11

Implementation of Dynamic Logic Algorithm for Detection of EM Fields Scattered by Langmuir Soliton

V.I. Sotnikov, R.W. Deming and L. Perlovsky
Air Force Research Laboratory, Sensors Directorate, Hanscom AFB, MA, USA

1. Introduction

In plasma diagnostics, scattering of electromagnetic waves is widely used for identification of density and wave field perturbations. Very often difficulties with implementations of this type of diagnostics are connected with the fact that scattered signals have small amplitudes comparable with the noise level. Under such conditions it is very important to implement the algorithm which allows us to identify and separate the scattered wave spectrum from the noise pattern. In the present work we use a powerful mathematical approach, dynamic logic (DL) [1], to identify the spectra of electromagnetic (EM) waves scattered by a Langmuir soliton – a one-dimensional localized wave structure. Such wave structures are understood to be basic elements of strong plasma turbulence.

We are interested in the implementation of the Dynamic Logic algorithm for diagnostics of electromagnetic waves scattered by Langmuir soliton in the presence of random noise with amplitudes comparable with the amplitudes of the scattered waves. Scattering of EM waves on Langmuir soliton was previously analyzed by many authors (see for example Sinitzin, 1978, 1979; Mendoca, 1983).

For our purposes we will use the results for electromagnetic wave (EM) scattering on Langmuir soliton obtained in Mendonca, 1983, where scattering on low frequency part of density perturbations (density well) as well as scattering on high frequency part of density perturbations was examined and corresponding expressions for the amplitudes of the scattered waves were obtained.

2. EM scattering by Langmuir soliton

Langmuir soliton is a solution to the well known Zakharov equations (Zakharov, 1972):

$$2i\frac{\partial E}{\partial t}+3\omega_{pe}r_{De}^2\frac{\partial^2 E}{\partial x^2}-\omega_{pe}\frac{\delta n}{n_0}E \tag{1}$$

$$\frac{\partial^2 \delta n}{\partial t^2}-c_s^2\frac{\partial^2 \delta n}{\partial x^2}=\frac{1}{32\pi M}\frac{\partial^2 |E|^2}{\partial x^2} \tag{2}$$

Here n_0 is the plasma density, δn the density of perturbation, $c_s = \sqrt{\frac{T_e}{M}}$ the ion-sound velocity (T_e is electron temperature and M is ion mass), E the complex amplitude of the Langmuir waves.

Fundamental object of the theory of strong Langmuir turbulence, Langmuir soliton (Rudakov, 1972), can be obtained from the equations (1)-(2) when we are looking for traveling wave solutions in the form:

$$\tilde{\mathbf{E}}_{LS}(\mathbf{r},t) = E_{LS}(x-ut)\mathbf{e}_x \exp[i(k_{LS}x - \delta\omega_{LS}t)] + c.c \tag{3}$$

with imposed boundary conditions $E(x \to \pm\infty) \to 0$ and with $\delta\omega_{LS} = \omega - \omega_{pe}$. In (3) $\mathbf{e}_x$ is the unit vector along the x-axis.

Below we will use a well known solution to (1)-(2) which is given by:

$$\tilde{\mathbf{E}}_{LS}(\mathbf{r},t) = E_{LS}\mathbf{e}_x \sec h(\frac{x-ut}{\Delta})\exp(ik_{LS}x - i\delta\omega_{LS}t) + c.c. \tag{4}$$

Where Δ is the soliton half-width, which can be expressed through the electrostatic energy density normalized to the plasma thermal energy density $W_{LS} = {|E_{LS}|^2}/{4\pi n_0 T_e}$:

$$\Delta = \frac{r_{De}}{\sqrt{3W_{LS}}} \tag{5}$$

Frequency $\delta\omega_{LS}$ and wave number k_{LS} are defined as:

$$\delta\omega_{LS} = \omega_{pe}[1 + \frac{1}{2}(W_{LS} - \frac{u^2}{3\mathrm{v}_{Te}^2})] \tag{6}$$

$$k_{LS} = \frac{u}{3\mathrm{v}_{Te}}\frac{1}{r_{De}} \tag{7}$$

where v_{Te} and r_{De} are electron thermal speed and Debye radius.

Electric field of an EM wave interacting with a Langmuir soliton can be written in the following form:

$$\tilde{\mathbf{E}}_0(\mathbf{r},\mathbf{t}) = E_0\mathbf{e}_p e^{i(\mathbf{k}_0\mathbf{r}-\omega_0 t)} + c.c. \tag{8}$$

where $\mathbf{e}_p$ is the polarization unit vector, ω_0 and $\mathbf{k}_0$ are the frequency and the wave vector of the EM wave. These are connected through the dispersion relation:

$$\omega_0^2 = k_0^2c^2 + \omega_{pe}^2 \tag{9}$$

In the process of scattering the EM wave will interact with electron density perturbations associated with a Langmuir soliton. These density oscillations consist of high and low frequency parts:

$$\delta n = \delta n_1 + \delta n_2 \tag{10}$$

The high frequency oscillations can be presented as:

$$\delta n_1 = i(\frac{2}{3}W_{LS})^{1/2}\frac{u}{\mathrm{v}_{te}}n_0 \sec h(\frac{x-ut}{\Delta})e^{i(k_{LS}x-\omega_{LS}t)} + c.c. \tag{11}$$

For the low frequency part (density well) we can use:

$$\delta n_2 = -W_{LS}n_0 \sec h^2(\frac{x-ut}{\Delta}) \tag{12}$$

It is well known that the ratio $\delta n_1 / \delta n_2$ is of the order of unity and we need to take into account contributions to the scattered EM fields from both high and low frequency parts of density perturbations. As a result far away from the soliton it is possible to obtain the scattered wave field in the following asymptotic form (Mendonca, 1983):

$$E_f(x) = 2\pi A(\omega)E_{LS}E_0 e^{iqx}\delta(\omega-\omega_{s1}) + 2\pi B(\omega)W_{LS}E_0 e^{iqx}\delta(\omega-\omega_{s2}) \tag{13}$$

In (13) $A(\omega)$ and $B(\omega)$ are given by:

$$A(\omega) = \pm\pi\Delta\frac{\omega}{\omega_0}\frac{\omega_{pe}^2}{c^2}\frac{u}{\mathrm{v}_{Te}}(\frac{\varepsilon_0}{3n_0T_e})^{1/2}(\mathbf{e}_0\cdot\mathbf{e}_{s1})\sec h(\frac{\pi}{2}\frac{\Omega_1}{u}\Delta) \tag{14}$$

$$B(\omega) = -\pi\Delta\frac{\omega}{\omega_0}\frac{\omega_{pe}^2}{c^2}(\mathbf{e}_0\cdot\mathbf{e}_{s1})co\sec h(\frac{\pi}{2}\frac{\Omega_2}{u}\Delta) \tag{15}$$

where Ω_1 and Ω_2 are defined as:

$$\Omega_1 = \omega - \omega_0 \pm \omega_{LS} \tag{16}$$

$$\Omega_2 = \omega - \omega_0$$

As it follows from (13)-(16) at large distances from the soliton it is possible to detect four scattered waves. Below for illustrative purposes we are presenting these four scattered waves which appear due to interaction of the incident EM wave, propagating at the 45^0 angle in x-y plane with the moving along the x - axis soliton.

The pattern of the total normalized electric field amplitudes of scattered EM waves given in (13) is presented in Fig. 2.

The Fourier spectrum of scattered EM waves is presented in Figure 3. Below we will use this spectrum in implementation of the Dynamic Logic algorithm.

We are interested in implementation of Dynamic Logic algorithm for diagnostics of electromagnetic waves scattered by Langmuir soliton in the presence of random noise with amplitudes comparable with the amplitudes of the scattered waves. Imposing the random

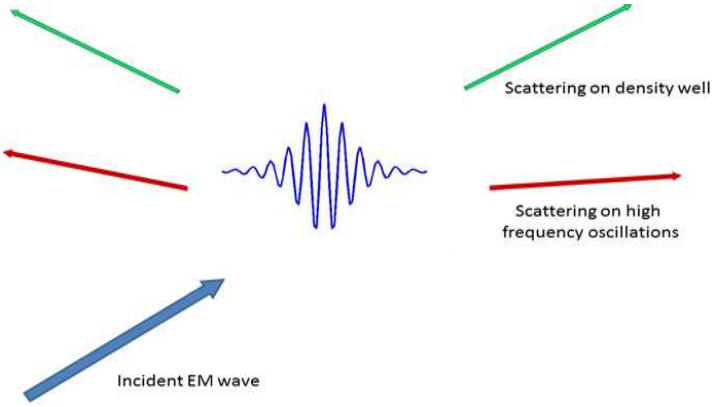


Fig. 1. Scattered EM waves produced due to interaction of the incident EM wave with Langmuir soliton.

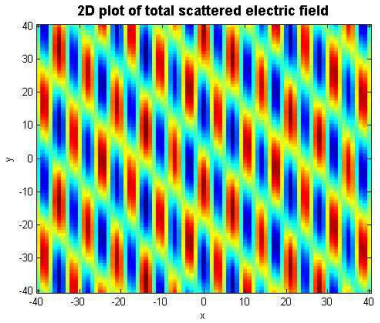


Fig. 2. The pattern of the total normalized electric field amplitudes of scattered EM waves.

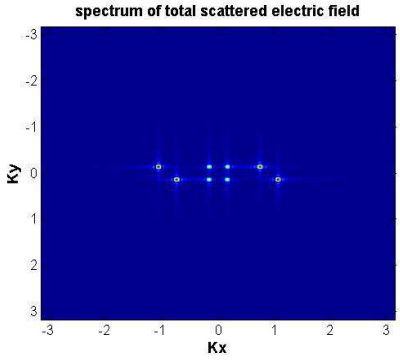


Fig. 3. The Fourier spectrum of scattered EM waves from (13).

noise pattern on top of the Fourier spectrum in Figure 3 with amplitudes comparable with that of the scattered EM waves we will obtain the following 2D pattern in the Fourier space:

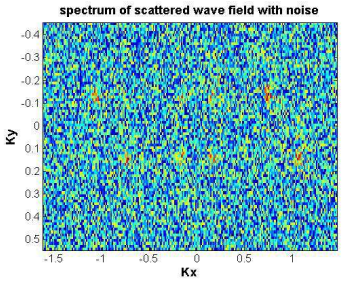


Fig. 4. Combined Fourier spectrum, consisting from the spectrum of scattered EM waves and random noise pattern with comparable amplitudes.

As it clearly seen in Figure 4, scattered EM wave spectrum is hidden in the noise pattern. Our aim is to apply Dynamic Logic algorithm in order to extract information on scattered EM waves from the pattern in Figure 4.

3. Implementation of the dynamic logic algorithm

In order to demonstrate the basics of the DL algorithm we will present below one illustrative example. Suppose we have two classes ω_1 and ω_2. We also have samples $\mathbf{X}_k$ where samples with k=1 belong to class ω_1 and with k=2 to class ω_2. Every sample k has two features $\mathbf{X}_k = (x_{1k}, x_{2k})$. In Figure 5 two groups of samples are presented. Red dots define samples from class k=1 and blue dots – from class k=2.

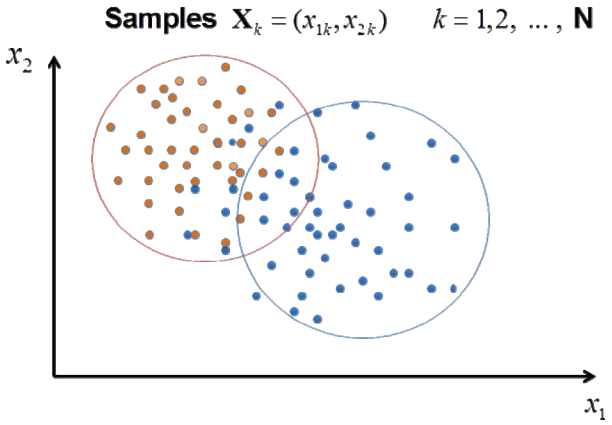


Fig. 5. Two groups of samples. Samples defines as red dots belong to the class k=1 and samples defined as blue dots to class k=1.

Typically, due to random variations there is good reason to expect the features in every class to be distributed normally (that is, according to the well-known Gaussian function, or "bell

curve"). A normal distribution is completely specified by its set of parameters $\Theta = \{\mathbf{M}, \Sigma\}$, where $\mathbf{M}$ is the mean vector and Σ is the covariance matrix (related to the width, or "spread", of the distribution). The equation for a normally-distributed probability density function (pdf) for the class k=1 is:

$$p(\mathbf{X}_n, \Theta_1) = \frac{1}{\sqrt{(2\pi)^d \mid \Sigma_1 \mid}} \exp[-\frac{1}{2}(\mathbf{X}_n - \mathbf{M}_1)^T \Sigma_1^{-1} (\mathbf{X}_n - \mathbf{M}_1)] \tag{17}$$

In (17) $p(\mathbf{X}_n, \Theta_1)$ is the probability that sample n with the feature vector $\mathbf{X}_n = (X_{n,1}, X_{n,2})$ belongs to process ω_1 with $\Theta_1 \equiv (\mathbf{M}_1, \Sigma_1)$. In (17) $\mathbf{M}_1$ is the mean vector for the class ω_1 and components of the mean vector are defined as:

$$M_{1,k} = \int_{-\infty}^{\infty} X_k p(\mathbf{X} \mid \Theta_1) d\mathbf{X} \tag{18}$$

Components of the covariance matrix Σ_1 for the class ω_1 are defined as:

$$\sigma_{1,ij} = \int_{-\infty}^{\infty} (X_i - M_{1,i})(X_j - M_{1,j}) p(\omega_1 \mid \mathbf{X}, \Theta_1) d\mathbf{X} \tag{19}$$

When we have N training samples $\mathbf{X}_{1,n}$ which belong only to class ω_1 the procedure to find the parameters of the Gaussian distribution $\Theta_1 = \{\mathbf{M}_1, \Sigma_1\}$ is as follows [6]. We need to introduce the likelihood function $L(\Theta_1)$:

$$L(\Theta_1) = \prod_{n=1}^{N} p(\mathbf{X}_{1,n} \mid \Theta_1)$$

Logarithm of the likelihood function :

$$LL(\Theta_1) = \ln L(\Theta_1) = \sum_{n=1}^{N} \ln p(\mathbf{X}_{1,n} \mid \Theta_1) \tag{20}$$

can be maximized by setting to zero its partial derivatives with respect to the parameters $\Theta_1 = \{\mathbf{M}_1, \Sigma_1\}$. Resulting maximum likelihood values for the mean and covariance in the class ω_1 are:

$$\hat{\mathbf{M}}_1 = \frac{1}{N} \sum_{n=1}^{N} \mathbf{X}_{1,n} \tag{21}$$

Components of the covariance matrix Σ for the class ω_1 are defined as:

$$\hat{\Sigma}_1 = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{X}_{1,n} - \hat{\mathbf{M}}_1)(\mathbf{X}_{1,n} - \hat{\mathbf{M}}_1)^T \tag{22}$$

Now we will return to the problem formulated at the beginning of this section. Suppose we have two processes (classes) running ω_1 and ω_2. We also have identified N samples $\{\mathbf{X}_{n,1}\}$ which belong to class ω_1 and another N samples $\{\mathbf{X}_{n,2}\}$ which belong to class ω_2. Our aim is to design a classifier for sorting out the remaining unlabelled samples which belong either to the class k=1 or to the class k=2, assuming that we know the features $\mathbf{X}_n = (x_{1n}, x_{2n})$ of remaining samples, but do not know to which class do they belong. The procedure we should follow is as follows. First we gather a set of training samples N for each of the two classes k=1 and k=2. Next we assign feature values $\mathbf{X}_n = (x_{1n}, x_{2n})$ to each sample. Then for each class, separately, we estimate the mean and covariance using (22), to give us a set of parameters $\Theta_k = \{\mathbf{M}_k, \Sigma_k\}$ with k = 1,2. Knowing these parameters we can compute the pdf $p(\mathbf{X} \mid \Theta_k)$ of the data according to equation (17). Then we can compute, for each unlabelled sample, the probability

$$P(k \mid n) = \frac{p(\mathbf{X}_n \mid \Theta_k) P(k)}{p(\mathbf{X}_n)} \tag{23}$$

where $p(\mathbf{X}_n) = \sum_{k=1}^{k=2} p(\mathbf{X}_n \mid \Theta_k) P(k)$

Equation (23) gives the probability $P(k \mid n)$ that sample *n* belongs to class *k*. To classify the unlabelled sample n, we simply choose the class k', for which probability is the highest.

Equations (23) is just an expression of Baye's rule. Note that in equation (23) it is assumed we know the a priori probabilities P(k) for each class. Obviously, for each sample the probabilities must add up to one, i.e.

$$\sum_{k=1}^{k=2} P(k \mid n) = 1 \tag{24}$$

After we described how to estimate the mean and covariance of a single normal (Gaussian) distribution from a collection of data samples, below we will describe how to estimate parameters when the pdf is a weighted mixture, or summation, of Gaussian functions. Gaussian mixtures can be used to model complicated data sets. In such a case, each source corresponds to a class k which is parameterized by its own mean and covariance which we need to estimate. In addition to the mean and covariance, we also need to estimate the a priori probabilities for each class, i.e. what fraction of the samples belong to each class. The estimation of parameters for mixture models is complicated by the fact that the data samples are unlabelled, i.e. we don't know which data samples correspond to which class. Because of this inconvenience, it will be impossible to compute direct analytical expressions for the maximum-likelihood parameter estimates. However, we will be able to derive a convergent, iterative process to estimate the parameters. To do so we will perform the following basic procedure. First, a generic model for the data is defined in which the parameters for each class k are assumed to be the only unknowns. Next, a metric is selected which will be used to evaluate the "goodness of fit" between the model and the data for certain set of parameter values. Finally, an algorithm is developed to maximize the "goodness of fit" with respect to the parameters. Below we will describe maximum likelihood (ML) parameter estimation, in

which we seek to maximize the log-likelihood of the parameters with respect to a collection of data samples. This approach is appropriate for standard data in which all samples are weighted equally in terms determining model parameters.

Next, we introduce an alternative known as maximum entropy (ME) parameter estimation. Here the parameters are chosen to maximize the Einsteinian log-likelyhood. This metric is appropriate for processing digitized image data, in which samples are weighted according to their pixel values.

4. Maximum likelihood (ML) estimation

We now describe maximum-likelihood (ML) parameter estimation for a Gaussian Mixture [1, 6]. Mathematically, the mixture model for two classes k=1 and k=2 is represented by the equation:

$$p(\mathbf{X} \mid \Theta) = \sum_{k=1}^{k=2} r_k p(\mathbf{X} \mid , \Theta_k) \tag{25}$$

where r_k is the mixture weight for class k and $\Theta = \{\Theta_1, r_1, \Theta_2, r_2\}$ is the total set of parameters for the mixture. Also $\Theta_k = (\mathbf{M}_k, \Sigma_k)$ is the subset of mean and covariance parameters for class k. Each class k has the normal (Gaussian) distribution:

$$p(\mathbf{X}, \Theta_k) = \frac{1}{\sqrt{(2\pi)^d \mid \Sigma_k \mid}} \exp[-\frac{1}{2}(\mathbf{X} - \mathbf{M}_k)^T \Sigma_k^{-1} (\mathbf{X} - \mathbf{M}_k)]$$

Since r_k denotes the fraction of the total samples that belong to class k, it is, by definition, equivalent to the *a priori* probability $P(k)$ for class k. It is a well-known fact that these probabilities should sum to one, i.e.:

$$\sum_{k=1}^{K} P(k) \equiv \sum_{k=1}^{K} r_k = 1 \tag{26}$$

Our goal is to compute the maximum likelihood (ML) parameter estimates $\widehat{\Theta} = \{\widehat{\Theta}_1, \widehat{r}_1, \widehat{\Theta}_2, \widehat{r}_2\}$ given the additional constraint in equation (25). As in the previous Section, this is accomplished by a process that incorporates the derivatives of the log-likelihood with respect to each parameter:

$$LL(\Theta) = \sum_{n=1}^{N} \ln p(\mathbf{X}_n \mid \Theta) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} r_k p(\mathbf{X}_n \mid \Theta_k) \tag{27}$$

The maximum-likelihood parameter estimates exactly satisfy the equations:

$$\widehat{\mathbf{M}}_k = \frac{\sum_{n=1}^{N} P(k \mid n) \mathbf{X}_n}{\sum_{n=1}^{N} P(k \mid n)} \tag{28}$$

$$\hat{\Sigma}_k = \frac{\sum_{n=1}^{N} P(k|n)(\mathbf{X}_n - \hat{\mathbf{M}}_k)(\mathbf{X}_n - \hat{\mathbf{M}}_k)^T}{\sum_{n=1}^{N} P(k|n)} \tag{29}$$

$$\hat{r}_k = \frac{1}{N}\sum_{n=1}^{N} P(k|n) \tag{30}$$

$$P(k|n) = \frac{r_k p(\mathbf{X}_n|\Theta_k)}{p(\mathbf{X}|\Theta)} = \frac{r_k p(\mathbf{X}_n|\Theta_k)}{\sum_{j=1}^{K} r_j p(\mathbf{X}|\Theta_j)} \tag{31}$$

Equation (31) is actually a form of Baye's and can be interpreted as the association probability, i.e. the probability that sample n belongs to class k. Equations (28) and (29) are intuitively pleasing in the sense that they are identical to their counterparts Eqs. (21) and (22) for the normal distribution, except that here each sample is weighted by the association probability $P(k/n)$.

Although Eqs. (28)-(31) look simple enough, they are actually a set of coupled nonlinear equations, because the quantity $P(k/n)$ on the right-hand sides is a function of all the parameters $\hat{\boldsymbol{\Theta}}$, as can be seen from equations (31) and (25). If a person could solve these equations directly, they would yield the exact values for the maximum-likelihood parameters - unfortunately a direct solution is intractable. However, we can compute the parameters indirectly, by iterating back and forth between the parameter estimation equations (28)-(30) and the probability estimation equation (31). In other words, the parameter estimates $[\mathbf{M}_k^{(I)}, \Sigma_k^{(I)}, r_k^{(I)}]$ in iteration I are computed using the association probabilities $P^{(I-1)}(k/n)$ computed from the previous iteration (I -1). Next, the association probabilities $P^{(I)}(k/n)$ in iteration I are computed using the parameter estimates $[\mathbf{M}_k^{(I)}, \Sigma_k^{(I)}, r_k^{(I)}]$ from iteration I. Mathematically, this procedure corresponds to an approximate set of equations, analogous to Eqs. (28)-(31), which are iteration procedure to compute parameter estimates:

$$\hat{\mathbf{M}}_k^{(I)} = \frac{\sum_{n=1}^{N} P^{(I-1)}(k|n)\mathbf{X}_n}{\sum_{n=1}^{N} P^{(I-1)}(k|n)} \tag{32}$$

$$\hat{\boldsymbol{\Sigma}}_k^{(I)} = \frac{\sum_{n=1}^{N} P^{(I-1)}(k|n)(\mathbf{X}_n - \hat{\mathbf{M}}_k^{(I)})(\mathbf{X}_n - \hat{\mathbf{M}}_k^{(I)})^T}{\sum_{n=1}^{N} P^{(I-1)}(k|n)} \tag{33}$$

$$\hat{r}_k^{(I)} = \frac{1}{N}\sum_{n=1}^{N} P^{(I-1)}(k|n) \tag{34}$$

and iteration procedure to compute probabilities :

$$P^{(l)}(k \mid n) = \frac{r_k^{(l)} p(\mathbf{X}_n \mid \Theta_k^{(l)})}{p(\mathbf{X} \mid \Theta^{(l)})} = \frac{r_k^{(l)} p(\mathbf{X}_n \mid \Theta_k^{(l)})}{\sum_{j=1}^{K} r_j^{(l)} p(\mathbf{X} \mid \Theta_j^{(l)})} \tag{35}$$

To initialize the procedure, parameter values are assigned at random, or by other means. Then the above set of equations is repeated until convergence, typically on the order of 100 iterations.

Below in Figures 6 – 7 results of implementation of the DL algorithm for the problem of identification of samples which belong to two different classes ω_1 and ω_2 described above and illustrated in Figure 5, are presented.

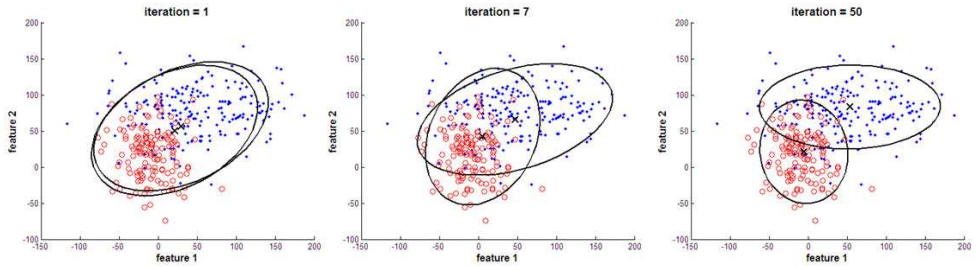


Fig. 6. With increasing iterations, the Dynamic Logic (DL) model adapts to segment data samples drawn from multiple random processes. In this example, the data samples are drawn from two Gaussian random processes, and the DL model consists of two components. The dots and circles represent the data samples, while the black ellipses represent the 2-sigma extent of the DL covariance, centered at the mean for each DL component. (a) Initially, the mean parameters for the components are assigned random initial values, and the covariance parameters are initialized in order to roughly span the sample data. (b) After several iterations the DL components begin to partition the data in a fuzzy manner as their evolving mean and covariance parameters cause them to "shrink" to fit each sample cluster. (c) Upon convergence, the DL model correctly segments the samples, and the mean and covariance parameters of the DL components provide good estimates for the mean and covariance of the underlying random processes.

In practice, one can decide whether convergence has been obtained by plotting the log-likelihood vs. iteration number and looking for a leveling of in the curve. The procedure is guaranteed to converge, in the sense that *LL* will never increase between successive iterations. However, convergence to an extraneous local minimum is sometimes possible. If this error occurs, and can be detected, the user must repeat the procedure after choosing a new set of initial conditions. The iterative procedure described above was pioneered by Perlovsky [1] and, independently, by others. In fact, the procedure has been shown to be an application of the well-known *expectation-maximization* (or EM) algorithm which is described, for example, in [20] and the references therein.

5. Maximum entropy (ME) estimation

In ML estimation, discussed above, we utilized the *log-likelihood* to determine the "goodness of fit" between the model and the data. This well-known metric is appropriate for standard data in which all samples are weighted equally in terms determining model parameters. However for some data sets, for example those involving digitized images, we require a different metric that will incorporate the natural weighting of samples corresponding to their pixel values [1,6]. For these problems we now introduce the metric known as the *Einsteinian log-likelihood*:

$$LL_E(\Theta) = \sum_{n=1}^{N} p_0(\mathbf{X}_n) \ln p(\mathbf{X}_n \mid \Theta) \tag{36}$$

Here, the function $p_0(\mathbf{X}_n)$ denotes the distribution of pixel values as a function of pixel position within the image. Note that $\mathbf{X}_n$ is no longer a vector of classification features, as has been the case up until now, but rather a vector giving the position $\mathbf{X}_n = [x_{1n}, x_{2n}]^T$ of the n^{th} pixel in the image.

The Einsteinian log-likelihood LL_E, which was introduced by Perlovsky[1], has some interesting physical interpretations relating to photon ensembles, etc., as discussed in [1]. It is also related to some standard measures of similarity used in the information theory arena. In fact, maximizing LL_E with respect to the parameters is equivalent to minimizing the *relative entropy*, or *Kullback-Leibler distance* [5].

$$D(p_0 \mid p) = \sum_x p_0(x) \ln[\frac{p_0(x)}{p(x)}] \tag{37}$$

In any case, it is easy to see that if $p_0(\mathbf{X}_n) = 1$, then the Einsteinian log-likelihood LL_E in equation (36) reduces to the standard log-likelihood LL given by equation (27).

The equation describing our mixture distribution is the same as equation (25):

$$p(\mathbf{X} \mid \Theta) = \sum_{k=1}^{K} r_k p(\mathbf{X} \mid , \Theta_k) \tag{38}$$

Where

$$p(\mathbf{X}, \Theta_k) = \frac{1}{\sqrt{(2\pi)^d \mid \Sigma_k \mid}} \exp[-\frac{1}{2}(\mathbf{X} - \mathbf{M}_k)^T \Sigma_k^{-1} (\mathbf{X} - \mathbf{M}_k)]$$

Also, as before, $\Theta = (\Theta_k, r_k)$, k = 1; 2; ::: ;K , is the total set of parameters for the mixture, while $\Theta_k = \{\mathbf{M}_k, \Sigma_k\}$ is the subset of mean and covariance parameters for class k, and r_k is the mixture weight for class k. Substituting equation (38) into equation (37), we obtain the following expression for the Einsteinian log-likelihood:

$$LL_E(\Theta) = \sum_{n=1}^{N} p_0(\mathbf{X}_n) \ln \sum_{k=1}^{K} r_k p(\mathbf{X}_n \mid \Theta_k) \tag{39}$$

For convenience we will assume that the data $p_0(\mathbf{X}_n)$ is normalized, i.e.:

$$\sum_{n=1}^{N} p_0(\mathbf{X}_n) = 1 \tag{40}$$

Therefore, to insure that also:

$$\sum_{n=1}^{N} p(\mathbf{X}_n \mid \Theta) = 1 \tag{41}$$

so that the data and model have the same total energy, the r_k must sum to one, i.e.:

$$\sum_{k=1}^{K} r_k = 1 \tag{42}$$

Our goal is to compute the maximum entropy (ME) parameters $\hat{\Theta}$ which maximize LL_E given by equation (39) subject to the constraint in equation (42). The mathematical derivation for the ME parameter estimates is almost identical to the mathematics governing the ML estimates. The result is that the ME parameters exactly satisfy the equations:

$$\hat{\mathbf{M}}_k = \frac{\sum_{n=1}^{N} P(k \mid n) p_0(\mathbf{X}_n) \mathbf{X}_n}{\sum_{n=1}^{N} P(k \mid n) p_0(\mathbf{X}_n)} \tag{43}$$

$$\hat{\Sigma}_k = \frac{\sum_{n=1}^{N} P(k \mid n) p_0(\mathbf{X}_n)(\mathbf{X}_n - \hat{\mathbf{M}}_k)(\mathbf{X}_n - \hat{\mathbf{M}}_k)^T}{\sum_{n=1}^{N} P(k \mid n) p_0(\mathbf{X}_n)} \tag{44}$$

$$\hat{r}_k = \frac{1}{N} \sum_{n=1}^{N} P(k \mid n) p_0(\mathbf{X}_n) \tag{45}$$

$$P(k \mid n) = \frac{r_k p(\mathbf{X}_n \mid \Theta_k)}{\sum_{j=1}^{K} r_j p(\mathbf{X} \mid \Theta_j)} \tag{46}$$

As described in Section IV, the parameter estimation equations, above, are actually a set of coupled nonlinear equations, due to the implicit dependence within the bracket notation on the unknown parameters. However, we can compute the ME parameter estimates using a convergent, iterative process, identical to the process described in Section IV for the ML parameter estimates. In this process we alternate, in each iteration, between updating the probabilities $P(k \mid n)$ and the parameters $\{\mathbf{M}_k,\ \Sigma_k,\ r_k\}$.

Applying maximum entropy estimation method to the problem of detection of scattered by a Langmuir soliton EM waves hidden in a noise pattern (see Fig. 5) we will obtain the following result, presented in Fig. 7.

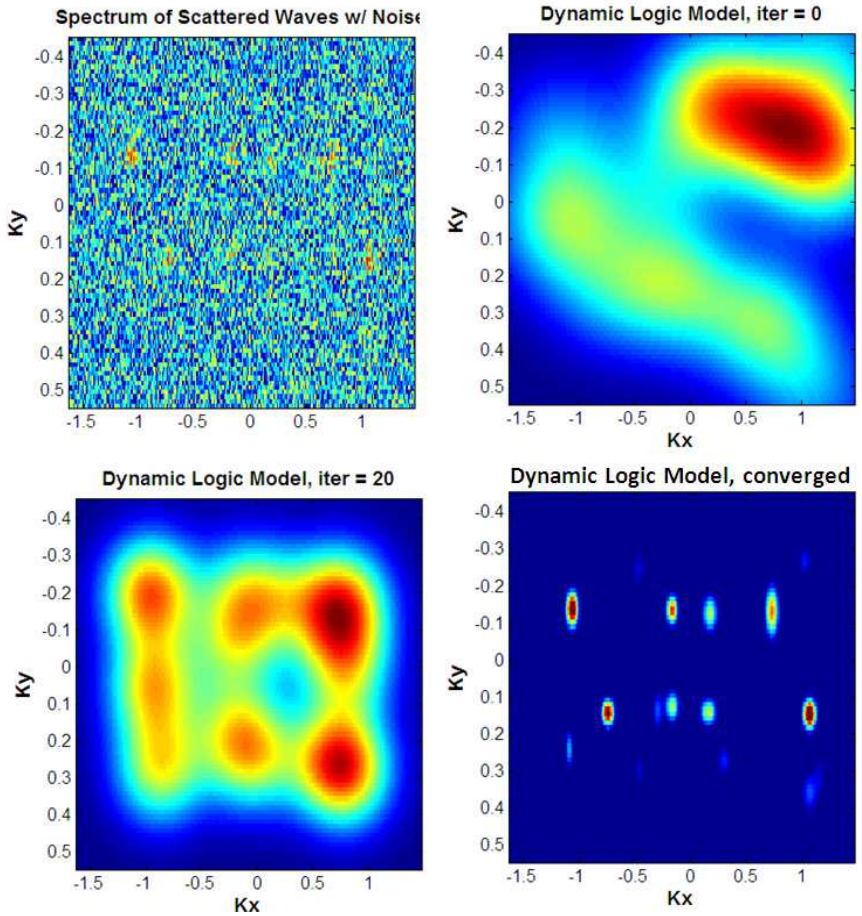
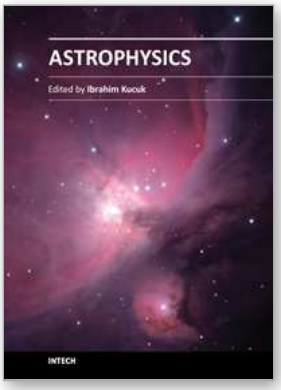


Fig. 7. (upper left) The combined measured spectra of several soliton waves are completely obscured by measurement noise. (upper right) The DL model is initialized with random model parameters. (lower left) After a few iterations the DL model begins to learn the structure of the data, and model components begin to adapt to fit the soliton spectra. (lower right) Upon convergence, the DL model is able to adaptively segment the data into background noise vs. multiple soliton spectra, despite the low signal-to-clutter ratio.

6. References

[1] Perlovsky, L.I. (2001). Neural Networks and Intellect: using model-based concepts. Oxford University Press, New York, NY (3rd printing). [2] Rudakov, L.I., DAN Sov. Acad. Sci, vol. 207, p. 821-823, 1972.

[2] Rudakov, L.I., Langmuir soliton, DAN Sov. Acad. Sci, vol. 207, p. 821-823, 1972.
[3] Mendonca J.T., Scattering of waves by Langmuir soliton, J. Plasma Physics, Vol.30, part 1, pp. 65-73, 1983.
[4] Sotnikov V. J.N. Leboeuf, S. Mudaliar, Scattering of electromagnetic waves in the presence of wave turbulence excited by a flow with velocity shear, Trans. Plasma Sci. IEEE, Vol. 38, No. 9, September 2010.
[5] MacKay, D.J.C., *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, Cambridge, U.K., 2003.
[6] R.W. Deming, Parameter estimation for mixtures, August, 2005.
[7] Moon, T.K., The Expectation-Maximization Algorithm, IEEE Signal Processing Magazine, November, 1996, pp. 47-60.
[8] Duda, R. O. and Hart, P. E., Pattern Classification and Scene Analysis, section 6.4, New York: Wiley, 1973.
[9] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum Likelihood from incomplete data via the EM algorithm", Journal of Royal Statistical Society, Series B, vol. 39, (1977), 1-38.
[10] McLachlan, G. J., and Krishnan, T., The EM Algorithm and Extensions, New York: Wiley, 1997.
[11] Redner, R. A., and Walker, H. F., "Mixture densities, maximum likelihood and the EM algorithm", SIAM Review, vol. 26, (1984), 195-239.



Astrophysics

Edited by Prof. Ibrahim Kucuk

ISBN 978-953-51-0473-5
Hard cover, 398 pages
Publisher InTech
Published online 30, March, 2012
Published in print edition March, 2012

This book provides readers with a clear progress to theoretical and observational astrophysics. It is not surprising that astrophysics is continually growing because very sophisticated telescopes are being developed and they bring the universe closer and make it accessible. Astrophysics Book presents a unique opportunity for readers to demonstrate processes do occur in Nature. The unique feature of this book is to cover different aspects in astrophysics covering the topics: • Astronomy • Theoretical Astrophysics • Observational Astrophysics • Cosmology • The Solar System • Stars • Planets • Galaxies • Observation • Spectroscopy • Dark Matter • Neutron Stars • High Energy Astrophysics

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821