Recursive approach for constructing the $q=1/2$ maximum entropy distribution from redundant data

L. Rebollo-Neira
NCRG, Aston University, Birmingham B4 7ET, United Kingdom

A. Plastino
Instituto de Física La Plata (IFLP), Universidad Nacional de La Plata and CONICET, Casilla de Correo 727, 1900 La Plata, Argentina

(Received 5 April 2002; revised manuscript received 10 June 2002; published 27 September 2002)

A recursive approach for computing the $q=1/2$ nonextensive maximum entropy distribution of the previously introduced formalism for data subset selection is proposed. Such an approach is based on an iterative biorthogonalization technique, which allows for the incorporation of the Lagrange multipliers that determine the distribution to the workings of the algorithm devised for selecting relevant data subsets. This technique circumvents the necessity of inverting operators and yields a recursive procedure to appropriately modify the Lagrange multipliers so as to account for each new constraint.

DOI: 10.1103/PhysRevE.66.032102 PACS number(s): 05.20.−y, 02.50.Tt

I. INTRODUCTION

In a recent publication [1] we have introduced a method for data subset selection, which is based on the nonextensive maximum entropy formalism [2–9] by considering the case $q=1/2$ previously discussed in other contexts [10–13]. The method evolves iteratively by selecting, at each iteration, the measure yielding a $q=1/2$ distribution capable of making predictions minimizing the Euclidean distance to the available piece of data. During the selection process, however, such a distribution is not actually computed, as doing so would involve computing an inverse operator at each iteration. Instead, we use a convenient orthogonalization which avoids operator inversion at each step of the iterative process. The inversion is performed at the end of the selection process, so as to obtain the Lagrange multipliers which eventually determine the $q=1/2$ distribution.

In this Brief Report we show that, by means of an adequate biorthogonalization technique, one can include the computation of the $q=1/2$ distribution, at every step of the selection process, without the need of inverting operators. By recourse to the present approach, the $q=1/2$ distribution is recursively ''adapted'' at each iteration, at low computational cost. This new algorithm is based on the use of biorthogonal vectors for representing orthogonal projections (rather than using orthogonal vectors as proposed in [1]). The possibility of implementing this proposal at low computational cost lies in the existence of a recursive approach, to be discussed here, for computing biorthogonal vectors yielding the abovementioned orthogonal projectors.

The paper is organized as follows: Section II summarizes the formalism for data subset selection proposed in [1]. In Sec. III, a new approach for computing the corresponding Lagrange multipliers is introduced, based on a biorthogonal technique for constructing orthogonal projectors. The conclusions are drawn in Sec. IV.

II. SETTING UP THE PROBLEM

Along the lines of [1], let us consider M pieces of data $f_1^o, f_2^o, \ldots, f_i^o, \ldots, f_M^o$, each of which is the expectation value of a random variable. This variable adopts the possible values $f_{i,n}$; $n=1,...,N$. The expectation values are computed using a (generalized, see [1]) probability distribution $p_n^{1/2}$; $n=1,...,N$. Thus, the data model is expressed in terms of M equations of the form

$$f_i^o = \sum_{n=1}^{N} p_n^{1/2} f_{i,n}, \quad i=1....,M \tag{1}$$

that, adopting a Dirac's vectorial notation, are recast as

$$|f^o\rangle = \hat{A}|p^{1/2}\rangle, \tag{2}$$

where $|p^{1/2}\rangle$ is represented in terms of the *standard basis* $|n\rangle$, $n=1,...,N$ of $\mathcal{R}^N$,

$$|p^{1/2}\rangle = \sum_{n=1}^{N} |n\rangle\langle n|p^{1/2}\rangle = \sum_{n=1}^{N} p_n^{1/2}|n\rangle, \tag{3}$$

while the data vector $|f^o\rangle$ is represented in terms of the standard basis $|i\rangle$, $i=1,...,M$ of $\mathcal{R}^M$,

$$|f^o\rangle = \sum_{i=1}^{M} |i\rangle\langle i|f^o\rangle = \sum_{i=1}^{M} f_i^o|i\rangle. \tag{4}$$

The operator $\hat{A}:\mathcal{R}^N \to \mathcal{R}^M$ in Eq. (2) is given by the matrix elements $\langle i|\hat{A}|n\rangle = f_{i,n}$; $i=1,...,M$; $n=1,...,N$. Thus, by defining vectors $|f_n\rangle \in \mathcal{R}^M$ in such a way that $\langle i|f_n\rangle = f_{i,n}$, the operator $\hat{A}$ is expressed as

$$\hat{A} = \sum_{n=1}^{N} |f_n\rangle\langle n|. \tag{5}$$

By considering as constraints a subset of k Eq. (1) labeled by indexes l_j, $j=1,...,k$, the resultant maximum entropy $|p^{1/2}\rangle$ distribution adopts the form [1]

$$|p^{1/2(k)}\rangle=\left(\frac{1}{N}-\frac{1}{N}\sum_{j=1}^{k}\langle g|l_j\rangle\langle l_j|\lambda^{(k)}\rangle\right)\sum_{n=1}^{N}|n\rangle+\sum_{j=1}^{k}\hat{A}^{\dagger}|l_j\rangle\langle l_j|\lambda^{(k)}\rangle \tag{6}$$

with

$$|g\rangle=\sum_{n=1}^{N}|f_n\rangle\equiv\sum_{n=1}^{N}\hat{A}|n\rangle. \tag{7}$$

We have introduced here the superscript k to explicitly indicate that the $|p^{1/2(k)}\rangle$ distribution is built out of an optimizing process involving k constraints. The Lagrange multiplier vector $|\lambda^{(k)}\rangle$ is determined in such a way that, using the associated probability distribution, one may be in a position to make ''sensible'' predictions. By this we mean that $|p^{1/2(k)}\rangle$ enables one to predict a complete data vector $|f^p\rangle=\hat{A}|p^{1/2(k)}\rangle\in\mathcal{R}^M$ that *minimizes the distance to the observed vector* $|f^o\rangle$. Such a requirement entails

$$|\lambda^{(k)}\rangle=(\hat{F}_k^{\dagger}\hat{F}_k)^{-1}\hat{F}_k^{\dagger}|\tilde{f}^o\rangle, \tag{8}$$

where $|\tilde{f}^o\rangle=|f^o\rangle-|g\rangle/N$ and

$$\hat{F}_k=\sum_{j=1}^{k}|\alpha_{l_j}\rangle\langle l_j|, \tag{9}$$

with

$$|\alpha_{l_j}\rangle=\sum_{n=1}^{N}|f_n\rangle\langle f_n|l_j\rangle-\frac{1}{N}|g\rangle\langle g|l_j\rangle. \tag{10}$$

The subindices l_j, $j=1,...,k$ are iteratively selected as follows: given k subindexes l_j, $j=1,...,k$, the corresponding l_{k+1} is obtained by the requirement that the predicted vector $|f^p\rangle$ minimizes the distance to the observed vector $|f^o\rangle$. This is equivalent to selecting the vector $|\alpha_{l_{k+1}}\rangle$ minimizing

$$\||\tilde{f}^o\rangle-\hat{F}_{k+1}(\hat{F}_{k+1}^{\dagger}\hat{F}_{k+1})^{-1}\hat{F}_{k+1}^{\dagger}|\tilde{f}^o\rangle\|^2. \tag{11}$$

The observation that $\hat{F}_{k+1}(\hat{F}_{k+1}^{\dagger}\hat{F}_{k+1})^{-1}\hat{F}_{k+1}^{\dagger}$ is the orthogonal projector operator onto the subspace $V_{k+1}=V_k\oplus|\alpha_{l_{k+1}}\rangle$, where V_k is spanned by $|\alpha_{l_j}\rangle$, $j=1,...,k$, has led in [1] to conclude that minimization of (11) is tantamount to maximization of functionals e_i, $i=1,...,M$ given by

$$e_i=\frac{b_i}{d_i}=\frac{|\langle\alpha_i|\Delta f\rangle|^2}{\langle\alpha_i|\alpha_i\rangle-\Sigma_{l=1}^{k}|\langle\tilde{\psi}_l|\alpha_i\rangle|^2},\quad b_i>0, \tag{12}$$

where $|\tilde{\psi}_{k+1}\rangle=|\psi_{k+1}\rangle/\||\psi_{k+1}\rangle\|$ and $|\psi_{k+1}\rangle$ are orthogonal vectors arising, from $|\psi_1\rangle=|\alpha_{l_1}\rangle$, as

$$|\psi_{k+1}\rangle=|\alpha_{l_{k+1}}\rangle-\hat{P}_{V_k}|\alpha_{l_{k+1}}\rangle. \tag{13}$$

Maximization of Eq. (12) yields an effective strategy for minimization of (11), with much lower computational effort than that involved in tackling directly the latter. Such a procedure, however, does not provide a direct way of computing the distribution $|p^{1/2(k)}\rangle$. Indeed, given k subindexes l_j, $j=1,...,k$, the corresponding distribution is to be obtained from Eq. (6) and, since the vector $|\lambda^{(k)}\rangle$ is given by Eq. (8), we need to compute the inverse operator $(\hat{F}_k^{\dagger}\hat{F}_k)^{-1}$ in order to determine $|p^{1/2(k)}\rangle$.

The goal of this effort is to avoid the need for such an inversion so as to be able to introduce the calculation of $|p^{1/2(k)}\rangle$ into the iterative process that selects the subindices l_j; $j=1,...,k$. We tackle the issue in the forthcoming section.

A recursive approach for constructing $|p^{1/2(k)}\rangle$

We introduce here an iterative procedure which allows us to quickly modify the $|p^{1/2(k)}\rangle$ distribution each time a new subindex, say l_{k+1}, is selected.

The key idea for achieving such a goal is to make use of the fact that the orthogonal projector $\hat{F}_{k+1}(\hat{F}_{k+1}^{\dagger}\hat{F}_{k+1})^{-1}\hat{F}_{k+1}^{\dagger}$ admits a representation in terms of biorthogonal vectors which are computed in an iterative fashion. Indeed, given a set of vectors $|\alpha_{l_n}\rangle$, $n=1,...,k+1$, let us define vectors $|\tilde{\tilde{\psi}}_{k+1}\rangle$ as $|\tilde{\tilde{\psi}}_{k+1}\rangle=|\overline{\psi}_{k+1}\rangle/\||\psi_{k+1}\rangle\|=|\psi_{k+1}\rangle/\||\psi_{k+1}\rangle\|^2$, with $|\psi_{k+1}\rangle$ as given in Eq. (13). Then, the dual vectors $\langle\tilde{\alpha}_{l_n}^{k+1}|$, $n=1,...,k+1$ which are obtained by recourse to the recursive relations

$$\langle\tilde{\alpha}_{l_n}^{k+1}|=\langle\tilde{\alpha}_{l_n}^{k}|-\langle\tilde{\alpha}_{l_n}^{k}|\alpha_{l_{k+1}}\rangle\langle\tilde{\tilde{\psi}}_{k+1}|,\quad n=1,...,k,$$

$$\langle\tilde{\alpha}_{l_{k+1}}^{k+1}|=\frac{\langle\psi_{k+1}|}{\langle\psi_{k+1}|\alpha_{l_{k+1}}\rangle}=\frac{\langle\psi_{k+1}|}{\langle\psi_{k+1}|\psi_{k+1}\rangle}=\langle\tilde{\tilde{\psi}}_{k+1}|, \tag{14}$$

with $\langle\psi_{l_1}|=\langle\alpha_{l_1}|/\langle\alpha_{l_1}|\alpha_{l_1}\rangle$, satisfy the following properties [14].

(a) Biorthogonality with respect to vectors $|\alpha_{l_n}\rangle$, $n=1,...,k+1$, i.e.,

$$\langle\tilde{\alpha}_{l_n}^{k+1}|\alpha_{l_m}\rangle=\delta_{l_n,l_m},$$

$$n=1,...,k+1;\quad m=1,...,k+1. \tag{15}$$

(b) They provide a representation of the orthogonal projection operator onto V_{k+1} as given by

$$\hat{P}_{V_{k+1}}=\sum_{n=1}^{k+1}|\alpha_{l_n}\rangle\langle\tilde{\alpha}_{l_n}^{k+1}|=\hat{P}_{V_{k+1}}^{\dagger}=\sum_{n=1}^{k+1}|\tilde{\alpha}_{l_n}^{k+1}\rangle\langle\alpha_{l_n}|. \tag{16}$$

From properties (a) and (b) it immediately follows that the vectors $\langle\tilde{\alpha}_{l_n}^{k+1}|$ given in Eqs. (14) give rise to a recursive formula that yields the Lagrange multipliers involved in minimizing the distance to the observed data vector $|f^o\rangle$.

Proposition 1. The Lagrange multipliers that minimize the distance to the observed data $|f^o\rangle$ are amenable to be recursively adapted, when a new constraint is introduced, according to the recursive relation

$$\langle l_n|\lambda^{(k+1)}\rangle = \langle l_n|\lambda^{(k)}\rangle - \langle \tilde{\alpha}^k_{l_n}|\alpha_{l_{k+1}}\rangle\langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle,$$

$$n=1,...,k,$$

$$\langle l_{k+1}|\lambda^{(k+1)}\rangle = \langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle, \tag{17}$$

with $\langle l_1|\lambda^{(1)}\rangle = \langle\alpha_{l_1}|\tilde{f}^o\rangle/\||\alpha_{l_1}\rangle\|^2$.

Proof. As discussed above [cf. Eq. (8)], given $k+1$ constraints the Lagrange multiplier vector minimizing the distance to the observed data $|f^o\rangle$ is a solution to the equation

$$|\lambda^{(k+1)}\rangle = (\hat{F}^\dagger_{k+1}\hat{F}_{k+1})^{-1}\hat{F}^\dagger_{k+1}|\tilde{f}^o\rangle. \tag{18}$$

Multiplying both sides of Eq. (18) by $\hat{F}_{k+1}$, we obtain

$$\hat{F}_{k+1}|\lambda^{(k+1)}\rangle = \hat{F}_{k+1}(\hat{F}^\dagger_{k+1}\hat{F}_{k+1})^{-1}\hat{F}^\dagger_{k+1}|\tilde{f}^o\rangle \tag{19}$$

and, since operator $\hat{F}_{k+1}(\hat{F}^\dagger_{k+1}\hat{F}_{k+1})^{-1}\hat{F}^\dagger_{k+1}$ is the orthogonal projector onto V_{k+1}, by using Eqs. (16) in (19) and expressing $\hat{F}_{k+1}$ as given in Eq. (9) we have

$$\sum_{n=1}^{k+1}|\alpha_{l_n}\rangle\langle l_n|\lambda^{(k+1)}\rangle = \sum_{n=1}^{k+1}|\alpha_{l_n}\rangle\langle\tilde{\alpha}^{k+1}_{l_n}|\tilde{f}^o\rangle. \tag{20}$$

We proceed now to performing the inner product of the two sides with each of the vectors $\langle\tilde{\alpha}^{k+1}_{l_n}|$, $n=1,...,k+1$. The biorthogonality property (15) then gives rise to the set of equations

$$\langle l_n|\lambda^{(k+1)}\rangle = \langle\tilde{\alpha}^{k+1}_{l_n}|\tilde{f}^o\rangle, \quad n=1,...,k+1 \tag{21}$$

so that, after using Eq. (14) in these equations, the recursive formula (17) follows. ■

In order to write in a convenient form the corresponding formula yielding $|p^{1/2(k+1)}\rangle$, let us first define an operator $\hat{\tilde{F}}_{k+1}$ in the fashion

$$\hat{\tilde{F}}_{k+1} = \sum_{n=1}^{k+1}|\tilde{\alpha}^{k+1}_{l_n}\rangle\langle l_n|, \tag{22}$$

which, by means of Eq. (14), can be recursively computed as

$$\hat{\tilde{F}}_{k+1} = \hat{\tilde{F}}_k + |\tilde{\tilde{\psi}}_{k+1}\rangle\langle l_{k+1}| - |\tilde{\tilde{\psi}}_{k+1}\rangle\langle\alpha_{l_{k+1}}|\hat{\tilde{F}}_k, \tag{23}$$

with $\tilde{\tilde{F}}_1 = |\alpha_{l_1}\rangle\langle l_1|/\||\alpha_{l_1}\rangle\|^2$. The recursive formula (17) for the Lagrange multipliers adopts, thereby, the form

$$\langle l_n|\lambda^{(k+1)}\rangle = \langle l_n|\lambda^{(k)}\rangle - \langle l_n|\hat{\tilde{F}}^\dagger_k|\alpha_{l_{k+1}}\rangle\langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle,$$

$$n=1,...,k,$$

$$\langle l_{k+1}|\lambda^{(k+1)}\rangle = \langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle, \tag{24}$$

with $\langle l_1|\lambda^{(1)}\rangle = \langle\alpha_{l_1}|\overline{f^o}\rangle/\||\alpha_{l_1}\rangle\|^2$.

Now, from Eqs. (6) and (24), we finally obtain the recursive formula for $|p^{1/2(k+1)}\rangle$ as given by

$$\langle n|p^{1/2(k+1)}\rangle = \langle n|p^{1/2(k)}\rangle - \frac{1}{N}\langle g|l_{k+1}\rangle\langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle + \frac{1}{N}\sum_{j=1}^{k}\langle g|l_j\rangle\langle l_j|\hat{\tilde{F}}^\dagger_k|\alpha_{l_{k+1}}\rangle\langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle + \langle n|\hat{A}^\dagger|l_{k+1}\rangle\langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle$$
$$-\sum_{j=1}^{k}\langle n|\hat{A}^\dagger|l_j\rangle\langle l_j|\hat{\tilde{F}}^\dagger_k|\alpha_{l_{k+1}}\rangle\langle\tilde{\tilde{\psi}}_{k+1}|\tilde{f}^o\rangle, \quad n=1,...,N. \tag{25}$$

Let us recall that the method for selecting the relevant data (indexes l_j) advanced in [1] is also able to yield both the vectors $|\psi_{k+1}\rangle$ and the vectors $|\alpha_{l_{k+1}}\rangle$ (see [1] for a sketch of the pertinent algorithm). Here we simply make use of the availability of these vectors so as to iteratively construct the operator $\hat{\tilde{F}}_k$ [given in Eq. (23)] in order to recursively (that is, at each and every stage of the iterative algorithm) compute the all-important Lagrange multipliers. As a final remark, we would like to stress that the recursive formula (24) for the Lagrange multipliers yields an iterative procedure to encode the data vector $|\tilde{f}^o\rangle \in \mathcal{R}^M$ into a vector of lower dimension $|\lambda^{(k)}\rangle \in \mathcal{R}^k$, where k is the number of relevant data (usually, we have $k \ll M$). The $|p^{1/2(k)}\rangle$ distribution provides us then with the corresponding decoding tool, via Eq. (2).

III. CONCLUSIONS

We have introduced a maximum entropy procedure for data selection that represents a significant improvement in reducing the computational cost for evaluating the associated maximum entropy Lagrange multipliers.

We build up the present approach upon the foundations developed in [1]. The data selection criterion of [1] is not affected in any way. Accordingly, when applying this methodology one must expect the results to be identical to those

produced by the algorithm developed in [1].

However, we are now in a position to avoid, at each iteration stage, the need of inverting an operator so as to obtain the Lagrange multipliers. This makes the technique especially appropriate when dealing with a large number of data. Additionally, we are now in a position to tackle a very important question: as a matter of fact, extremizing a nonextensive entropy does not guarantee the positiveness of the ensuing probability distribution. In particular, we cannot guarantee the positiveness of the $|p^{1/2}\rangle$ distribution given in Eq. (6). Actually, even if, using Eq. (6), one does obtain a non-negative distribution *from noiseless data*, the introduction of noise may certainly affect the positiveness property. The importance of the present approach for computing $|p^{1/2}\rangle$ lies in the fact that it allows one to remedy such a situation. Indeed, since we can *now* recursively modify the distribution when a new datum (constraint) is selected, we are in a position to disregard, at each iteration, constrains yielding a $|p^{1/2(k)}\rangle$ distribution which is not endowed with the property of positiveness.

From the above remarks we conclude that the new approach considerably widens the possible range of applications of the $q=1/2$ distributions for data selection and data compression.

ACKNOWLEDGMENT

Support from EPSRC (GR/R86355/01) is acknowledged.

[1] L. Rebollo-Neira and A. Plastino, Phys. Rev. E **65**, 011113 (2002).
[2] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[3] C. Tsallis, Fractals **6**, 539 (1995), and references therein.
[4] E. M. F. Curado and C. Tsallis, J. Phys. A **24**, L69 (1991); **24**, 3187 (1991); **25**, 1019 (1992).
[5] A. R. Plastino and A. Plastino, Phys. Lett. A **177**, 177 (1993).
[6] A. R. Plastino and A. Plastino, Phys. Lett. A **174**, 384 (1993).
[7] A. R. Plastino and A. Plastino, Phys. Lett. A **193**, 140 (1994).
[8] A. Plastino and A. R. Plastino, Braz. J. Phys. **29**, 50 (1999).
[9] C. Tsallis, Braz. J. Phys. **29**, 1 (1999), and references therein. An updated bibliography can be found in http://tsallis.cat.cbpf.br/biblio.htm
[10] B. M. R. Boghosian, Phys. Rev. E **53**, 4754 (1996).
[11] L. Rebollo-Neira, A. Plastino, and J. Fernandez-Rubio, Physica A **258**, 458 (1998).
[12] L. Rebollo-Neira, J. Fernandez-Rubio, and A. Plastino, Physica A **261**, 555 (1998).
[13] B. R. La Cour and W. C. Schieve, Phys. Rev. E **62**, 7494 (2000).
[14] L. Rebollo-Neira, e-print math-ph/0209026.