the bright (red) double component spot in the right turns out to be a vehicle parked on the grass. Shade and cinder show in figure (c). Playa is also showed in (d), rhyolite shows in figure (d) and (e), and vegetation shows in (d), (e) and (f). The output of Lagrangian-constrained Neural Network (LCNN) shows that the black color means no contribution to the specific object-class geometry-statistics. The color migrates from picture to picture indicating different statistics of soil, tree, and volcano rocks. The circle points are the unknown man-made objects detected within the Rock Mountain volcano area.

The *a priori MaxEnt* H(**S**) with the measurement constraint is powerful to solve the unknown independent classes of irradiation source from 200 component vector data **X** =[A]**S**, and 100% of total but unknown number of components that varies from pixel-to-pixel. However, in contrast, the BSAO ICA assumes the *a-posteriori MaxEnt* H(**V**) of neuron post-processing output components **V** =sigmoid([W]**X**) where **X**=[A]**S**, and the required ICA pixel-ensemble average does not permit running 200 images, each 200x200 pixels on PC at the same time. Seldom is the case that a real world application, e.g. optical remote sensing, satisfies the linear and instantaneous ICA matrix vector model.

6. CONCLUSION

The important realization of the statistical mechanical approach to the independent classification analysis (ica) namely LCNN based on *a priori MaxEnt* H(**S**) is to circumvent the tough mathematics challenge to be able to uniquely factorize the unknown joint-pdf from multi-channel mixture data alone, namely ICA based on *a posteriori MaxEnt* H(**V**), if only we wish to know is at the first the classes, but not yet the underlying pdf to characterize/identify the classes to be done subsequently with supervision once we have them sorted. We can then compute the ica images their histogram and moments for solving the inverse pdf problem from moments. To demonstrate the versatility of Lagrangian constraint calculus, we consider nonlinear measurement constraints:

$$H(S_j) = -\sum_{j=1}^{N} S_j \ln(S_j) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i (F(A_{ij}S_j) - X_i^{(known)}) - (\lambda_0 - 1)(\sum_{j=1}^{N} S_j - 1)$$

ACKNOWLEDGEMENT

The author would like to thank Charles Hsu and Shane Ren for conducting the Landsat and hyperspectral numerical experiments respectively.

REFERENCES

(Standard ICA references of BSAO & JHCC, see ICA e-mail Web. Here are only the remote sensing**)**

Adams, J.B. and M.O. Smith, Spectral mixture modeling: a new analysis of rock and soil types at the Viking lander 1 suite, *J. Geophysical Res*., vol. 91, no. B8, pp. 8098-8112, July 10, 1986.

Settle, J.J. and Drake, N.A., Linear mixing and estimation of ground cover proportions, *International Journal of Remote Sensing*, vol. 14, pp. 1159-1177, 1993.

Szu, H. and C. Hsu, Landsat Spectral Unmixing à la Superresolution of Blind Matrix Inversion by Constraint MaxEnt Neural Nets, in Wavelet Applications IV, *Proc. SPIE*, 3078, pp. 147-160, 1997a.

Szu, H. and C. Hsu, Blind De-mixing with Unknown Sources, *Proc. of ICNN*, vol. 4, pp 2518-2523, Houston, 1997

Szu, H., Progresses in unsupervised artificial neural networks of blind image demixing, IEEE Indus. Elec.. Soc.. News., pp. 7-12, June 1999a.

Szu, H., ICA-an enabling tech for Intelligent Sensory, IEEE Cir. Sys. Soc.News, pp. 14-41, Dec. 1999b

Szu, H. and Ren,H., Unified Lagrangian Neural Network Method for Subpixel Classification in Hyperspectral Imagery , SPIE Proc. V. 4391, Wavelet Applications , Orlando, April 2001

.

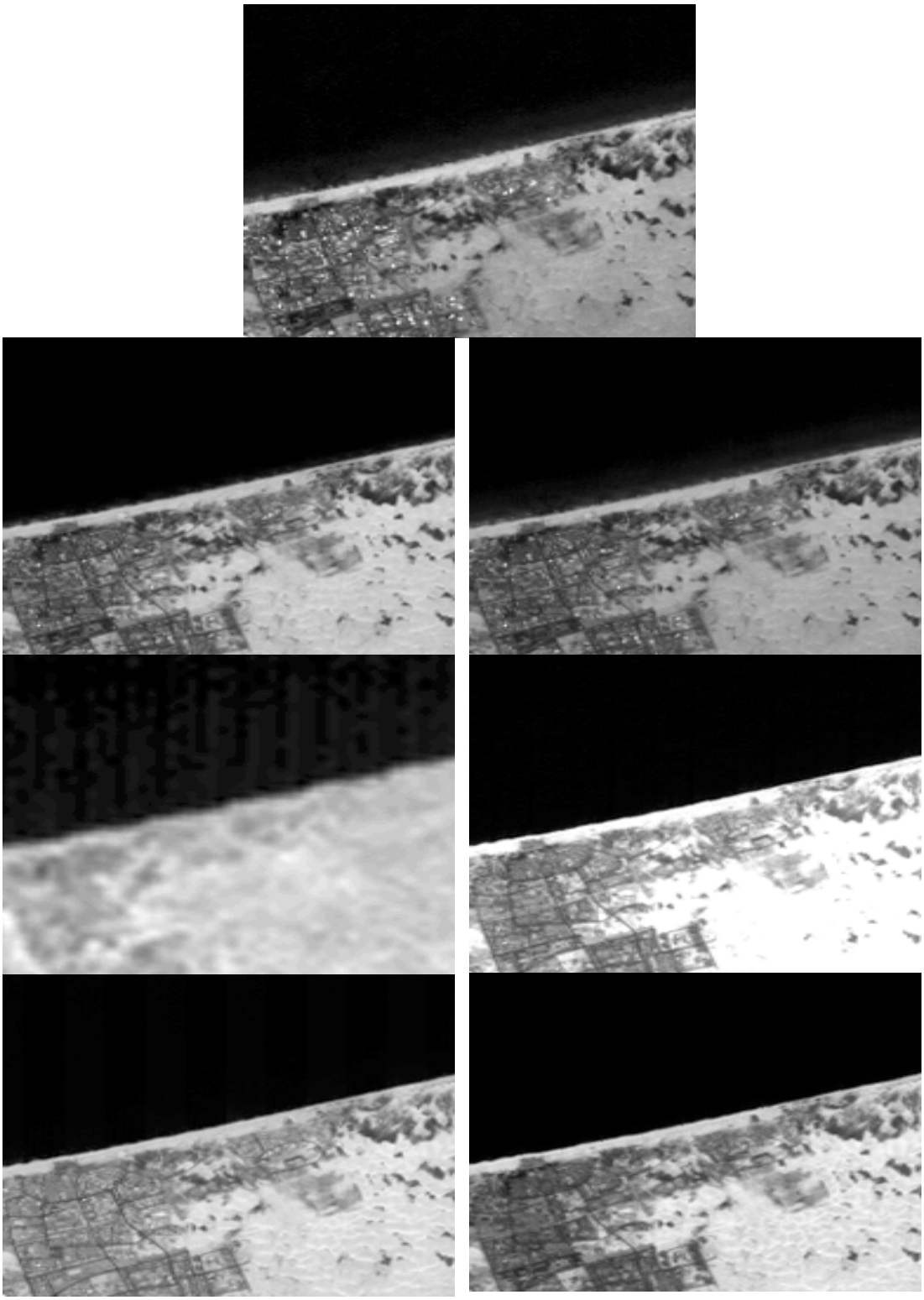


Fig. 3 Original NASA Landsat multi-spectral images (visible to near, mid, long IR) over a Mediterranean town with some cloud covering over desert area (lower right). No sensor pattern noise or fine structure or hidden objects are revealed in the energy spectrum domain. Visual classification however prefers the geometry texture statistics than energy spectrum.

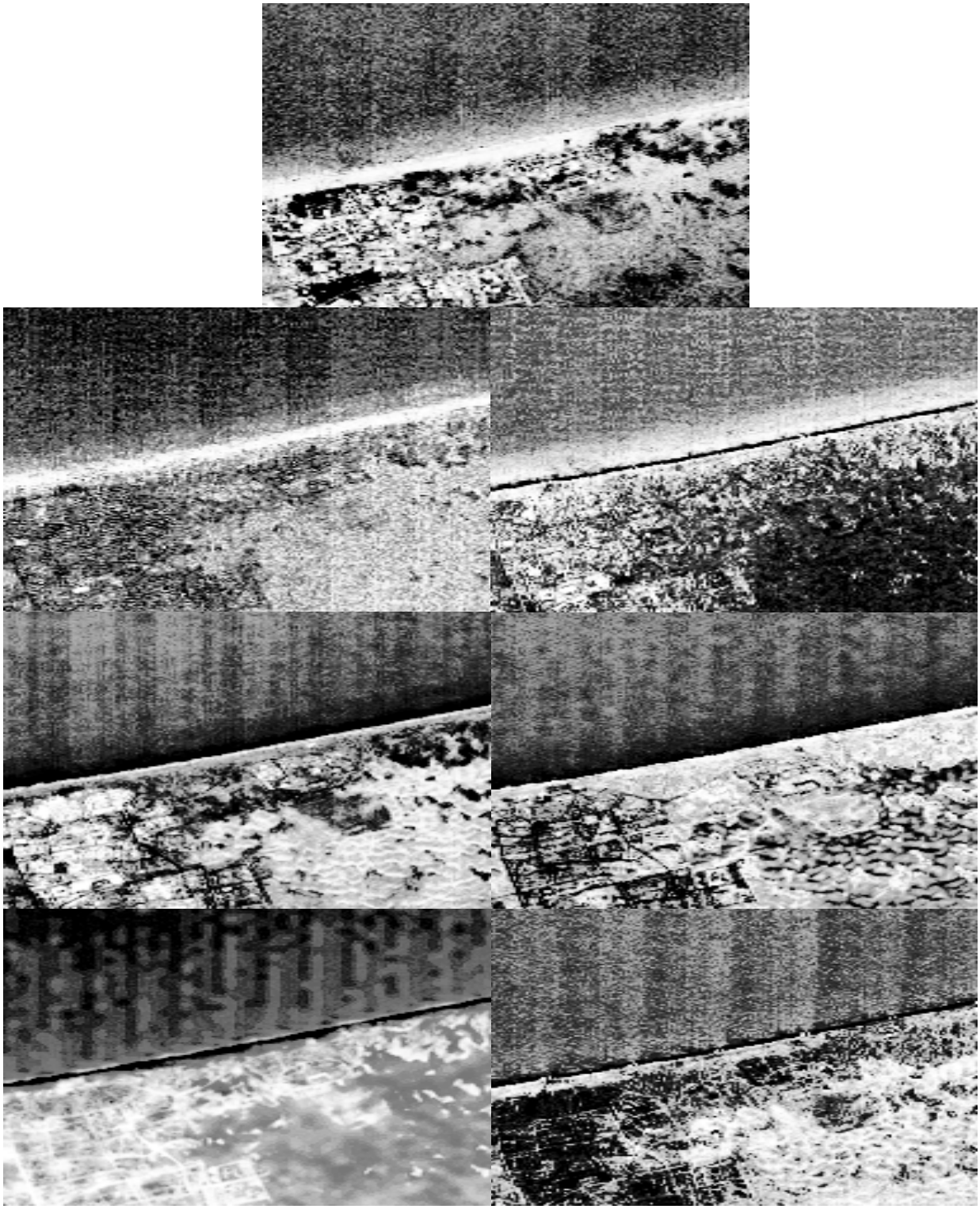


Fig. 4 *A Posteriori MaxEnt* **H**(σ([W]**X**) **ICA** (BSAO algorithm) reveals vividly in the pixel-ensemble average method the strip pattern sensor noise in the Mediterranean sea (upper half) due to unsettled cooling pattern (as

NASA concurred after our discovery in 1997), fine structure along shoreline, Tel-Avis City in the lower left hand side); cloud, sand pattern & interesting non-periodic structure in the desert (lower right hand side)**.**

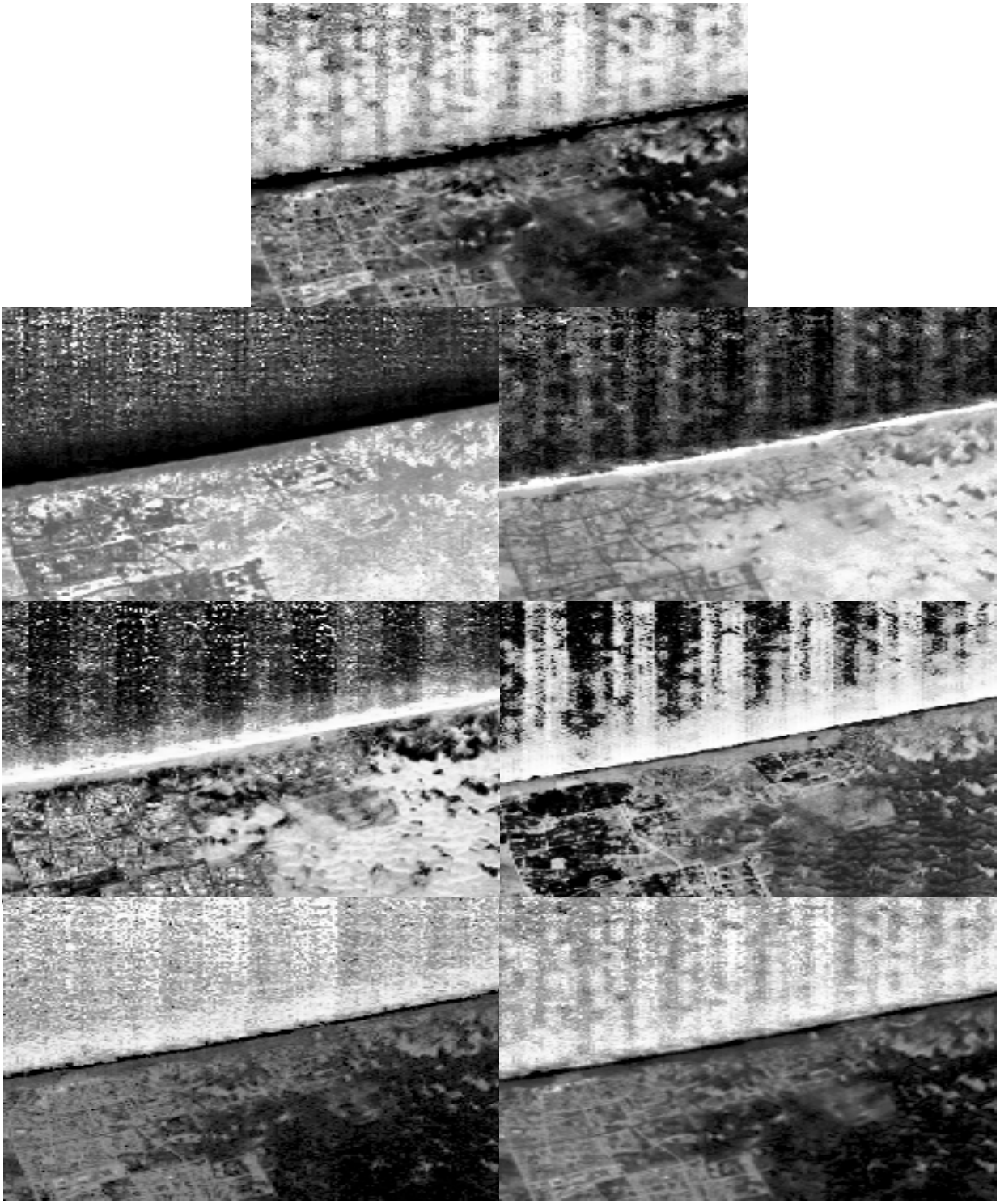


Fig. 5 *A Priori MaxEnt* H(**S**) independent class analysis (ica) is to find classes not yet the pdf of classes by means of statistical-mechanics Lagrangain Constraint Neural Network (LCNN) approach that allow us to solve pixel-by-pixel the decomposition of Landsat 7 spectral images (Fig.3) into density-images for the percentage of independent density (100%=bright, 0%=black) without neighborhood average. (A Priori MaxEnt ica is less ambitious than *a posteriori MaxEnt* ICA for the factorization of the underlying joint-pdf that must utilize the pixel-ensemble average over all images shown in Fig. 4). Results are similar to Fig.4 that gives us the confidence of applying the pixel-by-pixel LCNN to the hypersepctral of 200 channel per pixel that BSAO is intractable for 200 images of 200x200 pixels.

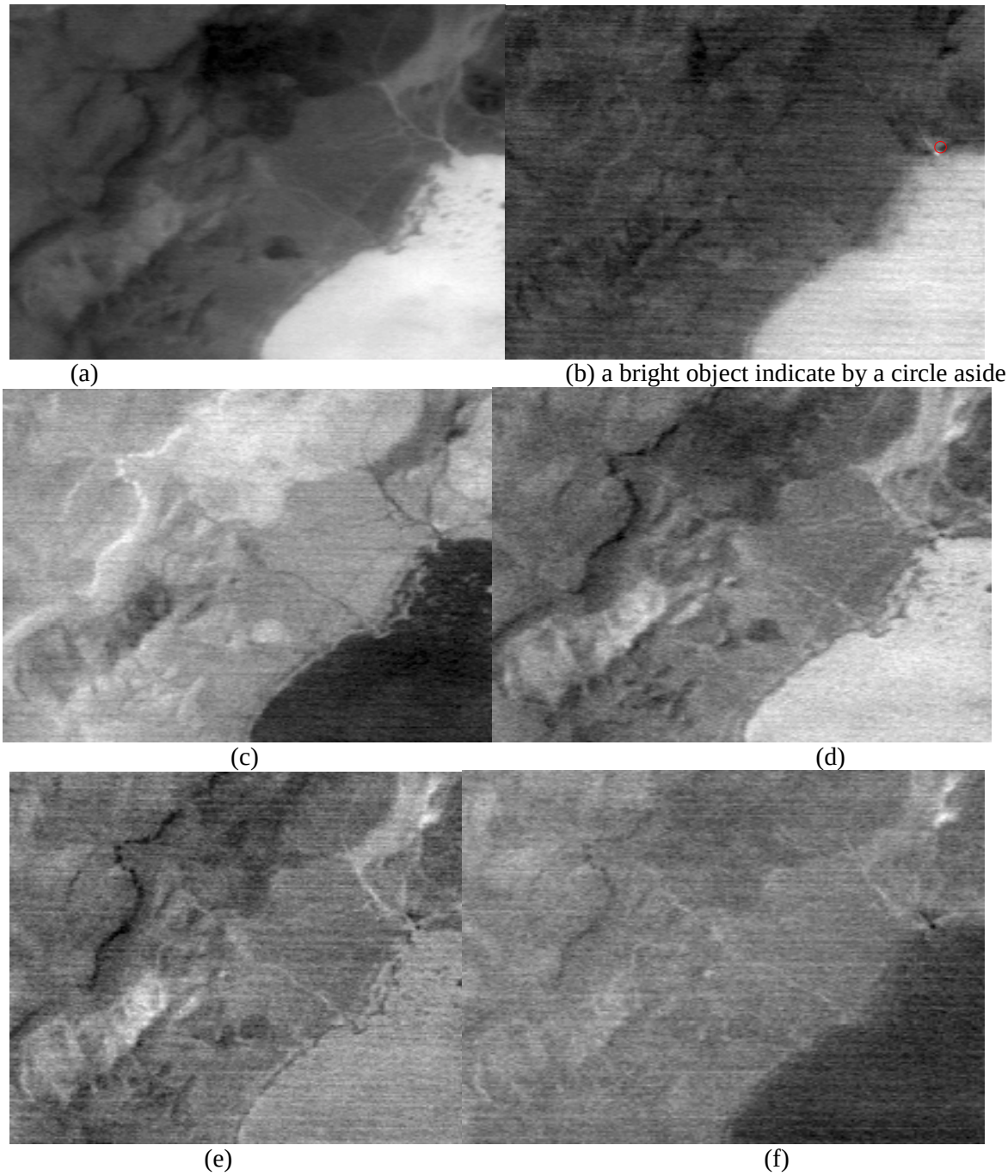


Fig. 6 Sample results of LCNN of AVIRIS hyperspectral 200 channels which are unsupervised with unknown [A] using Lagrangian Constraint Neural Network (LCNN) approach. Figure (a) shows playa and also in figure (b), but there is also a very strong interference in (b), the bright spot in the right. Shade and cinder show in figure (c). Playa is also showed in (d), rhyolite shows in figure (d) and (e), and vegetation shows in (d), (e) and (f). The full 200 channel results is given by Szu & Ren in SPIE 2001 Orlando Wavelet Applications

APPENDIX Lagrangian Constraint Neural Network Algorithm

MATLAB code for unsupervised Lagrangian Constraint NN algorithm

```
 function maxent
% Lagrangian Constraint NN
solving
```

$$MaxH(S_j) = -\sum_{j=1}^{N} S_j \ln(S_j) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i (A_{ij} S_j - X_i^{(known)}) - (\lambda_0 - 1)(\sum_{j=1}^{N} S_j - 1)$$

```
% Pixel-by-pixel divide-and-conquer non-ensemble
approach to large data set
% H. Szu and C. Hsu, “Landsat Spectral Unmixing à
la Superresolution of Blind Matrix Inversion by
% Constraint MaxEnt Neural Nets,” in Wavelet
Applications IV, Proc. SPIE, 3078, 1997, 147-160.

 [X,Y]=meshgrid(1:7,1:7);

% initialization
lambda = [0.6228, 0.6337, 0.4577, 0.1095, 0.7252,
0.01752, 0.4128];% Lagranging force randomly
initialized
X0 = [0.5382, 0.1023, 0.6404, 0.4358, 0.0278,
0.2425, 0.3299];  % data given per single pixel


A = X0'*sigmoid(X0) + rand(7,7)*0.5;          % add
small noise to avoid rank 1 singularity
A = normalize(A);              % normalize at unit hyper
sphere
figure(1);mesh(A);title('Initial weights');

cnt=40; %iteration steps

for i=1:cnt
            lambda0 = log(sum2(exp(-A'*lambda')));
            %based on unit constraint (Eq.(9))
   S = exp(-A'*lambda'-lambda0);    %Endmember

   % update Lagranging forces
   X = A*S; % Linear mixing constraint
   dX = X0' - X;
   z=X0'*X0-A*((S*ones(size(S'))).*A);
   if abs(det(z))<=1e-10
      dlambda = zeros(size(dX));
   else
      dlambda = inv(z)*dX;
   end
   lambda = lambda + dlambda';

   % update weight a la Hopfield-like ( dXi/dt =
dH/dSi )
   A = (dX+log(S)-
lambda0)*(lambda./sum2(lambda.^2));  % Inverse
AM matrix,

% e.g. if y=[A]x then [A]=y*x'/|x|^2
   A = normalize(A);          % normalize each row
vector producing unit hyper sphere

   % for display purpose
   SS(:,i)=S;
end
x=(1:cnt);
figure(2);
plot(1,X0(1),'x',1,X0(2),'x',1,X0(3),'x',1,X0(4),'x',1,X
0(5),'x',1,X0(6),'x',.1,X0(7),'x',x,SS(1,:),'b',x,SS(2,:),'k
',x,SS(3,:),'r',x,SS(4,:),'g',x,SS(5,:),'m',x,SS(6,:),'c',x,S
S(7,:),'y');title('Endmember in percentage');
figure(3);
mesh(A); title('Normalized spectral reflectance
matrix');
function y = sigmoid(x)
y = 1./(1+exp(-x));
function A=normalize(A)
for i=1:size(A,1)
   A(1,:) = A(1,:)./sqrt(sum2(A(1,:).^2));
end
function Y=sum2(X)
Y=sum(sum(X));
```