AN ALGORITHM TO PROCESS SIGNALS CONVEYING STATISTICAL INFORMATION

L. Rebollo Neira
Department of Electrical and Electronic Engineering
Imperial College
Exhibition Road, London SW7 2BT, England

A. Plastino
Departamento de Física, Universidad Nacional de La Plata, c.c
67, 1900 La Plata, Argentina

ABSTRACT. A method for processing signals containing information about the state distribution of a physical system is presented. The concomitant algorithm is specifically devised to suitably adapt lineal restrictions so as to take into account the presence of noise due to experimental errors.

1. Introduction

We call Statistical Signals the ones which convey information about systems that consist of subsystems of known properties whose relative proportions we want to find. We shall adopt a vectorial representation denoting a signal f as a vector $|f\rangle$ and a measurement as a mapping that assigns to it a real number.

For the sake of definiteness we assume that the system S we are interested in consists of a number M of subsystems S_n. Our purpose is that of finding out the relative population of S, assuming that the one corresponding to S_n is $C_n \geq 0$ (unknown). We take the view [1] that in order to study S one interacts with it by means of an input signal $|I\rangle$, the interaction between the signal $|I\rangle$ and S resulting in a response signal $|f\rangle$. The corresponding process is represented according to

$$\hat{W}|I\rangle = |f\rangle, \tag{1}$$

where the linear operator $\hat{W}$ portrays the effect that the system produces upon the input signal and can be decomposed in the following fashion

$$\hat{W} = \sum_{n=1}^{M} C_n \hat{W}_n, \tag{2}$$

where $\hat{W}_n|I\rangle = |n\rangle$. We work under the hypothesis that we know the response $|n\rangle$ evoked by S_n and that this set of vectors gives rise to a linear space U_M of dimension M. From

67

(1) and (2) it is clear that the response $|f\rangle$ is contained whiting U_M and carries information concerning the numbers C_n we are tying to find out. In order to accomplish such a goal one needs to perform observations upon $|f\rangle$. The corresponding measurement procedure provides numbers $\{f_1, \ldots, f_N\}$ out of $|f\rangle$ which can be regarded as the numerical representation of the signal.

2. Treatment of a Numerical Representation

Let's suppose that the numerical representation of $|f\rangle$ is obtained in such a way that measurements are performed as a function of a parameter x which adopts the values x_i with $(i = 1, \ldots, N)$. If the measurements are performed independently, we can regard the x_i as defining an (orthogonal) set of vectors $|\hat{x}_i\rangle$ that span an N-dimensional linear space E. We associate this space with the measurement instrument [2].

Let us $|f\rangle_p$ be the projection of $|f\rangle$ in E, i.e.

$$|f\rangle_p = \sum_{i=1}^{N} \langle \hat{x}_i | f \rangle | \hat{x}_i \rangle. \tag{3}$$

The expressions $\langle \hat{x}_i | f \rangle$ in a general case represent bilinear forms [2] and they are supposed to be given by experimental observations, so what we really have are numbers f_i^o affected by uncertainties Δf_i^o. Thus, instead of (3) we have, for the representation of $|f\rangle$ in E

$$|f^o\rangle_p = \sum_{i=1}^{N} f_i^o | \hat{x}_i \rangle. \tag{4}$$

The problem we face is that of building up a vector $|f^\star\rangle \in U_M$

$$|f^\star\rangle = \sum_{n=1}^{M} C_n^\star |n\rangle \tag{5}$$

out of the $\{f_i^o, i = 1, \ldots, N\}$-set, such that the $C_n^\star$ constitute a good approximation to the "true" C_n. For this purpose we construct the representatives in E of $|n\rangle$ and $|f^\star\rangle$

$$|n\rangle_p = \sum_{i=1}^{N} \langle \hat{x}_i | n \rangle | \hat{x}_i \rangle \quad ; \quad |f^\star\rangle_p = \sum_{n=1}^{M} C_n^\star |n\rangle_p. \tag{6}$$

The nearest vector $|f^\star\rangle_p$ to $|f^o\rangle_p$ that can be built is the one that fulfills the least distance equations [2]. These equations can be written in the form

$$F_n = \sum_{k=1}^{M} C_k^\star a_{n,k} \quad ; \quad n = 1, \ldots, M. \tag{7}$$

where the $a_{n,k}$ are constructed out of the projections of vector $|n\rangle$ in E while the F_n contains the experimental data

$$a_{n,k} = \sum_{i=1}^{N} \langle n | \hat{x}_i \rangle \langle \hat{x}_i | k \rangle \quad ; \quad F_n = \sum_{i=1}^{N} \langle \hat{x}_i | n \rangle f_i^o. \tag{8}$$

Of course, as the f_i^o are affected by the experimental uncertainties Δf_i^o so will the F_n be subjected to corresponding uncertainties ΔF_n. Furthermore, the set of conditions (7) do not restrict the $C_n^\star$ to the domain of the non-negative real numbers, so we will adopt an algorithm to obtain a non-negative set of C_n that fulfills the set of equations (7), *within the margin allowed by the uncertainties ΔF_n.*

3. A Maximum Entropy Algorithm

We start by writing the equations (7) in the form

$$F_k = A \sum_{n=1}^{M} p_n a_{k,n} \quad ; k = 1, \dots, M, \tag{9}$$

where $A \geq 0$ is a constant such that $\sum_{n=1}^{M} p_n = 1$. We can now think of the weights p_n as defining a probability space over a discrete set of M events whose informational content is given by

$$H = - \sum_{n=1}^{M} p_n \ln p_n. \tag{10}$$

We regard each F_k in (9) as proportional to the mean value of a random variable that adopts the values $a_{k,n}$;$(n = 1, \dots, M)$ with a probability distribution given by the $\{p_n\}$-set. As A is an unknown constant, we employ one of the equations , say the l-th one, to determine it and are now in a position to solve the set of equations in an iterative fashion. We start our iterative process, by employing the Maximum Entropy Principle [3] in each step in order to construct an "optimal conjecture", that improves upon the results obtained in the previous step .

The zeroth-order approximation (first step) is devised by requiring that the zeroth-order weights p_n maximize H. This entails $p_n^{(0)} = 1/M$ so that we *predict* a zeroth-order value for the F_k. The quality of our conjecture can be measured by defining the "predictive error" ϵ_k as

$$\epsilon_k = \frac{|F_k - F_k^{(0)}|}{|F_k|} \quad ; k = 1, \dots, M. \tag{11}$$

In order to construct our first order approximation we select, among the ϵ_k, the largest one, let us call it ϵ_{l1}. We shall then obtain the first-order weights $p_n^{(1)}$ by requiring that they maximize H with a constraint that ensures that the $l1$-th equation in (9) be fulfilled. We evaluate now the $F_k^{(1)}$ and the concomitant (new) set of ϵ_k. After selecting the largest one, ϵ_{l2}, say, we obtain the $p_n^{(2)}$ by maximizing H with the constraint that *both* the equations (9) for $k = l1$ and $k = l2$ be fulfilled, etc ,etc. The j-th order approximation is given by

$$p_n^{(j)} = \frac{exp(-\sum_{i=1}^{j} \lambda_i [F_l a_{li,n} - F_{li} a_{l,n}])}{\sum_{s=1}^{M} exp(-\sum_{i=1}^{j} \lambda_i [F_l a_{li,s} - F_{li} a_{l,s}])}, \tag{12}$$

where the j Lagrange multipliers λ_i are obtained by solving the j equations

$$\sum_{n=1}^{M} p_n^{(j)} [F_l a_{li,n} - F_{li} a_{l,n}] = 0 \quad ; i = 1, \dots, j. \tag{13}$$

The iterative process is to be stopped when

$$\epsilon_k \leq \frac{\Delta F_k}{|F_k|} \quad ; \quad \Delta F_k = \sum_{i=1}^{N} \Delta f_i^o |\langle k|\hat{x}_i\rangle| \quad ; \quad k = 1 \ldots, M. \tag{14}$$

Let us assume that the "convergence" (14) is attained at the L-th iteration. With this solution we can evaluate the numerical values

$$\langle \hat{x}_i|f^{(L)}\rangle = \sum_{n=1}^{M} C_n^{(L)} \langle \hat{x}_i|n\rangle \quad ; \; i = 1, \ldots, N. \tag{15}$$

If these conjectures are such that

$$|\langle \hat{x}_i|f^{(L)}\rangle - f_i^o| \geq \Delta f_i^o \quad for \; some \; i, \tag{16}$$

the number of iterations can be augmented until the direction of the inequality is reversed. However, there is no guarantee that this type of convergence will *always* be achieved. Even more, in the realistic cases where we only can guess same estimations for the errors, to require that the direction of the inequality be reversed for all i it becomes a too stringent requirement. Although in the application we will discuss this type of convergence can be achieved, we wish to keep the discussion open so as to suitably adapt the "stop" point to the errors concomitant to any given particular model.

4. Numerical Text

Consider that we have a mixture of $M = 11$ different rare earth elements which satisfies a simple paramgnetic model [4], their respective proportions in the mixture being denoted by p_n. For any given n we list the corresponding quantum number S_n, L_n and J_n in Table I and set $|n\rangle \equiv |S_n L_n J_n\rangle$. We take a series of $N =40$ values of the magnetic field at the temperature T, which generates the parameters $x_i = H_i/T$;$(i = 1, \ldots, 40)$ The projection of vector $|n\rangle$ for a given value x_i is given by the magnetization of the ion n in Table I

$$\langle \hat{x}_i|n\rangle = g_n J_n \mu_B B_n(x_i), \tag{17}$$

where μ_B is the Bohr magneton, g_n is the spectral factor for the ion n, and $B_n(x_i)$ the appropriate Brillouin function[4]

$$g_n = 1 + \frac{J_n(J_n + 1) + S_n(S_n + 1) - L_n(L_n + 1)}{2J_n(J_n + 1)}, \tag{18}$$

$$B_n(x_i) = \frac{2J_n + 1}{2J_n} cotgh[\frac{2J_n + 1}{2J_n}x_i] - \frac{1}{2J_n} cotgh[\frac{x_i}{2J_n}]. \tag{19}$$

Two sets of "weights" $\{p_n\}$ and $\{p'_n\}$ are listed in Table I which correspond to two hypothetical mixtures S and S'. By recourse to these sets we have numerically simulated a series of measurements of the magnetization and have randomly distorted them within a 3% range (bars in Fig 1) where a) correspond to S and b) to S'.

n	Ion	S_n	L_n	J_n	p_n	$p_n^\star$	$p_n^{(1)}$	p'_n	$p'^\star_n$	$p'^{(2)}_n$
1	Ce^{3+}	$\frac{1}{2}$	3	$\frac{5}{2}$	0.003	2661	0.006	0.549	535	0.457
2	Pr^{3+}	1	5	4	0.004	-5966	0.003	0.329	-1246	0.296
3	Nd^{3+}	$\frac{1}{2}$	6	$\frac{4}{2}$	0.005	6886	0.003	0.076	1432	0.110
4	Pm^{3+}	2	6	4	0.006	-2979	0.005	0.042	-598	0.137
5	Gd^{3+}	$\frac{7}{2}$	0	$\frac{7}{2}$	0.055	496	0.600	0.000	93	0.000
6	Tb^{3+}	3	3	6	0.156	1606	0.163	0.000	310	0.000
7	Dy^{3+}	$\frac{5}{2}$	5	$\frac{15}{2}$	0.338	-8984	0.334	0.000	-1653	0.000
8	Ho^{3+}	2	6	8	0.301	9517	0.296	0.000	1751	0.000
9	Er^{3+}	$\frac{3}{2}$	6	$\frac{15}{2}$	0.104	-2510	0.106	0.000	-477	0.000
10	Tm^{3+}	1	5	6	0.019	-34	0.019	0.000	-4	0.000
11	Yb^{3+}	$\frac{1}{2}$	3	$\frac{7}{2}$	0.005	-6124	0.004	0.000	-124	0.000

Table I. For each rare-earth ion the pertinent quantum numbers are given. The relative properties in the rare-earth mixture are denoted by p_n (system S) and p'_n (system S'). $p_n^{(1)}$ and $p_n^\star$ are theoretical results for system S that allow for a 3% error in the input data (see fig 1a)). The $p_n^\star$ correspond to a least-square treatment and the $p_n^{(1)}$ to a first-order version of the present approach. $p'^\star_n$ and $p'^{(2)}_n$ are theoretical results for system S'.

The algorithm of section 3 gives the weights $p_n^{(1)}$ (up to first order) for the mixture S and the weights $p_n'^{(2)}$ (up to second order) for mixture S'. If we employ a Marquardt's algorithm [5] in order to get a real least-distance solution (it is known as a least-square approximation), we obtain the $p_n^\star$ and coefficients for the systems S and S' respectively. (All the pertinent figures are listed in Table I)

6. Conclusions

We have discussed a method, based upon the optimum conjecture derived from the Maximum Entropy Principle, that on the basis of measurements performed on the signal the system produces, allows one to find out its state distribution. The main idea is to build-up a signal belonging to a subspace determined by the system whose projection in a subspace, determined by the measurements instrument, is close to a vector constructed by experimental observations. The algorithm presented has been specifically devised so as to deal with lineal restrictions. It was achieved by a Maximum Entropy criterion and the requirement that the distance between observed measurements and predicted measurements be bounded by experimental errors. In order to compare the present approach with a least square approximation a low dimension numerical test has been performed. A larger dimension realistic situation involving X ray diffraction has also been tackled but is not included in this paper [2]. In both cases the algorithm has reached a rapid convergence, and few parameters are needed.
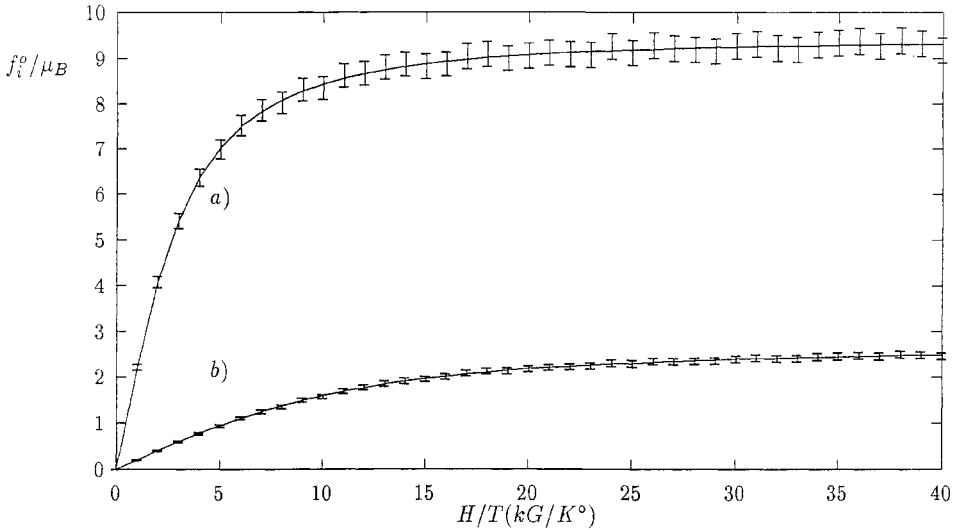
Fig. 1. Magnetization vs external applied field at the temperature T.The error bars are the input data of the numerical test and allow for a 3% distortion. The continuous curve represents both the predictions obtained with the present approach and with the least-squares approximation. Curve $a)$ correspond to system S and curve $b)$ to system S'

REFERENCES

[1] A. Plastino, L. Rebollo Neira, A. Alvarez, *Phys. Rev.* **A40** (1989) 1644.

[2] A. Plastino, L. Rebollo Neira, F. Zyserman, A. Alvares, R. Bonetto, H. Viturro.(To be published) (1992).

[3] E.T. Jaynes, *Phys. Rev.* **160** (1957) 620.

[4] C. Kittel.'Introduction to Solid State Physics'. John Wiley and sons, Inc, New York (1959).

[5] D. W. Marquardt, *J. Soc. Indust. Appl. Math.* **11** (1963) 2.