Statistics & Decisions 13, 141 – 165 (1995)

Rates of Convergence for Best Entropy Estimates

Dominikus Noll[1]

Received: Revised version: November 4, 1994

Abstract

Maximum entropy spectral density estimation is a technique for reconstructing an unknown density function from a set of observed data by maximizing a given measure of entropy of the estimate. It is widely used in the analysis of stationary time series, and is successfully applied in various areas of physical sciences and engineering. We obtain convergence rates for the corresponding moment matching programs based on the Fisher information measure and we discuss the stability of this best entropy device in the presence of noise.

1 Introduction

A common problem in various areas of physical sciences consists in trying to estimate an unknown density function $\bar{x}(t) \geq 0$ by measuring some of its moments

$$b_k = \int_T a_k(t)\,\bar{x}(t)\,dt, \qquad k = 0, \ldots, n. \tag{1.1}$$

Typically, the b_k might be known Fourier coefficients or known Hausdorff moments of the unknown function $\bar{x}(t)$, in which case the weight functions $a_k(t)$ are trigonometric resp.

[1]Research supported in part by Deutsche Forschungsgemeinschaft (DFG) and Natural Sciences and Engineering Research Council of Canada (NSERC).
AMS 1990 subject classification: 62 M 15, 41 A 25
Keywords and phrases: Spectral density estimation, asymptotic estimates, stochastic convergence

algebraic polynomials, i.e., $a_k(t) = \exp\{ikt\}$ on $T = [-\pi, \pi]$, resp. $a_k(t) = t^k$, T any interval, and similarly for their multidimensional analogues.

Given only a finite number of moments or Fourier coefficients, this estimation problem is clearly underdetermined. One methodology for selecting an estimate among the functions $x(t)$ satisfying (1.1) is to choose it to maximize some given measure of entropy $H(x)$. This approach, known as *maximum entropy density estimation*, has been widely and successfully used in such divers areas as astronomy, crystallography, speech and image processing, geophysics, and others. For a survey see [18, 19, 20, 25, 26, 32, 33, 38, 39, 20], and also [1, 4, 5, 17, 23, 36, 37, 41]. As an application of particular interest, let us consider the analysis of a stationary time series.

Let $x_1, \ldots, x_n$ be a realization of an unknown complex valued stationary time series $(X_t)_{t\in\mathbb{Z}}$ with mean 0, that is,

$$E(X_t) = 0, \quad E|X_t|^2 < \infty,$$
$$\gamma(h) := \mathrm{Cov}(X_{t+h}, X_t) = E(X_{t+h}X_t^*) \quad \text{independent of } t,$$

where $\gamma(\cdot)$ is the *autocovariance function* of the process (X_t), and where * denotes complex conjugation. The sample autocovariance function

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})^*, \qquad h = 0, \ldots, n-1 \tag{1.2}$$

($\bar{x}$ the sample mean), is used to estimate the values $\gamma(0), \ldots, \gamma(n-1)$. (Here, for practical purposes, only the estimates $h \leq n/4$ are considered reliable). In order to forecast and control the observed phenomenon, it is necessary to supplement in some reasonable way the values of the autocovariance function at times h where it may not be obtained directly from the sample. In contrast with the widely used techniques of modelling the unknown process (X_t) using autoregressive, moving average, or mixed ARMA processes, see [12], spectral density estimation is based on the following nonparametric methodology.

Let us regard the unknown process (X_t) as a stochastic integral over an orthogonal increment process (Z_s). This means there exists a spectral distribution function $F(t)$, $F(-\pi) = 0$, F nondecreasing and right-continuous, such that

$$\gamma(h) = \int_{-\pi}^{\pi} \exp\{iht\}\, dF(t), \qquad h = 0, 1, 2, \ldots \tag{1.3}$$

Usually, $dF(t) = f(t)\,dt$ for a *spectral density* $f(t) \geq 0$, which in the engineering literature is often referred to as the *power spectrum* of the process (X_t), usually under normalization, that is after replacing covariances with correlations. Instead of modelling $\gamma(\cdot)$ directly, spectral density estimation tries to exhibit a spectral density function $f(t) \geq 0$ on $[-\pi, \pi]$ under the constraints

$$\hat{\gamma}(h) = \int_{-\pi}^{\pi} \exp\{iht\} f(t)\, dt, \qquad h = 0, \ldots, n-1, \tag{1.4}$$

and this is precisely a problem of type (1.1). Maximum entropy spectral density estimation now chooses the estimate $f(t) \geq 0$ satisfying the constraints (1.4) which maximizes a given measure of entropy $H(f)$.

There have been various debates over the correct choice or rather the relative merits of the different entropy measures $H(x)$, which are typically integral functionals of the form

$$I(x) = -H(x) = \int_T \phi(x(t))\, dt. \tag{1.5}$$

The classical choices are the *Boltzmann-Shannon* and the *Burg entropy*, defined respectively by

$$\phi(x) = \begin{cases} x \log x & , x > 0 \\ 0 & , x = 0 \\ +\infty & , x < 0 \end{cases}, \qquad \phi(x) = \begin{cases} -\log x & , x > 0 \\ +\infty & , x \leq 0 \end{cases}, \tag{1.6}$$

the debate between the two being controversial (see e.g. [27, 25, 13]). Various other entropy measures have been used, see [1, 17, 22, 23, 41] and [37, 8, 9]. In [8, 9], we have studied extended entropy/information models of the form

$$I(x) = -H(x) = \int_T \phi(x(t), x'(t))\, dt \tag{1.7}$$

which attempt to control derivative values of the unknown densities. This includes in particular the *Fisher information*, whose integrand is

$$\phi(x, x') = \begin{cases} \frac{x'^2}{x} & , \text{for } x > 0 \\ 0 & , \text{for } x = x' = 0 \\ +\infty & , \text{else} \end{cases} \tag{1.8}$$

The use of the Fisher information for the inference problems of type (1.1) was proposed in [37] and so far is mainly motivated by practical aspects. The integrand (1.8) has a smoothing effect (see Appendix I in [8]) which in practice is often desired. Numerical results [9, 10] indicate that the Fisher information often performs better than the Burg entropy. Parallel to these more practical aspects, the present paper is intended to give a foundation for the Fisher best entropy estimation from a more stochastic point of view.

Various information theoretic and probabilistic reasons for selecting the Boltzmann-Shannon entropy are known (see [25, 26]), and they are generally based on a principle of uncertainity. Similarly, the use of the Burg entropy may be justified by means of a statistical uncertainity principle, which has a particularly nice interpretation in time seris analysis. In fact, among all mean zero stationary time series (Y_t) with autocovariance function $\gamma_Y(\cdot)$ satisfying $\gamma_Y(h) = \hat{\gamma}(h)$, $h = 0, \ldots, n-1$, or rather, with spectral density f_Y satisfying (1.4), the variational problem based on the Burg entropy selects the process (Y_t) which is the least predictable in the sense that the corresponding one-step mean square prediction error

$$E|Y_{t+1} - \hat{Y}_{t+1}|^2 = 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_Y(t)\, dt\right\} \tag{1.9}$$

(Kolmogoroff's formula) is maximal. It is known (see [12, 30]), that for a real-valued process (X_t), this least predictable process (Y_t) is precisely the stationary autoregressive process $AR(n)$ with parameters determined by the Yule-Walker equations $\gamma_Y(h) = \hat{\gamma}(h)$, $h = 0, \ldots, n-1$. In other words, the quite sophisticated Burg estimation principle leads to a technically simple and appealing formulation. The latter fact may be used to show that the Burg estimator has nice asymptotic statistical properties, at least in a special subclass of underlying processes (X_t); see for instance [12, Theorem 8.1.1] or [16]. Here we present a parallel result for the Fisher estimates, which are shown to converge in probability for a fairly general class of underlying stationary processes (X_t); (cf. Theorem 4.1).

Let $a_k(t)$ from now on represent algebraic or trigonometric polynomials. The mathematical model for the best entropy density estimation problem is the following:

$$\begin{aligned} &\text{minimize} \quad I(x) = -H(x) = \int_T \phi(x(t), x'(t))\, dt \\ (P_n) \quad &\text{subject to} \quad x \geq 0, x \in \mathcal{A}(T), \\ &\int_T a_k(t)x(t)\, dt = b_k \qquad \text{for } k = 0, \ldots, n, \end{aligned}$$

where $\mathcal{A}(T)$ denotes the space of absolutely continuous functions on $T = [t_0, t_1]$.

There are a variety of questions on the deterministic models (P_n), whose answers should have some impact on the choice of the underlying entropy measures, and our present results emphasize some of the advantages of the Fisher information (1.8). Firstly, one would certainly expect the existence and uniqueness of a solution x_n for (P_n). Secondly, one should have convergence $x_n \to \bar{x}$ of the estimates to the unknown true density $\bar{x}$ as the number of known moments increases. For the Boltzmann-Shannon and Burg entropies, these problems have been discussed in a series of papers by J.M. Borwein and A.S. Lewis [3, 4, 5, 6, 7], and also in [1, 17, 21, 22, 28, 28, 11, 41]. In [3] and [41], general convex objectives $\phi(x)$ have been discussed.

Besides the more practical questions concerning a numerically tractable formulation – see [6, 2, 28, 1, 21, 23] and the surveys [38, 39, 40, 20] or [18] for the Boltzmann-Shannon and Burg cases, and [37, 8, 9, 10] for the Fisher case – one of the main questions on the probabilistic models (P_n) is about stochastic convergence $x_n \to \bar{x}$ of the estimates x_n, possibly including rates of convergence. In the case of a stationary time series, the estimates $\hat{\gamma}(0),\dots,\hat{\gamma}(n-1)$ are usually asymptotically normally distributed with mean $\gamma(0),\dots,\gamma(n-1)$ and known covariance matrix, and one would like to know the asymptotic distribution of the estimates x_n, or the asymptotic error $\|x_n - \bar{x}\|_\infty$. In contrast with the Burg case where the dependence of the estimate x_n on the Fourier coefficients $b_k = \hat{\gamma}(k)$ is linear and therefore relatively easy to analyse, the Fisher estimation causes more technical problems. We obtain a weak consistency type result giving stochastic convergence of the estimates (Theorem 4.1) for a fairly general class of underlying processes (X_t).

Another interesting aspect of the models (P_n) is the phenomenon of noisy data b_k, which is often addressed in applied literature. In practice, this often leads to relaxations of the model (P_n) involving penalty techniques (see [18] for a discussion), or to models with tolerance. Here we shall consider the following *relaxed moment matching* programs

$$\begin{array}{lll} & \text{minimize} & I(x) = -H(x) = \int_T \phi(x(t), x'(t))\,dt \\ (P_{n,\epsilon}) & \text{subject to} & x \geq 0, x \in \mathcal{A}(T), \\ & & \|A_n x - b^n\| \leq \epsilon, \end{array}$$

where $A_n : \mathcal{A}(T) \to \mathbb{C}^{n+1}$ denotes the operator $A_n x = \left(\int_T a_0 x, \dots, \int_T a_n x\right)$, and $b^n = (b_0, \dots, b_n)$, and where $\|\cdot\|$ is some fixed norm on $\mathbb{C}^{n+1}$. Denoting the unique solution of $(P_{n,\epsilon})$ by $x_{n,\epsilon}$, we will show that in the case of the Fisher information measure (1.8), an asymptotic estimate of the form

$$\|x_{n,\epsilon} - \bar{x}\|_\infty \sim \mathcal{O}(\epsilon^{1/2}) \qquad (n \to \infty) \tag{1.10}$$

can be obtained (Theorem 3.9(b)). Similarly, if $(P_{n,\epsilon,b})$ and $(P_{n,\epsilon,c})$ denote the above program $(P_{n,\epsilon})$ with different data b^n resp. c^n, tolerance $\epsilon > 0$ and with $\|b^n - c^n\|_\infty \leq \epsilon$, then the corresponding optimal solutions $x_{n,\epsilon,b}$ and $x_{n,\epsilon,c}$ satisfy an asymptotic estimate of the form

$$\|x_{n,\epsilon,b} - x_{n,\epsilon,c}\|_\infty \sim \mathcal{O}(\epsilon^{1/2}) \qquad (n \to \infty) \tag{1.11}$$

(Theorem 3.9(c)), which indicates in particular that a moderate distortion of the data b_k should not lead to a drastic change of the estimates x_n. This kind of sensitivity analysis will hopefully give some insight into the question of robustness of the estimation programs (P_n) in the presence of noise.

The structure of the paper is as follows. In Section 2 we recall some of the results in [8] needed to analyse the Fisher best entropy estimation program (P_n). Section 3 is devoted to the analysis of the deterministic programs (P_n) and $(P_{n,\epsilon})$. Here we obtain rates of convergence for the solutions x_n and $x_{n,\epsilon}$, including the asymptotic estimates (1.10), (1.11). Section 4 then presents the main probabilistic result, which combines a Central Limit Theorem for stationary processes and the deterministic estimates obtained in Section 3.

2 Duality

In this Section we recall the relevant duality results for the Fisher moment matching programs (P_n) and $(P_{n,\epsilon})$ from [8]. The first program has been dealt with in full detail in [8], while the relevant changes for dealing with $(P_{n,\epsilon})$ are indicated in Section 6 of that paper. Throughout the following, $I(\cdot)$ will always denote the Fisher information (1.8).

The basic idea for analysing program $(P_{n,\epsilon})$ is to view it as an infinite dimensional convex optimization problem and apply convex duality methods. This resembles techniques used in convex optimal control problems, see for instance [24, 34, 35]. For this, we define a Lagrangian function

$$
\begin{aligned}
L(x,y,e;v,\lambda,\rho) &= I(x,y) + \langle v, x'-y\rangle + \langle \lambda, A_n x - b^n - e\rangle + \rho(\|e\|-\epsilon) \qquad (2.1)\\
&= \int_T \frac{y^2}{x} + \int_T v(x'-y) + \sum_{i=0}^{n} \lambda_i \left(\int_T a_i x - b_i - e_i\right) + \rho(\|e\|-\epsilon),
\end{aligned}
$$

where $x \in \mathcal{A}(T)$, $y \in \mathcal{L}_1(T)$, $e = (e_i) \in \mathbb{R}^{n+1}$, $v \in \mathcal{A}(T)$, $\lambda \in \mathbb{R}^{n+1}$, $\rho \geq 0$, and where L is assumed to take on the value $+\infty$ if either y^2/x is not integrable or if $\|e\| > \epsilon$. The program $(P_{n,\epsilon})$ may then be given the equivalent formulation

$$(P_{n,\epsilon}) \qquad \inf_{x,y,e} \sup_{v,\lambda,\rho} L(x,y,e;v,\lambda,\rho),$$

and the associated dual program is

$$(P^*_{n,\epsilon}) \qquad \sup_{v,\lambda,\rho} \inf_{x,y,e} L(x,y,e;v,\lambda,\rho).$$

Indeed, with the dummy variables y, e, an optimal solution x for the original program $(P_{n,\epsilon})$ gives rise to an optimal solution x, y, e for the Lagrangian form of $(P_{n,\epsilon})$ satisfying $y = x'$, while conversely an optimal solution x, y, e for the latter program means $y = x'$ and x optimal for the original $(P_{n,\epsilon})$. We then have the following

Theorem 2.1 *Suppose program $(P_{n,\epsilon})$ with the Fisher information measure is feasible. Then it has a unique optimal solution $x_{n,\epsilon}$ (with corresponding optimal $x_{n,\epsilon}, x'_{n,\epsilon}, e^{n,\epsilon}$ in the*

Lagrangian formulation). Dually, program $(P^*_{n,\epsilon})$ *has a unique optimal solution* $v_{n,\epsilon}, \lambda^{n,\epsilon}, \rho_{n,\epsilon}$, *and the values of both programs are identical. Moreover,* $x_{n,\epsilon}$ *is an entire function which is stictly positive on* $[t_0, t_1]$. *It may be recovered from the dual optimal solution via the (complementary slackness type) formulae*

$$\frac{1}{2}v_{n,\epsilon} = \frac{x'_{n,\epsilon}}{x_{n,\epsilon}}, \qquad A_n x_{n,\epsilon} = b^n + e^{n,\epsilon}, \qquad \|e^{n,\epsilon}\| = \epsilon. \tag{2.2}$$

Finally, the dual program $(P^*_{n,\epsilon})$ *has the equivalent formulation*

$$(P^*_{n,\epsilon}) \quad \begin{array}{ll} \text{maximize} & -\sum_{i=0}^{n} \lambda_i b_i - \epsilon\|\lambda\|_* \\ \text{subject to} & v' + \frac{1}{4}v^2 = \sum_{i=0}^{n} \lambda_i a_i, \\ & v \text{ entire}, \ v(t_0) = v(t_1) = 0 \end{array}$$

where $\|\cdot\|_*$ *is the dual norm of* $\|\cdot\|$ *on* $\mathbb{R}^{n+1}$.

Proof. The case $\epsilon = 0$ was given in [8], Theorems 2.1, 3.2(1), Propositions 4.1, 5.5. Notice here that the constraint qualification (CQ_F) used in [8] (and formulated in Example 3.1) is automatically satisfied for $a_k(t)$ either algebraic or trigonometric polynomials. Indeed, (CQ_F) then reduces to mere feasibility of the program (cf. [8],[29]). The relaxed program $(P_{n,\epsilon})$ is now treated in the same way, with the necessary changes indicated in [8, Example 6.3]. □

Concerning the convergence of the optimal solutions x_n of (P_n), we will need the following result from [8]:

Theorem 2.2 *Let* $x_\infty \in C^1(T)$ *be strictly positive on* T, *and let* $b_k = \int_T a_k x_\infty$ *for* $k = 0, 1, \ldots$ *be its moments. Let* x_n *resp.* $x_{n,\epsilon}$ *be the optimal solutions of the programs* (P_n) *resp.* $(P_{n,\epsilon})$. *Then*

1. $\|x'_n - x'_\infty\|_2 \to 0$, *and* $\|x_n - x_\infty\|_\infty \to 0$ *as* $n \to \infty$,
2. *There exists a function* $x_{\infty,\epsilon} \in \mathrm{dom}I(\cdot)$ *such that* $\|x'_{n,\epsilon} - x'_{\infty,\epsilon}\|_2 \to 0$, $\|x_{n,\epsilon} - x_{\infty,\epsilon}\|_\infty \to 0$ *as* $n \to \infty$.

Proof. Statement (1) is precisely Theorem 2.2 in [8]. As for statement (2), the argument in the above reference shows that every minimizing sequence has a subsequence which converges to some $\tilde{x} \in \mathrm{dom}I(\cdot)$, and one has but to show that there is only one such limit

$\tilde{x}$, which is then $x_{\infty,\epsilon}$. The latter follows from the fact that the programs $(P_{n,\epsilon})$ converge to the limiting program

$$(P_{\infty,\epsilon})\quad \begin{array}{ll} \text{minimize} & I(x) \\ \text{subject to} & x \geq 0,\ x \in \mathcal{A}(T) \\ & \|A_n x - b^n\| \leq \epsilon \quad \text{for all } n = 1, 2, \ldots \end{array}$$

which has a feasible solution, x_∞, and therefore by the strict convexity type argument for $I(\cdot)$ given in [8] has a unique optimal solution $x_{\infty,\epsilon}$. Concerning the notion of convergence of programs involved we refer to [6]. □

The quoted results from [8] are mainly functional analytic in nature. Our present impetus, however, is to determine the rates of convergence in (1) and (2) of Theorem 2.2, and this requires techniques from approximation theory. As will be seen, we shall need some additional assumptions on the unknown function x_∞ such as a certain degree of smoothness and, in the Fourier case, periodicity, which, as we indicate, are not needed in Theorem 2.2. This point is emphasized since it shows an at least theoretical advantage of the Fisher information measure over other objectives like the Boltzmann-Shannon or Burg entropy measures (1.6).

Some questions in Theorem 2.2 remain. For instance we would expect $x_{n,\epsilon}$ to be close to x_n. An answer to this question will be provided in the next Section (Corollary 3.10).

Notation

We keep the notations $(P_n), (P_{n,\epsilon})$ for the programs above. If the data vector $b = (b_0, \ldots, b_n)$ deserves special mentioning, we use the notation $(P_{n,\epsilon,b})$. Let $x_n, x_{n,\epsilon}$ be the optimal solutions of $(P_n), (P_{n,\epsilon})$, and $x_{\infty,\epsilon}$ the optimal solution of $(P_{\infty,\epsilon})$. Let x_∞ be the unknown density with moments

$$\int_T a_k(t) x_\infty(t)\, dt = b_k \qquad k = 0, 1, 2, \ldots. \tag{2.3}$$

We will assume that x_∞ is strictly positive on $T = [t_0, t_1]$, and that $x_\infty \in \mathcal{C}^k(T)$ for some $k \geq 3$, while in the trigonometric case this condition will be replaced by $x_\infty \in \mathcal{C}^k_{\text{per}}(-\pi, \pi)$. Also, we will need an extra assumption which, however, is natural in the light of the applications we have in mind (see Section 4): $x'_\infty(t_0) = x'_\infty(t_1) = 0$.

The corresponding solutions of the dual programs (P_n^*), $(P_{n,\epsilon}^*)$ are v_n, $v_{n,\epsilon}$. This means that $\frac{1}{2} v_n = x'_n / x_n$, $\frac{1}{2} v_{n,\epsilon} = x'_{n,\epsilon} / x_{n,\epsilon}$, and that v_n (and similarly $v_{n,\epsilon}$) satisfy the Riccati equation with boundary conditions

$$v'_n + \frac{1}{4} v_n^2 = \sum_{k=0}^{n} \lambda_k^n a_k, \qquad v_n(t_0) = v_n(t_1) = 0,$$

(see Theorem 2.1). Here λ_k^n are the Lagrange multipliers coming along with the dual solutions. We also need $\frac{1}{2}v_\infty = x'_\infty/x_\infty$, which by our assumption on x_∞ is of class $\mathcal{C}^{k-1}$ on T, and similarly $\frac{1}{2}v_{\infty,\epsilon} = x'_{\infty,\epsilon}/x_{\infty,\epsilon} \in \mathcal{A}(T)$.

3 Convergence Rates

In this Section we obtain rates of convergence of the estimates x_n to the unknown density x_∞. Here our models are deterministic. To begin with, let us consider the following approximation constants which depend completely on $x_\infty(t)$:

$$F_n = \inf\left\{ \left\|v'_\infty + \frac{1}{4}v_\infty^2 - p\right\|_\infty : v' + \frac{1}{4}v^2 = p = \sum_{k=0}^n \lambda_k a_k,\ v \text{ entire}, \right. \\ \left. v(t_0) = v(t_1) = 0,\ \lambda \in \mathbb{R}^{n+1} \right\} \tag{3.1}$$

and

$$F_n^* = \inf\left\{ \left\|v'_\infty + \frac{1}{4}v_\infty^2 - p\right\|_\infty : v' + \frac{1}{4}v^2 = p = \sum_{k=0}^n \lambda_k a_k,\ v \text{ entire}, \right. \\ \left. v(t_0) = v(t_1) = 0,\ \|\lambda\|_* \le M \right\}, \tag{3.2}$$

where $M > 0$ is a constant depending on x_∞ and the choice of the norm $\|\cdot\|_*$, which will be specified later. Here for the trigonometric polynomials a_k we intend $\lambda \in \mathbb{C}^{2n+1}$, and $p = \sum_{|k|\le n} \lambda_k \exp ikt$, which by $\lambda_{-k} = \lambda_k^*$ is still real.

Lemma 3.1 *Under the assumption $x'_\infty(t_0) = x'_\infty(t_1) = 0$, we have the following estimates:*

(a) $I(x_n) \ge I(x_\infty) - b_0 F_n$;

(b) $I(x_{n,\epsilon}) \ge I(x_\infty) - b_0 F_n^* - M\epsilon$.

Proof. Let us prove statement (b). The proof for (a) is similar but easier. Pick $\lambda \in \mathbb{R}^{n+1}$ such that $\|\lambda\|_* \le M$, and for some entire function v, $v' + \frac{1}{4}v^2 = \sum \lambda_k a_k$, $v(t_0) = v(t_1) = 0$, and such that the infimum in the definition of F_n^* is attained, that is, $\|v'_\infty + \frac{1}{4}v_\infty^2 - \sum \lambda_k a_k\|_\infty = F_n^*$. Since $x_\infty > 0$, the latter gives the estimate

$$-v'_\infty x_\infty - \frac{1}{4}v_\infty^2 x_\infty + \sum_{k=0}^n \lambda_k a_k x_\infty \le F_n^* x_\infty. \tag{3.3}$$

On integrating (3.3) over $T = [t_0, t_1]$, using (2.3), and $a_0 \equiv 1$, we obtain

$$\int_T \left(-2\frac{x_\infty x''_\infty - x'^2_\infty}{x^2_\infty} x_\infty - \frac{x'^2_\infty}{x^2_\infty} x_\infty \right) + \sum_{k=0}^{n} \lambda_k b_k \le F^*_n b_0. \tag{3.4}$$

Due to the assumption $x'_\infty(t_0) = x'_\infty(t_1) = 0$, the first term on the left hand side of (3.4) vanishes, an so, on adding $\epsilon\|\lambda\|_*$ on both sides, and using (1.8), we obtain

$$I(x_\infty) + \sum_{k=0}^{n} \lambda_k b_k + \epsilon\|\lambda\|_* \le b_0 F^*_n + \epsilon\|\lambda\|_* \le b_0 F^*_n + M\epsilon. \tag{3.5}$$

But notice that the pair (v, λ) is feasible for the dual program $(P^*_{n,\epsilon})$ as presented in Theorem 2.1. Hence

$$-\sum_{k=0}^{n} \lambda_k b_k - \epsilon\|\lambda\|_* \le \max(P^*_{n,\epsilon}) = \min(P_{n,\epsilon}) = I(x_{n,\epsilon}),$$

and this in tandem with (3.5) implies statement (b). □

The following result will be needed for the estimates with noisy data.

Lemma 3.2 *Under the assumption $x'_\infty(t_0) = x'_\infty(t_1) = 0$, let $c^n = (c_0, \dots, c_n)$ be such that $\|c^n - b^n\| \le \epsilon$. Let $x_{n,\epsilon,c}$ be the solution of program $(P_{n,\epsilon,c})$. Then*

(c) $I(x_{n,\epsilon,c}) \ge I(x_\infty) - b_0 F^*_n - 2M\epsilon$

Proof. Choosing v, λ as in the proof of Lemma 3.1, the same reasoning gives

$$I(x_\infty) + \sum_{k=0}^{n} \lambda_k b_k + \epsilon\|\lambda\|_* \le b_0 F^*_n + M\epsilon. \tag{3.6}$$

Now let $b_k = c_k + e_k$ for $e = (e_k)$ satisfying $\|e\| \le \epsilon$. Then

$$\left| \sum_{k=0}^{n} \lambda_k e_k \right| \le \|\lambda\|_* \|e\| \le M\epsilon,$$

and combining this with the estimate (3.6) gives

$$I(x_\infty) + \sum_{k=0}^{n} \lambda_k c_k + \epsilon\|\lambda\|_* \le b_0 F^*_n + 2M\epsilon.$$

Again, since (v, λ) is feasible for $(P^*_{n,\epsilon,c})$, we have

$$-\sum_{k=0}^{n} \lambda_k c_k - \epsilon\|\lambda\|_* \le I(x_{n,\epsilon,c}),$$

and this implies estimate (c). □

The following result provides the error terms for the dual solutions v_n and $v_{n,\epsilon}$ with respect to the norm $\|\cdot\|_2$, and involving the constants F_n, F_n^*. On integrating these estimates, we will finally be in the position to obtain $\|\cdot\|_\infty$- norm error terms involving $x_n, x_{n,\epsilon}$.

Lemma 3.3 *Let $x'_\infty(t_0) = x'_\infty(t_1) = 0$, and $\delta := \inf\{x_\infty(t) : t_0 \le t \le t_1\} > 0$. Then we have the estimates*

(a) $\|v_\infty - v_n\|_2^2 \le b_0\delta^{-1}F_n$;

(b) $\|v_\infty - v_{n,\epsilon}\|_2^2 \le b_0\delta^{-1}F_n^* + 2M\delta^{-1}\epsilon$.

Proof. We prove statement (b). The proof for (a) is similar but easier. Notice that x_∞ is feasible for program $(P_{n,\epsilon})$. Since $x_{n,\epsilon}$ is the optimal solution for this program, we must have

$$0 \le \frac{1}{t}\Big(I(x_{n,\epsilon} + t(x_\infty - x_{n,\epsilon})) - I(x_{n,\epsilon}) \Big) < +\infty \tag{3.7}$$

for all $t > 0$. Since $x_{n,\epsilon}$ is strictly positive (Theorem 2.1), the integrand (1.8) is strictly positive for the terms in (3.7) for $t > 0$ small enough. Therefore we obtain the estimate

$$0 \le \frac{1}{t}\int_{t_0}^{t_1} \frac{2tx'_{n,\epsilon}(x'_\infty - x'_{n,\epsilon})x_{n,\epsilon} + t^2(x'_\infty - x'_{n,\epsilon})^2 x_{n,\epsilon} - tx'^2_{n,\epsilon}(x_\infty - x_{n,\epsilon})}{x_{n,\epsilon}\big(x_{n,\epsilon} + t(x_\infty - x_{n,\epsilon})\big)} < +\infty. \tag{3.8}$$

Now due to the convexity of the Fisher information $I(\cdot)$, the integrand in (3.8) is monotonically increasing in t. Therefore the Monotone Convergence Theorem allows for passing to the limit $t \to 0^+$ under the integral sign, and this implies

$$0 \le \int_{t_0}^{t_1} \left(\frac{2x'_{n,\epsilon}x_{n,\epsilon}x'_\infty - x'^2_{n,\epsilon}x_\infty}{x^2_{n,\epsilon}} - \frac{x'^2_{n,\epsilon}}{x_\infty} \right) < +\infty. \tag{3.9}$$

On rearranging (3.9) and subtracting $I(x_\infty)$ on both sides we obtain

$$\begin{aligned} I(x_{n,\epsilon}) - I(x_\infty) &\le \int_{t_0}^{t_1} \left(\frac{2x'_{n,\epsilon}x_{n,\epsilon}x'_\infty - x'^2_{n,\epsilon}x_\infty}{x^2_{n,\epsilon}} - \frac{x'^2_\infty}{x_\infty} \right) \qquad (3.10)\\ &= -\int_{t_0}^{t_1} x_\infty \left(\frac{x'_{n,\epsilon}x_\infty - x'_\infty x_{n,\epsilon}}{x_{n,\epsilon}x_\infty} \right)^2 \\ &= -\int_{t_0}^{t_1} x_\infty \big(v_{n,\epsilon} - v_\infty\big)^2. \end{aligned}$$

Hence by Lemma 3.1(b),

$$\delta\|v_{n,\epsilon}-v_\infty\|_2^2 \le I(x_\infty)-I(x_{n,\epsilon}) \le b_0F_n^* + M\epsilon,$$

as desired. This proves statement (b). □

We have the following analoguous result which applies to the noisy data case and uses Lemma 3.2.

Lemma 3.4 *Let* $\delta = \inf\{x_\infty(t) : t_0 \le t \le t_1\} > 0$. *Let* $c^n \in \mathbb{R}^{n+1}$ *satisfy* $\|c^n - b^n\| \le \epsilon$. *Let* $x_{n,\epsilon,c}$ *be the optimal solution of program* $(P_{n,\epsilon,c})$, *with corresponding dual optimal solution* $v_{n,\epsilon,c}$. *Then*

(c) $\|v_\infty - v_{n,\epsilon,c}\|_2^2 \le b_0\delta^{-1}F_n^* + 2M\delta^{-1}\epsilon,$

Proof. Notice that program $(P_{n,\epsilon,c})$ is feasible since x_∞ is admitted. We may then go through the same calculations as in the proof of Lemma 3.3, with $v_{n,\epsilon}$ now replaced by $v_{n,\epsilon,c}$. Using estimate (c) in Lemma 3.2 finally gives the estimate

$$\delta\|v_{n,\epsilon,c}-v_\infty\|_2^2 \le I(x_\infty)-I(x_{n,\epsilon,c}) \le b_0F_n^* + 2M\epsilon,$$

hence the result. □

Our next step is to get estimates for the 'dual' approximation constants F_n, F_n^* in terms of the following more natural approximation constants E_n, E_n^*. With the same assumptions on x_∞ as used before, we define

$$E_n = \inf\left\{\|(\log x_\infty)'' - p\|_\infty : p = \sum_{k=0}^{n}\lambda_k a_k \text{ for some } \lambda \in \mathbb{R}^{n+1}\right\}, \tag{3.11}$$

and

$$E_n^* = \inf\left\{\|(\log x_\infty)'' - p\|_\infty : p = \sum_{k=0}^{n}\lambda_k a_k \text{ for some } \|\lambda\|_* \le N\right\}, \tag{3.12}$$

where $N > 0$ is a constant depending only on x_∞ and the norm $\|\cdot\|_*$, which is to be specified later. Again we intend $p = \sum_{|k|\le n}\lambda_k \exp\{ikt\}$ with $\lambda_{-k} = \lambda_k^*$ in $\mathbb{C}$ in the trigonometric case. Also recall the constant M, which has the same meaning as before. With these definitions, we have the following

Lemma 3.5 *Let* $\|\cdot\|_*$ *be any of the* p*-norms,* $1 \le p \le +\infty$. *Let* N, M *be such that* $6N \le M$. *For the algebraic polynomials* $a_k(t) = t^k$, *there exist constants* $K > 0$ *and* $K^* > 0$ *depending only on* x_∞ *such that we have* *(a)* $F_{2n+2} \leq K \cdot E_n$;

(b) $F^*_{2n+2} \leq K^* \cdot E^*_n$.

Proof. Again we give an explicit proof in the case (b), while case (a) is similar but easier. Pick an algebraic polynomial $p_0 = \sum_{k=0}^{n} \rho_k a_k$ of degree $\leq n$ such that the infimum in (3.12) is attained. In particular, $\|\rho\|_* \leq N$. Let q be the primitive of p_0 having $q(t_0) = 0$. Then on integrating the estimate (3.12), we find $\|(\log x_\infty)' - q\|_\infty \leq (t_1 - t_0)E^*_n$, since $x'_\infty(t_0) = 0$. But then

$$|q(t_1)| \leq \left| \int_{t_0}^{t_1} p_0(s)\, ds \right| = \left| \int_{t_0}^{t_1} (p_0 - (\log x_\infty)'')ds \right| \leq (t_1 - t_0)E^*_n.$$

Now let $s(t)$ be the line segment satisfying $s(t_0) = 0$, $s(t_1) = q(t_1)$, then $|s'| \leq |q(t_1)/(t_1 - t_0)| \leq E^*_n$, hence

$$\left\|(\log x_\infty)'' - (p_0 - s')\right\|_\infty \leq 2E^*_n.$$

Therefore, on setting $p' = q - s$, we get an algebraic polynomial of degree $\leq n+2$ satisfying $p'(t_0) = p'(t_1) = 0$ and $\|(\log x_\infty)'' - p''\|_\infty \leq 2E^*_n$, and with the choice of $p(t_0)$ still free. Now define $x(t) = \exp\{p(t)\}$ with $p(t_0)$ chosen so that $x(t_0) = x_\infty(t_0)$, and let as usual $\frac{1}{2}v(t) = x'(t)/x(t)$. Then we have

$$v'(t) + \frac{1}{4}v(t)^2 = 2p''(t) + p'(t)^2,$$

and $v(t_0) = v(t_1) = 0$ by the above construction of p. The polynomial

$$2p'' + p'^2 =: \sum_{k=0}^{2n+2} \lambda_k a_k =: r$$

is of degree $\leq 2n+2$, and the pair (v, λ) is therefore admitted in the definition of F^*_{2n+2}. Indeed, since $v(t_0) = v(t_1) = 0$ is clear, we have but to check $\|\lambda\|_* \leq M$. This is established in Lemma 3.6 below.

Suppose now this has been shown. The definition of F^*_{2n+2} then implies

$$F^*_{2n+2} \leq \|v'_\infty + \frac{1}{4}v^2_\infty - r\|_\infty = \|v'_\infty + \frac{1}{4}v^2_\infty - (v' + \frac{1}{4}v^2)\|_\infty. \tag{3.13}$$

Consider the positive function $y_\infty = x^{1/2}_\infty$, and let $y(t) = \exp\{\frac{1}{2}p(t)\}$, so that $y = x^{1/2}$. Then the last term in (3.13) equals $\|(y''/y_\infty) - (y''/y)\|_\infty$, so we are led to find an estimate for this expression. Now observe that

$$\left\|\frac{y''_\infty}{y_\infty} - \frac{y''}{y}\right\|_\infty = \left\|\frac{1}{y_\infty y}(y''_\infty y - y'' y_\infty)\right\|_\infty \leq 2\delta^{-1}\|y''_\infty y - y'' y_\infty\|_\infty \tag{3.14}$$

for n large enough. Indeed, we have $x \geq x_\infty/2$ from some index n depending on x_∞ on, since the polynomial p_0 approximates $(\log x_\infty)''$, an so, by the choice of $p(t_0)$, $x = \exp\{p\}$ approximates x_∞ as $n \to \infty$. By the definition of y_∞, y this also gives $y \geq y_\infty/2$. We continue

$$\|y''_\infty y - y'' y_\infty\|_\infty \leq \|y''_\infty\|_\infty \|y - y_\infty\|_\infty + \|y_\infty\|_\infty \|y''_\infty - y''\|_\infty. \tag{3.15}$$

Here the first term $\|y - y_\infty\|_\infty$ may be estimated as follows. We know that $2\|\log y - \log y_\infty\|_\infty = \|\log x - \log x_\infty\|_\infty \leq (t_1 - t_0)^2 E_n^* =: K_1 E_n^*$ by twice integrating (3.12). Now we use Lemma 4.4(b) in [4], which gives an estimate of $\|\exp\{f\} - \exp\{g\}\|_\infty$ in terms of $\|f - g\|_\infty$. This gives

$$\|y - y_\infty\|_\infty \leq K_2 \exp\{E_n^*\} E_n^* \leq 2K_2 E_n^*$$

for n large enough, since $E_n^* \to 0$. Therefore, it remains to estimate the term $\|y''_\infty - y''\|_\infty$ in (3.15). Here we use the formula $f'' = f\big((\log f)'' - (\log f)'^2\big)$, which implies

$$\begin{aligned}\|y'' - y''_\infty\|_\infty \;\leq\; & \|y_\infty\|_\infty \|(\log y)'' - (\log y_\infty)''\|_\infty + \|y_\infty\|_\infty \|(\log y)'^2 - (\log y_\infty)'^2\|_\infty \\ & + \|(\log y)'' - (\log y)'^2\|_\infty \|y - y_\infty\|_\infty.\end{aligned}$$

Here the term $\|(\log y)'' - (\log y)'^2\|_\infty$ is bounded by a constant $K_3 > 0$ depending on x_∞, since by the argument already used above, the function $(\log x)''$ approximates $(\log x_\infty)''$ as $n \to \infty$ since $E_n^* \to 0$, and hence $(\log y)''$ approximates $(\log y_\infty)''$. On integrating, one gets the same observation for $(\log y)'^2$ and $(\log y_\infty)'^2$. On the other hand, the term involving $\|y - y_\infty\|_\infty$ has already been estimated above. Hence $\|y'' - y''_\infty\|_\infty = \mathcal{O}(E_n^*)$, and the proof is complete. □

Lemma 3.6 *Let $\|\cdot\|_*$ and the constants M, N, $6N \leq M$ be as above. Let $p = \sum_{k=0}^n \rho_k a_k$ with $\|\rho\|_* \leq N$, and let $r = \sum_{k=0}^{2n+2} \lambda_k a_k$ be the polynomial $r = 2p'' + p'^2$. Then $\|\lambda\|_* \leq M$.*

Proof. Let $p'^2 = \sum_{k=0}^{2n+2} \mu_k a_k$, then $\|\lambda\|_* \leq 2\|\rho\|_* + \|\mu\|_*$, hence it suffices to show that $\|\mu\|_* \leq 4N$, for then $\|\lambda\|_* \leq 6N \leq M$.

Fix $\rho = (\rho_k)$ satisfying $\|\rho\|_* \leq N$. Consider the infinite matrix $A(\rho)$ whose nth row is

$$\frac{2\rho_n}{n}, \frac{2\rho_{n-1}}{(n-1)\cdot 1}, \frac{2\rho_{n-2}}{(n-2)\cdot 2}, \ldots, \frac{\rho_n/2}{(n/2)^2}, 0, 0, \ldots \quad \text{for } n \text{ even}$$

$$\frac{2\rho_n}{n}, \frac{2\rho_{n-1}}{(n-1)\cdot 1}, \frac{2\rho_{n-2}}{(n-2)\cdot 2}, \ldots, \frac{2\rho_{(n+1)/2}}{(n+1)/2\cdot(n-1)/2}, 0, 0, \ldots \quad \text{for } n \text{ odd.}$$

Then $\mu = A(\rho)\cdot\rho$ for the sequence μ of coefficients of p'^2, if both μ and ρ are now considered as infinite sequences. Therefore we are left to prove that $A(\rho)$ has operator norm $\leq 4\|\rho\|_* \leq$

$4N$ for any of the norms $\|\cdot\|_*$ involved. One readily checks that $\|A(\rho)\|_{\infty,\infty} \le 2\|\rho\|_\infty$ and $\|A(\rho)\|_{1,1} \le 2\|\rho\|_1$. For $1 < p < \infty$, we obtain $\|A(\rho)\|_{p,p} \le 2\|\rho\|_p\|\{\frac{1}{n}\}\|_{p'} \le 4\|\rho\|_p$, as desired $(1/p + 1/p' = 1)$. This proves the claim. □

Our next step is to prove an analogue of Lemma 3.5 for the trigonometric case.

Lemma 3.7 *Let $\|\cdot\|_*$ be any of the p–norms, $1 \le p \le \infty$. Let N, M be constants such that $6N \le M$. For the trigonometric polynomials $a_k(t) = \exp\{ikt\}$, and with $x_\infty \in C^k_{\text{per}}(-\pi,\pi)$ for some $k \ge 3$, there exist constants L, L^* depending only on $x_\infty(t)$ such that*

(a) $F_{2n} \le L \cdot E_n$;

(b) $F^*_{2n} \le L^* \cdot E^*_n$.

Proof. We prove statement (b). Choose a trigonometric polynomial p_0 of degree $\le n$ such that E^*_n is attained. Let $p_0 = \sum_{|k|\le n} c_k \exp\{ikt\}$, with $\|\{c_k\}\|_* \le N$ as in (3.12). Let $p_1 = p_0 - c_0$, which is real since c_0 is real. Then $\|p_1 - (\log x_\infty)''\|_\infty \le 2E^*_n$, since

$$2\pi|c_0| = \left| \int_{-\pi}^{\pi} (p_0 - (\log x_\infty)'') \right| \le 2\pi\|p_0 - (\log x_\infty)''\|_\infty = 2\pi E^*_n. \tag{3.16}$$

Let p be the trigonometric polynomial satisfying $p'' = p_1$ and $p(-\pi) = \log x_\infty(-\pi)$. Then we have $\|(\log x_\infty)' - p'\|_\infty \le 8\pi E^*_n$ and $\|\log x_\infty - p\|_\infty \le 16\pi^2 E^*_n$. Indeed, let $q(t) = \sum_{|k|\le n} \gamma_k \exp\{ikt\} = \gamma_0 + \sum_{0<|k|\le n} \frac{c_k}{ik} \exp\{ikt\}$ be the primitive of p_1 satisfying $q(-\pi) = (\log x_\infty)'(-\pi) = 0$, then $\|q - (\log x_\infty)'\|_\infty \le 4\pi E^*_n$ on integrating. But now an estimate similar to (3.16) gives $|\gamma_0| \le 4\pi E^*_n$, so with $p' = q - \gamma_0$, we get the estimate $\|p' - (\log x_\infty)'\|_\infty \le 8\pi E^*_n$ as desired. Integrating and using $p(-\pi) = \log x_\infty(-\pi)$ then implies $\|p - \log x_\infty\|_\infty \le 16\pi^2 E^*_n \to 0$.

As in the proof of Lemma 3.5, we define $x = \exp\{p\}$, $v = 2x'/x$, and let r be the trigonometric polynomial $r = 2p'' + p'^2 =: \sum_{k=0}^{2n} \lambda_k a_k$ of degree $\le 2n$. Then (v, λ) is admitted in the definition of F^*_{2n} (trigonometric case) as soon as we prove that for $p'' = \sum_{0<|k|\le n} c_k \exp\{ikt\}$ with $\|\{c_k\}\|_* \le N$, the coefficients λ of $r(t)$ satisfy $\|\lambda\|_* \le M$. This is now checked with an argument similar to the one given in Lemma 3.6, and it uses $6N \le M$. We leave the details to the reader.

Suppose the result has been shown. Then $F^*_{2n} \le \|v'_\infty + \frac{1}{4}v^2_\infty - (v' + \frac{1}{4}v^2)\|_\infty$, and the proof now proceeds completely as in Lemma 3.5. □

We shall now obtain explicit estimates for the constants E_n, E^*_n depending on the norms $\|\cdot\|_*$. Clearly here the most interesting case for the programs $(P_{n,\varepsilon,c})$ is when $\|\cdot\|$ denotes the supremum norm, because this captures the worst case deviation $\|b^n - c^n\|_\infty$.

Proposition 3.8 *Let $a_k(t) = \exp\{ikt\}$ on $T = [-\pi, \pi]$. Let $x_\infty \in C^k_{\text{per}}(-\pi, \pi)$ for some $k \geq 3$. Suppose x_∞ is strictly positive on T and satisfies $x'_\infty(-\pi) = x'_\infty(\pi) = 0$. Let $c^n = (c_0, \ldots, c_n)$ be fixed. Then*

(a) For $k \geq 3$ we have $\|(\log x_\infty)' - (\log x_n)'\|_2^2 \leq K_1 n^{-k+2}$,

(b) For $k \geq 3$ we have $\|(\log x_\infty)' - (\log x_{n,\epsilon})'\|_2^2 \leq K_2(n^{-k+2} + \epsilon)$,

(c) For $k \geq 3$, $\|\cdot\|_* = \|\cdot\|_\infty$, and with $\|b^n - c^n\|_1 \leq \epsilon$ we have $\|(\log x_\infty)' - (\log x_{n,\epsilon,c})'\|_2^2 \leq K_3(n^{-k+2} + \epsilon)$,

(d) For $k \geq 3$, $\|\cdot\|_* = \|\cdot\|_p$, $1 < p < \infty$ and with $\|b^n - c^n\|_{p'} \leq \epsilon$ $(1/p + 1/p' = 1)$, we have $\|(\log x_\infty)' - (\log x_{n,\epsilon,c})'\|_2^2 \leq K_4(n^{-k+2} + \epsilon)$,

(e) For $k \geq 4$, $\|\cdot\|_* = \|\cdot\|_1$ and with $\|b^n - c^n\|_\infty \leq \epsilon$, we have $\|(\log x_\infty)' - (\log x_{n,\epsilon,c})'\|_2^2 \leq K_5(\log n \cdot n^{-k+3} + \epsilon)$,

for certain constants K_1, K_2, K_3, K_4, K_5 depending completely on x_∞.

Proof. Let us first prove statement (a). Lemma 3.3(a) implies $\|v_\infty - v_n\|_2^2 = \|(\log x_\infty)' - (\log x_n)'\|_2^2 \leq L_1 F_n \leq L_2 E_{r(n)}$, where $r(n) = n/2$ for n even and $r(n) = (n-1)/2$ for n odd, and with L_1, L_2 depending on x_∞. By Jackson's Theorem ([14]), $(\log x_\infty)'' \in C^k_{\text{per}}(-\pi, \pi)$ implies $E_r = \mathcal{O}(r^{-k+2})$, and since $r(n) = \mathcal{O}(n)$, we get $\|(\log x_\infty)' - (\log x_n)'\|_2^2 = \mathcal{O}(n^{-k+2})$, proving statement (a).

Notice that estimate (b) follows from (c), since program $(P_{n,\epsilon})$ is $(P_{n,\epsilon,c})$ with $c^n = b^n$. Let us now prove statement (c). Here we have to specify the constants M, N for the case $\|\cdot\|_* = \|\cdot\|_\infty$. We take $N = 2\|(\log x_\infty)''\|_2$ and $M = 6N$. Then we find $E_n = E_n^*$ with this choice of N. Indeed, let $p = \sum_{|k| \leq n} \lambda_k \exp\{ikt\}$ be a trigonometric polynomial for which the infimum in the definition of E_n is attained. Then

$$\begin{aligned} 2\pi|\lambda_k| &= \left| \int p \cdot \exp\{-ikt\} \right| \leq \int |(p - (\log x_\infty)'')| + \left| \int (\log x_\infty)'' \exp\{-ikt\} \right| \\ &\leq 2\pi\|p - (\log x_\infty)''\|_\infty + 2\pi|c_k|, \end{aligned} \tag{3.17}$$

where the c_k are the Fourier coefficients of $(\log x_\infty)''$. Then $|\lambda_0| \leq E_n$, since $c_0 = 0$. Furthermore, since $|c_k| \leq \|(\log x_\infty)''\|_2$, we have $|\lambda_k| \leq E_n + N \leq 2N$, since $E_n \to 0$ by the differentiability assumption on x_∞, proving $\|\{\lambda_k\}\|_* = \|\{\lambda_k\}\|_\infty \leq N$ as desired. Hence the claim $E_n = E_n^*$.

Now Lemma 3.4(c) and Lemma 3.5(b) imply $\|v_\infty - v_{n,\epsilon,c}\|_2^2 \leq L_4(F_n^* + \epsilon) \leq L_5(E^*_{r(n)} + \epsilon) = L_5(E_{r(n)} + \epsilon)$ for certain L_4, L_5 depending on x_∞, and using the assumption that $\|b^n - c^n\| = \|b^n - c^n\|_1 \leq \epsilon$. Here $r(n)$ has the same meaning as in the first part of the proof. As above the Jackson Theorem provides the estimate $E^*_{r(n)} = E_{r(n)} = \mathcal{O}(n^{-k+2})$, and therefore $\|(\log x_\infty)' - (\log x_{n,\epsilon,c})'\|_2 \le L_6(n^{-k+2} + \epsilon)^{1/2}$, which completes the proof of statement (c).

The proof of statement (d) being similar, let us now consider statement (e) with the norm $\|\cdot\|_* = \|\cdot\|_1$. This is the most interesting case giving $\|\cdot\| = \|\cdot\|_\infty$. In this situation we need the stronger assumption $k \ge 4$. We have to specify N, M. Let

$$N := 4\|(\log x_\infty)'''\|_2 + \|(\log x_\infty)''\|_\infty$$

and $M = 6N$. As before, we know from Lemma 3.4 that $\|v_\infty - v_{n,\epsilon,c}\|_2^2 \le K(F_n^* + \epsilon)$, with K depending on x_∞, and further, $F_n^* \le K' E^*_{r(n)}$ by Lemma 3.5(b). Hence in view of $r(n) = \mathcal{O}(n)$, it suffices to show that $E_n^* = \mathcal{O}(\log n \cdot n^{-k+3})$.

Let $s_n(t) = \sum_{0<|k|\le n} c_k \exp\{ikt\}$ be the nth partial sum of the Fourier-series of $(\log x_\infty)'''$. Then we have

$$\|(\log x_\infty)''' - s_n\|_\infty \le K'' \cdot \log n \cdot n^{-k+3} \to 0$$

(see [14, p. 104/105]). We consider its primitive $S_n(t) = \gamma_0 + \sum_{0<|k|\le n} \frac{c_k}{ik} \exp\{ikt\} =: \sum_{|k|\le n} \gamma_k \exp\{ikt\}$, where we choose γ_0 such that $S_n(-\pi) = (\log x_\infty)''(-\pi)$. Then integrating gives

$$\|(\log x_\infty)'' - S_n\|_\infty \le 2\pi K'' \cdot \log n \cdot n^{-k+3},$$

and the coefficients γ of S_n satisfy

$$\|\{\gamma_k\}\|_1 \le |\gamma_0| + \frac{\pi}{\sqrt{3}} \|\{c_k\}\|_2 \le |\gamma_0| + 2\|(\log x_\infty)'''\|_2. \tag{3.18}$$

It remains to estimate $|\gamma_0|$. We have $(\log x_\infty)''(-\pi) = \gamma_0 + \sum_{0<|k|\le n} -\frac{c_k}{ik}$, hence we find

$$\begin{aligned} |\gamma_0| &\le \|(\log x_\infty)''\|_\infty + \left| \sum_{0<|k|\le n} -\frac{c_k}{ik} \right| \\ &\le \|(\log x_\infty)''\|_\infty + \frac{\pi}{\sqrt{3}} \|\{c_k\}\|_2 \\ &\le \|(\log x_\infty)''\|_\infty + 2\|(\log x_\infty)'''\|_2 \end{aligned}$$

using Parseval's identity. This readily implies $\|\{\gamma_k\}\|_1 \le N$. The proof is complete. □

Theorem 3.9 *Let $x_\infty \in C^k_{\text{per}}(T)$ with $k \ge 3$ be strictly positive on $T = [-\pi, \pi]$ and satisfy $x'_\infty(-\pi) = x'_\infty(\pi) = 0$. Let $b^n = (b_0, \dots, b_n)$ be the first $n+1$ Fourier coefficients of x_∞, and let $c^n = (c_0, \dots, c_n)$ be any vector satisfying $|b_k - c_k| \le \epsilon$ for $k = 0, 1, \dots, n$. Let $x_n, x_{n,\epsilon}$ and $x_{n,\epsilon,c}$ be the optimal solutions of the moment matching programs (P_n), $(P_{n,\epsilon})$ and $(P_{n,\epsilon,c})$ respectively. Then we have the estimates*

(a) *For* $k \geq 3$, $\|x_\infty - x_n\|_\infty = \mathcal{O}(n^{-k/2+1})$,

(b) *For* $k \geq 3$, $\|x_\infty - x_{n,\epsilon}\|_\infty = \mathcal{O}((n^{-k+2} + \epsilon)^{1/2})$,

(c) *For* $k \geq 4$, $\|x_\infty - x_{n,\epsilon,c}\|_\infty = \mathcal{O}((\log n \cdot n^{-k+3} + \epsilon)^{1/2})$.

Proof. Consider statement (a). Integrating estimate (a) in Proposition 3.8 implies $\|\log x_\infty - \log x_n\|_\infty \leq L_1 n^{-k/2+1}$. Now we apply Lemma 4.4(b) from [4], which gives an estimate of $\|\exp f - \exp g\|_\infty$ in terms of $\|f - g\|_\infty$. Namely, we obtain $\|x_\infty - x_n\|_\infty \leq L_1 n^{-k/2+1}(1 + \exp\{L_1 n^{-k/2+1}\} \cdot L_1 n^{-k/2+1})\|x_\infty\|_\infty \leq L_2 n^{-k/2+1}$.

Statement (b) is a consequence of (b) in Proposition 3.8, while statement (c) will follow from (e) in 3.8. We present the argument in the latter case, the first one being similar. Integrating (e) in Proposition 3.8 gives the estimate $\|\log x_\infty - \log x_{n,\epsilon,c}\|_\infty \leq L_3((\log n \cdot n^{-k+3} + \epsilon)^{1/2}) =: \beta$. Now again we apply Lemma 4.4(b) in [4], which gives the estimate $\|x_\infty - x_{n,\epsilon,c}\|_\infty \leq \beta(1 + e^\beta \beta/2)\|x_\infty\|_\infty$. Since $1 + e^\beta\beta/2$ is bounded, the right hand side is $\mathcal{O}(\beta) = \mathcal{O}((\log n \cdot n^{-k+3} + \epsilon)^{1/2})$ as claimed. □

Remark. Notice that statements (b) and (c) in Theorem 3.9 are the asymptotic estimates for the deterministic programs mentioned in the introduction. Namely, for $\|b^n - c^n\|_\infty \leq \epsilon$ we have

$$\|x_\infty - x_{n,\epsilon,c}\|_\infty \sim \mathcal{O}(\epsilon^{1/2}) \tag{3.19}$$

for large n. In particular, we get the following improvement of Theorem 2.2(2):

Corollary 3.10 *Let* $x_\infty \in C^k_{\text{per}}(-\pi,\pi)$ *for some* $k \geq 4$, *and suppose* x_∞ *is strictly positive with* $x'_\infty(-\pi) = x'_\infty(\pi) = 0$. *Let* $x_{\infty,\epsilon,c}$ *be the optimal solution of the limiting program* $(P_{\infty,\epsilon,c})$, *where* $\|b^n - c^n\|_\infty \leq \epsilon$ *for every* n. *Then*

$$\|x_\infty - x_{\infty,\epsilon,c}\|_\infty = \mathcal{O}(\epsilon^{1/2}).$$

□

Naturally, $k \geq 3$ would be sufficient here if we had $\|b^n - c^n\|_{p'} \leq \epsilon$ for all n and some $1 \leq p < \infty$. Equally, for $b^n = c^n$, that is, for the limiting program $(P_{\infty,\epsilon})$, the assumption $k \geq 3$ is sufficient.

We end this Section with a brief outline of how to prove the analogue of Proposition 3.8 and Theorem 3.9 for algebraic moments $a_k(t) = t^k$.

Theorem 3.11 *Let $a_i(t) = t^i$. Suppose $x_\infty \in C^k(T)$ -for some $k \geq 3$- is strictly positive on $T = [t_0, t_1]$ and satisfies $x'_\infty(t_0) = x'_\infty(t_1) = 0$. Let $b^n = (b_0, \ldots, b_n)$ be its first $n+1$ Hausdorff moments, and let $c^n = (c_0, \ldots, c_n)$ be a fixed vector. Let $x_n, x_{n,\epsilon}$ and $x_{n,\epsilon,c}$ be the optimal solutions of the moment matching programs $(P_n), (P_{n,\epsilon})$ and $(P_{n,\epsilon,c})$ respectively. Then*

1. *For $k \geq 3$, $\|x_\infty - x_n\|_\infty = \mathcal{O}(n^{-k/2+1})$;*
2. *For $k \geq 3$, $\|x_\infty - x_{n,\epsilon}\|_\infty = \mathcal{O}((n^{-k+2} + \epsilon)^{1/2})$;*
3. *For $k \geq 4$, $\|\cdot\|_* = \|\cdot\|_1$ and with $\|b^n - c^n\|_\infty \leq \epsilon$, we have $\|x_\infty - x_{n,\epsilon,c}\|_\infty = \mathcal{O}((\log n \cdot n^{-k+3} + \epsilon)^{1/2})$.*

Proof. Similar to Proposition 3.8, the proof of statement (1) follows from Lemma 3.3(a) and Lemma 3.5(a) when combined with Jackson's Theorem governing the rate of convergence for polynomial approximation (see for instance [15]).

Let us prove statement (3). We follow the pattern of Proposition 3.8(e). Let us normalize $T = [t_0, t_1] = [-1, 1]$. On replacing the monomials a_k by the Tchebysheff polynomials T_k on $[-1, 1]$, (cf. [15] for a definition), the results in Lemmas 3.3 and 3.5 are not affected, since $\text{lin}\{a_0, \ldots, a_n\} = \text{lin}\{T_0, \ldots, T_n\}$.

Let $s_n = \sum_{k=0}^n c_k T_k$ be the nth partial sum of the expansion of $f := (\log x_\infty)'''$ in Tchebysheff polynomials. Then we have an estimate of the form

$$\|f - s_n\|_\infty \leq K \cdot \log n \cdot n^{-k+3}. \tag{3.20}$$

Indeed, this may be seen from the fact that under the transform $t = \cos\theta$, $s_n(t) = \tilde{s}_n(\theta)$ is just the nth partial sum of the Fourier series of $\tilde{f}(\theta) = f(\cos\theta)$. But f is of class C^{k-3} and $k \geq 4$, hence [14, p. 106, Cor. 2.4.6] gives the claimed rate of convergence (3.20) for the Fourier series $\tilde{s}_n \to \tilde{f}$. Notice that the quoted result relies on Jackson's Theorem and the fact that the Dirichlet kernel grows logarithmically.

Observe next that $\|\{c_k\}\|_2 \leq \sqrt{2}E_1 \leq \sqrt{2}\|f\|_\infty$ by [15, p. 131, Thm. 5(ii)]. Now consider the primitive $S_n = \sum_{k=0}^{n+1} d_k T_k$ of s_n. The formula $\int T_n = T_{n+1}/(2n+2) - T_{n-1}/(2n-2)$ (see [15, p. 63]) shows that $d_0 = c_1/4 + \delta_0$, $d_1 = c_0 - c_2/2$, $d_k = (c_{k-1} - c_{k+1})/2k$ for $k \geq 2$, with the choice of δ_0 still free. Choosing the latter such that $S_n(-1) = (\log x_\infty)''(-1)$, we obtain $\|(\log x_\infty)'' - S_n\|_\infty = \mathcal{O}(\log n \cdot n^{-k+3})$, that is, $E_n^* = \mathcal{O}(\log n \cdot n^{-k+3})$ if N is chosen such that $\|\{d_k\}\|_1 \leq N$. Now $\|\{d_k\}\|_1 \leq \frac{\pi}{\sqrt{3}}\|\{c_k\}\|_2 \leq \pi\sqrt{2/3}\|(\log x_\infty)'''\|_\infty$ by the above, so the choice $N := \pi\sqrt{2/3}\|(\log x_\infty)'''\|_\infty$ and $M = 6N$ will do. For the rest of the argument we follow the pattern of Proposition 3.8 and Theorem 3.9, and this completes the proof of statement (3). For statement (2) we follow the proof of Proposition 3.8, with the changes as indicated above. □

Clearly a result analoguous to Corollary 3.10 may be obtained. Also, the analogues of statements (c), (d) from Proposition 3.8 may be established in the algebraic moment case using the same reasoning as above. Since the case $\|\cdot\|_* = \|\cdot\|_1$ is the most interesting one, we skip the results. Notice here that our arguments apply in much the same way to other choices of norms $\|\cdot\|_*$. One has but to prove a result like Lemma 3.6 in these situations.

Remarks. 1) In general, the smoother x_∞, the faster E_n, E_n^* tend to 0. If x_∞ is analytic, then $E_n, E_n^* \to 0$ linearly, that is $E_n^* = \mathcal{O}(\rho^n)$ for some $0 \leq \rho < 1$. If x_∞ is an entire function, convergence is even superlinear (cf. [31]).

2) Notice that as an alternative to our present approach, the question of noisy data b_k may be formulated as a problem of sensitivity analysis of the finite dimensional convex program

$$(P_{n,\epsilon}^*)\quad \begin{array}{ll} \text{maximize} & -\sum_{k=0}^{n} \lambda_k b_k - \epsilon\|\lambda\|_* \\ \text{subject to} & \lambda \in \mathbb{R}^{n+1} \\ & k(\lambda) = 0 \end{array}$$

where k is defined as $k(\lambda) = v(\lambda, t_1)$, whenever $v(\lambda, \cdot)$ denotes the unique solution of the initial value problem $v' + \frac{1}{4}v^2 = \sum \lambda_k a_k$ with $v(t_0) = 0$. The usual techniques of sensitivity analysis are based on the implicit function theorem, and they would require some knowledge about the Hessian of $k(\lambda)$ or rather, about the curvature of the level curve $k(\lambda) = 0$. Due to the complicated nature of the function $k(\lambda)$, the latter shows that this approach is not very promising.

3) In [3] J.M. Borwein and A.S. Lewis have obtained convergence results for the deterministic programs (P_n) using the Boltzmann-Shannon entropy (1.6). By an appropriate modification of their results, one could obtain estimates of the form $\mathcal{O}(\epsilon)$ for the corresponding relaxed programs $(P_{n,\epsilon})$, $(P_{n,\epsilon,c})$, and with respect to the norm $\|\cdot\|_1$ (see Theorem 4.7 in [3]). This is not very satisfactory, however, since a small deviation in $\|\cdot\|_1$ -norm may still cause a drastic change in supremum norm. Notice that the authors of [3] also obtain convergence results in the supremum norm for their programs (P_n) by a clever use of $L_{p,\infty}$ distortion theorems (see [3]). For the programs $(P_{n,\epsilon})$, this technique seems to give only errors of size $\mathcal{O}(n\epsilon)$ (instead of $\mathcal{O}(\epsilon^{1/2})$ in the Fisher case). Numerical experiments for the relaxed Boltzmann-Shannon programs $(P_{n,\epsilon})$ suggest, however, that the solution is in fact more robust with respect to moderate perturbations of the data. Increasing the tolerance ϵ in $(P_{n,\epsilon})$ usually leads to mollified solutions $x_{n,\epsilon}$ which do not differ drastically from the ideal solutions x_n. Here, of course, numerically, a small tolerance ϵ is always allowed implicitely due to the presence of rounding errors.

4) We have not tried too carefully to limit the size of the constants occurring in the estimates, since they depend on the unknown density x_∞. In a Bayesian approach, however,

where we might have some a priori information on x_∞, one might wish to calculate explicit numerical constants, and would then try to get more stringent estimates.

4 Stochastic Convergence Results

In this Section we obtain stochastic convergence results for the time series model as presented in the Introduction. Starting with a mean zero stationary time series (X_t), we use $\hat{\gamma}(\cdot)$ given in (1.2) to estimate the autocovariance function $\gamma(\cdot)$. It is known that (1.2) is biased, but its asymptotic distibution (as $n \to \infty$) has mean $\gamma(h)$ e.g. under the structural assumption that (X_t) is a moving average process of the form

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}. \tag{4.1}$$

Here (Z_t) is an indepentent and identically distributed sequence with mean zero and variance $\sigma^2 > 0$, and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $EZ_t^4 = \eta\sigma^4 < \infty$ (or alternatively $\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty$, see [12, §7.2]). Under any of these hypotheses, one can say more: In fact, for any fixed h, the random vector $(\hat{\gamma}(0), \ldots, \hat{\gamma}(h))$ is asymptotically normally distributed with mean $(\gamma(0), \ldots, \gamma(h))$ and covariance matrix $\frac{1}{n}V_h$, where $V_h = (v_{ij})_{i,j=0,\ldots,h}$ is given by the Bartlett formula

$$v_{ij} = (\eta - 3)\gamma(i)\gamma(j) + \sum_{k=-\infty}^{\infty} \left(\gamma(k)\gamma(k-i+j) + \gamma(k+j)\gamma(k-i)\right). \tag{4.2}$$

Now observe that the assumption $\sum |\psi_j| < \infty$ guarantees that $\sum |\gamma(k)| < \infty$, which in particular shows that (X_t) has a spectral density x_∞ satisfying $x'_\infty(-\pi) = x'_\infty(\pi) = 0$. It also implies that the columns of the matrices V_h are uniformly bounded in ℓ_1-norm. Since this property is preserved under conjugation with an orthogonal matrix, we derive that the eigenvalues of the matrices V_h are uniformly bounded by some constant.

Theorem 4.1. *Let (X_t) be a mean zero moving average process (4.1) satisfying $\sum_{k=-\infty}^{\infty} |\psi_k| < \infty$ and $EZ_t^4 = \eta\sigma^2 < \infty$ (or alternatively, $\sum \psi_j^2 |j| < \infty$). Suppose the spectral density x_∞ of (X_t) is strictly positive on $[-\pi, \pi]$ and of class C^k_{per} for some $k \geq 4$. Let $h \in \mathbb{N}$ be fixed. Let $b_{0,n} = \hat{\gamma}(0), \ldots, b_{h,n} = \hat{\gamma}(h)$ be the first $h+1$ values of the sample autocovariance function (1.2) based on n observations $X_1, \ldots, X_n$. Now let $x_{h,\epsilon,b(n)}$ denote the solution of program $(P_{h,\epsilon,b(n)})$ based on the data $b(n) = (b_{0,n}, \ldots, b_{h,n})$. Then*

$$\|x_\infty - x_{h,\epsilon,b(n)}\|_\infty = \mathcal{O}_P\left((\log h \cdot h^{-k+3} + \epsilon)^{1/2}\right) \qquad n \to \infty \tag{4.3}$$

in probability.

Proof. As noticed above the random vector $n^{1/2}(\hat{\gamma}(0)-\gamma(0),\dots,\hat{\gamma}(h)-\gamma(h))$ converges in distribution to a normal random vector Y with mean zero and covariance matrix V_h given by (4.2), see [12, Prop. 7.3.4]. This means

$$P\left\{n^{1/2}\|b(n)-b\|_\infty \le n^{1/2}\epsilon\right\} - \int_{-n^{1/2}\epsilon}^{n^{1/2}\epsilon}\cdots\int_{-n^{1/2}\epsilon}^{n^{1/2}\epsilon} \phi_{0,V_h}(t)\,dt_1\cdots dt_h \to 0, \tag{4.4}$$

as $n\to\infty$, where ϕ_{0,V_h} denotes the density of the normal law with mean zero and covariance matrix V_h, and b is the vector of the first $h+1$ true autocovariances $b_0=\gamma(0),\dots,b_h=\gamma(h)$. Clearly the right hand term in (4.4) converges to 1 as $n\to\infty$, and hence so does the left hand term. Since by Theorem 3.9(c) we have $\|x_\infty - x_{h,\epsilon,b(n)}\|_\infty \le K\left((\log h\cdot h^{-k+3}+\epsilon)^{1/2}\right)$ for the events $\|b(n)-b\|_\infty\le\epsilon$, and with a constant $K>0$ depending only on x_∞, the claimed estimate (4.3) follows. □

It seems a realistic device that even for a long time observation $x_1,\dots,x_n$ with large n, we would limit the number h of Fourier coefficients used for the density estimation program $(P_{h,\epsilon,b(n)})$ to a moderate size in order to have a managable numerical problem. Nevertheless, one might wish to simultaneously increase the number h_n of coefficients used for the $(P_{h_n,\epsilon,b(n)})$ as the number n of observations increases. It is intuitively clear that n should grow much faster than h_n in order to eventually allow for an asymptotic estimate of the form

$$\|x_\infty - x_{h_n,\epsilon,b(n)}\|_\infty \sim \mathcal{O}_{\mathcal{P}}(\epsilon^{1/2}), \qquad (n\to\infty).$$

Once this is established, one would in fact obtain stochastic convergence $\|x_\infty - x_{h_n,\epsilon_n,b(n)}\|_\infty \xrightarrow{\mathcal{P}} 0$ on choosing appropriate tolerances $\epsilon_n\to 0^+$. We do not pursue this idea here since it does not seem to have any practical relevance.

Conclusion

We have calculated convergence rates for the deterministic best entropy spectral density estimation programs (P_n) based on the Fisher information measure (1.8), which might typically be applied in the analysis of stationary time series. Allowing for tolerances in the moment matching problem, we have shown that the programs (P_n) are stable under moderate changes of the program data (Theorem 3.9(c), Theorem 3.11(c)) and therefore may be expected to exhibit a reasonable robustness with regard to data corrupted by noise. A large sample Theorem guarantees the same qualitative results for the stochastic programs (Theorem 4.1).

References

[1] A. Ben-Tal, J.M. Borwein, M. Teboulle, A dual approach to multidimensional L_p spectral estimation problems, SIAM J. Control Optimization 26, 1988, 985 - 996.

[2] J.M. Borwein, W.Z. Huang, A heuristic method for polynomial moment problems with Boltzmann-Shannon entropy, SIAM J. Optimiz., to appear.

[3] J.M. Borwein, A.S. Lewis, Convergence of best entropy estimates, SIAM J. Optimization 1, 1991, 191 - 205.

[4] J.M. Borwein, A.S. Lewis, On the convergence of moment problems, Trans. Amer. Math. Soc. 325, 1991, 249 - 271.

[5] J.M. Borwein, A.S. Lewis, Partially finite programming in L_1 and the existence of maximum entropy estimates, SIAM J. Optimiz. 3 (1993), 248 - 267.

[6] J.M. Borwein, A.S. Lewis, Duality relationships for entropy-like minimization problems, SIAM J. Control Optimiz. 29, 1991, 325 - 338.

[7] J.M. Borwein, A.S. Lewis, Strong rotundity and optimization, SIAM J. Optimiz. 4 (1994), 146 - 158.

[8] J.M. Borwein, A.S. Lewis, D. Noll, Maximum entropy reconstruction using derivative information part I: Fisher information and convex duality, Math. Oper. Research, to appear.

[9] J.M. Borwein, A.S. Lewis, M.N. Limber, D. Noll, Maximum entropy reconstruction using derivative information part II: Computational results, Numerische Mathematik, to appear.

[10] J.M. Borwein, M.N. Limber, D. Noll, Fast heuristic methods for function reconstruction using derivative information and applications, to appear.

[11] P. Borwein, A.S. Lewis, Moment-matching and best entropy estimation, J. Math. Anal. Appl., to appear.

[12] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, 2nd ed. Springer Verlag 1991.

[13] J.P. Burg, Maximum entropy spectral analysis, 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, 1967.

[14] P.L. Butzer, R.J. Nessel, Fourier Analysis and Approximation I, Birkhäuser Verlag, 1971.

[15] E.W. Cheney, Introduction to Approximation Theory, McGraw-Hill, 1966.

[16] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, 1991.

[17] D. Dacunha-Castelle, F. Gamboà, Maximum d'entropie et problème des moments, Ann. Inst. Henri Poincaré 26, 1990, 567 - 596.

[18] A. Decarreau, D. Hilhorst, C. Lemaréchal, J. Navaza, Dual methods in entropy maximization: Applications to some problems in crystallography, SIAM J. Optimiz. 2, 1992, 173 - 197.

[19] T.S. Durrani, C.E. Goutis, Optimization techniques for digital images reconstructed from their projections, IEEE Proc. Part E, 1990, 161 - 169.

[20] G.J. Erickson, C.R. Smith (ed.), Maximum-entropy and Bayesian methods in science and engineering, Kluwer, Dordrecht, The Netherlands, 1988.

[21] F. Gamboa, Méthode du maximum d'entropie sur la moyenne et applications, Thèse Univ. Paris Sud, 1989.

[22] E. Gassiat, Problème sommatoire par maximum d'entropie, C.R. Acad. Sci. Paris, Sér. I, 303, 1986, 675 - 680.

[23] R.K. Goodrich, A. Steinhardt, L_2 spectral estimation, SIAM J. Applied Math. 46, 1986, 417 428.

[24] W.W. Hager, S.K. Mitter, Lagrange duality theory for convex control problems, SIAM J. Control Optimiz. 14, 1976, 843 - 856.

[25] E.T. Jaynes, Prior probabilities, IEEE Transactions on Systems Science and Cybernetics, SSC-4, 1968, 227 - 241. Reprinted in: Papers on prob., statistics and stat. Physics, Synthese Library vol. 158, 1983, Reidel, Dordrecht.

[26] E.T. Jaynes, On the rationale of maximum entropy methods, Proc. IEEE, Special issue on spectral estimation, vol. 70, 1982, 939 - 952.

[27] R.W. Johnson, J.E. Shore, Which is the better entropy expression for speech processing: $S \log S$ or $\log S$? IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-32, 1984, 129 - 137.

[28] A.S. Lewis, Pseudo-Haar functions and partially finite programming, to appear.

[29] A.S. Lewis, The convergence of entropy-based approximations for moment problems, Optimization 28 (1994), 383 - 395.

[30] F. Liese, I. Vajda, Convex statistical distances, Teubner, Leipzig 1987.

[31] G.G. Lorentz, Approximation of functions, 2nd ed., Chelsea, New York, 1986.

[32] J.H. McClellan, S.W. Lang, Multi-dimensional MEM spectral estimation, Spectral Analysis and its use in underwater acoustics, 10.1 - 10.8, Imperial College, London, U.K., 1982, Institute of Acoustics.

[33] J. Navaza, The use of non-local constraints in maximum entropy electron density reconstruction, Acta Crystallographica 1, 1985, 93 - 102.

[34] R.T. Rockafellar, Existence and duality theorems for convex problems of Bolza, Trans. Amer. Math. Soc. 159, 1971, 1 - 40.

[35] R.T. Rockafellar, State constraints in convex control problems of Bolza, SIAM J. Control Optimiz. 10, 1972, 691 - 715.

[36] A. Seghier, Reconstruction de la densité spectrale par maximum d'entropie, cas d-dimensionel, C.R. Acad. Sci. Paris, Sér. I, 305, 1987, 517 - 520.

[37] R.N. Silver, Quantum statistical inference, IEEE Proc. 1992.

[38] J. Skilling (ed.), Maximum-entropy and Bayesian methods, Kluwer, Dordrecht, The Netherlands, 1989.

[39] J. Skilling, S.F. Gull, The entropy of an image, SIAM-AMS Proc. 14, 1984, 167 - 189.

[40] C.R. Smith, G.J. Erickson (ed.), Maximum-entropy and Bayesian analysis and estimation problems, Kluwer, Dordrecht, The Netherlands, 1987.

[41] M. Teboulle, I. Vajda, Convergence of best ϕ-entropy estimates, IEEE Trans. Inform. Theory, 1992.

Dominikus Noll, Universität Stuttgart, Mathematisches Institut B, Pfaffenwaldring 57, 70550 Stuttgart, Germany