Spectrum Estimation Using Multirate Observations

Omid S. Jahromi, Bruce A. Francis and Raymond H. Kwong

Department of Electrical and Computer Engineering,

University of Toronto,

Toronto, Ontario, Canada M5S 3G4

Phone: (416) 978-6343 Fax: (416) 978-0804

Email: {omidj, francis, kwong}@control.toronto.edu

EDICS: 2-SPEC, 2-MWAV, 2-INFO

Corresponding author: Omid S. Jahromi

Abstract

This article considers merging the statistical information gained in low-rate measurements of a non-observable high-rate signal. We consider a model where a wide-sense stationary random signal $x(n)$ is being observed indirectly using several linear multirate sensors. Each sensor outputs a measurement signal $v_i(n)$ whose sampling rate is only a fraction of the sampling rate assumed for the original signal. We pose the following problem: Given certain autocorrelation coefficients of the observable signals $v_i(n)$, estimate the power spectral density of the original signal $x(n)$.

It turns out that this problem is ill-posed. We suggest to resolve this issue by using the principle of Maximum Entropy. We address technical difficulties associated with the Maximum Entropy solution and then devise a practical algorithm for its approximate computation. We demonstrate the viability of this algorithm through simulation examples.

I. INTRODUCTION

Spectral estimation is concerned with determining the distribution in frequency of the power of a random process. Questions such as "Does most of the power of the signal reside at low or high frequencies?" or "Are there resonance peaks in the spectrum?" are often answered as a result of a spectral analysis. Spectral analysis finds frequent and extensive use in many areas of physical sciences. Examples abound in oceanography, electrical engineering, geophysics, astronomy and hydrology.

Let $x(n)$ denote a zero-mean Gaussian wide-sense stationary (WSS) random process. It is well-known that a complete statistical description of such a process is provided by its *autocorrelation sequence* (ACS)

$$R_x(k) \stackrel{\triangle}{=} E\{x(n)x(n+k)\}$$

or, equivalently, by its *power spectral density* (PSD)

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_x(k)e^{-j\omega k}.$$

The autocorrelation sequence is a time-domain description of the second order statistics of a random process. The power spectral density provides a frequency domain description of the same statistics.

An issue of practical importance is how to *estimate* the power spectral density of a time series given a finite-length data record. This is not a trivial problem as reflected in a bewildering array of power spectrum estimation procedures, with each procedure claimed to have or show some optimum property[1]. The reader is referred to the excellent texts [3], [4], [5] and [6] for analysis of emperical spectrum estimation methods.

[1]The controversy is rooted in the fact that power spectral density is a probabilistic quantity and probabilistic quantities cannot be

In this paper, we are interested in estimating the power spectral density of a WSS random signal when the signal itself is not available but some low-resolution (low-sampling-rate) signals derived from it are observable (Fig. 1). This situation may arise in many practical applications, for instance, in photonics where it is expensive or even impossible to build a measurement sensor with high enough sampling rate [7]. Other potential areas of application are radar, non-uniform phased-array antenna, remote sensing and diversity wireless receivers.

This paper is organized as follows. The problem of multirate spectrum estimation is introduced and formulated in Section II. Section III introduces the Maximum Entropy method and describes its application to the problem of multirate spectral estimation. Section IV illuminates the geometry of the Maximum Entropy formulation through the use of operator theory notation and analogous examples in the Euclidean space. Section V is devoted to calculating the Maximum Entropy solution. Existence, uniqueness and stability properties of the Maximum Entropy solution are studied in Section VI. Computational issues associated with the Maximum Entropy solution are addressed in Section VII. Specifically, a practical algorithm called the Maximum Entropy Inference Engine is devised for computing an approximate solution. The performance of this algorithm is demonstrated by presenting two simulated examples in Section VIII. The paper ends with the concluding remarks in Section IX.

Notation: Vectors are denoted by capital letters. Boldface capital letters are used for matrices. Elements of a matrix $\mathbf{A}$ are referred to as $[\mathbf{A}]_{ij}$. We denote the set of real M-tuples by $\mathbb{R}^M$ and use the notation $\mathbb{R}_+$ for positive real numbers. The linear convolution operator is denoted by $\star$. The spaces of Lebesgue-measurable functions are represented by $\mathbf{L}^1(a,b)$, $\mathbf{L}^2(a,b)$, etc. The end of an example is indicated using the symbol $\diamondsuit$.

II. Formulating the problem

Consider the setup shown in Fig. 1. This setup models a *multirate observer* or *multirate sensor* system which outputs low-rate observations or *measurements* $v_i(n)$ of a non-observable signal $x(n)$. The linear filters in Fig. 1 model the bandwidth limitations of each sensor. Each filter is followed by a down-sampler device which models the (possible) difference in sampling rate among the sensors.

It is straightforward (e.g., [8]) to show that the observations $v_0(n)$ to $v_{M-1}(n)$ in Fig. 1 are also WSS processes whose ACS are given by

$$R_{v_i}(k) = R_{x_i}(N_i k) \tag{1}$$

constructed using finite-size sample records. Indeed, neither the axiomatic theory [1] nor the frequency theory [2] of probability specify a constructive way for building probability measures from empirical samples.

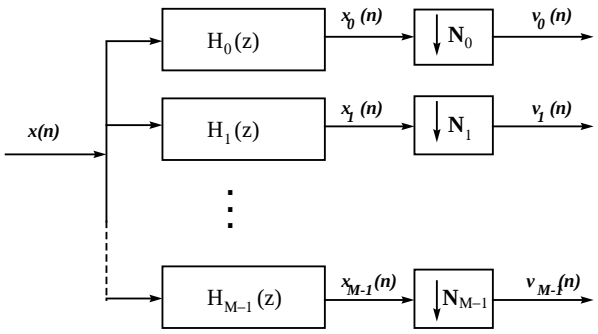


Fig. 1. An M-channel multirate observer system.

where

$$R_{x_i}(k) = (h_i(k) \star h_i(-k)) \star R_x(k), \tag{2}$$

and $h_i(k)$ denotes the impulse response of $H_i(z)$. We can express $R_{v_i}(k)$ as a function of the input PSD as well. Let's define $G_i(z) \stackrel{\triangle}{=} H_i(z)H_i(z^{-1})$. Now, (2) can be written as

$$R_{x_i}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jk\omega}d\omega. \tag{3}$$

From (1) and (3), we can then write

$$R_{v_i}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jN_ik\omega}d\omega. \tag{4}$$

Thus, knowledge of $P_x(e^{j\omega})$ uniquely specifies $R_{v_i}(k)$ for all values of i and k.

One might apply standard statistical methods and estimate some autocorrelation coefficients of the observable low-rate signals. Inferring statistics of the non-observable original signal is associated with the following inverse problem:

Problem 1 (Estimating $P_x(e^{j\omega})$ given statistics of $v_i(n)$) *Let a set of numbers $\{\rho_i(k),\ 0 \le i \le M-1,\ 0 \le k \le L-1\}$ be given. Knowing that these numbers are autocorrelation coefficients of the observable signals, estimate the power spectral density of the input signal. In other words, find $P_x(e^{j\omega})$ given that $R_{v_i}(k) = \rho_i(k)$ for $i = 0, \ldots, M-1$ and $k = 0, 1, \ldots, L-1$.*

It is easy to observe that, in general, knowing $R_{v_i}(k)$ for some limited values of i and k is not sufficient for characterizing $P_x(e^{j\omega})$ uniquely. That is, given a finite set of ACS values $R_{v_i}(k)$ there usually exist infinitely many $P_x(e^{j\omega})$ which can generate those values. Thus, Problem 1 is an ill-posed problem[2].

[2]A problem of mathematical physics is called well-posed (more precisely, well-posed in the sense of Hadamard) if i) for all

III. The Maximum Entropy principle

In any application of probability theory to the problems of science one meets the question of how to assess the probability of the occurrence of some event or the validity of a certain hypothesis. Although the mathematical formalism of probability theory serves as a powerful tool for addressing such problems, *it cannot by itself answer this question*. In fact, the formalism necessarily remains silent on this issue since its goal is only to provide theorems valid for all probability assignments allowed by its axioms. Recourse is thus necessary to an additional *inference rule* which tells us in which case one ought to assign which values to probabilities.

The Maximum Entropy (ME) principle, introduced in 1957 by E. T. Jaynes in the field of thermodynamics, provides an elegant way for resolving a wide class of statistical ill-posed problems [11], [12]. According to this principle, one should choose *the most random* solution that satisfies the known constraints. For the case of Problem 1, the ME principle asserts that $P_x(e^{j\omega})$ should be chosen such that (i) it is consistent with the known ACS values $R_{v_i}(k)$, and (ii) it associates with the most random input signal[3].

Applying the ME principle to spectral estimation is credited to J. P. Burg [17] who used it to estimate PSD of random signals given a limited number of their autocorrelation coefficients. Burg's solution is equivalent to an extrapolation of the autocorrelation function of the available time series which maximizes the entropy of the process. Here, we use the ME principle in a way similar to the way Burg did. However, Problem 1 is more complex than the standard autocorrelation extrapolation problem.

In the ME literature [12], [18], randomness is usually quantified using Shannon's entropy. The notion of entropy defined by Shannon for a scalar random variable is generalized to the notion of *entropy rate* for WSS random processes. If $x(n)$ is a Gaussian WSS process, its entropy rate is given by [19]

$$H(P_x) = \frac{1}{2}\ln 2\pi + \frac{1}{2} + \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) d\omega. \tag{5}$$

By definition, any WSS process $x(n)$ has finite power, that is, $E\{\hat{x}^2(n)\} < \infty$. This implies that $P_x(e^{j\omega})$

admissible data it has a solution, ii) the solution is unique and iii) the solution depends on the data continuously. Otherwise, the problem is considered ill-posed. For a detailed analysis of ill-posed problems and the corresponding *rgularization methods* see, e.g., [9] or [10].

[3] A thorough analysis of the pros and cons of the Maximum Entropy principle as *the* just rule of assigning probabilities is beyond the scope of this paper. For our purpose, it suffices to emphasize that this principle, as used in the context of resolving Problem 1, is not a physical principle in the proper sense, but should be interpreted as a guideline for statistical inference. The reader is referred to the excellent papers by Jaynes [11], [12], Csiszár [13], [14] and Uffink [15], [16] where the authors analyze the Maximum Entropy principle from opposing viewpoints.

is a (real-valued, even and non-negative) function in $\mathbf{L}^1(-\pi,\pi)$. Thus, the set of all power spectral density functions is a subset of

$$\mathcal{M} \stackrel{\triangle}{=} \{P_x(e^{j\omega}) :\ P_x \in \mathbf{L}^1(-\pi,\pi),\ P_x(e^{j\omega}) = P_x(e^{-j\omega}),\ P_x(e^{j\omega}) \geq 0\}. \tag{6}$$

We denote by $\mathcal{S}_i$ the set of all input PSDs which are consistent with the autocorrelation values obtained from the signal $v_i(n)$. That is

$$\mathcal{S}_i \stackrel{\triangle}{=} \left\{ P_x(e^{j\omega}) : \begin{array}{l} \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega}) G_i(e^{j\omega}) e^{jN_i k\omega} d\omega = \rho_i(k),\ 0 \leq k \leq L-1; \\ \\ P_x(e^{j\omega}) \in \mathcal{M} \end{array} \right\}. \tag{7}$$

We also define $\mathcal{Q}_i$ as the set of all input PSDs consistent with autocorrelation values known for *multiple* low-rate signals $v_0(n)$ to $v_i(n)$:

$$\mathcal{Q}_i \stackrel{\triangle}{=} \bigcap_{j=0}^{i} \mathcal{S}_j. \tag{8}$$

Thus, $P_x(e^{j\omega}) \in \mathcal{Q}_i$ means that it is consistent with the data (ACS values) available from *the first i channels*. Now, according to the ME principle, we should seek a solution for Problem 1 by solving the following *constrained optimization* problem:

Problem 2 (ME formulation of Problem 1) *Find* $P_x^*(e^{j\omega}) = \arg\max\ H(P_x)$ *subject to* $P_x \in \mathcal{Q}_{M-1}$.

The Maximum Entropy principle is closely related to another induction principle known as the *Minimum Cross-Entropy (MCE)* principle [20], [21]. Based on this latter principle, one starts with a prior guess and chooses, among all feasible solutions, the one that has minimum cross-entropy from the prior. Cross-entropy, also known as Kullback-Leibler divergence, is a popular quantity in statistical inference used to measure the degree by which two probability measures may be discriminated from each other. Let $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ be two candidate PSDs for the zero-mean Gaussian WSS random process $x(n)$. If $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ are essentially bounded from below[4], then the Kullback-Leibler divergence of $P_1(e^{j\omega})$ from $P_2(e^{j\omega})$ exists and is given by [19, Theorem 10.5.1] [22, Proposition 8.29]

$$D(P_1 \parallel P_2) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \left(\frac{P_1(e^{j\omega})}{P_2(e^{j\omega})} - \ln \frac{P_1(e^{j\omega})}{P_2(e^{j\omega})} - 1 \right) d\omega. \tag{9}$$

[4]This is to say, there is a positive constant ϵ such that $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ are both greater than ϵ for almost all ω. Note that if $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ are rational with no poles or zeroes on the unit circle, then this condition is satisfied.

Kullback-Leibler divergence has important distance-like properties which can be used to introduce interesting geometries on the space of PSD functions [23, Chapter 5]. However, it is not a metric since it does not treat $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ symmetrically, nor does it satisfy the triangle inequality.

The prior guess is important in the viability of MCE as an induction principle. When no information is available on the form of a PSD function, it is "reasonable" to assume that it is uniform. Thus, the uniform or flat spectrum $P_{flat}(e^{j\omega}) \stackrel{\triangle}{=} c$, $c > 0$, is a rational choice for the prior. However, there is no "reasonable way" to infer the value of c. Because of this ambiguity, the MCE principle cannot resolve Problem 1 unless we know the input signal's variance $E\{x^2(n)\}$ or, equivalently, $\|P_x\|_{\mathbf{L}^1}$ in advance. Once c is fixed, Problem 1 is converted to the following minimization problem:

Problem 3 (MCE formulation of Problem 1) *Find* $P_x^{\dagger}(e^{j\omega}) = \arg\min\ D(P_x \parallel P_{flat})$ *subject to* $P_x \in \mathcal{Q}_{M-1}$ *and* $\|P_x\|_{\mathbf{L}^1} = c$.

Note that, when $\|P_x\|_{\mathbf{L}^1}$ is fixed, minimizing $D(P_x \parallel P_{flat})$ is equivalent to maximizing $H(P_x)$, that is, the MCE and ME will give the same solution. Throughout this paper, we use the Maximum Entropy as our main induction principle and, hence, avoid the ambiguity that exists in the choosing a prior PSD for MCE. Never the less, we take advantage of the equivalence between MCE and ME under the condition mentioned above when we discuss properties of the ME solution in Section VI.

We will derive a parametric expressions for $P_x^*(e^{j\omega})$ later in Section V. Before that, however, we will examine the geometry of Problem 2 in order to provide more insight into the nature of the Maximum Entropy solution $P_x^*(e^{j\omega})$.

IV. Geometry of Problem 2

Let us re-formulate Problem 2 as a linear operator equation of Fredholm type. Define

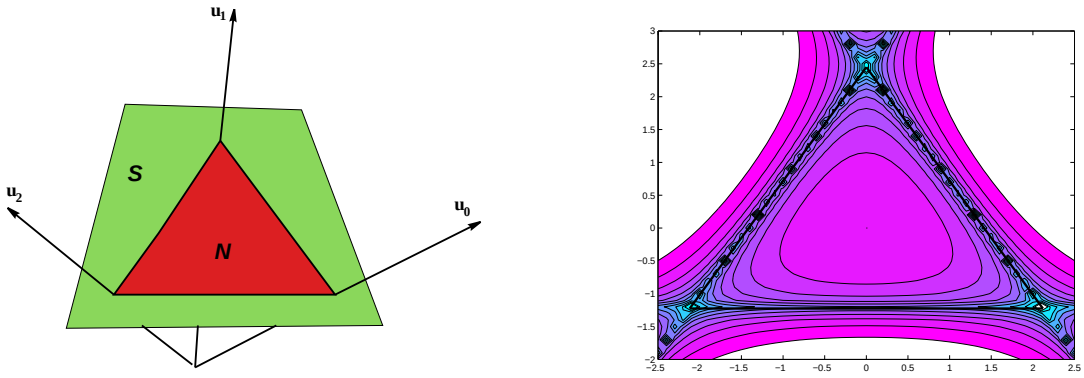
$$R \stackrel{\triangle}{=} [\rho_0(0), \cdots, \rho_0(L-1), \rho_1(0), \cdots \rho_1(L-1), \cdots \cdots, \rho_{M-1}(L-1)]^T$$

as the (long) vector containing all the known ACS values for the observable signals. Then, the constraint that $P_x(e^{j\omega})$ must be consistent with these given data, might be written as

$$AP_x = R, \tag{10}$$

where A is the linear integral operator from $\mathcal{M}$ to $\Sigma \subset \mathbb{R}^{ML}$ defined by

$$A \stackrel{\triangle}{=} [A_{0,0}, \cdots, A_{0,L-1}, A_{1,0}, \cdots A_{1,L-1}, \cdots \cdots, A_{M-1,L-1}]^T, \tag{11}$$



(a) (b)

Fig. 2. (a) Geometric visualization of the set $\mathcal{N}$ as a triangular region in the 3 dimensional (Euclidean) space. (b) Contours generated on this region by the Entropy function $\mathcal{H}(X) = \sum_{k=0}^{2} \ln x_k$. Those contours that lie outside the triangular region are irrelevant.

in which

$$A_{i,k}P_x \stackrel{\triangle}{=} \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jN_ik\omega}d\omega. \tag{12}$$

Using the above notation, we can write the constraint set $\mathcal{Q}_{M-1}$ that appears in Problems 2 and 3 as

$$\mathcal{Q}_{M-1} = \{P_x(e^{j\omega}) :\ P_x \in \mathcal{M},\ AP_x = R\}. \tag{13}$$

It is clear from the above definition that $\mathcal{Q}_{M-1}$ is a convex subset of $\mathbf{L}^1(-\pi,\pi)$ which, in turn, is a Banach space. The PSDs which are in $\mathcal{Q}_{M-1}$ can be pictured as points on the intersection of the linear manifold $\{P_x(e^{j\omega}) :\ P_x \in \mathbf{L}^1(-\pi,\pi), P_x(e^{j\omega}) = P_x(e^{-j\omega}),\ AP_x = R\}$ and the closed convex cone $\{P(e^{j\omega}) :\ P \in \mathbf{L}^1(-\pi,\pi),\ P(e^{j\omega}) \geq 0\}$. The following toy examples illustrate the above concepts using their analogs in the Euclidean space $\mathbb{R}^3$:

Example 1 Let $X = [x_0\ x_1\ x_2]^T$ be a vector in the Euclidean space $\mathbb{R}^3$ and define $\mathcal{N} = \{X \mid X_0^T X = 3, X \geq 0\}$ where $X_0 = [1\ 1\ 1]^T$ is fixed. The equation $X_0^T X = 3$ specifies a plane surface that we call $\mathcal{S}$ (see Fig. 2(a)). The set $\mathcal{N}$ simply represents those vectors (points) in $\mathcal{S}$ which are in the first quadrant. Note that the first quadrant is, in fact, a convex cone representing the condition $X \geq 0$. $\diamondsuit$

Example 2 Consider the set $\mathcal{N} \subset \mathbb{R}^3$ and the associated geometry introduced in the previous example. The contours generated on this set by the entropy function $\mathcal{H}(X) = \sum_{k=0}^{2} \ln x_k$ are shown in Fig. 2(b). Clearly, entropy increases as we approach the centre of the triangle that represents $\mathcal{N}$ and decreases as we approach its borders. The Entropy is $-\infty$ on the side of this triangle.

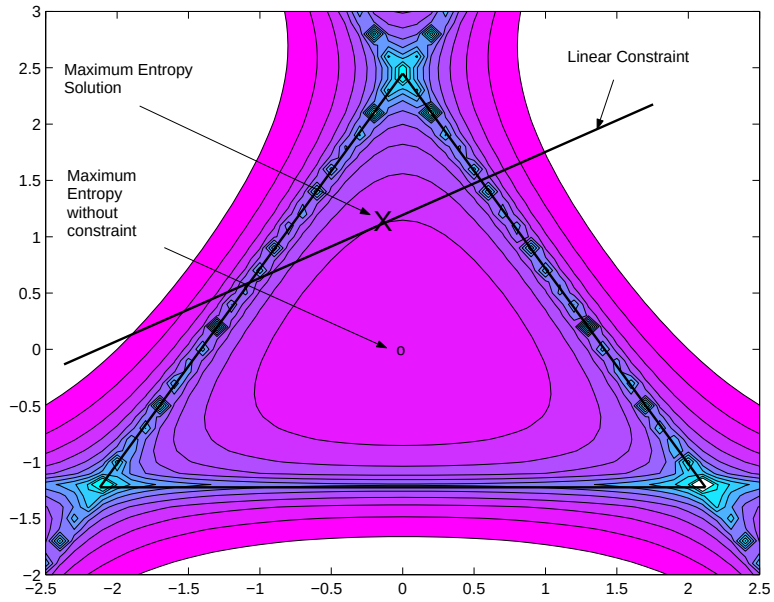


Fig. 3. A linear constraint will specify a line in $\mathbb{R}^3$. The set of feasible solutions is represented by the segment of this line which intersects with the triangular region that represents $\mathcal{N}$. The Maximum Entropy principle chooses a point on this segment which has the highest entropy. As entropy increase for points close the centre of the triangle $\mathcal{N}$, the selected point will be "close" to the centre.

Any linear constraint will specify a line in $\mathcal{N}$. The Maximum Entropy principle chooses a point on this line which i)lies on the triangle $\mathcal{N}$ and ii)has the highest entropy (Fig. 3). $\diamondsuit$

V. Solving Problem 2

Theorem 1 *If Problem 2 has a continuous solution, it is in the form*

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} G_i(e^{j\omega}) F_i(e^{jN_i\omega})}, \tag{14}$$

where $F_i(z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$. *The coefficients* λ_{ik} *of the transfer functions* $F_i(z)$ *in (14) are specified such that*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{G_i(e^{j\omega}) e^{jN_i k\omega}}{\sum_{j=0}^{M-1} G_j(e^{j\omega}) F_j(e^{jN_j\omega})} d\omega = \rho_i(k) \tag{15}$$

holds for all values of $i = 0, 1, \cdots M-1$ *and* $k = 0, 1, \cdots L-1$.

Proof: Using standard techniques from calculus of variations, we can convert Problem 2 to an unconstrained

optimization problem. To do so, we form the functional

$$J \stackrel{\triangle}{=} \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln P_x^*(e^{j\omega})d\omega + \sum_{i=0}^{M-1}\sum_{k=-(L-1)}^{L-1} \lambda_{ik}\left[\rho_i(k) - \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x^*(e^{j\omega})G_i(e^{j\omega})e^{jN_ik\omega}d\omega\right]$$

where λ_{ik} are the Lagrange multipliers. Note that since both $\rho_i(k)$ and $R_{v_i}(k)$ are symmetric in k, we have $\lambda_{ik} = \lambda_{i(-k)}$. Taking the variation of J with respect to P_x^* we get

$$\delta J = \frac{1}{4\pi}\int_{-\pi}^{\pi} \delta P_x^*(e^{j\omega})\frac{1}{P_x^*(e^{j\omega})}d\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\delta P_x^*(e^{j\omega})\left[\sum_{i=0}^{M-1}\sum_{k=-(L-1)}^{L-1}\lambda_{ik}G_i(e^{j\omega})e^{jN_ik\omega}\right]d\omega.$$

Putting δJ equal to zero[5] leads to

$$\int_{-\pi}^{\pi}\left[\frac{1}{P_x^*(e^{j\omega})} - 2\sum_{i=0}^{M-1}G_i(e^{j\omega})\left(\sum_{k=-(L-1)}^{L-1}\lambda_{ik}e^{jN_ik\omega}\right)\right]\delta P_x^*(e^{j\omega})d\omega = 0.$$

Since $\delta P_x^*(e^{j\omega})$ is arbitrary, the quantity in the square brackets must be zero. Therefore, we obtain

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1}G_i(e^{j\omega})\left(\sum_{k=-(L-1)}^{L-1}2\lambda_{ik}e^{jN_ik\omega}\right)}. \tag{16}$$

It's convenient to define $F_i(z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1}2\lambda_{ik}z^{-k}$ such that (16) can be written in the compact form

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1}G_i(e^{j\omega})F_i(e^{jN_i\omega})}. \tag{17}$$

Finally, λ_{ik} are obtained by the requirement that $P_x^*(e^{j\omega})$ must satisfy the constraints, i.e. $P_x^*(e^{j\omega}) \in \mathcal{Q}_{M-1}$. Thus, in view of (7) and (8), $F_i(z)$ must be chosen such that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{G_i(e^{j\omega})e^{jN_ik\omega}}{\sum_{j=0}^{M-1}G_j(e^{j\omega})F_j(e^{jN_j\omega})}d\omega = \rho_i(k) \tag{18}$$

holds for all values of $i = 0, 1, \cdots M-1$ and $k = 0, 1, \cdots L-1$. **Q.E.D.**

Remark 1 *The transfer function $F_i(e^{j\omega})$ is a real function of ω (since $P_x^*(e^{j\omega})$ is real). Thus, we might use the simpler expression $F_i(e^{j\omega}) = \sum_{k=0}^{L-1}\lambda_{ik}\cos(k\omega)$ in (14) and (15).*

[5] We'll show in the next section that the functional $H(P_x)$ is strictly concave on its domain of definition. Thus, putting the first variation δJ of the functional J is sufficient for obtaining a global maximizer.

VI. Properties of the solution of Problem 2

Studying various measures of entropy for regularizing ill-posed density estimation problems has been an active area of research in applied mathematics. See, for example, [24], [25], [26], [27], [9] and [28]. Such problems usually lead to an optimization problem of the form

$$\begin{aligned} &\text{minimize} \int_a^b \phi(P(\omega))d\omega,\ P(\omega) \in \mathbf{L}^1(a,b) \\ &\text{subject to} \int_a^b G_i(\omega)P(\omega)d\omega = \alpha_i, \end{aligned} \tag{19}$$

where $i = 0, 1, \cdots, N-1$ indexes the constraints, $G_i(\omega)$ are known functions and α_i are the measured moments. The function $\phi(\cdot)$ reflects the choice of the entropy measure. The two classical choices correspond to the (negative) Shannon entropy

$$\phi(P) \stackrel{\triangle}{=} P \log P$$

and (negative) entropy rate

$$\phi(P) \stackrel{\triangle}{=} -\log P$$

which is sometimes called the Burg entropy.

Well-posedness of the minimization problem (19) when ϕ is the Shannon entropy has been established in several works including [24], [29], [25], [27] and [30]. Well-posedness results exist for certain classes of functions ϕ as well [30], [28]. Unfortunately, the entropy rate $\phi(P) = -\log P$ which is the one used in this paper does not belong to these classes. The reason is that a crucial property of the classes of regularizing functions ϕ studied in both [30] and [28] is $\lim_{P\to\infty} \frac{\phi(P)}{P} = \infty$, which is not satisfi ed by $\phi(P) = -\log P$.

While entropy rate (5) lacks theoretical support as a regularizing functional, it has been used extensively and successfully in many areas of application including spectral estimation and image restoration (see, for example, [18] and references therein). The very important feature of using entropy rate in the context of our work is that its minimization (subject to linear constraints $P_x^*(e^{j\omega}) \in \mathcal{Q}_{M-1}$) leads to a rational PSD function. This is very desirable since the theory of WSS processes with rational spectra is well known.

In the rest of this section we present some results on uniqueness, existence and stability of the solution of Problem 2.

Lemma 2 *When it exists, the solution to Problem 2 is unique.*

Proof: The proof is based on the fact that the entropy functional $H(\cdot)$ is concave. To see this, let $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ be two distinct PSDs in the constraint set $\mathcal{Q}_{M-1}$. Defi ne $P_\alpha(e^{j\omega}) = \alpha P_1(e^{j\omega}) + (1-\alpha)P_2(e^{j\omega})$.

It is easy to check that $\mathcal{Q}_{M-1}$ is a convex set. Thus $P_\alpha(e^{j\omega}) \in \mathcal{Q}_{M-1}$ for $0 \leq \alpha \leq 1$ as well. Now, we define

$$H(\alpha) \stackrel{\triangle}{=} H(P_\alpha) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln(\alpha P_1(e^{j\omega}) + (1-\alpha)P_2(e^{j\omega}))d\omega.$$

It is easy to check that

$$\frac{\partial^2 H(\alpha)}{\partial \alpha^2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{-(P_1(e^{j\omega}) - P_2(e^{j\omega}))^2}{(P_\alpha(e^{j\omega}))^2} d\omega < 0.$$

The above expression shows that $H(\cdot)$ is a (strictly) concave functional which, in turn, proves that its maximum over $\mathcal{Q}_{M-1}$, when it exists, is unique. **Q.E.D.**

An extensive analysis of the problem of existence of a Maximum Entropy solution is given in [31]. The results in [31] show that, if the Lagrange multipliers λ_{ik} are specified such that $P^*(e^{j\omega})$ given in (14) is both consistent with the constraints and positive, then it *is* the solution. However, if $P^*(e^{j\omega})$ obtained in Theorem 1 has any zeros, then the solution to Problem 2 may have a singular part as well. We do not know if the solution given by Theorem 1 is positive for all feasible data $\rho_i(k)$. Because of this and other difficulties, we shall use a modified parameterization of $P_x^*(e^{j\omega})$ in our computational algorithms. This modified parameterization is an approximation but guarantees positivity of the solution (see Section VII-A).

Another important issue is the stability of $P_x^*(e^{j\omega})$ with respect to small perturbations in the data $\rho_i(k)$. Unfortunately, continuous dependence of $P_x^*(e^{j\omega})$ on the data R is not formally established. Nonetheless, there has been some progress in finding bounds on the magnitude of perturbations of the MCE solution. For completeness, we discuss some of the contributions here.

Consider the functional

$$W(P_x) \stackrel{\triangle}{=} \|AP_x - R\| + \alpha^2 D(P_x \parallel P_{flat}), \tag{20}$$

where α is a small constant. Minimizing this functional can be regarded as an approximation to Problem 3 which, in turn, is equivalent to Problem 2 when the input signal's variance is known.

Let R_δ denote a perturbed data vector (which might, for example, be obtained by a *noisy measurement* of the true ACS values that constitute the data vector R) and assume that $\|R_\delta - R\| \leq \delta$. The following theorem, essentially due to Eggermont [27], establishes a bound on the Kullback-Leibler divergence between the solutions of the original problem and the perturbed one:

Theorem 3 *Let R and R_δ be two admissible data vectors such that $\|R_\delta - R\| \leq \delta$. Also, let P, P_δ and P_0 denote three PSD functions in $\mathcal{M}$. If P minimizes the functional $W(P) = \|AP - R\| + \alpha^2 D(P \parallel P_0)$ over*

$\mathcal{M}$ and P_δ minimizes the functional $W_\delta(P) = \|AP - R_\delta\| + \alpha^2 D(P \parallel P_0)$ again over $\mathcal{M}$, then

$$\|A(P - P_\delta)\| + \alpha^2 D(P \parallel P_\delta) \leq 4\delta^2. \tag{21}$$

Proof: The above result follows from Theorem 3.3 in [27] after the following observations:

(i) The constraint set $\mathcal{M}$ is closed and convex and hence satisfi es the conditions mentioned in [27, Section 3].

(ii) The Kulback-Leibler divergence $D(P_1 \parallel P_2)$ possesses the decomposition required by equation (3.2) in [27] provided that we replace $d(\cdot)$ in that equation by the convex functional $d(P) = \int_{-\pi}^{\pi} \ln\cosh P(e^{j\omega}) d\omega$. **Q.E.D.**

The above theorem shows that the Kullback-Leibler divergence between the solution of the perturbed problem and that of the original problem is bounded by $\frac{4}{\alpha^2}\|R_\delta - R\|^2$. Since the Kullback-Leibler divergence is not a metric, Theorem 3 alone does not imply continuous dependence of the solution of Problem 2 on the data. To obtain a formal continuity result, we have to show that the distance (in the metric of some valid topology) between the solution of the perturbed problem and that of the original problem is bounded when $\|R_\delta - R\|$ is bounded. The following theorem shows that this can be done provided that we specify a bound on the peak value of $P_x(e^{j\omega})$.

Theorem 4 *Let $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ denote two continuous PSD functions. Then, $\|P_1 - P_2\|_2^2 \leq cD(P_1 \parallel P_2)$ where $c = \frac{8\pi}{3}(\|P_2\|_\infty^2 + 2\|P_1\|_\infty\|P_2\|_\infty)$.*

Proof: We start with the elementary inequality of Kemperman [27], [32, Page 63]

$$(x - y)^2 \leq \left(\frac{2}{3}x + \frac{4}{3}y\right)\left(x\log(\frac{x}{y}) + y - x\right).$$

Using the above inequality and noticing that $P_2(e^{j\omega}) > 0$ almost everywhere, we get

$$(P_1(e^{j\omega}) - P_2(e^{j\omega}))^2 \leq \left(\frac{2}{3}P_2^2(e^{j\omega}) + \frac{4}{3}P_1(e^{j\omega})P_2(e^{j\omega})\right)\left(\frac{P_1(e^{j\omega})}{P_2(e^{j\omega})} - \log(\frac{P_1(e^{j\omega})}{P_2(e^{j\omega})}) - 1\right).$$

Then, by using the inequality $|\int f(t)g(t)dt| \leq \|f\|_\infty\|g\|_1$, we may write

$$\|P_1(e^{j\omega}) - P_2(e^{j\omega})\|_2^2 \leq \frac{2}{3}\|P_2^2(e^{j\omega}) + 2P_1(e^{j\omega})P_2(e^{j\omega})\|_\infty \int_{-\pi}^{\pi}\left(\frac{P_1(e^{j\omega})}{P_2(e^{j\omega})} - \log(\frac{P_1(e^{j\omega})}{P_2(e^{j\omega})}) - 1\right)d\omega.$$

Finally, we note that the integral in the right-hand side is 4π times $D(P_1 \parallel P_2)$ and that continuity of $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ allows changing the order of $\mathbf{L}^\infty$ norm and multiplication operators. **Q.E.D.**

Combining the results of Theorems 3 and 4, we conclude that if an upper bound M exists such that $\|P_\delta\|_\infty \leq M, \forall \delta$, then

$$\|R - R_\delta\| \to 0 \implies \|P - P_\delta\|_2 \to 0. \tag{22}$$

In other words, the Maximum Entropy solution is stable provided that $\|P_\delta\|_\infty$ is bounded a priori.

The Maximum Entropy PSDs given by (17) are continuous rational functions of $e^{j\omega}$. Hence, the $\mathbf{L}^\infty$ norm of P (the solution of the exact problem) is known. Also, P_δ is in $\mathbf{L}^\infty(-\pi, \pi)$ provided that it has no poles on the unit circle. However, this is not enough for the implication in (22) to be valid. What we need is an a priori bound on the $\mathbf{L}^\infty$ norm of P_δ. Whether such a bound can be actually specifi ed remains an open problem for research.

VII. Algorithms for computing an approximate solution

We showed in Section V that the continuous solution to Problem 2, when it exists, is given by

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} |H_i(e^{j\omega})|^2 F_i(e^{jN_i\omega})}, \tag{23}$$

where $F_i(z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$ and λ_{ik} are Lagrange multipliers[6]. The Lagrange multipliers embedded in $F_i(z)$ need to be specifi ed such that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|H_i(e^{j\omega})|^2 e^{jN_i k\omega}}{\sum_{j=0}^{M-1} |H_j(e^{j\omega})|^2 F_j(e^{jN_j\omega})} d\omega = \rho_i(k) \tag{24}$$

holds for all values of $i = 0, 1, \cdots M-1$ and $k = 0, 1, \cdots L-1$.

In practice, it is quite likely that the set of feasible power spectra $\mathcal{Q}_{M-1}$ is empty for a collection of $\rho_i(k)$ found or estimated empirically. In this case, the system of equations (24) does not have a solution. To get around this diffi culty, we generalize our notion of a "Maximum Entropy solution" by identifying it with the solution to the following least-squares problem:

Problem 4 (Least-squares method for specifying the parameters of $P_x(e^{j\omega})$)

Let $\Lambda \stackrel{\triangle}{=} [\lambda_{00}\ \lambda_{01}\ \ldots\ \lambda_{(M-1)(L-1)}]^T$. *Find* Λ *that minimizes*

$$J(\Lambda) = \sum_{i=0}^{M-1} \sum_{k=0}^{L-1} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|H_i(e^{j\omega})|^2 e^{jN_i k\omega}}{\sum_{j=0}^{M-1} |H_j(e^{j\omega})|^2 F_j(e^{jN_j\omega})} d\omega - \rho_i(k) \right)^2, \tag{25}$$

where $F_i(z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$.

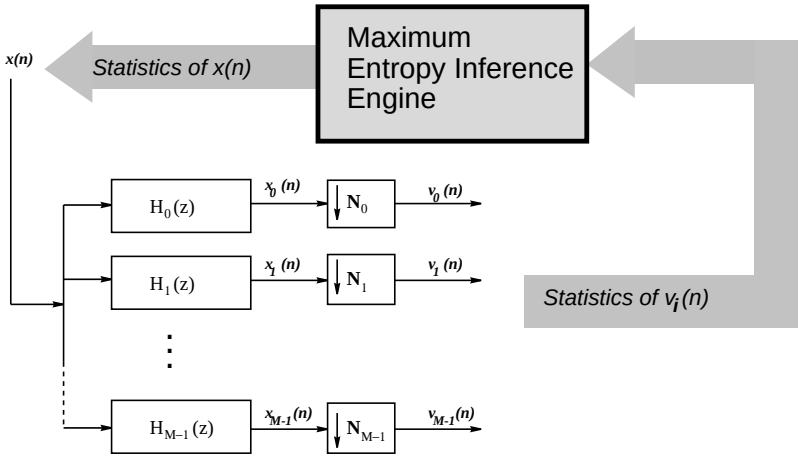


Fig. 4. MEIE is an algorithm that calculates an approximation to the Maximum Entropy estimate of the input PSD given autocorrelation coeffi cients of the low-rate observations.

If the system of equations (24) has a solution, it will be obtained by solving Problem 4. Even when (24) doesn't have a solution, solving Problem 4 specifi es a PSD which approximately satisfi es the constraints in the least-squares sense. Thus, in practice, we specify λ_{ik} by solving Problem 4 instead of the system of equations in (24).

Applying any standard optimization routine to solve Problem 4 requires calculating the integrals involved in the objective function J. It turns out that numerical calculation of these integrals is very diffi cult, mainly because the integrand tends to blow-up whenever any $H_j(e^{j\omega})$ gets close to zero. In the next subsection, we suggest a modifi ed way of parameterizing $P_x(e^{j\omega})$. This enables us to calculate J without having to perform ill-conditioned numerical integrations.

A. Avoiding numerical integrations

It is well known that every rational PSD function $P_x(e^{j\omega})$ admits a spectral factorization of the form

$$P_x(e^{j\omega}) = |Q(e^{j\omega})|^2, \tag{26}$$

or, if we replace $e^{j\omega}$ by the complex variable z,

$$P_x(z) = Q(z)Q(z^{-1}), \tag{27}$$

[6]Recall from Section V that $\lambda_{ik} = \lambda_{i(-k)}$. This means there are only L coeffi cients to be specifi ed for each $F_i(z)$.

where $Q(z)$ represents the transfer function of a causal and stable system [5], [4]. Let us define the parameterized family of PSD candidates

$$P_x(\Lambda; e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} |H_i(e^{j\omega})|^2 F_i(\Lambda; e^{jN_i\omega})},$$

where the parameter vector $\Lambda \stackrel{\triangle}{=} [\lambda_{00}\ \lambda_{01}\ \ldots\ \lambda_{(M-1)(L-1)}]^T$ is embedded in the (non-causal) FIR transfer function $F_i(\Lambda; z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$. Also, let us use the notation $Q(\Lambda; z)$ for the parameterized family of spectral factors associated with $P_x(\Lambda; e^{j\omega})$ when the spectral factors exist.

For any Λ, if $Q(\Lambda; z)$ exists, it can be identified with the modeling filter of an ARMA process whose order depends on the degree and the nature of the analysis filters $H_i(z)$, the down-sampling ratios N_i and the number of ACS values available as data. WSS processes can be represented by their modeling filter driven by (unit-variance) white noise. If we apply this fact to the input signal $x(n)$ in Fig. 1, then we can calculate the ACS values $R_{v_i}(k)$ of the observable signals $v_i(n)$:

Define $q(\Lambda; k)$ as the impulse response of the modeling filter $Q(\Lambda; z)$ and $h_i(k)$ as the impulse response of the analysis filter $H_i(z)$. Then, we have

$$R_{x_i}(k) = (h_i(k) \star h_i(-k)) \star (q(\Lambda; k) \star q(\Lambda; -k)), \tag{28}$$

from which we can calculate

$$R_{v_i}(k) = R_{x_i}(N_i k). \tag{29}$$

The integrals in (25) are nothing but $R_{v_i}(k)$ calculated for a certain value of the parameter Λ. Thus, (28) and (29) enable one to calculate J without integration. There is a price, however, to be paid for this convenience: In order to use (28), we have to know the functional form of the spectral factors $Q(\Lambda; z)$.

In general, $P_x(\Lambda; e^{j\omega})$ is not positive definite for all Λ. Thus, $Q(\Lambda; z)$ does not exist for all Λ. To circumvent this difficulty, we modify the functional form of $F_i(\Lambda; z)$ such that

$$F_i(\Lambda; z) = A_i(z) A_i(z^{-1}) \text{ where } A_i(z) \stackrel{\triangle}{=} \sum_{k=0}^{L-1} \lambda_{ik} z^{-k}. \tag{30}$$

In other words, we make the transfer functions $F_i(\Lambda; e^{jN_i\omega})$ positive definite a priori. This ensures that $P_x(\Lambda; e^{j\omega}) \geq 0$ for all Λ[7].

Although the number of parameters remains the same, the parameterizations (30) is not equivalent to the original one where $F_i(\Lambda; z) = \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$. Nevertheless, backed by the results of our numerical

[7]Note that $F_i(\Lambda; e^{jN_i\omega}) > 0$ implies $P_x(\Lambda; e^{j\omega}) > 0$ but the converse need not be true.

simulations, we believe that the new parameterization (30) has enough degrees of freedom to capture the essence of the original case for practical applications. We leave it as an open problem to investigate, theoretically, how well this new parameterization approximates the original one and/or find some bounds on the error caused by (possible) lack of enough degrees of freedom.

B. The Maximum Entropy Inference Engine

We use the name *Maximum Entropy Inference Engine (MEIE)* for the algorithm that calculates $P_x^*(e^{j\omega})$ using the procedure outlined in the previous section (Fig. 4). The steps involved are summarized in Algorithms 1 and 2 below.

Algorithm 1 (Maximum Entropy Inference Engine)

Input:

1. $\rho_i(k) = E\{v_i(n)v_i(n+k)\}$ `for` $i = 0, \cdots, M-1$ `and` $k = 0, \cdots L-1$.
2. `Transfer functions` $H_i(z)$ `and down-sampling ratios` N_i.

Output:

`An estimate` $P_x^*(e^{j\omega})$ `of the input PSD.`

Procedure:

1. `Let` $\Lambda \stackrel{\triangle}{=} [\lambda_{00} \;\; \lambda_{01} \;\; \cdots \;\; \lambda_{(M-1)(L-1)}]^T$.
2. `Import` $g_{i,k}(\Lambda)$ `from Algorithm 2.`
3. `Form` $J(\Lambda) = \sum_{i=0}^{M-1} \sum_{k=0}^{L-1} \left(g_{i,k}(\Lambda) - \rho_i(k)\right)^2$.
4. `Find` $\Lambda^* \stackrel{\triangle}{=} \arg\min J(\Lambda)$.
5. `Form` $A_i(z) = \sum_{k=0}^{L-1} \lambda_{ik} z^{-k}$.
6. `Return` $P_x^*(e^{j\omega}) = \dfrac{1}{\sum_{i=0}^{M-1} |H_i(e^{j\omega})|^2 A_i(e^{N_i\omega}) A_i(e^{-N_i\omega})}$.

Remark 2 *Several steps in Algorithm 2 require operations on infinite impulse-response sequences. Such operations, of course, can only be performed approximately after the sequence is truncated to a reasonable length.*

Remark 3 *We use the MATLAB function* `fminu` *for performing the optimization in Step 4, Algorithm 1. The spectral factorization required in Algorithm 2, Step 4, may be performed, after truncating* $e(n)$ *to a reasonable length, by the MATLAB function* `ac2poly`.

Algorithm 2 (calculates $g_{ik}(\Lambda)$ for Algorithm 1)

Input:

1. Two integers i and k; $0 \le i \le M-1$, $0 \le k \le L-1$.
2. The parameter vector $\Lambda \stackrel{\triangle}{=} [\lambda_{00} \;\; \lambda_{01} \;\; \ldots \;\; \lambda_{(M-1)(L-1)}]^T$.
3. Transfer functions $H_i(z)$ and down-sampling ratios N_i.

Output:

The functions $g_{i,k}(\Lambda)$ used in Algorithm 1.

Procedure:

1. For $0 \le j \le M-1$, form

$$A_j(z) = \sum_{l=0}^{L-1} \lambda_{jl} z^{-l},$$

$$B_j(z) = H_j(z) A_j(z^{N_j}),$$

$$C_j(z) = B_j(z) B_j(z^{-1}).$$

2. Form $D(z) = \sum_{j=0}^{M-1} C_j(z)$.
3. Find the impulse response $d(n)$ of $D(z)$.
4. Find $e(n)$ such that $d(n) = e(n) \star e(-n)$.
5. Form $E(z) = \dfrac{1}{\sum_{n=0}^{\infty} e(n) z^{-n}}$ then calculate $U_i(z) = E(z) H_i(z)$.
6. Find the impulse response $u_i(n)$ of $U_i(z)$.
7. Return $g_{i,k}(\Lambda) = \displaystyle\sum_{n=-\infty}^{\infty} u_i(n) u_i(N_i k + n)$.

VIII. EXAMPLES

In this section we demonstrate the performance of the Maximum Entropy Inference Engine by providing two simulated examples[8].

Example 3 Consider a 4-channel multirate observer system of the form shown in Fig. 1. Assume that the down-sampling ratio is equal to 4 in all channels. Thus, $N_0 = N_1 = N_2 = N_3 = 4$. The transfer functions of the measurement filter $H_0(z)$ to $H_3(z)$ are chosen as follows:

$$H_0(z) = \frac{0.0753 + 0.1656z^{-1} + 0.2053z^{-2} + 0.1659z^{-3} + 0.0751z^{-4}}{1.0000 - 0.8877z^{-1} + 0.6738z^{-2} - 0.1206z^{-3} + 0.0225z^{-4}}$$

$$H_1(z) = \frac{0.4652 - 0.1254z^{-1} - 0.3151z^{-2} + 0.0975z^{-3} - 0.0259z^{-4}}{1.0000 - 0.6855z^{-1} + 0.3297z^{-2} - 0.0309z^{-3} + 0.0032z^{-4}}$$

$$H_2(z) = \frac{0.3732 - 0.8648z^{-1} + 0.7139z^{-2} - 0.1856z^{-3} - 0.0015z^{-4}}{1.0000 - 0.5800z^{-1} + 0.5292z^{-2} - 0.0163z^{-3} + 0.0107z^{-4}}$$

$$H_3(z) = \frac{0.1931 - 0.4226z^{-1} + 0.3668z^{-2} - 0.0974z^{-3} - 0.0405z^{-4}}{1.0000 + 0.2814z^{-1} + 0.3739z^{-2} + 0.0345z{-}3 - 0.0196z^{-4}}$$

The frequency response of the measurement filters are shown in Fig. 5. We emphasize that the above transfer functions were designed to show typical low-pass, band-pass and high-pass characteristics and were obtained using standard filter design techniques. In other words, they were not specifically designed to

[8]MATLAB codes for the simulation examples presented in this section are available online at `http://www.multirate.org`.

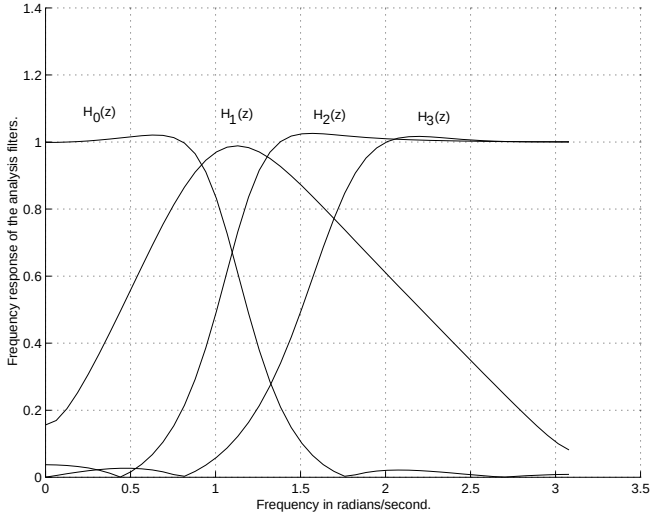


Fig. 5. Frequency response of the analysis filters used in Example 3.

belong to certain classes of filters (e.g., Quadrature Mirror Filters, Power Complementary Filters, Perfect Reconstruction Filters, Paraunitary Filters, etc.) used in classical filter bank designs [33].

The input signal $x(n)$ is chosen to be a low-pass Gaussian ARMA process whose modelling filter is of the form

$$Q_x(z) = \frac{\sum_{i=0}^{10} a(i) z^{-i}}{\sum_{j=0}^{10} b(j) z^{-j}}.$$

The coefficients $a(i)$ and $b(i)$ were calculated using the MATLAB command

```
[a,b]=YULEWALK( 10, [0 .5 .8 1], [1 1 0 0])
```

which implements the Yule-Walker filter design algorithm. For the PSD described above, we can calculate the ACS of the observable signals using (28) and (29). By calculating ACS values up to lag 4 (that is, calculating $\rho_i(k)$ for all i and $0 \leq k \leq 4$), we get the numbers in Table I. These numbers will be used as data by the MEIE to estimate the input PSD.

The actual input PSD and the PSD estimated by the MEIE are shown in Fig. 6. As seen in this figure, the MEIE has estimated $P_x(e^{j\omega})$ very accurately using the low-rate ACS values supplied. For the sake of comparison, we have also calculated intermediate estimates of the input PSD using a partial number of ACS values given in Table I. The resulting PSDs are plotted in Fig. 7. As seen in Fig. 7, the quality of estimation increases as more data are provide to the MEIE. This, however, is not a general rule as we will demonstrate in the next example. ♢

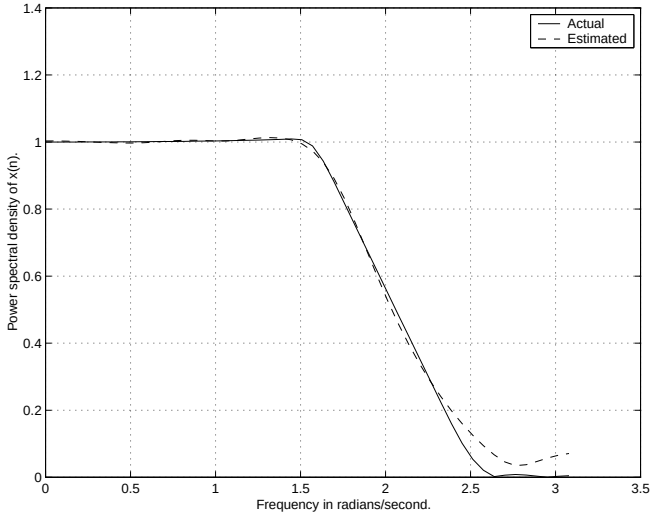


Fig. 6. PSD of the input signal used in Example 3 along with its estimate obtained by the MEIE.

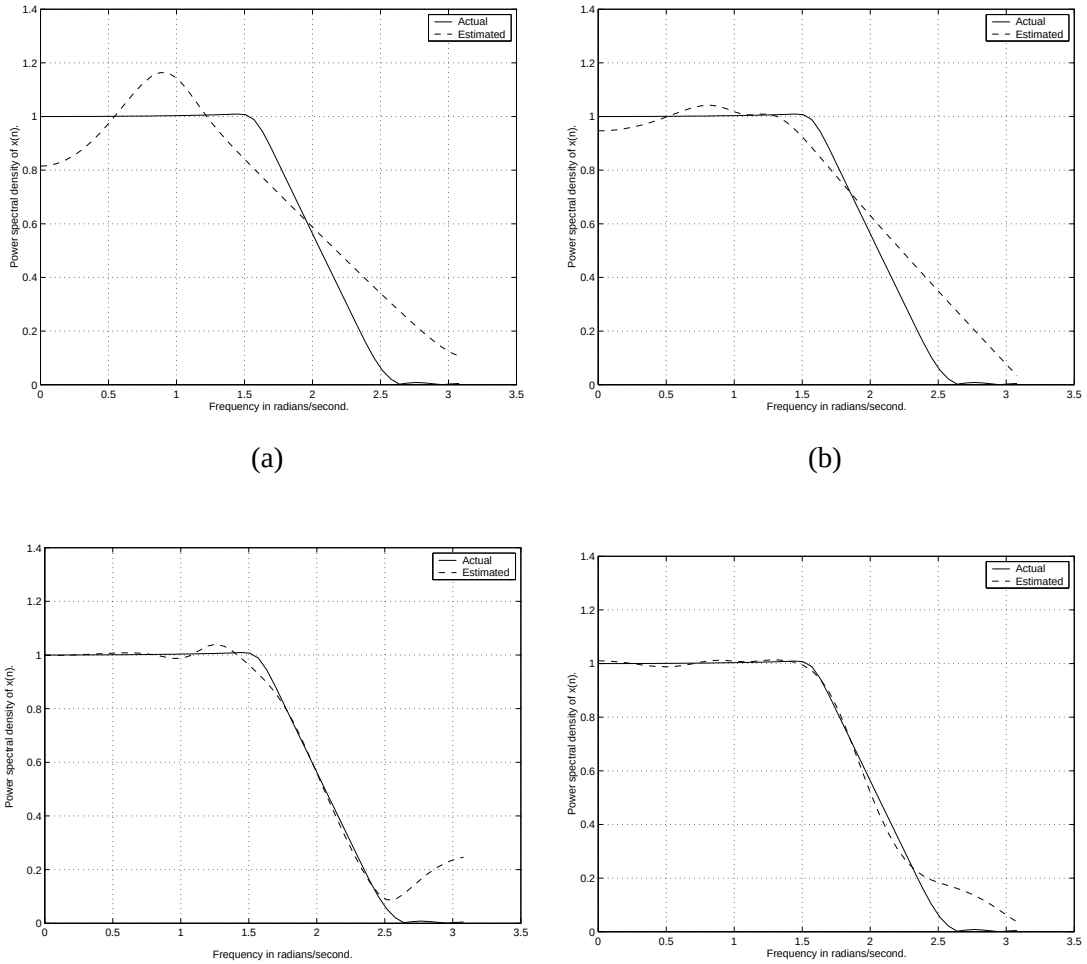


Fig. 7. Intermediate PSD estimates obtained using only part of the ACS data given in Table I. Each plot shows the estimated PSD using $\rho_i(k)$ data up to and including the kth lag: (a) $k = 0$, (b) $k = 1$, (c) $k = 2$, (d) $k = 3$. Note that the estimation quality increases as k increases.

TABLE I

CORRELATION COEFFICIENTS USED AS DATA IN EXAMPLE 3.

$\rho_i(k)$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$i=0$	0.3536	-0.0649	0.0143	0.0004	-0.0010
$i=1$	0.3395	-0.0059	0.0052	0.0028	0.0012
$i=2$	0.2618	0.1180	0.0035	0.0003	0.0024
$i=3$	0.0947	0.0378	0.0002	0.0003	0.0002

Example 4 Consider the 3-channel analysis system shown in Fig. 8. The down-sampling ratios for this system are $N_0 = 2$, $N_1 = 2$ and $N_2 = 4$. Therefore, it is a non-uniform over-sampled analysis filter bank. The filters $H_0(z)$ to $H_2(z)$ are chosen to be FIR. The transfer function of each filter is as follows:

$$H_0(z) = 0.0753 + 0.1656z^{-1} + 0.2053z^{-2} + 0.1659z^{-3} + 0.0751z^{-4}$$

$$H_1(z) = 0.4652 - 0.1254z^{-1} - 0.3151z^{-2} + 0.0975z^{-3} - 0.0259z^{-4}$$

$$H_2(z) = 0.1931 - 0.4226z^{-1} + 0.3668z^{-2} - 0.0974z^{-3} - 0.0405z^{-4}$$

The frequency response $|H_i(e^{j\omega})|$ of these filters are shown in Fig. 9. Again, we emphasize that the above transfer functions were designed simply to show low-pass, band-pass and high-pass characteristics. They may or may not belong to any specific classes of filters used in classical filter bank designs. The non-observable input signal $x(n)$ is the same low-pass PSD as in Example 3.

The correlation coefficients associated with the low-rate observable signals $v_i(n)$ are shown in Table II. The PSD estimates obtained by the MEIE using these correlation coefficients are shown in Fig. 10. It's observed that the estimates quickly capture the shape of the actual PSD and approximate it more closely as k increases. The approximation, however, does not improve uniformly (compare, e.g., Fig. 10(d) with Fig. 10(e)) . ♢

In standard (i.e. single-rate) spectrum estimation, it is possible, at least in principle, to achieve arbitrarily good estimates of $P_x(e^{j\omega})$ if a sufficiently large number of the ACS values $R_x(k)$ are available. However, in the multirate case, the estimate $P_x^*(e^{j\omega})$ might or might not converge to the true PSD $P_x(e^{j\omega})$ depending on the nature of the filters $H_i(z)$ and the down-sampling rations N_i. Recall from Section III that observing low-rate autocorrelation coefficients $\rho_i(k)$, $0 \le i \le M-1$, $0 \le k \le L-1$, effectively reduces the set $\mathcal{M}$ of all possible $P_x(e^{j\omega})$ to a "smaller" set $\mathcal{Q}_{M-1} \subset \mathcal{M}$.

It is possible for $\mathcal{Q}_{M-1}$ to reduce to a single power spectrum as $k \to \infty$ in which case $P_x^*(e^{j\omega})$ will converge to the true PSD $P_x(e^{j\omega})$. However, in general, the set $\mathcal{Q}_{M-1}$ may retain infinitely many elements even when $k \to \infty$. In this case, $P_x^*(e^{j\omega})$ will converge to the PSD in $\mathcal{Q}_{M-1}$ which has the highest entropy and not necessarily to the actual PSD $P_x(e^{j\omega})$. Further details on such issues along with an analysis of the *quantity of information* contained in individual low-rate measurements will be provided in our companion paper [34]. See also [35] and [36, Chap. 4].

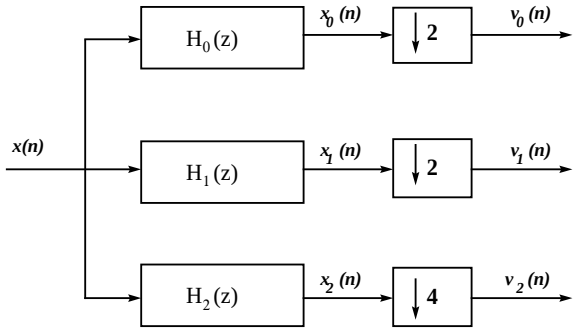


Fig. 8. The 3-channel nonuniform analysis system used in Example 4.

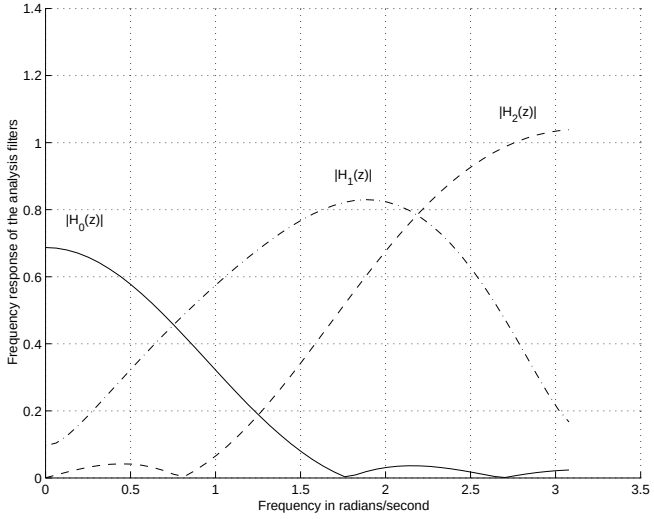
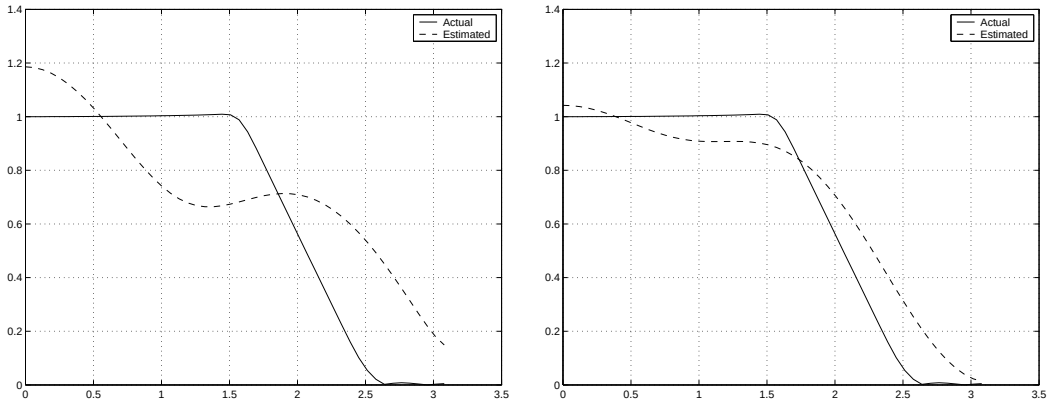


Fig. 9. Frequency response of the analysis filters used in Example 4.

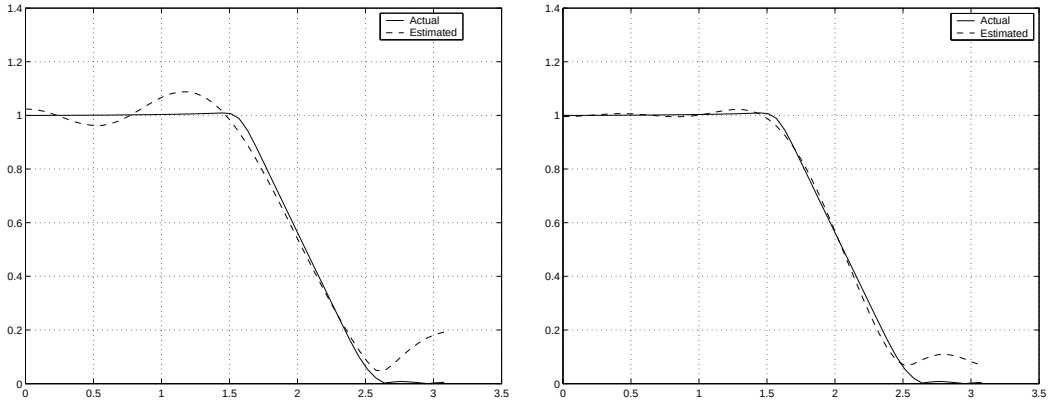
TABLE II

CORRELATION COEFFICIENTS USED AS DATA IN EXAMPLE 4.

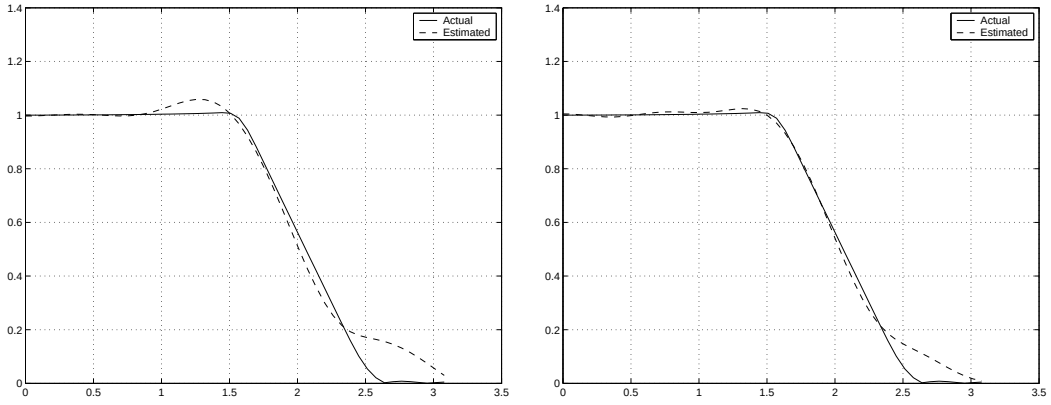
$\rho_i(k)$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$i=0$	0.1084	0.0583	0.0056	-0.0001	-0.0000	0.0000
$i=1$	0.1974	-0.1260	0.0423	-0.0169	0.0063	-0.0038
$i=2$	0.0438	0.0141	-0.0002	0.0008	0.0002	0.0001



(a) (b)



(c) (d)



(e) (f)

Fig. 10. Intermediate and final PSD estimates using the ACS data given in Table II. Each plot shows the estimated PSD using $\rho_i(k)$ data up to and including the kth lag: (a) $k = 0$, (b) $k = 1$, (c) $k = 2$, (d) $k = 3$, (e) $k = 4$, (f) $k = 5$,. Observe that the estimation quality improves (although not uniformly) from (a) to (f).

IX. Concluding Remarks

A. Why is the cross-correlation information not used?

It is possible to take the cross-correlation between the low-rate measurements into account while estimating the PSD of the high-rate signal. Technically, this would improve the accuracy of the estimates. However, we chose not to incorporate cross-correlation coefficients in our Maximum Entropy inference. There are several reasons for this choice. The most important one is that, in practice, cross-correlation estimates are far more erroneous than autocorrelation estimates. For completeness, the reason why this is the case is briefly explained below. The reader is referred to [37, Section 10.6] for details.

Imagine that we want to estimate the covariance matrix $\mathbf{C}_{XX}$ of a zero mean Gaussian random vector X of dimension N from L realizations $\{X_k\}_{0\leq k<L}$. If we try to do this using the sample-mean estimator

$$\bar{\mathbf{C}}_{XX} = \frac{1}{L}\sum_{k=0}^{L-1} X_k X_k^T, \tag{31}$$

it turns out [37, Page 509] that the expected squared error for each element $[\bar{\mathbf{C}}_{XX}]_{lm}$ of $\bar{\mathbf{C}}_{XX}$ will be

$$E\{([\bar{\mathbf{C}}_{XX}]_{lm} - [\mathbf{C}_{XX}]_{lm})^2\} = \frac{1}{L}(([\mathbf{C}_{XX}]_{lm})^2 + [\mathbf{C}_{XX}]_{ll}[\mathbf{C}_{XX}]_{mm}) \tag{32}$$

and the total error variance is given by

$$E\{\|\bar{\mathbf{C}}_{XX} - \mathbf{C}_{XX}\|_{HS}^2\} = \frac{\|\mathbf{C}_{XX}\|_{HS}^2}{L} + \frac{E^2\{\|X\|^2\}}{L} \tag{33}$$

where $\|.\|_{HS}^2$ denotes the squared Hilbert-Schmidt norm[9]. The formula (32) shows that for off-diagonal elements, $E\{([\bar{\mathbf{C}}_{XX}]_{lm} - [\mathbf{C}_{XX}]_{lm})^2\}$ depends not only on $([\mathbf{C}_{XX}]_{lm})^2$ but also on the amplitude of the diagonal coefficients $[\mathbf{C}_{XX}]_{ll}$ and $[\mathbf{C}_{XX}]_{mm}$. Thus, even though $[\mathbf{C}_{XX}]_{lm}$ may be small, the estimation error is large if the diagonal coefficients are large:

$$E\{([\bar{\mathbf{C}}_{XX}]_{lm} - [\mathbf{C}_{XX}]_{lm})^2\} \geq \frac{[\mathbf{C}_{XX}]_{ll}[\mathbf{C}_{XX}]_{mm}}{L}. \tag{34}$$

Other reasons for ignoring the statistical information contained in cross-correlation coefficients include tremendous increase in the complexity of the MEIE and the necessity of synchronization between the measurement signals $v_i(n)$.

[9]For a matrix A we have $\|A\|_{HS}^2 \triangleq Tr(AA^T)$.

B. Summary and future work

We considered merging the information contained in a set of low-rate measurements about a non-observable high-rate signal. The problem posed in its original form (Problem 1) is ill-posed since it usually has infi nitely many solution. It also might happen that the autocorrelation coeffi cients that are used as data are noisy estimates in which case there might be no solution.

We used the Maximum Entropy principle for selecting a unique solution among all feasible ones. While uniqueness of the Maximum Entropy solution is guaranteed (Lemma 2), the issues of existence and stability are not resolved completely.

Another important challenge in applying the Maximum Entropy principle is how to compute the solution. We addressed some of the computational issues in Section VII and devised a practical algorithm called the Maximum Entropy Inference Engine for computing an approximate solution. We illustrated the performance of this algorithm through simulated examples in Section VIII.

Step 4 in Algorithm 1 requires the use of some optimization package which, in turn, might need considerable computational resources. This, and other computationally demanding steps (such as the spectral factorization required in Step 4 of Algorithm 2) prohibit using the Maximum Entropy Inference Engine in large-scale or real-time applications. Seeking improved numerical methods in terms of both the quality of approximation and computational speed are in the focus of our current research.

Acknowledgements

The authors are very thankful to Professor Paul Eggermont, Department of Food and Resource Economics, University of Delaware, USA, and Professor Jonathan Borwein, Centre for Experimental and Constructive Mathematics, Simon Fraser University, Canada, for helpful correspondence on Maximum Entropy regularization. They would also like to thank the anonymous reviewers for their helpful comments which resulted in an improved presentation.

REFERENCES

[1] H. Jeffreys, *Theory of Probability*, Oxford University Press, London, 3rd edition, 1967.

[2] R. von Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, New York, 1964.

[3] S. M. Kay, *Modern Spectrum Estimation: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1988.

[4] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*, Cambridge University Press, 1993.

[5] M. H. Hayes, *Statistical Signal Processing and Modeling*, Wiely, New York, NY, 1996.

[6] B. Buttkus, *Spectral Analysis and Filter Theory in Applied Geophysics*, Springer-Verlag, Berlin, 2000.

[7] B. L. Shoop, *Photonic Analog-to-digital conversion*, Springer-Verlag, Berlin, 2001.

[8] V. Sathe and P. P. Vaidyanathan, "Effect of multirate systems on the statistical properties of random signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 41, no. 1, pp. 131–146, January 1993.

[9] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[10] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola, *Nonlinear ill-posed problems*, Chapman and Hall, London, 1998, (2 volumes).

[11] E. T. Jaynes, "On the rationale of maximum entropy methods," *Proceedings of the IEEE*, vol. 70, pp. 939–952, 1982.

[12] E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, 1983, (Editted by R. D. Rosenkrantz).

[13] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approch to inference for linear inverse problems," *The Annals of Statistics*, vol. 19, pp. 2032–2066, 1991.

[14] I. Csiszár, "Maxent, mathematics and information theory," in *Maximum entropy and Bayesian methods*, K. M. Hanson and R. N. Silver, Eds., The Netherlands, 1996, pp. 35–50, Kluwer Academic Publishers.

[15] J Uffi nk, "Can the Maximum Entropy principle be explained as a consistency requirement?," *Studies on the History and Philosophy of Modern Physics*, vol. 26, no. 3, pp. 223–261, 1995.

[16] J Uffi nk, "The constraint rule of the Maximum Entropy principle," *Studies on the History and Philosophy of Modern Physics*, vol. 27, no. 1, pp. 47–79, 1996.

[17] J. P. Burg, "Maximum entropy spectral analysis," in *37th Annual International Meeting of Society of Exploration Geophysicists*, Oklahoma City, Oklahoma, 1967.

[18] N. Wu, *The Maximum Entropy Method*, Springer, Berlin, 1997.

[19] M. S. Pinsker, *Information and information stability of random variables and processes*, Holden-Day, San Francisco, 1964.

[20] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Transactions on Information Theory*, vol. 26, no. 1, pp. 26–37, 1980.

[21] J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 472–482, 1981.

[22] I. Vajda, *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, Boston, 1989.

[23] S. I. Amari and H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society and Oxford University Press, 2000.

[24] M. Klaus and R. T. Smith, "A Hilbert space approach to maximum entropy requlrarization," *Mathematical Methods in Applied Sciences*, vol. 10, pp. 397–406, 1988.

[25] J. M. Borwein and A. S. Lewis, "Convergence of best maximum entropy estimates," *SIAM Journal of Optimization*, vol. 1, pp. 191–205, 1991.

[26] H. W. Engl and G. Landl, "Convergence rates for maximum entropy regularization," *SIAM Journal on Numerical Analysis*, vol. 30, no. 5, pp. 1509–1536, 1993.

[27] P. P. B. Eggermont, "Maximum entropy regularization for Fredholm integral equations of the fi rst kind," *SIAM J. Math. Anal.*, vol. 24, no. 6, pp. 1557–1576, 1993.

[28] A. S. Leonev, "A generalization of the maximal entropy method for solving ill-posed problems," *Siberian Mathematical Journal*, vol. 41, no. 4, pp. 716–724, 2000.

[29] U. Amato and W. Hughes, "Maximum entropy regularization of Fredholm integral equations of the first kind," *Inverse Problems*, vol. 7, pp. 793–808, 1991.

[30] M. Teboulle and I. Vajda, "Convergence of best ϕ-entropy estimates," *IEEE Transactions on Information Theory*, vol. 39, no. 1, January 1993.

[31] J. M. Borwein and A. S. Lewis, "Partially-finite programming in $\mathbf{L}_1$ and the existence of maximum entropy estimates," *SIAM Journal of Optimization*, vol. 3, pp. 248–267, 1993.

[32] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization, Theory and Examples*, Springer, New York, NY, 2000.

[33] P. P. Vaidyanathan, *Multirate Sytems and Filter Banks*, Prentice-Hall, Upper Saddle River, NJ, 1993.

[34] O. S. Jahromi, B. A. Francis, and R. H. Kwong, "Information content of multirate measurements," *IEEE Transactions on Signal Processing*, manuscript, to be submitted.

[35] O. S. Jahromi, R. H. Kwong, and B. A. Francis, "Information theory of multirate systems," in *Proceedings of the International Symposium on Information Theory (ISIT)*, Washington, DC, June 2001.

[36] O. S. Jahromi, *Theory of Multirate Statistical Signal Processing*, PhD Thesis, University of Toronto, Toronto, Ontario, April 2002, available online at `http://www.multirate.org`.

[37] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 2nd edition, 1999.