Thermodynamic Learning Rule

Harold Szu
Catholic University of America, USA
Corresponding author: szuharoldh@gmail.com

Abstract

The 2nd law of thermodynamic governs an open dynamical system at an isothermal equilibrium. Helmholtz proved that such a system operates at the Minimum Free-Helmholtz Energy (MFE) min $H \equiv E - T_o S$, much like an efficient car engine. An MFE engine has internal combustion energy *E*, with an exhaust entropy *S* operating at an optimum engine temperature T_o. We propose to model the human brain's learning process according to the isothermal equilibrium, assigning an MFE cost function to associated input vector time series', with unknown output features. We further examine the implication of modeling synaptic ion currents among neurons of various inter-connection sizes for 'grey matter boxes' of arbitrary emissivity. We compare these with the normal modes of a single back box Planck black-body of an ideal emissivity radiation curve. As a result, we derive a *Hebbian learning rule* that is consistent with Donald Hebb's original observation in neurophysiology a half century ago. Such an Artificial Neural Network (*ANN)* enjoys a self-referenced *unsupervised learning* process known as regularized *Lagrange Constraint Neural Network (LCNN).* We rigorously solve the space-variant, ill-posed inverse imaging problem called Blind Separation *of Equivalent Planck Source's (EPS's).*

1. Introduction

On Earth, the animals which can learn by experience seem to be equipped with: (1) warm blooded brains that provide steady kinetic transport for efficient cellular operations, and (2) the 'power of paired sensors' (pops) which gather vector time series data $\vec{X}_{S_i}(t)$ for self-referenced unsupervised learning. Likewise, humans have symmetrical vector time series sensors (ears, eyes, nostrils, olfactory bulbs, taste buds, limbs extremities) which communicate with each other through the nervous system. The nervous system and the brain must be kept at isothermal equilibrium, $T_o = 37^oC$. These are necessary but not sufficient conditions for intelligent beings. A higher temperature does not necessarily imply smarter or quicker learning. For example, the chicken's brain is in equilibrium at 40^oC but they lack the hands and tools necessary to be smarter than humans.

Almost all imaging at a distance produces imagery having an unknown mixture of Equivalent Planck Sources (EPSs) of arbitrary emissivity [*proved in Sect. 1.2*]. Solving the inverse imaging problem requires thermodynamic learning rules.

It turns out that physics dealt with efficient measurement that took into account of robust and reliable results. The linear programming used in compressive sensing [1,2,3] turns out to be a linear approximation of the MFE cost function [*proved in Remark #2*]:

$$min. \quad H = E - T_o S, \tag{1}$$

The MFE is like an efficient car engine which has an internal combustion energy *E*, an exhaust entropy *S*, which operates at the optimum engine temperature T_o.

Eq.(1) included CRT&D CS as a special case when (1) the negentropy, –S, was the convex hull L1 minimization $-T_o S \approx T_o \sum_i^k < logs_i >_{S_i}$, because the class entropy must be real positive ($-logs_i \geq 0$; $s_i \leq 1$); and (2) the internal energy *E* is analytic function of which the first order Taylor series expansion assumes the linear error slope:

$$\mu_j(\vec{s}^{(o)}) \equiv \frac{\partial E}{\partial s_j}\Big]_{\vec{s}=\vec{s}^{(o)}} \approx \left(s_j - s_j^{(o)}\right), \tag{2a}$$

which becomes the LMS L2 similarity.

$$\begin{aligned} E &= E_o + \sum_{i=1}^k \frac{\partial E}{\partial s_i}(s_i - s_i^{(o)}) \\ &= E_o + \sum_{i=1}^k (s_i - s_i^{(o)})(s_i - s_i^{(o)}) \\ &= E_o + ||s_i - s_i^{(o)}||^2 \end{aligned} \tag{2b}$$

In this paper, we have generalized Eq.(2) with a non-linear 2nd order Lagrange Constrained Neural Network (LCNN).

1.1 Application of Planck Law to the"Brain in Box"

Planck discovered a unique irradiance distribution which peaked at a unique wavelength for each isothermal temperature of a perfectly emissive black body, at unit emissivity $\varepsilon = 1$. We compare Planck's fixed black body resonator cavity at arbitrarily constant temperature with our brain. Our brain is kept at a fixed isothermal equilibrium at 37^oC, but consists of different sized grey-matter interconnection boxes having arbitrary emissivity $0 < \varepsilon \leq 1$. Despite our grey matter box is not an ideal black body, we suspected this unique peak might be true as long as an arbitrary grey matter body box has a unknown but fixed emissivity [proved in Theorem 1]. In addition to Planck, our inspiration comes from the 'big-bang' perspective of

the universe; expanding incessantly without an outside boundary to reflect back any outgoing electromagnetic waves. The Chinese philosopher Lao-tze once said that the largest has no outside while the smallest has no inside. Therefore, the universe cannot be ideal black body but at best a grey body. Nobel Laureates Arno Penzias and David Wilson of Bell Lab measured the comic background radiation and found it is indeed peaked at the twelve hundred microns wavelength (1.2 mm at 160 GHz) having the apparent brightness temperature at $\varepsilon T_K = 0.91x3^o = 2.73^o\ Kelvin$ at the non-ideal emissivity ε~0.91[Remark #4] Our EEG brainwaves are mediated by synaptic ion currents within different sizes of functional grey matter boxes of arbitrary emissivity filled with neurons. The largest box is the Left logical & Right artistic Hemispheres of the size of 7 cm; associative memory Hippocampuses at 3 cm; the 'emotional' Amygdala at 2 cm; and the smallest may be the Pituitary gland grey matter box clock cycles at 1cm. Consequently, our brains support mixture of different modes: delta (0-4 Hz); theta (4-7 Hz); alpha (8-12Hz); beta (13-30 Hz). Another interesting observation is that all *EEG* waves have peak frequencies which are separated equally by 4 Hz intervals. This could correspond to fundamental topological structures among sub-grey-matter boxes.

Our mental activity is thought to be an unknown mixture of *EEG* waves which co-exist at an isothermal equilibrium that might generate like *EPS's* with various peaks mixture along the full *em* spectrum. Human brains can now be non-invasively monitored by a wireless baseball hat, equipped with dry nano-electrodes which utilize a compressive sensing algorithm of sparse linear combinations of *EEG* signals[4]. However, no one has systematically analyzed brainwaves from the inter-nested, "Pandora's brain" made of arbitrarily sized grey matter and fixed emissivity viewpoints. Thus, the myth of telepathy as a 'super-resonance' among different "Pandora's brains" remains. The simple physics is given for the collaboration with neurophysiologists.

Planck's law described a quantized set of simple normal modes of *electromagnetic (em)* waves that oscillate within a black-body resonator cavity, realizing the vanishing *em*-amplitudes at a constant wall temperature T_K, kept by an outside large heat reservoir. The Planck heat source is an ideal black box resonator. It supports all positive integer numbers $n = I_+$ of harmonic wavelengths $\lambda = \frac{c_o}{n\nu}$ like a violin string vibrating at the fundamental frequency ν and the constant speed c_o: $\left\{\lambda = \frac{c_o}{n\nu}\middle| n = I_+\right\}$. Use was made of *Einstein n*-photon energy formula: $E \Longrightarrow E_n = nh\nu$ to compute *Maxwell-Boltzmann* probability: $\exp\left(-\frac{E}{K_B T_K}\right) \equiv z^n$, resulted in an infinite geometric series of $z \equiv \exp(-\frac{h\nu}{K_B T_K})$:

$$z^1 + z^2 + z^3 + \cdots = z(1 + z + z^2 + ..) = \frac{z}{1-z} = \frac{1}{z^{-1}-1} \quad (1)$$

where the fluctuating vacuum state $n = 0$ was intentionally excluded to result in z^{-1} other than z in Eq.(1) [cf. Remark 6]. Multiplying the density of states $\frac{2h\nu^3}{c_o^2}$, *Planck* derived rigorously and reproduced early laws:

$$I_\nu(T_K) = \frac{2h\nu^3}{c_o^2}\frac{1}{\exp\left(+\frac{h\nu}{K_B T_K}\right) - 1}$$

$$= \begin{cases} \frac{2h\nu^3}{c_o^2}\exp\left(-\frac{h\nu}{K_B T_K}\right); h\nu \gg K_B T_K;\ Wien\ law \\ 2\nu^2 K_B T_K;\ h\nu \ll K_B T_K\ ;\ Rayleigh - Jeans\ Law \end{cases} \quad (2)$$

(*UV catastrophe Ehrenfest*) Q.E.D.

where *Planck*'s constant h =4.1 × 10^{-15} *eV sec;* $K_B T_K = \frac{1}{40} eV$ at room $T_K = 300$ and thus *Boltzmann* $K_B \approx 0.1\ meV$.

Theorem 1: Image Processing by Equivalent Planck Sources (EPRs)

If we define the observed apparent irradiance in terms of an apparent brightness temperature T_{B_ν}, then the apparent irradiance I_{B_ν} is a percentage of true blackbody irradiance that is reduced by incomplete thermal accommodation of absorbed and re-emitted radiation. This causes a non-zero boundary condition and non-ideal emissivity $0 < \varepsilon_\nu \leq 1$:

$$I_{B_\nu}(T_{B_\nu}) \equiv \varepsilon_\nu I_\nu(T_K) \quad (3)$$

Given that Planck sources have unit emissivity, then grey bodies have an arbitrary emissive source which is uniquely defined by the *EPS of apparent brightness temperature in* proportional to the grey body emissivity multiplied with the ideal Blackbody Temperature:

$$T_{B_{\nu_o}} \cong \varepsilon_{\nu_o} T_K \quad (4)$$

PROOF:

Use was made of Planck's law Eq.(2) and the definition of apparent irradiance for $0 < \varepsilon_{\nu_o} \leq 1$ Eq.(3),

$$\varepsilon_{\nu_o}\left(\exp\left(\frac{h\nu_o}{K_B T_{B_{\nu_o}}}\right) - 1\right) = (\exp(\frac{h\nu_o}{K_B T_K}) - 1);$$

$$\frac{T_{B_{\nu_o}} K_B}{h\nu_o} = \frac{1}{Log\left[1 + \frac{\exp\left(\frac{h\nu_o}{K_B T_K}\right) - 1}{\varepsilon_{\nu_o}}\right]}.$$

Fig.1 plotted the apparent brightness temperature $T_{B_{\nu_o}}$ against the black body Kelvin temperature T_K at the peak emissivity ε_{ν_o} in 10% increments. Q.E.D.

1.2 Mixing Matrix

We measured day color triplets per pixel $\vec{X}_s$=(*Blue: 0.2~0.3 μm; Green*: 0.3~0.5*μm* and *Red:* 0.6~0.8 *μm)* or night infrared (IR) triplets $\vec{X}_s$=(*MWIR*: 3~5 *μm, LWIR I: 8~10 μm* and *LWIR II: 10~12 μm).* We derived the unknown mixture of discrete heat sources associated with discrete temperatures in the percentage vector $S \Rightarrow S_j \Longrightarrow \vec{S}$; j = 1,2,3, whose probability is related to the Boltzmann entropy formula derived in R,G,B color entities [*Remark #3*]. Typically, we had illuminating source S_1, object body

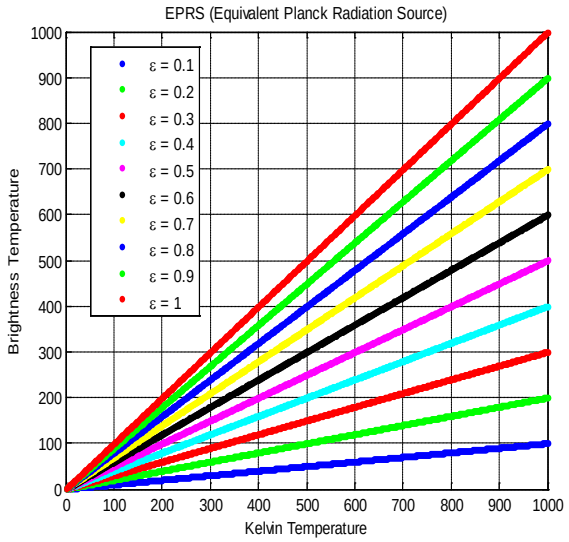


Fig.1 Defining EPS, we plotted the Brightness temperature $T_{B_{\nu_o}}$ versus the Kelvin Temperature T_K. Setting unit $h\nu_o/K_B$=1, we plot: $y = 1/Log(1 + \frac{1}{\varepsilon_{\nu_o}}(\exp(\frac{1}{x})-1))$ from 1 to 1000, as we step emissivity $\varepsilon_{\nu_o} \leq 1$ in 0.1 increments.

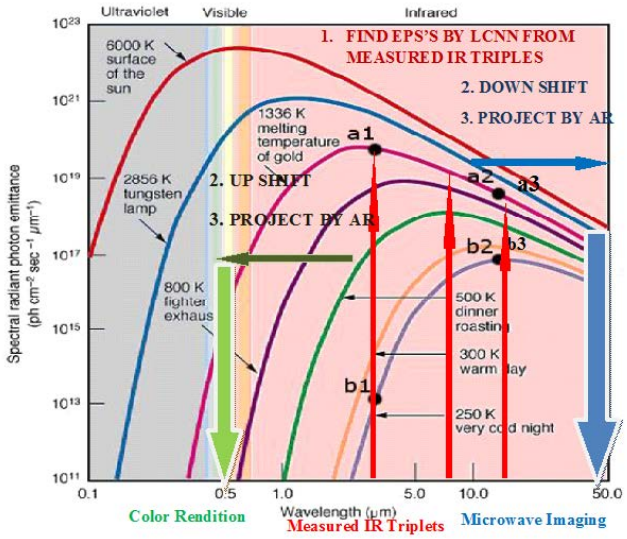


Fig.2 Compressive Sensing by means of Equivalent Planck Sources (EPSs): These virtual upper-down spectral extrapolations were possible because the peak wavelength associated one, and only one Kelvin Temperature in a single monotonic curve, where the left-shoulders are up-left-shifted to a shorter sub-micron wavelength toward X-rays. Wilhelm Wien's displacement law ~T^4 followed due to monotonic single peaks per Kelvin temperature, according to the dimensionality analysis of the integrated Eq.(2a).

hot spot source S_2, and other sensor coolant source S_3, etc. in terms of Boltzmann entropy probability normalization $1=S_1+S_2+S_3$. The mixing matrix mapped j-temperature sources to spectral *i*-components:

$$\vec{X}_{s_i} = [A_{i,j}]\vec{S} = S_1\,\vec{a}(T_1) + S_2\,\vec{b}(T_2) + S_3\vec{c}(T_3) \qquad (5)$$

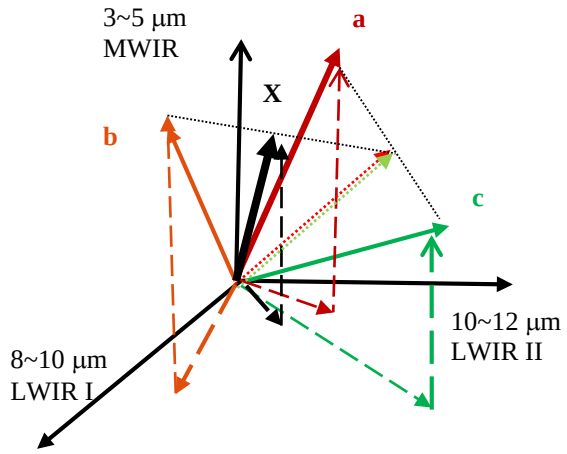


Fig.3 Three column vectors corresponding to three EPS associated equivalent brightness temperatures.

The *EPS* temperature sources of mixing matrix [A] are usually not known for the generation of ground spectral data. The *IR triplets* per pixel as transposed (*Tr*) row vector $\vec{X}^{Tr} \equiv (X_1 \quad X_2 \quad X_3)$ were measured at the center of each band value, i.e., 4 μm, 9 μm, and 11 μm, respectively, in **Fig.2.** The mixing matrix had three column vectors $\vec{a}(T_1), \vec{b}(T_2), \text{and } \vec{c}(T_3)$, corresponding to three EPS associated equivalent brightness temperatures $\vec{a}(330^oK), \vec{b}(265^oK), \text{and } \vec{c}(200^oK)$ shown in **Fig.3**. For example,

$$\mathbf{X}(\text{pixel}) \equiv \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = [\boldsymbol{a}(T_1) \quad \boldsymbol{b}(T_2) \quad \boldsymbol{c}(T_3)]\boldsymbol{S}(pixel)$$

$$\equiv \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}$$

$$= 10^{-5} \begin{pmatrix} 67.46 & 4.65 & 0.05 \\ 502.82 & 152.39 & 21.41 \\ 449.86 & 168.13 & 33.61 \end{pmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}.$$

Each source S_j might have different percentage values for each pixel such as $\vec{S}^{Tr} = (S_1, S_2, S_3) = (30\%, 50\%, 20\%)$. The space-variant propagation defines an unknown mixing matrix that can vary in space from a group of pixels to others known as different isoplanar or Komogorov regions in astronomy imaging; while the unknown object sources could vary from sub-pixel to pixel for a high definition picture.

Mathematically, Blind EPS Separation (BSS) is challenging because the mixing matrix [*A*] in Eq.(5), mapping the temperature sources to the spectral band values, was unknown for the case of remote sensing, and the inverse weight matrix is typically ill-conditioned: **c1**

$$\vec{S}_j = [W_{j,i}]\vec{X}_{s_i}\text{, where } [W_{j,i}] = [A_{i,j}]^{-1} \qquad (6)$$

The following theorems met the challenge.

2. Thermodynamic Learning Rule

The key enabler of our new approach is resolving the 'Mexican Standoff' slowdown, or in critical slow down during thermodynamic phase transition phenomena.

LEMMA: Due to unsupervised learning, the energy cost function is unknown for image processing at remote sensing. The first order Taylor series becomes 2nd order in the smallness, requiring the second order Taylor series expansion, curvature C^k, to determine the Lagrange error slope vector μ_α together with the estimation error which converges self-consistently.

$$\boldsymbol{E} \cong \boldsymbol{E}^{(o)} + \mu_\alpha\left([W_{\alpha,\beta}]X_\beta - S_\alpha^{(o)}\right) + \tfrac{1}{2}C^k|[W_{i,\beta}]X_\beta - S_i^{(o)}|^2 \Longrightarrow O(0)^2; \quad (7)$$

2.1 Critical Slowdown Phenomena

When the state function is known, then the 1st order Taylor series derivative is of 1st order smallness. However, Lagrange knew that the critical slowdown phenomena in thermodynamic phase transition occurred because the state function is unknown which required other expansion to determine it, causing the 1st order to become 2nd order in smallness. For example in BSS, the minimum of Helmholtz free energy, $min.H = E - T_oS$, is expanded in a Taylor series with respect to unknown number of heat sources in terms of associated class entropy per pixel, $\boldsymbol{S} \Longrightarrow \{\boldsymbol{S}_i \leftrightarrow \vec{\boldsymbol{S}} = [W]\vec{\boldsymbol{X}}_s\}$. We applied the 1st order Taylor series of MFE of the unknown cost function *E*

$$\left(\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{S}_\alpha}\right)\left(\boldsymbol{S}_\alpha - \boldsymbol{S}_\alpha^{(o)}\right) \equiv \boldsymbol{\mu}_\alpha \boldsymbol{c}_\alpha(\vec{\boldsymbol{S}}) \Longrightarrow O(0)^2, \quad (8)$$

where the inner product involves the 2nd order smallness and it is difficult to determine the iteration involving both the data and the slope of data. Simultaneously, the *unknown Lagrange energy slope vector* flattens near the minimum with a zero slope

$$\boldsymbol{\mu}_\alpha\left(\vec{\boldsymbol{s}}^{(o)}\right) \equiv \frac{\partial \boldsymbol{E}}{\partial \boldsymbol{s}_\alpha^{(o)}} \Longrightarrow O(0)$$

together with the weighted learning of the ***s***-spectral vector measurement data $\vec{\boldsymbol{X}}_{\boldsymbol{s}}$ generating an learning error vector.

$$c_i\left(\vec{\boldsymbol{s}}^{(o)}\right) \equiv \left(\boldsymbol{s}_i - \boldsymbol{s}_i^{(o)}\right) = \left[W_{\alpha,\beta}\right]\boldsymbol{X}_\beta - \boldsymbol{s}_\alpha^{(o)} \Longrightarrow O(0). \quad (9)$$

Consequently, the 1st order of MFE involves 2nd order smallness in a product which becomes too small to determine whether it is the cause or the consequence about which one of the product to take the next step that approaches zero. Such double loops of iterations will suffer slow convergence, known in unsupervised learning of the artificial neural network (ANN) community with the nickname 'Mexican standoff', from wild western Hollywood movies. Taylor expansion of the *MFE* Eq.(7) involved unknown internal energy *E* defines the linear slope as the Lagrange vector constraint of the error slope $\vec{\mu}_j$ and the second order of smallness curvature as the Karush-like penalty function:

$$E = E_o + \sum_{i=1}^{K}\frac{\partial E}{\partial \boldsymbol{s}_i^{(o)}}(\boldsymbol{s}_i - \boldsymbol{s}_i^{(o)}) + \frac{1}{2}\frac{\partial^2 E}{\partial \boldsymbol{s}_\alpha^{(o)}\partial \boldsymbol{s}_\beta^{(o)}}\left(\boldsymbol{s}_\alpha - s_\alpha^{(o)}\right)\left(\boldsymbol{s}_\beta - \boldsymbol{s}_\beta^{(o)}\right)$$

$$\cong E_o + \boldsymbol{\mu}_\alpha\left(\vec{\boldsymbol{s}}^{(o)}\right)\left(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}\right) + \tfrac{1}{2}C^k|\vec{\boldsymbol{s}} - \vec{\boldsymbol{s}}^{(o)}|^2 \ ,$$

$$\frac{\partial^2 E}{\partial s_i^{(o)}\partial s_j^{(o)}} \cong C^k\,\delta_{i,j};\quad C^k = \beta_o C^{k-1};$$

$$\beta_0 > 0; k = 1,2,3, etc. \quad (10)$$

Theorem 2: Hebb learning Rule of Neural Network Weight Matrix

Given a measured ***s***-spectral band vector per pixel location, $\vec{\boldsymbol{X}}_s$: we solve unknown heat sources $\vec{\boldsymbol{S}} = [W]\vec{\boldsymbol{X}}_s$ by Artificial Neural Network unsupervised leaning weight matrix update.

$$\vec{S} = [A]^{-1}\vec{X} \equiv [W]\vec{X}. \quad (11a)$$

$$\left[\mathrm{W_{i,j}}\right]^{\mathrm{k+1}} = \left[\mathrm{W_{i,j}}\right]^{\mathrm{k}} - \frac{1}{\mathrm{C^k}}\frac{<\vec{\mu}_i^k\vec{X}_j>}{<[\vec{X}\vec{X}^T]>} \quad (11b)$$

PROOF:

$$H^{(2)} = E_o + \frac{\partial E}{\partial \boldsymbol{s}_\alpha^{(o)}}\left(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}\right) + \frac{1}{2}\frac{\partial^2 E}{\partial \boldsymbol{s}_\alpha^{(o)}\partial \boldsymbol{s}_\beta^{(o)}}\left(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}\right)\left(s_\beta - s_\beta^{(o)}\right) + T_o\,K_B\sum_{i=1}^{k}\boldsymbol{s}_i\,\log \boldsymbol{s}_i + (\mu_o - T_o\,K_B)\left(\sum_{i=1}^{k}\boldsymbol{s}_i - 1\right).$$

We may consistently vary $H^{(2)}$ w.r.t. Artificial Neural Network learning weight matrix:

$$\frac{\delta H^{(2)}}{\delta \mathrm{W_{j,i}}} = \mu_j X_i + C^k\{\left[W_{j,\alpha}\right]X_\alpha - s_j\}\mathrm{X_i} = 0$$

We assume the Ergodic hypothesis that the temporal average of high frequency turbulent fluctuations is equivalent to the equilibrium ensemble average with the nearest neighbor 3x3 average, denoted with the angular brackets to make sure the outer product of pixel spectral vector to be full rank for non-singular inverse $<\left[\vec{X}\vec{X}^T\right]>^{-1}$.This was consistent with our image resolution assumption, a 3x3 neighborhood resolvable as a single space-invariant macro-pixel.

$$<\vec{\mu}\vec{X}^T> + C^k[W] <\vec{X}\vec{X}^T> - C^k <\vec{s}\vec{X}^T> = 0;$$

$$[W]^{k+1} = \left\{<\vec{s}\vec{X}^T> - \frac{1}{C^k}<\vec{\mu}^k\vec{X}>\right\} <\left[\vec{X}\vec{X}^T\right]>^{-1}, \quad (12)$$

Use is further made of the definition of weight matrix: $\vec{s} = [W]\vec{X}$, and we can simply the first term of Eq.(12) and derived the learning rule Eq.(11b). Q.E.D.

2.2 Remarks on Some Fundamental Concepts

Hebbian Product Rule: Donald Hebb discovered that the neuro-biological synaptic junction learning rule is similar to a pipeline flow, that is proportional to how much goes in and how much comes out. The Hebbian product learning rule:

$$W_{i,j} \propto \vec{X}_i \vec{\mu}_j.$$

We demanded a proper normalization, i.e., $\frac{<\vec{\mu}_i^k \vec{X}_j>}{<[\vec{X}\vec{X}^T]>} \rightarrow 1$, if $\vec{\mu}_i^k \sim \vec{X}_i$.

What is the thermodynamic learning rule? It's systematic way to guess the most probable inverse source solution by directly computing the maximum probability. By systematic trial and errors, we can de-mix the local mixtures by the MFE principle. There is a finite number of ways that the positive sum of a photon counts can be made. Among them, we choose the lowest energy cases, e.g., giving Beethoven first 3 notes " 5, 5, 1…." , we split the sum 5 = (0+5; 1+4; 2+3; 3+2; 4+1; 5+0) in the unit of energy at temperature K_BT=1/40eV for T=300°; and find hidden source tones 2+3 and 3+2 occurring twice that have the highest canonical probability 2 $exp(-2/K_BT)exp(-3/K_BT)$. In MFE, we might wish to rule out the rare *high energy* cases (0+5 and 1+4) in favor with lower energy, but *higher chances* in equilibrium (twice 2+3) unless other summations involve also these specific pixels. Given a set of vector measurements of multiple spectral bands, we applied the thermodynamic equilibrium theory to find hidden object sources at MFE by LCNN.

Helmholtz MFE: Helmholtz assumed such an open dynamic sub-system within the heat reservoir closed system where Boltzmann heat death at maximum entropy was assumed.

$$\Delta H_{object} \equiv \Delta E_{object} - T_0 \Delta S_{object} \leq 0;$$
$$\min.\ H \equiv E - T_o S \qquad (13)$$

PROOF:
Let S_{Total} denote the total entropy of a closed system. Then S_{Total} is the sum of entropy of reservoir and object,

$$S_{Total} = S_{Reservoir} + S_{object}.$$

If the object takes ΔE_{object} energy from its surroundings, the entropy change of $S_{Reservoir}$ will be $\Delta S_{Reservoir} = -\Delta E_{object}/T_0$, and the total entropy change is

$$\begin{aligned} \Delta S_{Total} &= \Delta S_{Reservoir} + \Delta S_{object} \\ &= -\frac{\Delta E_{object}}{T_0} + \Delta S_{object} \\ &= -\frac{\Delta E_{object} - T_0 \Delta S_{object}}{T_0} = -\frac{\Delta H_{object}}{T_0} \end{aligned} \qquad (14)$$

where $\Delta H_{object} \equiv \Delta E_{object} - T_0 \Delta S_{object}$ is the change of the object's Helmholtz free energy, which is an analytic state function defined by $H = E - T_o S$. Note that $\Delta S_{Total} > 0$ since the total entropy of a closed system is always increasing, and $\Delta H_{object} \leq 0$ given a positive T_0. Q.E.D.

Boltzmann Entropy: Ludwig Boltzmann inscribed on his tomb headstone the entropy formula (cf. a picture of his Math Genealogy)

$$S = K_B \, Log \, W; \quad `$$

where W is the aforementioned phase space trajectory volume that represents all possibility which an identical macroscopic system can be prepared and realized; and K_B is the Boltzmann constant, i.e., 0.1 *meV*. To be explicit for remote sensing, we considered 3 kinds of identical entities; R denotes the number of red balls/molecules/photons, likewise G & B in a closed system. Thus, the chance of realizing total N balls is $N!$ divided by identical colors $R!\,G!\,B!$, because of the over-counting of permutations with identical particles.

$$W = \frac{N!}{R!G!B!}.$$

Sterling approximation of logarithmic factorial was valid when $N > 10$.

$$Log\ N! \cong N\ Log\ N - N;$$
$$\frac{N}{N} = \frac{R}{N} + \frac{G}{N} + \frac{B}{N};$$
$$1 = S_1 + S_2 + S_3.$$

Then, Boltzmann discrete entropy formula follows:

$$S = -K_B \sum_{i=1}^{K} S_i Log\ S_i - \text{const.}\,(\sum_{i=1}^{K} S_i - 1);$$
$$const. = \frac{\mu_o}{T_o} - K_B$$

where the minus sign was derived due to $Log\ S_i \leq 0$ for $S_i \leq 1$ and the Lagrange scalar constraint of the probability norm $\sum_{i=1}^{K} S_i - 1 = 0$ was chosen to be $\left(\frac{\mu_o}{T_o} - K_B\right)$ that insured the normalization and a simple slope:

$$\begin{aligned} \frac{\partial S}{\partial S_i} &= -K_B \left(1 + \sum_{i=1}^{K} \log S_i\right) - \left(\frac{\mu_o}{T_o} - K_B\right) \\ &= -K_B \log S_i - \frac{\mu_o}{T_o} \end{aligned}$$ Q.E.D.

The maximum entropy in a closed system corresponds to the minimum free energy in open sub-systems.

Grey-Body Planck Law: Planck's law was the triumph of modern quantum physics during 1900~1919. Max Planck received the Nobel Prize in Physics in 1918. A remarkable result which we exploited theoretically in this paper was that the spectral irradiance $I_\nu(T_K)$, leaking out of a small hole of the black body cavity's opaque walls kept at a constant temperature, peaked at a single wavelength monotonically $\lambda_o = c_o/\nu_o$ and that uniquely determines the associated Kelvin temperature T_K once and only once in Fig.2. We were not the first one either. Astronomers applied the apparent measured brightness temperature T_B related

by the unique peak spectrum value to an equivalent black body Kelvin temperature T_K. It turned out roughly $T_B \sim \varepsilon T_K$. Cosmic Background Radiation is not a blackbody for a large expanding universe having no outside, yet the estimation of an approximated reflectivity $\gamma \sim 0.1$ suggests by the conservation of energy, the equivalent emissivity is about $\varepsilon \sim 0.9$. Nevertheless, the universe was cooled down after the Big Bang happened at 13B years; 380K years ago, the universe reached 3000K~0.25 eV, which is the time of hydrogen atoms were formed with 13.6 eV ionization energy (5% light mater, 27% dark mater, 68% dark energy). Thus, the background light did not have enough energy to become de-coupled from the hydrogen matter, and the universe became transparent. As a result, the Cosmic Microwave Background Radiation (CMBR) can be observed and may be called the "time of last (inelastic) scattering." The decoupled photons from matter are continuously cooled down 1000 times at now 2.7 ^oK ~0.23meV corresponding to microwave range frequency of 160.2 GHz(1.9 mm wavelength). Robert Dicke, George Gamow, Ralph Alpher, and Robert Herman conjectured that CMBR was the inflationary Big Bang theory. 1978 Nobel Laureates Arno Penzias and David Wilson applied Dicke radiometry of 15 meter horn antenna to measure the peak at $T_B \sim 2.73^oK \sim 0.9x3^0K$. Subsequently, NASA's Cosmic Background Explorer (COBE) satellite using differential microwave instruments confirmed an anisotropic CMBR (George Smoot & John Mather, Nobel Prize 2006).

Quantum Statistics: G.E. Uhlenbeck & S. Goudsmidt discovered in 1910 in the Stern & Gerlach experiment an electron beam split under an inhomogeneous magnetic field into 2 beans: spin up or spin down: that the spin fine structure constant $2s+1=2$ that implies the e-spin quantum number s=1/2. It suggested the e-wave function with the phase factor $e^{i(n-1)\pi} = (-1)^{n-1}$ generates alternation signs in the Fermi-Dirac distribution function

$$z - z^2 + z^3 - z^4 + \cdots = z(1 - z + z^2 - z^3 + \cdots)$$
$$= \frac{Z}{1+Z} = \frac{1}{Z^{-1}+1} = \frac{1}{\exp\left(+\frac{E}{K_B T}\right)+1} \leq 1$$

for odd integer electron spin Fermion the Pauli's exclusion principle, one pigeon per hole, and led to a finite Fermi surfaces as the Band gap phenomena in the semiconductors. Bose-Einstein condensation of integer spin Bosons is due to the friendship principle: the condensed ground state becomes divergent $1/[\exp(+E/K_B T) - 1] \to \infty$, where $E/K_B T \to 0$. BCS theory of superconductor of electron spin ½ was due to the lattice vibration of bounding two Fermions together called Cooper pairs, electrons or positrons which become a spin-1 Boson. Paul Chu et al. discovered higher temperature superconductor made of ceramic $Y_1 Ba_2 Cu_3 O_x$ material whose lattice defects, positron holes, enjoyed a larger internal pressure due to a larger replaced Ba_2 molecule, which bounded two Fermions together, by phonon exchange energy, into a spin-1 Boson and sustained a disruptive thermal noise $\gtrsim 77^oK$.

Vacuum Fluctuation: Paul Ehrenfest wrote the corresponding principle between classical mechanics and quantum mechanics. The Poison Bracket is related to Heisenberg uncertainty principle commutator between position $\hat{P}$ and momentum $\hat{Q}$ operators. It represented the effects of two different sequences of measurements. It helped us quantized the Hamiltonian $\hat{H}$ of simple harmonic oscillator in the quantum field theory in terms of 2nd quantization operators $\hat{a}^\dagger$: $\hat{H} = \hat{P}^2 + \hat{Q}^2 = \left(\hat{P} + i\hat{Q}\right)\left(\hat{P} - i\hat{Q}\right) = \hat{a}^\dagger \hat{a}$ that computes the vacuum fluctuation due to the commutator uncertainty principle generating the non-zero vacuum energy $\hat{H}\left|0>=\frac{1}{2}h\nu\right|0>$. This zero-point vacuum fluctuation existed everywhere helped Higgs, following the Anderson phase transition model: $\hat{\phi}^4 \approx (\hat{\phi}-\phi_+)^2(\hat{\phi}-\phi_-)^2$ (having a lower potential well at symmetric ground state $|\phi_+>$ associated with a non-zero order parameter) condensing the energy into the mass $m = E/c_o^2$. CERN experiments seemed to have verified the Higgs boson phase transition mechanism.

3. Nonlinearly Regularized Lagrange Constraint Neural Network (LCNN)

The Mexican standoff will be regularized by **Karush, Kuhn-Tucker (KKT) 2nd order** penalty by steepening an isotropic sphere. CRT&D CS assumed that the Lagrange slope was the estimation error itself, and no longer is an unknown. However, when the cost function is unknown, we have demonstrated the double iteration at the linear order is at 2nd order smallness, and therefore cannot consistently determined at the 1st order LCNN. We identified the penalty as the 2nd order Taylor series of MFE. This generates a linearly decoupled closed set of 3 equations for solving sources from spectral vector data, as a fast LANCELOT algorithm, called regulated LCNN (Szu, Miao, Qi, SPIE 2007). Given input s-spectral vector data $\boldsymbol{X}_s$ per pixel we give double iteration superscript index $k = I_+ \equiv \{1,2,3,etc.\}$.

Theorem 3: NL Regularized Lagrange Constraint Neural Network (LCNN) is a fast LANCELOT algorithm of nonlinear optimization. Given lemma on ANN learning matrix $\left[W_{j,i}\right]^k$ and the j-component $\boldsymbol{c_j}$ of the EPS sources estimation error vector together, we determine the slope j-component $\boldsymbol{\mu}_j^k$ of the Lagrange error energy by iteration as;

Hebbian rule:

$$\left[\mathrm{W}_{i,j}\right]^{\mathrm{k}+1} = \left[\mathrm{W}_{i,j}\right]^{\mathrm{k}} - \frac{1}{\mathrm{C}^{\mathrm{k}}} \frac{<\vec{\mu}_i^{\mathrm{k}} \vec{\mathrm{X}}_j>}{<[\vec{\mathrm{X}}\vec{\mathrm{X}}^{\mathrm{T}}]>}; \quad (15)$$

Lagrange error slope rule:

$$\vec{\boldsymbol{\mu}}_j^{k+1} = \vec{\boldsymbol{\mu}}_j^k + C^k\{\left[W_{j,\alpha}^{k+1}\right]\vec{\boldsymbol{X}}_\alpha - \vec{\boldsymbol{S}}_j^{k+1}\} \; ; \quad (16)$$

Unknown object sources:

$$T_o \, K_B \log \vec{\mathrm{S}}_j^{\,\mathrm{k}+1} + C^k \vec{\mathrm{S}}_j^{\,\mathrm{k}+1} = C^k \left[\mathrm{W}_{j,\alpha}{}^{\mathrm{k}}\right]\vec{\mathbf{X}}_\alpha + \vec{\boldsymbol{\mu}}_j^{\,\mathrm{k}} - \mu_0{}^{\mathrm{k}}; \quad (17)$$

Curvature Penalty:

$$C^k = \beta_o C^{k-1};\ \beta_0 > 0; k = 1,2,3, etc. \quad (18)$$

PROOF:

The tradeoff between minimum energy and maximum entropy for the most probable configuration requires the 1st and 2nd order Taylor series expansions:

$$H^{(1)} = E_o + \frac{\partial E}{\partial \boldsymbol{s}_\alpha^{(o)}}\left(s_\alpha - \boldsymbol{s}_\alpha^{(o)}\right) + T_o\, K_B \sum_{i=1}^{k} \boldsymbol{s}_i \log \boldsymbol{s}_i + (\mu_o - T_o\, K_B)\left(\sum_{i=1}^{k} \boldsymbol{s}_i - 1\right) \quad (19a)$$

$$H^{(2)} = E_o + \frac{\partial E}{\partial \boldsymbol{s}_\alpha^{(o)}}\left(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}\right) + \frac{1}{2}\frac{\partial^2 E}{\partial \boldsymbol{s}_\alpha^{(o)} \partial \boldsymbol{s}_\beta^{(o)}}\left(\boldsymbol{s}_\alpha - \boldsymbol{s}_\alpha^{(o)}\right)\left(s_\beta - s_\beta^{(o)}\right) + T_o\, K_B \sum_{i=1}^{k} \boldsymbol{s}_i \log \boldsymbol{s}_i + (\mu_o - T_o\, K_B)\left(\sum_{i=1}^{k} \boldsymbol{s}_i - 1\right) \quad (19b)$$

We take variation calculus to set flattening extremes to be zero:

$$\delta E^{(1)} = \frac{\delta E^{(1)}}{\delta \boldsymbol{s}_\beta{}^{(o)}} \boldsymbol{c}_\beta\left(\vec{\boldsymbol{s}}^{(o)}\right) \equiv \boldsymbol{\mu}_\beta\{\left[W_{\beta,\alpha}\right]\boldsymbol{X}_\alpha - \boldsymbol{s}_\beta{}^{(o)}\} \quad (19c)$$

$$\delta H^{(1)} = \frac{\delta H^{(1)}}{\delta \boldsymbol{s}_j} = -\boldsymbol{\mu}_\beta \frac{\delta \boldsymbol{c}_\beta}{\delta \boldsymbol{s}_j} + T_o\, K_B \log \boldsymbol{s}_j + \mu_o \quad (19d)$$

$$\delta H^{(2)} = \frac{\delta H^{(2)}}{\delta s_j} = -(\mu_\beta + C^k\{\mathrm{W}_{\beta,\alpha}\, X_\alpha - s_\beta\}\frac{\delta \boldsymbol{c}_\beta}{\delta s_j} + T_o\, K_B \log s_j + \mu_o \quad (19e)$$

where use is made of isotropic curvature in an increasing constant penalty term for k+1 iteration: $C^{k+1} = \beta_o C^k$; $\beta_o \approx 4$, and the simple component selection: $\frac{\delta c_i}{\delta \boldsymbol{s}_j} = -\delta_{\mathrm{i,j}}$.

From the variation of 2nd order (19e) :

$$\delta \mathrm{H}^{(2)} = 0:$$

$$\mathrm{T_o\, K_B} \log \mathbf{s}_\mathrm{j} + \mathrm{C^k}\mathbf{s}_\mathrm{j} = \mathrm{C^k}\left[\mathrm{W_{j,\alpha}}\right]\mathbf{X}_\alpha + \boldsymbol{\mu}_\mathrm{j} - \mu_0 \quad (20a)$$

From the variation of 1st order (19d):

$$\delta \mathrm{H}^{(1)} = 0:$$

$$\boldsymbol{\mu}_\mathrm{j} + \mathrm{T_o\, K_B} \log \mathbf{s}_\mathrm{j} + \mu_\mathrm{o} = 0 \quad (20b)$$

We obtain the consistency condition, after cancelling the entropy related common terms: $\mathrm{T_o\, K_B} \log \mathbf{s}_\mathrm{j} + \mu_\mathrm{o}$, the next iteration of the 1st order Lagrange error energy slope $\boldsymbol{\mu}_\mathrm{j}^{\mathrm{k+1}}$ is determined by the 2nd order variation of MFE:

$$\boldsymbol{\mu}_\mathrm{j}^{\mathrm{k+1}} = \boldsymbol{\mu}_\mathrm{j}{}^{\mathrm{k}} + \mathrm{C^k}\mathbf{c}_\mathrm{j}^{\mathrm{k+1}};\ \mathbf{c}_\mathrm{j}^{\mathrm{k+1}} = \left[\mathrm{W_{j,\alpha}^{k+1}}\right]\mathbf{X}_\alpha - \mathbf{s}_\mathrm{j}^{\mathrm{k+1}} \quad (20c)$$

This linear decoupled Lagrange error slope equation is historically called LANCELOT in FORTRAN massive database optimization. The blind sources estimation error provided the next gradient descent of Lagrange energy error slope vector $\boldsymbol{\mu}_\mathrm{j}{}^{\mathrm{k}}$. De-mixing weight matrix Eq.(15) was proved in Theorem 2. Q.E.D.

4. Conclusion

This thermodynamics learning rule may be a paradigm shift for dealing with spectral image processing with thermodynamics. Various applications have been developed and reported in different journals. It might allow us to consider virtually crossing the full electromagnetic spectrum. Compressive modeling and simulation based on NL LCNN will be published in Optical Engineering (Krapels, Cha, Espinola, Szu). IR triplets for seeing through hot fire and cold dust will be published in IEEE Tran IT (Cha, Abbott, Szu). Thermodynamics physics laws and modern applications will be published in Journal of Modern Physics (Szu, Willey, Cha, Espinola, Krapels). Lots more can happen with your participation in Appendix A and Appendix B. MATLAB pseudo source code is given with benchmarked results.

References

[1].E. J. Candes, J. Romberg, and T. Tao, "Robust Uncertainty Principle: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Trans IT*, 52(2), pp.489-509, Feb. 2006.

[2].E. J. Candes, and T. Tao, "Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies," *IEEE Trans IT* 52(12), pp.5406-5425, 2006.

[3].D. Donoho, "Compressive Sensing," *IEEE Trans IT*, 52(4), pp. 1289-1306, 2006.

[4].H. Szu, C. Hsu, G.Moon, T. Yamakawa, B.Q. Tran, T.P. Jung, J. Landa, "Smartphone household wireless Electroencephalogram hat," *Applied Computational Intelligence and Soft Computing*, 2012 (Hindawi Internet Publ.).

[5].H. Szu, L. Miao, & H. Qi, "Unsupervised learning with minimum free energy," Proc. of SPIE, Vol. 6576, IN: *Independent Component Analyses, Wavelets, Unsupervised Nano-Biomimetic Sensors, and Neural Networks V* (edited by Harold H. Szu, Jack Agee), pp. 657605, 2007.

[6].H. Szu & C. Hsu, "Landsat spectral unmixing a la super-resolution of blind matrix inversion by constraint maxent neural nets," *Proc. of SPIE*, Vol. 3078, pp. 147–160, 1997.

[7].US PTO Patent, Harold Szu, 7355182 "Infrared multi-spectral camera and process of using infrared multi-spectral camera" April 8, 2008; US PTO 7366564, Harold H. Szu et al; "Nonlinear blind demixing of single pixel underlying radiation sources and digital spectrum local thermometer," April 29, 2008.

[8].H. Szu, C. Hsu, J. Jenkins, J. Willey, J. Landa, "Capture Significant Events with Neural Networks," *Neural Networks* 29-30(2012) pp. 1-7.

[9].R. Appleby, P. Coward, J. N. Sanders-Reed, "Evaluation of a Passive Millimeter Wave Imaging for wide detection in degraded visual condition," *SPIE Proc. 7309*, April 2009; P. Coward & R. Appleby, "Comparison of passive millimeter-wave and IR imagery in a nautical environment," *SPIE Proc. 7309*," pp. 04-1, 04-8. 2009.

[10].J.H. Cha, A. L. Abbott, and H.H. Szu, "Passive Ranging Redundancy Reduction in Diurnal Weather Conditions," *SPIE Proc. 8750*, "ICA, Compressive Sampling, Wavelets etc." (ed. Szu & Dai) May 1~3, 2013 (Baltimore).

[11]. C. Hsu, M.K. Hsu, J. Cha, T. Iwamura, J. Landa, C. Nguyen, H.Szu," Adaptive Compressive Sensing Camera, *SPIE Proc. 8750*, "ICA, Compressive Sampling, Wavelets etc." (ed. Szu & Dai) May 1~3, 2013 (Baltimore).

[11] S. Vahidinia, "Regolith Radiative Transfer Model; Applications to Saturn's Icy Rings," Ph.D. 2010, UC Santa Cruz.

[12]. J. F. Mustardm, J. E. Hays, " Effects of hyperfine particles on reflectance spectra from 0.3 to 25 microns," *ICARUS 125*, pp, 145-163 (1997).

[13] V. Payan and A. Royer, "Analysis of temperature emissivity separation (TES) algorithm applicability and sensitivity," *Int. J. Remote Sensing V.*25, pp. 15-27, Jan 2004.

[14] P. Vernazza, B. Carry, J. Emery, J.L. Hora, D. Cruikshank, R.P. Binzel, J. Jackson, J. Helbert, A. Maturilli, "Mid-IR spectral variability for compositionally similar asteroids; implications of asteroid particle size distribution," *ICARUS 207*, pp. 800-809 (2010).

[15] B. Hapke, "Bi-direction reflectance spectroscopy I. Theory," *Geophysical Rev*. Vol. 86, pp. 3039-3054,1981.

[16] H. Ren, H. Szu, and J. Buss," Lagrange Constraint Neural Network for Fully constrained Subpixel Classification in Hyperspectral imagery," *Proc. SPIE 4738*, pp. 184-190, 2002.

[17] I. Kopriva, and H. Szu, "Fast LCNN ica for Unsupervised Hyperspectral Image Classifier," *Proc. SPIE 4738*, pp. 169-183, 2002.

[18] H. Ren, and C.I. Chang," target-constrained interference-minimized approach to subpixel target detection for hyperspectral images," *Opt. Eng.* V.39, pp. 3138-3145, 2000.

[19] C.I. Chang and D.C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sesnign*, Vol. 38, pp. 1144-1158, 2000.

Appendix A: BSS by Engineering Filter Approach: Pixel Parallelism (at Maximum Output Entropy)

Bell, Sejnowski, Amari & Oja (BSAO) have systematically formulated an *unsupervised learning of ANN algorithm for unknown but identical for space-invariant mixing* by varying the unknown de-mixing weight matrix $[W_{i.j}]$ until nothing but the Max Entropy $S(y_i)$ of the output $y_i = [W_{i,\alpha}]x_\alpha$, where $x_j = [A_{j,\alpha}]s_\alpha$ with measured x_j and unknown $[A_{j,\alpha}]$ and s_α. Here, the repeated Greek indices represent the summation. ANN model used a monotonically sigmoid-squashed threshold output

$$y_i = \sigma(x_i) \equiv \{1 + \exp(-[W_{i,\alpha}]x_\alpha)\}^{-1}.$$

that is nonlinear analytic solution of Riccati equation $\frac{dy}{dx} = y(y-1)$, for asymptotically binary logic $y = 0\ or\ 1$ for no or yes. Since a single neuron learning rule turns out to be massively parallel to N neurons in tensor index notion, for simplicity, we derived for a single neuron to point out why the engineering filter does not follow Hebb's synaptic weight updates. A bona fide unsupervised learning did not have a desirable specific output entropy S(y) became maximized, the de-mixing filtering $[W_{i.j}]$ becoming the inverse of unknown mixing matrix $[A_{j,\alpha}]$. Thus, the filter weight adjustment is defined as:

$$\frac{\delta w}{\delta t} = \frac{\partial S(y)}{\partial w}, \quad S(y) = -\int f(y)\log f(y)dy$$

$$\Longrightarrow \delta w = \frac{\partial H(y)}{\partial w}\delta t = \{|w|^{-1} + (1-2y)x\}\delta t.$$

Derivation: From the normalized probability definitions:

$$\int f(y)dy = \int g(x)dx = 1\ ; f(y) = \frac{g(x)}{\left|\frac{dy}{dx}\right|};$$

$$H(y) \equiv -<\log f(y)>_f\ ,$$

we expressed the output pdf in terms of the input pdf with changing Jacobian variables. We exchanged the orders of operation of the ensemble average brackets and the derivatives to compute

$$\frac{\partial H(y)}{\partial w} = \frac{\partial -<\log\left|\frac{dy}{dx}\right|>_f}{\partial w} \cong |\frac{dy}{dx}|^{-1}\frac{\partial\left|\frac{dy}{dx}\right|}{\partial w};$$

Since Riccati equation was satisfied by the sigmoid: $y = [1+\exp(-wx)]^{-1}$; $\frac{dy}{d(wx)} = y(1-y)$, we readily derived the following results by chain rule

$$\frac{dy}{dx} = wy(1-y); \frac{dy}{dw} = xy(1-y).$$

Substituting these results into Max Ent learning rule, one obtains the Bell-Sejnowski equation

$$\frac{\partial H(y)}{\partial w} = [W]^{-1} - (2y-1). \qquad \text{Q.E.D.}$$

The first term computing the inverse matrix $|w|^{-1}$ is not scalable with increasing *N* nodes, while the second term satisfied the Hebbian product rule between bipolar output 2*y*-1 and input *x*. S. Amari et al. at RIKEN assumed the identity $[\delta_{i,k}] = [W_{i,j}][W_{j,k}]^{-1}$ and multiplied the identity through both sides of original non-biological algorithm

$$\frac{dH}{dW_{i,j}}[\delta_{i,k}] = \{[\delta_{i,j}] - (2\vec{y}-1)\vec{y}^T]\}[W_{i,j}]^{-1},$$

where use was made of $y_i = [W_{i,\alpha}]x_\alpha$ to change the input $x_j$to the synaptic gap by its weighted output y_i. Amari et al. derived a natural gradient ascend as the final BSAO algorithm in information geometry,

$$\frac{dH}{dW_{i,j}}[W_{i,j}] = \{[\delta_{i,j}] - (2\vec{y}-1)\vec{y}^T]\},$$

which was not in the original gradient direction $dH/dW_{i,j}$ and enjoyed a faster update without the inverse .

Fast ICA: Erkki Oja began his ANN learning of nonlinear PCA for pattern recognition in his Ph.D. study 1982.

$$<\vec{x}\vec{x}^T>\hat{e} = \lambda\hat{e};$$

$$w' - w = \vec{x}\sigma(\vec{x}^T\vec{w}) \cong <\vec{x}\vec{x}^T>\vec{w};$$

$$\frac{d\vec{w}}{dt} = <\vec{x}\vec{x}^T>\vec{w} \cong \sigma(\vec{x}^T\vec{w})\vec{x} \cong \frac{dK(u_i)}{du_i}\frac{\mathrm{d}u_i}{\mathrm{d}w_i} \equiv \mathrm{k}(\vec{x}^T\vec{w})\vec{x};$$

where Oja changed the unary logic to bipolar hyperbolic tangent logic as $v_i = \sigma(u_i) \approx u_i - \frac{2}{3}u_i^3 \cong \frac{dK(u_i)}{du_i}$; $u_i = w_{i,\alpha}x_\alpha$.

By derivation we obtain therefore the **BSAO** unsupervised learning collectively in a termination condition: It becomes similar to a Kurtosis slope, which suggested to Oja a new contrast function K. The following is the geometric basis of a stopping criterion of unsupervised learning. Taylor expansion of the normalization, and set $|\vec{w}|^2 = 1$:

$$|\vec{w}'|^{-1} = [(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x}))^T(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x}))]^{-\frac{1}{2}}$$

$$= 1 - \frac{\epsilon}{2}\, k(\vec{w}^T\vec{x})(\vec{x}^T\vec{w} + \vec{w}^T\vec{x}) + O(\epsilon^2).$$

$$\vec{w}'' \equiv \vec{w}'|\vec{w}'|^{-1}$$

$$= \left(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x})\right)\left(1 - \frac{\epsilon}{2}\, k(\vec{w}^T\vec{x})(\vec{x}^T\vec{w} + \vec{w}^T\vec{x})\right)$$

$$\Delta\vec{w}''=\vec{w}'' - \vec{w} = \epsilon\left[\delta_{\alpha,\beta} - w''_{\alpha}w''^{T}_{\beta}\right]\vec{x}_{\alpha}\frac{dK(u_\beta)}{d\vec{w}''_i}.$$

The ICA algorithm lets the joint probability density function be factorized equally according to Max Entropy requirement of white noise. Unfortunately, such a filter approach cannot work for the remote sensing imageries, because the atmospheric turbulence changes spatially rather quickly due to a large pixel footprint which is about a tenth of a squared kilometers (Landsat 30 km^2 per pixel). If one were insisting to apply ICA to mix together all spatial (x,y) ***s***-spectral vectors data $\vec{X}_s(x,y,t)$ in parallel, the assumption of spatial invariant mixing matrix will produce inaccurate sources at the **maximum entropy (Max Ent)**. Spatial invariant assumption will not work for a close-up dual infrared spectral band image for screening for cancer, because of the localization of cancer has a strongly space-variant physiology with and without the cancer. Solving the inverse of space-variant impulse response Green's function or optical point spread function (psf) we need new MFE approach to BSS. To improve it, we have explored space invariant imaging (in terms of 3x3 macro-pixels) for macro-pixel joint density factorization ICA method; (Du, Kopriva, Szu, by Non-negative Matrix Factorization IEEE 2005;& by JADE & Fast ICA Op Eng. 2006). Then, the comparison experience help us reformulated in 2007 a nonlinear LCNN, ala KKT penalty, to regularize the linear LCNN for assuming space-invariant result within 3x3 nearest-neighbor macro-pixels per tumor cluster cells, but space-varying among macro-pixels with or without tumor [4]. In remote sensing, the Mexican standoff product challenge was overcome by the 2nd order energy expansion curvature known as KKT nonlinear optimization penalty. Typical 80 spectral band images among 158 Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Kopriva, Szu, Fast ICA, SPIE ICA etc.2002) and corresponding BSS out of applying LCNN to 158 channels data sources maps where the color blue means no class, i.e. low probability while red means high probability.

Appendix B: BSS by Physics Source Approach: Pixel sequentially (Min. Free-Helmholtz Energy (MFE) Nonlinearlly regularized LCNN Matlab Code)

Pseudo-code of LCNN for 3-source model

Given multispectral data vector **x** = $(x_1\ x_2\ x_3)^T$ at a single pixel,

Initialize $\mu^0 = [0\ 0\ 0]^T$, $C^0 = 1$, $\beta_o = 4$, and W= A^{-1}.

Set iteration index $k = 1$, and maximum iteration number ITM

While $k \le ITM$

Calculate sources s^{k+1} subject to $0 \le s \le 1$ by solving Eq.(20a):

$$T_0K_B \ln(s^{k+1}) + C^k s^{k+1} = C^k W^k x + \mu^k - \mu^k{}_0$$

If s^{k+1} converges, return. Otherwise continue;

Update de-mixing matrix Eq.(15):

$$W^{k+1} = W^k - (1/C^k)\langle \mu^k x^T\rangle / \langle xx^T\rangle$$

Update Eq.(20c):

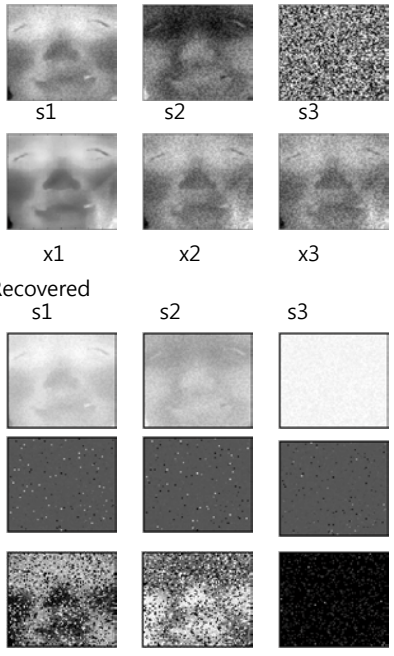
$$\mu^{k+1} = \mu^k + C^k(W^{k+1}x - s^{k+1})$$

Update $C^{k+1} = \beta_o C^k$

$k \leftarrow k + 1$

end

From multiple LWIR bands images we can find EPS from the ground truth data generated from original human face radiology map in the RHS assuming 2 temperature sources per pixel at 29% and 70%, and the third sensor coolant source 1% distributed spatial randomly. Then compute the Planck curve RHS projected on 3 spectral b*ands in the mean* value 4, 9, 11 microns.



Sources
s1 temp = 310 K
s2 temp < 310K
s3 temp = 77 K

Measurements
x1 at 4 um
x2 at 9 um

When the initial W is perfectly known

When the initial W is completely random

When the initial W is correlated with x1