MATHEMATICS OF OPERATIONS RESEARCH
Vol 21, No. 2, May 1996
Printed in U S A

MAXIMUM ENTROPY RECONSTRUCTION USING DERIVATIVE INFORMATION, PART 1: FISHER INFORMATION AND CONVEX DUALITY

J. M. BORWEIN, A. S. LEWIS AND D. NOLL

Maximum entropy spectral density estimation is a technique for reconstructing an unknown density function from some known measurements by maximizing a given measure of entropy of the estimate. Here we present a variety of new entropy measures which attempt to control derivative values of the densities. Our models apply among others to the inference problem based on the averaged Fisher information measure. The duality theory we develop resembles models used in convex optimal control problems. We present a variety of examples, including relaxed moment matching with Fisher information and best interpolation on a strip.

1. Introduction. We consider the problem setting of spectral density estimation, where we wish to reconstruct an unknown density function $\bar{x}(t) \geq 0$ from a set of known measurements

$$\int_T a_i(t)\bar{x}(t)\,dt = b_i, \qquad i = 1,\dots,N. \tag{1.1}$$

Here the b_i might be known Fourier coefficients or Hausdorff moments of $\bar{x}(t)$. Such problems occur in various applications such as time series analysis, problems of image reconstruction, speech processing, or in crystallography. See Jaynes (1982), Skilling (1989), Erickson and Smith (1988), or Navaza (1986) for background information. In analogy with the maximum entropy principle we give preference to a solution $x(t)$ of (1.1) which maximizes a given measure of entropy, $H(x)$, or equivalently, minimizes the corresponding information measure, $I(x) = -H(x)$, usually an integral of the form

$$I_\phi(x) = \int_T \phi(x(t))\,dt. \tag{1.2}$$

Here $\phi\colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semi-continuous convex function. The entropy/information measures most frequently encountered in practice are the Boltzmann-Shannon and the Burg entropy/information measures, defined respectively by

$$\phi(x) = \begin{cases} x\log x, & x \geq 0,\\ +\infty, & x < 0,\end{cases} \qquad \phi(x) = \begin{cases} -\log x, & x > 0,\\ +\infty, & x \leq 0.\end{cases} \tag{1.3}$$

Received July 12, 1993; revised January 5, 1995.
AMS 1991 subject classification. Primary: 49M39; secondary: 49M45, 90C34.
OR/MS Index 1978 subject classification. Primary: 658 Programming/Nonlinear/Duality; Secondary: 567 Probability/Entropy.
Key words. Partially finite convex programming, duality, Fisher information, generalized solutions, maximum entropy method, optimal control, spectral density estimation.

0364-765X/96/2102/0442/$01.25

We refer the reader to Borwein and Lewis (1991a), Borwein and Lewis (1992), Borwein and Lewis (1991b), resp., Decarreau, Hilhorst, Lemaréchal, and Navaza (1992), Erickson and Smith (1988), Lin and Wong (1990), Skilling (1989) for a presentation of the corresponding mathematical models.

The purpose of our present investigation is to discuss extended entropy/information models, which include entropies like (1.3), and at the same time allow for objectives that attempt to control derivative values of the densities $x(t)$. In particular, our aim was to include the *averaged Fisher information measure*, which is related to the Fisher information known in the realm of statistical decision making. This requires models of the form

$$I_\phi(x) = \int_T \phi(x(t), x'(t))\, dt, \tag{1.4}$$

where $\phi\colon \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ is now proper lower semi-continuous and convex on $\mathbb{R}^2$. For instance, the averaged Fisher information $I_F = I_{\phi_F}$ is then defined as

$$\phi_F(x, v) = \begin{cases} v^2/x, & x > 0, \\ 0, & x = v = 0, \\ +\infty, & \text{elsewhere}. \end{cases} \tag{1.5}$$

The Fisher information has been introduced in Fisher (1930) in the realm of maximum likelihood estimation, while its multidimensional version has first been considered by J. L. Doob (1934). We refer to the Appendix I for a brief outline of the origin of these information functions and their relation to what we call the averaged Fisher information. The idea of using the model (1.4), (1.5) for the inference type problems (1.1) has been proposed in Silver (1992).

The basic mathematical model we are discussing is the following:

$$(P) \qquad \begin{aligned} &\text{minimize} \quad I_\phi(x) = \int_T \phi(x(t), x'(t))\, dt \\ &\text{subject to} \quad x \in \mathscr{A}(T), \\ &\int_T a_i(t) x(t)\, dt = b_i, \qquad i = 1, \dots, N, \end{aligned}$$

where $\mathscr{A}(T)$ is the space of absolutely continuous functions on a finite interval $T = [t_0, t_1]$, and where $a_i \in \mathscr{L}_\infty(T)$. Notice that from a modelling viewpoint, it seems not entirely logical that the inclusion of the derivative values of the densities extends only to the objective, and not to the constraints. However, our approach applies equally well to more general constraints, some of which are discussed among the examples in §5. Mainly for the sake of simplicity, we restrict our general outline to constraints of the form $Ax = b$.

Similar to the case of the information models (1.3), the key idea in analyzing problem (P) lies in applying convex programming duality theory. The details are presented in the principal §§3 and 4. It turns out that the duality we obtain resembles models occurring in optimal control and variational problems such as discussed for instance in Rockafellar (1971, 1972, 1981), or Hager and Mitter (1976), and we therefore make our arguments as general as possible in order to indicate how to include these situations.

The principal aim of our presentation is to obtain explicit dual models for the spectral density estimation problem (1.1) which in particular allow for an easy numerical treatment. It should be emphasised that despite the different nature of the constraint structure and the fact that our objectives (1.4) are autonomous, the principal difference with more standard optimal control problems lies in the fact that the integral functions I_ϕ typically take on finite values only on very small subsets of the underlying space $\mathscr{A}(T)$. As we shall see, this has the effect that some of the standard techniques and results from optimal control theory do not apply directly in our situation.

As will be seen, duality eventually provides the clue to translating problem (P) into a numerically tractable formulation (see §4). Numerical results for the case of the Fisher information are presented in Borwein, Lewis, Limber and Noll (1995), and Borwein, Limber and Noll (1996).

Notation. Throughout the paper we will use the following notations. The interval T will be fixed as $[0,1]$. We denote the space of absolutely continuous functions having derivatives in $\mathscr{L}_p(T)$ by $\mathscr{A}_p(T)$, that is, $\mathscr{A}_p(T) = \{x \in \mathscr{L}_1(T)\colon x' \in \mathscr{L}_p(T)\}$, and with $\mathscr{A}(T) = \mathscr{A}_1(T)$. The function $\phi(x, v)$ will always be proper lower semi-continuous and convex, $\phi\colon \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$. In order to avoid pathological cases, we assume that $\operatorname{dom}(\phi)$ has nonempty interior. In particular, ϕ is then a normal convex integrand in the sense of Rockafellar (1968). We define the integral functional $I_\phi(\cdot)$ on the space $\mathscr{A}(T)$ by

$$I_\phi(x) = \int_T \phi(x(t), x'(t))\, dt, \qquad x \in \mathscr{A}(T),$$

where dt refers to Lebesgue measure. Similarly, on separating the variables x and x', we obtain the integral functional $J_\phi(x, y)$ defined by

$$J_\phi(x, y) = \int_T \phi(x(t), y(t))\, dt, \qquad x \in \mathscr{L}_\infty(T),\ y \in \mathscr{L}_1(T).$$

It follows from the results in Rockafellar (1968, I, §2) that $J_\phi(\cdot,\cdot)$ is a proper lower semi-continuous convex integral functional on the space $\mathscr{L}_\infty(T) \times \mathscr{L}_1(T)$. As $J_\phi(x, x') = I_\phi(x)$ for the $x \in \mathscr{A}(T)$, it is now routine to check that I_ϕ is proper lower semi-continuous and convex on the space $\mathscr{A}(T)$.

2. Existence and convergence. In this section we first consider the problem of existence and uniqueness of solutions and a Lagrangian duality theory for problem (P), formulated as follows:

$$(P) \qquad \text{minimize} \quad I_\phi(x) = \int_T \phi(x(t), x'(t))\, dt$$

$$\text{subject to} \quad Ax = b, \qquad x \in \mathscr{A}(T),$$

where $b = (b_i)_{i=1}^N$, and where A denotes the operator $A\colon \mathscr{L}_1(T) \to \mathbb{R}^N$, defined as $Ax = (\int_T a_i(t)x(t)\, dt)_{i=1}^N$, $a_i \in \mathscr{L}_\infty(T)$. Here $I_\phi(x) = +\infty$ outside the set

$$\operatorname{dom} I_\phi(x) = \{x \in \mathscr{A}(T)\colon \phi(x(\cdot), x'(\cdot)) \in \mathscr{L}_1(T)\}.$$

We assume throughout that problem (P) is feasible, that is, that there exists $x_0 \in \operatorname{dom} I_\phi$ fitting the data $Ax_0 = b$. In the standard cases where the a_i represent either algebraic or trigonometric moments, a method for testing feasibility of the data has been presented in Borwein and Lewis (1991). The value of (P) is defined as $V(P) = \inf\{I_\phi(x)\colon Ax = b\} \in \mathbb{R} \cup \{-\infty\}$.

In the case of the Fisher-information function we have the following result.

THEOREM 2.1. *Consider the problem (P) for the Fisher information $I_F(\cdot)$. Suppose 1 is in the linear hull of $a_1, \dots, a_N$, i.e., $1 \in \operatorname{lin}\{a_1, \dots, a_N\}$. Then (P) has a unique optimal solution $\bar{x} \in \mathscr{A}_2(T)$.*

PROOF. Instead of solving problem (P) directly, we consider the transformation $x = y^2, x' = 2yy'$, which turns (P) into the equivalent and more standard nonconvex problem

$$(\tilde{P}) \qquad \text{minimize} \quad \|y'\|_2^2 = \int_T y'(t)^2\, dt$$

$$\text{subject to} \quad y \in \mathscr{A}_2(T), \qquad Ay^2 = b.$$

Observe here that $x \in \operatorname{dom} I_F$, i.e., $x'^2/x \in \mathscr{L}_1(T)$ if and only if $y' \in \mathscr{L}_2(T)$, or rather, if $y \in \mathscr{A}_2(T)$, and that the transformation makes sense since $x'(t) = 0$ for almost all t in the set $\{t \in T\colon x(t) = 0\}$. Therefore, an optimal solution $\bar{y}$ for $(\tilde{P})$ gives rise to an optimal solution $\bar{x} = \bar{y}^2$ for (P).

The existence of a solution for $(\tilde{P})$ follows from standard techniques of variational calculus, once it becomes clear that any minimizing sequence $\{y_m\}$ must be bounded. To prove this, all we have to check is that the sequence $\{y_m(0)\}$ is bounded. Assume on the contrary that $|y_m(0)| \to +\infty$. By assumption we have $1 \in \operatorname{lin}\{a_1, \dots, a_N\}$, and we may therefore assume for simplicity that $a_1 \equiv 1$ on T. Then we have

$$|y_m(t)| \geq |y_m(0)| - \left| \int_0^t y_m'(s)\, ds \right|$$

$$\geq |y_m(0)| - \|y_m'\|_1 \to +\infty,$$

since $\|y_m'\|_1 \leq M < +\infty$ by uniform boundedness, and this contradicts

$$\int_T |y_m|\, dt \leq \left(\int_T y_m^2\, dt \right)^{1/2} \left(\int_T dt \right)^{1/2}$$

$$= b_1^{1/2} \cdot \operatorname{meas}(T)^{1/2} < +\infty.$$

So $y_m(0)$ must be bounded, and this provides the tool for proving the existence of an optimal solution. Since this is now a standard argument, we leave the details to the reader.

Returning to the original problem (P), we show that its solution is unique. Let x_1, x_2 be two optimal solutions of (P), then by convexity, $\frac{1}{2}(x_1 + x_2)$ is again an optimal solution of (P). This implies

$$\int_T \frac{x_1'^2}{x_1} + \frac{x_2'^2}{x_2} = \int_T \frac{(x_1' + x_2')^2}{x_1 + x_2},$$

so together with the fact that $x'(t) = 0$ almost everywhere on the set $\{t \in T: x(t) = 0\}$, we get

$$\int_T \frac{(x_1'x_2 - x_2'x_1)^2}{x_1x_2(x_1 + x_2)} = 0,$$

and this implies $x_1'x_2 = x_2'x_1$, a.e. Since $\int_T x_1 = \int_T x_2$ by assumption, we have $x_1 = x_2$. □

REMARK. The present technique for proving the existence of a solution for the Fisher problem (P) does not apply to more general situations. Equally, the usual control type existence proofs do not work since they either require an objective $\phi(x, v)$ which is everywhere defined, or at least need some form of directional Lipschitz behaviour of $\phi(x, v)$ or other types of regularity conditions which are typically violated for the type of functionals considered here; see Loewen (1993), or Clarke (1990), Clarke and Loewen (1989) for a state of the art discussion. Section 4 will present a method for proving existence for more general objectives.

Let us now consider the problem of convergence. Suppose the sequence $\{a_i\}$ is weak star densely spanning in $\mathscr{L}_\infty(T)$, or equivalently, that there is at most one function $\bar{x} \in \mathscr{L}_1(T)$ satisfying $\int_T a_i(t)\bar{x}(t)\,dt = b_i$ for $i = 1, 2, \dots$. For fixed $N \in \mathbf{N}$ let (P_N) denote the problem so far denoted by (P), and suppose each (P_N) has a unique optimal solution x_N. The convergence problem asks whether x_N converges to the unknown underlying density $\bar{x}$ as $N \to +\infty$, and if so, in which sense.

THEOREM 2.2. *For the Fisher information $I_F(\cdot)$, and with $a_1 \equiv 1$, suppose every (P_N) is feasible with unique optimal solution x_N. Suppose* (i) *the values $V(P_N)$ are bounded. Then there exists a unique $\bar{x} \in \operatorname{dom} I_F(\cdot)$ satisfying $\int_T a_i\bar{x} = b_i$ for all $i = 1, 2, \dots,$ and we have $\|x_N - \bar{x}\|_\infty \to 0$ and $\|x_N' - \bar{x}'\|_2 \to 0$ as $N \to \infty$. Conversely,* (ii), *if the values $V(P_N)$ tend to $+\infty$, there is no function $x \in \operatorname{dom} I_F(\cdot)$ satisfying $\int_T a_i x = b_i$ for all $i = 1, 2, \dots$.*

PROOF. Working in the transformed problems $(\tilde{P}_N)$ as formulated in the proof of Theorem 2.1, let y_N be the unique nonnegative solution of $(\tilde{P}_N)$ satisfying $x_N = y_N^2$. Clearly $0 \le V(\tilde{P}_N) \le V(\tilde{P}_{N+1}) < \infty$. Assume first (i) that $V(\tilde{P}_N) \le M < \infty$. Then the sequence (y_N') is weakly relatively compact, and therefore has a weakly convergent subsequence in $\mathscr{L}_2(T)$, denoted (y_N') again. But $y_N(0)$ is bounded by the argument presented in the proof of Theorem 2.1, so by Arzela-Ascoli, y_N has a $\|\cdot\|_\infty$-convergent subsequence. Say $y_N \to y_\infty$ in $\|\cdot\|_\infty$, with $y_N' \to y_\infty'$ weakly in $\mathscr{L}_2(T)$. It follows from norm convergence that $\int_T a_i y_\infty^2 = b_i$ for every i, so $y_\infty^2 =: \bar{x}$, and $y_\infty' \in \mathscr{L}_2(T)$ gives $\bar{x} \in \operatorname{dom} I_F$. It follows from the Kadec-Klee property of the norm $\|\cdot\|_2$ in $\mathscr{L}_2(T)$ that $y_N' \to y_\infty'$ weakly in tandem with $\limsup_{N\to\infty}\|y_N'\|_2 \le \|y_\infty'\|_2$ imply $y_N' \to y_\infty'$ in $\|\cdot\|_2$. This proves $x_N' = 2y_N y_N' \to \bar{x}' = 2y_\infty y_\infty'$ in $\|\cdot\|_2$. Notice here that y_∞ is feasible for (P_N), hence $\|y_N'\|_2 \le \|y_\infty'\|_2$. Thus the entire sequence converges as claimed.

On the other hand (ii), if $V(P_N) \to +\infty$, then no $x \in \operatorname{dom} I_F$ may satisfy all the moment conditions, for otherwise x would be feasible for all (P_N), giving $V(P_N) \le I_F(x) < \infty$. □

REMARKS. (1) As we will see later on, much more can be said about the optimal solution $\bar{x}$ for the Fisher moment matching problem. For instance, $\bar{x}$ will turn out to be an analytic function if the a_i are chosen as analytic, and $\bar{x}$ will be at least of class C^2 even when the a_i are only assumed continuous.

(2) For the algebraic or trigonometric moments a_i, and under an additional smoothness assumption on the solution $\bar{x} \in \operatorname{dom} I_F$, one may give rates of convergence for $\|x_N' - \bar{x}'\|_2 \to 0$ and $\|x_N - \bar{x}\|_\infty \to 0$. See Noll (1995).

The averaged Fisher-information (1.5) may be considered as a special case of a more general class of integrands of the form

$$\phi(x, v) = \begin{cases} x\psi(v/x) & \text{for } x > 0, \\ 0^+\psi(v) & \text{for } x = 0, \\ +\infty & \text{for } x < 0. \end{cases} \tag{2.1}$$

Here $\psi: \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous proper convex, its domain includes the half line $[0, +\infty)$, and $0^+\psi$ denotes its recession function (see Rockafellar (1970)). The class (2.1) was considered in Borwein and Lewis (1992) in a different context, and it was referred to as the *Csiszar-distances*. The case of the Fisher-information is recovered by choosing $\psi(t) = t^2$. In order to obtain results for the integrands (2.1) which extend Theorems 2.1 and 2.2, we need to impose the following assumptions on ψ:

(1) ψ is strictly convex on its domain;

(2) ψ coercive, that is $\psi(t)/|t| \to +\infty$ as $|t| \to +\infty$.

Notice that condition (2) here simplifies the definition of ϕ in (2.1) above, since we then have $0^+\psi(v) = +\infty$ for $v \neq 0, 0^+\psi(0) = 0$. As we will see in §4, Theorem 2.1 may be extended to the Csiszar class by means of the bidual approach.

In the second part of this section, we address the duality of problem (P) when considered as an infinite-dimensional convex optimization program. This requires introducing a Lagrangian formulation for (P). Let us consider the following *first Lagrangian*

$$\begin{aligned} L_1(x; \lambda) &= I_\phi(x) + \langle \lambda, Ax - b \rangle \\ &= \int_T \phi(x(t), x'(t))\, dt + \sum_{i=1}^{N} \lambda_i \left(\int_T a_i(t) x(t)\, dt - b_i \right), \end{aligned} \tag{2.2}$$

with $x \in \mathscr{A}(T)$, $\lambda \in \mathbb{R}^N$, taking on values in $\mathbb{R} \cup \{+\infty\}$. The duality arising from $L_1(x; \lambda)$ will be discussed presently. There is, however, a second possibility for a Lagrangian duality, which arises from separating the variables x and x'. The corresponding *second Lagrangian* is

$$\begin{aligned} L_2(x, y; w, \lambda) &= J_\phi(x, y) + \langle w, x' - y \rangle + \langle \lambda, Ax - b \rangle \\ &= \int_T \phi(x(t), y(t))\, dt + \int_T w(t)(x'(t) - y(t))\, dt \\ &\quad + \sum_{i=1}^{N} \lambda_i \left(\int_T a_i(t) x(t)\, dt - b_i \right), \end{aligned} \tag{2.3}$$

with $x \in \mathscr{A}(T)$, $y \in \mathscr{L}_1(T)$, $w \in \mathscr{L}_1(T)^* = \mathscr{L}_\infty(T)$, $\lambda \in \mathbb{R}^N$. The corresponding duality will be discussed in §3.

Let us start by considering the first Lagrangian $L_1(x; \lambda)$. Since the associated duality resembles more standard techniques, we shall be very brief here, pointing out only the major difference of our type of programs (P) with optimal control type situations.

Notice first that the primal program (P) admits the equivalent formulation

$$(P) \qquad \text{minimize} \quad \sup_{\lambda \in \mathbb{R}^N} L_1(x;\lambda) \quad \text{subject to} \quad x \in \mathscr{A}(T).$$

We define the corresponding dual program as

$$(P_1^*) \qquad \text{maximize} \quad \inf_{x \in \mathscr{A}(T)} L_1(x;\lambda) \quad \text{subject to} \quad \lambda \in \mathbb{R}^N.$$

It is clear that $\inf_x \sup_\lambda L_1(x;\lambda) \geq \sup_\lambda \inf_x L_1(x,\lambda)$, (weak duality) that is, $V(P) \geq V(P_1^*)$. We show that under the mild constraint qualification hypothesis given below we get a strong duality result, which tells us that the values of (P) and (P_1^*) are the same, and moreover, that (P_1^*) admits an optimal solution:

$$(CQ_1) \qquad b \in \operatorname{ri} A(\operatorname{dom} I_\phi).$$

Here $\operatorname{ri}(M)$ denotes the interior of M relative to the affine subspace it generates in $\mathbb{R}^N$. Equivalently, (CQ_1) means that $A(\operatorname{dom} I_\phi) - b$ is absorbing in the linear subspace it generates in $\mathbb{R}^N$.

THEOREM 2.3. *Suppose (CQ_1) is satisfied. Then problem (P_1^*) admits an optimal solution $\bar{\lambda}$, satisfying*

$$\inf_{x \in \mathscr{A}(T)} L_1(x;\bar{\lambda}) = V(P).$$

In particular, $V(P) = V(P_1^*)$.

PROOF. The proof is standard, see for instance Borwein (1983) or Noll (1991). Indeed, let S be the linear subspace of $\mathbb{R}^N$ generated by $A(\operatorname{dom} I_\phi) - b$, and define a convex function $f\colon S \to \mathbb{R} \cup \{+\infty\}$ by

$$f(\theta) = \inf\{I_\phi(x) : Ax - b = \theta\}, \qquad \theta \in S.$$

It follows that $\partial f(0) \neq \emptyset$ since f is lower semi-continuous and $\operatorname{dom} f$ is absorbing as a consequence of (CQ_1). Now $-\bar{\lambda} \in \partial f(0)$ gives the required Lagrange multiplier $\bar{\lambda}$. □

As (CQ_1) may not be easy to check directly, let us formulate the following condition, which is sufficient to imply (CQ_1), as some standard arguments will show:

$$\text{There exists } \hat{x} \in \mathscr{C}^1(T) \cap \operatorname{dom} I_\phi \text{ such that } A\hat{x} = b \text{ and}$$

$$(\hat{x}(t), \hat{x}'(t)) \in \operatorname{int}(\operatorname{dom} \phi) \text{ for every } t \text{ in some interval } (\alpha, \beta) \subseteq T.$$

This condition may be weakened considerably in many concrete examples. For instance in the case of the averaged Fisher information we have the following:

EXAMPLE 2.1. Assume for simplicity that the a_i form a pseudo-Haar system, which is to say that they are linearly independent on any set of positive measure (cf. Borwein and Lewis (1991)). Now consider the Fisher information I_F, or more generally any integrand $\phi(x, x')$ of the class (2.1). Then the following is sufficient to imply the constraint qualifications (CQ_1):

(CQ_F) There exists $\hat{x}(T), \hat{x} \in \mathscr{L}_1(T), \hat{x} \geq 0$, a.e., $\hat{x} \neq 0$, such that $A\hat{x} = b$.

Indeed, by Borwein and Lewis (1991), Theorem 2.9, (CQ_F) implies that there exists an $\bar{x} \in \mathscr{L}_\infty(T)$, $\bar{x} \geq \epsilon > 0$ for some $\epsilon > 0$, satisfying $A\bar{x} = b$. Now consider a sequence $\{x_n\}$ of positive C^1 functions which converges to $\bar{x}$ in $\mathscr{L}_1$ norm. So $Ax_n \to b$ $(n \to \infty)$. We may assume that $x_n \geq \epsilon/2 > 0$ on T. Due to the fact that A is open as an operator mapping $\mathscr{C}^1(T)$ onto $\mathbb{R}^N$, there exists $\eta > 0$ such that, given any $|\eta_i| \leq \eta$, $i = 1, \dots, N$, we find $v \in \mathscr{C}^1(T)$ such that $\|v\|_\infty \leq \epsilon/4$ and $Av = (\eta_i)_{i=1}^N$. Choose n so large that $\xi_i = b_i - (Ax_n)_i$ satisfies $|\xi_i| < \epsilon/8$. Now let $|\zeta_i| < \epsilon/8$ be fixed. We find $v \in \mathscr{C}^1$ such that $Av = (\xi_i) - (\zeta_i)$. Then $A(x_n + v) = b - (\xi_i) + Av = b + (\zeta_i)$, proving the desired $b + [-\epsilon/8, \epsilon/8] \subseteq A(\text{dom } I_F)$. Indeed, we have $x_n + v \in C^1$, $x_n + v \geq \epsilon/8$, hence $x_n + v \in \text{dom } I_F$ (and $x_n + v \in \text{dom } I_\phi$ for the integrands (2.1) correspondingly), and we use that $\zeta_i \in [-\epsilon/8, \epsilon/8]$ was chosen arbitrarily. □

Let us now consider the following consequences of the constraint qualification in the case of the Fisher information, which we state explicitly due to its relevance to the second part (Borwein, Lewis, Limber and Noll (1995)) of this paper.

THEOREM 2.4. *Let $I_F(\cdot)$ be the Fisher information* (1.5). *Suppose the constraint qualification (CQ_F) is satisfied and let $\bar{\lambda}$ be the dual optimal solution for (P_1^*). Suppose $1 \in \text{lin}\{a_1, \dots, a_N\}$, and $b \neq 0$, then*:

(1) *The unique optimal solution $\bar{x}$ for (P) is strictly positive on $T = [0, 1]$*;

(2) *For $\bar{x} \in \mathscr{A}_2(T)$ to be the unique optimal solution for (P) it is necessary and sufficient that $\bar{x}$ be strictly positive, fit the data $A\bar{x} = b$, and satisfy the Euler-Lagrange equation*

$$-2\bar{x}\bar{x}'' + \bar{x}'^2 + \sum_{i=1}^{N} \bar{\lambda}_i a_i \cdot \bar{x}^2 = 0, \tag{2.4}$$

with boundary conditions $\bar{x}'(0) = \bar{x}'(1) = 0$.

PROOF. Let us prove (1). Let $\bar{x}$ be the unique solution for (P) guaranteed by Theorem 2.1, and let $\bar{\lambda}$ be the Lagrange multiplier which exists by Theorem 2.3. First observe that $(\bar{x}, \bar{\lambda})$ is a saddle point for $L_1(x; \lambda)$, that is

$$0 \in \partial_x L_1(\bar{x}, \bar{\lambda}) \quad \text{and} \quad 0 \in \partial_\lambda L_1(\bar{x}; \bar{\lambda}). \tag{2.5}$$

(Notice here that ∂_x is a subderivative, ∂_λ a superderivative.)

The second condition in (2.5) simply means $A\bar{x} = b$, while the first condition is equivalent to

$$\begin{aligned} 0 &\leq \frac{1}{\tau}\left(L_1(\bar{x} + \tau h; \bar{\lambda}) - L_1(\bar{x}; \bar{\lambda})\right) \\ &= \frac{1}{\tau}(I_F(\bar{x} + \tau h) - I_F(\bar{x})) + \int_T \sum_{i=1}^{N} \bar{\lambda}_1 a_i \cdot h < +\infty \end{aligned} \tag{2.6}$$

for any $h \in \mathscr{A}(T)$ such that $\bar{x} + \tau h \in \text{dom } I_F$ for small $\tau > 0$.

Take $h \equiv 1$, which is certainly admitted in (2.6). Since $\bar{x}'(t) = 0$ for almost all t in the set $\{t \in T: \bar{x}(t) = 0\}$, we may restrict the integral over the set $\{t \in T: \bar{x}(t) > 0\}$.

Then (2.6) gives

$$0 \le \int_{\{\bar{x}>0\}} \frac{-\bar{x}'^2}{\bar{x}(\bar{x}+\tau)} + A\bar{\lambda} < +\infty.$$

The integrand is nondecreasing in $\tau > 0$, so monotone convergence allows us to pass to the limit $\tau \to 0^+$ under the integral sign, showing that $\bar{x}'^2/\bar{x}^2$, and hence $\bar{x}'/\bar{x} = (\log \bar{x})'$, must be integrable on $\{t: \bar{x}(t) > 0\}$. Now observe that by assumption $b \neq 0$, so $\bar{x} \equiv 0$ could not be optimal, and hence $\{\bar{x} > 0\}$ is nonempty. We show that this implies $\bar{x} > 0$ on all of T. Indeed, suppose for instance there exists an interval $(\alpha, \beta) \subset \{\bar{x} > 0\}$ is such that $\bar{x}(\alpha) = 0$ and $\bar{x}(\beta) > 0$. Then

$$\int_{\alpha+\delta}^{\beta} (\log \bar{x})'(t)\, dt = \log \bar{x}(\beta) - \log \bar{x}(\alpha+\delta) \to +\infty \qquad (\delta \to 0^+),$$

contradicting the integrability of $(\log \bar{x})'$ on $\{\bar{x} > 0\}$. This proves statement (1).

With the fact $\bar{x} > 0$ on T established, we are now back in a standard control type situation, the integrand $\phi(\bar{x}(t), \bar{x}'(t))$ now being locally Lipschitz in the first variable along the optimal path. The rest of statement (2) therefore follows via standard arguments in control theory (see e.g. Loewen (1993)). Notice that convexity as usual gives the sufficiency in statement (2). □

EXAMPLE 2.2. Consider problem (P) with the Fisher information and $N = 1$, $a_1 \equiv 1$, $b_1 = 1$. Then the primal optimal $\bar{x}$ is $\bar{x} \equiv 1$, which has $I_F(\bar{x}) = 0$. So the Lagrange multiplier $\bar{\lambda}_1$ must be $\bar{\lambda}_1 = 0$. Now under the transform $x = y^2$, equation (2.4) takes the form:

$$4y'' = \lambda y, \qquad y'(0) = y'(1) = 0.$$

This problem has the "negative" eigenvalues $\lambda_0 = 0$, $\lambda_k = -4k^2\pi^2$, $k = 1, 2, \dots$. The corresponding eigensolutions are $y_0 \equiv 1$ resp. $y_k(t) = c_k \cos k\pi t$, giving $x_0 = 1$, $x_k(t) = c_k^2 \cos^2 k\pi t$. Fitting the data $\int_T x_k = 1$ gives $c_k = \sqrt{2}$. But the x_k for $k = 1, 2, \dots$ are not the optimal solutions of (P) since $x_k(\frac{1}{2}) = 0$, and we know that x has to be strictly positive. This follows from $I_F(x_k) > 0$ as well as from the fact that $\lambda_1 \neq 0$ could not be optimal for the dual program.

This shows that a pair (x, λ) may be both, a solution of the boundary value problem and a feasible pair, but fail to be a saddle point for $L_1(x; \lambda)$, since $x > 0$ is violated. The reason is of course that without the condition $x > 0$, even though (x, λ) satisfies the Euler-Lagrange equation and $Ax = b$, we may not argue that the first condition in (2.5) is satisfied. □

REMARK. We have seen that the duality associated with the first Lagrangian led to a Euler-Lagrange equation in the case of the averaged Fisher information measure $I_F(\cdot)$. It may be seen from Example 5.1 that this need not be the case for other objectives $I_\phi(\cdot)$. In fact, since we restrict the domain of the $I_\phi(\cdot)$ to functions $x \ge 0$, it may happen that the class of h which we are allowed to use for our variation is not rich enough in order to apply the Dubois-Reymond Lemma. Typically, the method presented in this section will then only lead to a variational inequality. As we will see in the next two sections, the duality theory associated with the second Lagrangian is generally better suited to deal with this phenomenon.

3. Duality. In this section we consider the duality theory based on the second Lagrangian $L_2(x, y; w, \lambda)$ (see (2.3)) which we introduced by separating the variables x and x'. The formulation resembles the duality theory for convex control problems, as for instance presented in Rockafellar (1968, 1974, 1971), Hager and Mitter (1976), or Dacunha-Castelle and Gamboa (1990). The main difference to more standard optimal control type problems lies in the fact that in spectral density estimation the objectives J_ϕ are defined on small sets (compare with Rockafellar (1972)), where interiority type assumptions are not satisfied. Even more, standard results from nonsmooth optimal control theory will not always be applicable, as we already pointed out in the previous section. On the other hand, our models have the nice feature that the objectives are jointly convex, and this enables us to present a fairly concrete and explicit duality.

Let us start by observing that the primal problem may be stated in terms of the Lagrangian (2.3):

(P)

$$\text{minimize} \quad \sup_{w \in \mathscr{L}_\infty(T),\, \lambda \in \mathbb{R}^N} L_2(x, y; w, \lambda) \quad \text{subject to } x \in \mathscr{A}(T),\ y \in \mathscr{L}_1(T).$$

The corresponding dual program is then

(P_2^*)

$$\text{maximize} \quad \inf_{x \in \mathscr{A}(T),\, y \in \mathscr{L}_1(T)} L_2(x, y; w, \lambda) \quad \text{subject to } w \in \mathscr{L}_\infty(T),\ \lambda \in \mathbb{R}^N.$$

One immediately has $\inf_{x,y} \sup_{w,\lambda} L_2(x, y; w, \lambda) \geq \sup_{\lambda, w} \inf_{x,y} L_2(x, y; w, \lambda)$ (weak duality). In order to prove strong duality, we need a constraint qualification:

(CQ_2) $\quad (0, b)$ is an algebraic interior point of $(D \otimes A)(\mathrm{dom}\, J_\phi)$ in the closed affine subspace it generates in $\mathscr{L}_1(T) \times \mathbb{R}^N$.

Here the operator $D: \mathscr{A}(T) \times \mathscr{L}_1(T) \to \mathscr{L}_1(T)$ is defined as $D(x, y) = x' - y$, and A: $\mathscr{L}_1(T) \to \mathbb{R}^N$ is the usual $Ax = (\int_T a_i x)_{i=1}^N$, and $(D \otimes A)(x, y) = (x' - y, Ax)$.

Before discussing this condition in detail, let us formulate its main consequence.

PROPOSITION 3.1. *Suppose (CQ_2) is satisfied. Then there exist $\bar{w} \in \mathscr{L}_\infty(T)$ and $\bar{\lambda} \in \mathbb{R}^N$ such that*

$$\inf_{x \in \mathscr{A}(T),\, y \in \mathscr{L}_1} L_2(x, y; \bar{w}, \bar{\lambda}) = \inf(P). \tag{3.1}$$

In particular, the values of (P) and (P_2^*) are the same.

PROOF. We sketch the argument, which is standard and follows, e.g., from results in Borwein (1983) or Noll (1991). Let $S = \overline{\mathrm{lin}}((D \otimes A)(\mathrm{dom}\, J_\phi) - b) \subset \mathscr{L}_1(T) \times \mathbb{R}^N$. Define a convex function $f: S \to \mathbb{R} \cup \{+\infty\}$ by

$$f(z, \theta) = \inf\{J_\phi(x, y) : z = x' - y,\ \theta = Ax - b\}.$$

We have to show that f has a subgradient at $(0,0)$ in S. Since the constraint qualification guarantees that $(0,0)$ is an interior point of dom f, this follows either directly from Noll (1991), or from Borwein (1983) using the fact that J_ϕ is lower semi-continuous.

Let an element of $\partial f(0,0)$ be represented by $-(\overline{w}, \overline{\lambda}) \in \mathscr{L}_\infty \times \mathbb{R}^N$, then it is routine to check that $\overline{w}$ and $\overline{\lambda}$ are the desired Lagrange multipliers. □

EXAMPLE 3.1. We show that in the case of the Fisher information J_F, resp. the Csiszar objectives (2.1), the constraint qualification (CQ_F), with the $\{a_i\}$ being pseudo-Haar, implies (CQ_2).

Indeed, we have already seen in Example 2.1 that (CQ_F) provides $\hat{x} \in \mathscr{C}^1(T)$ such that $\hat{x} \geq \epsilon > 0$ and $A\hat{x} = b$. Now given any $z \in \mathscr{L}_1(T)$, $\theta \in \mathbb{R}^N$, we have to show that, for some $\rho > 0, (\rho z, b + \rho\theta) \in (D \otimes A)(\text{dom } J_F)$.

First observe that $A\colon \mathscr{C}^1(T) \to \mathbb{R}^N$ is surjective and hence open at $\hat{x}$, so for some $\delta > 0$, every $\mu \in B(0, \delta)$ may be represented as $\mu = Ax = A(\hat{x} + x) - b$ for some $x \in \mathscr{C}^1(T)$ having $\|x\|_\infty < \epsilon/4$, say. Let $v \in \mathscr{A}(T)$, with $v' = z$ fixed. Choose $\rho > 0$ so small that $\|\rho v\|_\infty < \epsilon/4$ and $\|\rho(Av - \theta)\|_\infty < \delta$. By the above we find $x_0 \in \mathscr{C}^1(T)$ having $\|x_0\|_\infty < \epsilon/4$ such that $\mu := \rho(\theta - Av) = A(\hat{x} + x_0) - b$. Let $x := \hat{x} + x_0 + \rho v$, then $x \geq \epsilon/2 > 0$. Let $y = \hat{x}' + x_0' \in \mathscr{C}(T) \subset \mathscr{L}_2(T)$. Then $x' - y = \rho v' = \rho z$ and $Ax - b = A\hat{x} - b + Ax_0 + \rho Av = \rho\theta$, as desired. Since $(x, y) \in \text{dom } J_F$, this proves (CQ_2). The same argument also shows that (CQ_F) implies (CQ_2) for any of the Csiszar distances (2.1). □

So far we know that under the constraint qualification hypothesis (CQ_2), problem (P^*) has an optimal solution $(\overline{w}, \overline{\lambda}) \in \mathscr{L}_\infty(T) \times \mathbb{R}^N$, a Lagrange multiplier, and that the values of (P) and (P_2^*) are the same. We shall now pursue two ideas.

We will show (I) that in many cases $\overline{w}$ is in fact an absolutely continuous function. This will require several steps, and will eventually be proved in §4, Propositions 4.1 and 4.2. Then (II) we will provide a method to recover the primal optimal solution $(\bar{x},\bar{y}) = (\bar{x}, \bar{x}')$ (if any) from the dual $(\overline{w}, \overline{\lambda})$. This will be established in Theorem 4.3. Let us start with our program (I). The first step is provided by the following:

PROPOSITION 3.2. *Suppose (CQ_2) is satisfied, and let $(\overline{w}, \overline{\lambda})$ be the dual optimal solution guaranteed by Proposition* 3.1. *Then $\overline{w} \in \mathfrak{BV}(T)$. More generally, let $(w, \lambda) \in \mathscr{L}_\infty(T) \times \mathbb{R}^N$ be any pair such that* $\inf\{L_2(x, y; w, \lambda)\colon x \in \mathscr{A}(T),\ y \in \mathscr{L}_1(T)\} > -\infty$. *Then $w \in \mathfrak{BV}(T)$.*

PROOF. Since $\overline{w}$, w are elements of $\mathscr{L}_\infty(T)$, the statement is to be understood in the sense that some realization of $\overline{w}$ or w is of bounded variation.

Recall our general assumption that dom ϕ has nonempty interior. Hence there exists a fixed $(\hat{x}, \hat{y}) \in \text{dom } J_\phi$ such that, for some $\epsilon > 0$, every $(\hat{x} + x, \hat{y})$ with $\|x\|_\infty \leq \epsilon$ is contained in the domain of J_ϕ. Since $L_2(\hat{x} + x, \hat{y}; w, \lambda) \geq c > -\infty$ by assumption, we deduce that $\langle w, x' \rangle$ is bounded on the set of $\{x \in \mathscr{A}(T)\colon \|x\|_\infty \leq \epsilon\}$. This means that the functional $\langle w, x' \rangle$ is a Radon measure, that is, there exists a finite Borel measure μ having

$$\langle w, x' \rangle = \int_T x(s)\, d\mu(s), \qquad x \in \mathscr{A}(T).$$

Let μ have the distribution function $v \in \mathfrak{BV}(T)$, then $v = -w$ a.e., up to a constant. This proves the statement. □

REMARK. A more elementary proof for this result could be obtained using the kind of argument used in Hager and Mitter (1976). If we wish to prove an analogous result for nonautonomous objectives $\phi(t, x, x')$, their technique seems to be better suited than the one presented above.

Let us keep the notation $(\overline{w}, \overline{\lambda})$ for the dual optimal pair. Since $\overline{w} \in \mathfrak{B}\mathscr{V}[0,1]$, we may now apply the integration by parts formula for functions of bounded variation (see Dunford and Schwartz (1963)) in order to write

$$(3.2) \quad \langle \overline{w}, x' \rangle = \int_0^1 \overline{w}(s) x'(s)\, ds = x(1)\overline{w}(1) - x(0)\overline{w}(0) - \int_0^1 x(s)\, d\overline{w}(s).$$

Here we may and will assume that $\overline{w}$ is continuous from the right, so that the meaning of $\overline{w}(0)$ is clear. But then, by choosing $x \equiv 1$, we obtain $\overline{w}(1) = \overline{w}(0) + \int_0^1 d\overline{w}$, which clarifies the meaning of $\overline{w}(1)$.

As we will see, it turns out that in many cases $\overline{w}$ is in fact absolutely continuous. This will then permit us to write the Riemann-Stieltjes integral on the right-hand side of (3.2) in the form

$$\int_0^1 x(s)\, d\overline{w}(s) = \int_0^1 x(s) \overline{w}'(s)\, ds.$$

DEFINITION 3.1. For a function $w \in \mathfrak{B}\mathscr{V}(T)$, let $w = v + u$ be the Lebesgue decomposition of $w \in \mathfrak{B}\mathscr{V}(T)$, that is $v \in \mathscr{A}(T), u' \equiv 0$ a.e., and $w(0) = v(0)$ (cf. Hewitt and Stromberg (1969), Rudin (1973)). We will occasionally use the notation $v = w_a, u = w_s$. Furthermore, let $u = u^+ - u^-$ for nondecreasing u^+, u^-, with $u^+(0) = u^-(0) = 0$, which is called the Hahn decomposition of u. □

With these definitions, and after some manipulations, the second Lagrangian (2.3) takes on the form

$$(3.3) \quad L_2(x, y; w, \lambda) = J_\phi(x, y) - \langle v' - A^t\lambda, x \rangle - \langle w, y \rangle + w(1)x(1) - w(0)x(0) - \langle du, x \rangle - \langle \lambda, b \rangle.$$

We need to calculate the conjugate of $L_2(x, y; w, \lambda)$, considered as a function of $x \in \mathscr{A}(T)$, $y \in \mathscr{L}_1(T)$.

THEOREM 3.3. *The Young-Fenchel conjugate $L_2^*(\cdot, \cdot; w, \lambda)$ of $L_2(\cdot, \cdot; w, \lambda)$: $\mathscr{A}(T) \times \mathscr{L}_1(T) \to \mathbb{R} \cup \{+\infty\}$ with respect to the incomplete dual pairing $\langle \mathscr{A}(T) \times \mathscr{L}_1(T), \mathscr{L}_1(T) \times \mathscr{L}_\infty(T) \rangle$ equals*

$$(3.4) \quad L_2^*(r, s; w, \lambda) = J_{\phi^*}\left(r + v' - \sum_{i=1}^N \lambda_i a_i, s + w \right) + \sum_{i=1}^N \lambda_i b_i - Pw(1) + Qw(0) + Mu^+(1) - mu^-(1),$$

$$r \in \mathscr{L}_1(T),\ s \in \mathscr{L}_\infty(T)$$

where $M = \sup\{\xi : \exists \eta\ \phi(\xi, \eta) < +\infty\}$, $m = \inf\{\xi : \exists \eta\ \phi(\xi, \eta) < +\infty\}$, *and*

$$(3.5) \qquad P = \begin{cases} M & \text{if } w(1) \le 0, \\ m & \text{if } w(1) > 0, \end{cases} \qquad Q = \begin{cases} M & \text{if } w(0) > 0, \\ m & \text{if } w(0) \le 0. \end{cases}$$

The proof of the theorem will be mainly based on the following result whose proof may be found in the Appendix II:

LEMMA 3.4. *The Young-Fenchel conjugate of* J_ϕ: $\mathscr{A}(T) \times \mathscr{L}_1(T) \to \mathbb{R} \cup \{+\infty\}$ *with respect to the dual pairing* $\langle \mathscr{A}(T) \times \mathscr{L}_1(T), \mathscr{L}_1(T) \times \mathscr{L}_\infty \rangle$ *equals* J_{ϕ^*}.

PROOF OF THEOREM 3.3. We have to show that

$$\sup_{x \in \mathscr{A}(T),\, y \in \mathscr{L}_1(T)} \langle r, x \rangle + \langle s, y \rangle - L_2(x, y; w, \lambda) \tag{3.6}$$

equals the right-hand side of (3.4). Observe that the $x \in \mathscr{A}(T)$ occurring in the supremum in (3.6) may be decomposed into two terms. Firstly, if the singular measure du is concentrated on the Lebesgue null set Ω, say, consider the part of $x \in \mathscr{A}(T)$ which vanishes on an open set G of arbitrarily small Lebesgue measure containing $\Omega \cup \{0, 1\}$. Then according to Lemma 3.4, the supremum over such x in (3.6) yields the expression

$$J_{\phi^*}\left(r + v' - \sum_{i=1}^{N} \lambda_i a_i, s + w\right) + \sum_{i=1}^{N} \lambda_i b_i,$$

leaving the other terms in (3.3) unaffected in the limit meas$(G) \to 0$. Secondly, consider the part of $x \in \mathscr{A}(T)$ which vanishes outside the set G above. Then the corresponding contribution to the term $J_\phi(x, y) - \langle r + v' - A^t\lambda, x \rangle$ may be made arbitrarily small by letting meas$(G) \to 0$, while on the other hand the contribution to

$$\sup_{x \in \mathscr{A}(T)} \left(-w(1)x(1) + w(0)x(0) + \langle du, x \rangle\right)$$

yields the term $-Pw(1) + Qw(0) + M\,du^+(T) - m\,du^-(T)$, where P, Q have the meaning (3.5). Now $du^+(T) = u^+(1)$, $du^-(T) = u^-(1)$ finally gives rise to (3.4). □

REMARK. Notice that we intend $0 \cdot (\pm\infty) = 0$ in formula (3.4), so the cases $M = +\infty, m = -\infty$ are not excluded. For instance, $M = +\infty$ simply implies that $u^+(1) = 0$, and so $u^+ = 0$, and similarly for the other terms occurring in (3.4). Let us mention that the typical case for the moment matching problem is $M = +\infty$, $m = 0$. This means for instance that the positive singular part u^+ vanishes, while the negative singular part u^- might be left over. Usually additional arguments are needed in order to show that u^- vanishes.

In summary, the results obtained so far allow us to state the dual problem (P_2^*) in the following form:

$$(P_2^*) \qquad \text{maximize} \quad -J_{\phi^*}\left(v' - \sum_{i=1}^{N} \lambda_i a_i, w\right) - \sum_{i=1}^{N} \lambda_i b_i$$

$$- Pw(1) + Qw(0) + M(w - v)^+(1) - m(w - v)^-(1)$$

$$\text{subject to} \quad w \in \mathfrak{BV}(T), v = w_a, v(0) = w(0), \lambda \in \mathbb{R}^N.$$

4. **Special cases.** In this section we show that the general dual scheme we obtained in §3 may be simplified in many cases. Before presenting the most typical case in spectral density estimation, let us recall the meaning of the Lebesgue and Hahn decompositions $w = v + u$ and $u = u^+ - u^-$ given in Definition 3.1.

PROPOSITION 4.1. *Let $(\overline{w}, \overline{\lambda})$ be the dual optimal solution for (P_2^*) guaranteed by (CQ_2). Suppose*
(a) $M = +\infty$. *Then* $\overline{u}^+ \equiv 0, \overline{w}(1) \geq 0, \overline{w}(0) \leq 0$. *In particular, if*
(b) $M = +\infty$ *and* $m = 0$, *then the conjugate Lagrangian equals*

$$(4.1) \qquad L_2^*(r, s; \overline{w}, \overline{\lambda}) = J_{\phi^*}(r + v' - A^t\lambda, s + \overline{w}) + \langle \overline{\lambda}, b \rangle.$$

PROOF. Since $M = +\infty$ implies $\overline{u}^+(1) = 0$, and since we always have $\overline{u}^+(0) = 0$, we obtain $\overline{u}^+ \equiv 0$. This implies $\overline{u} = \overline{u}^-$. Also $\overline{w}(1) < 0$ would imply $P = M = +\infty$, and then the finiteness of $P\overline{w}(1)$ would force $\overline{w}(1) = 0$, a contradiction. So $\overline{w}(1) \geq 0$, and similarly, $\overline{w}(0) \leq 0$.

Now according to (3.4), the conjugate Lagrangian equals

$$(4.2) \qquad L_2^*(r, s; \overline{w}, \overline{\lambda}) = J_{\phi^*}(r + \overline{v}' - A^t\overline{\lambda}, s + \overline{w}) + \langle \overline{\lambda}, b \rangle + m(\overline{w}(0) - \overline{w}(1) - \overline{u}(1))$$

and the last term cancels when $m = 0$. □

Similar reasoning shows that $\overline{u}^- \equiv 0$ and $\overline{w}(1) \leq 0, \overline{w}(0) \geq 0$ in case $m = -\infty$, so that we immediately deduce that $\overline{w} = \overline{v}, \overline{w}(0) = \overline{w}(1) = 0$ if both $M = +\infty, m = -\infty$ are satisfied. However, for the density reconstruction type problems we typically have $m = 0, M = +\infty$, so we cannot always deduce that the negative singular part $\overline{u}^-$ of $\overline{w}$ vanishes. This may be the case under some extra conditions, for instance if we know that the optimal solution $\bar{x}$ does not hit the lower boundary (that is $\bar{x}(t) > m$ for all t). Here the following result whose proof may be found in Appendix II, provides some help:

PROPOSITION 4.2. *Suppose the primal program (P) admits an optimal solution $\bar{x}$. Then the singular measure $d\overline{u}^-$ is supported on $\{t \in T: \bar{x}(t) = m\}$.*

The final step in our duality theory will give a method for reconstructing the primal optimal solution $\bar{x}$ from the Lagrange multipliers $(\overline{w}, \overline{\lambda})$.

THEOREM 4.3. *Suppose the constraint qualification (CQ_2) is satisfied, and let $(\overline{w}, \overline{\lambda})$ be a dual optimal solution guaranteed by (CQ_2). Suppose (P) admits an optimal solution $\bar{x}$. Then*

$$(4.3) \qquad (\bar{x}(t), \bar{x}'(t)) \in \partial\phi^*\left(\overline{w}_a'(t) - \sum_{i=1}^{N} \overline{\lambda}_i a_i(t), \overline{w}(t)\right) \quad \text{for almost all } t.$$

PROOF. (a) Let us first consider the following modified Lagrangian function (compare with (3.3)):

$$(4.4) \qquad \tilde{L}_2(x, y; w, \lambda) = J_\phi(x, y) - \langle v' - A^t\lambda, x \rangle - \langle w, y \rangle - \langle \lambda, b \rangle - Mu^+(1) + mu^-(1) + Pw(1) - Qw(0),$$

where $x \in \mathscr{A}(T)$, $y \in \mathscr{L}_1(T)$, $w \in \mathfrak{BV}(T), w = v + u$, $v \in \mathscr{A}(T), u' \equiv 0$ a.e., $v(0) = w(0)$, and where $u = u^+ - u^-, u^+(0) = u^-(0) = 0$. Here P, Q have the same meaning as in Theorem 3.3. From the definition of $\tilde{L}_2$ we obtain

$$\tilde{L}_2(x, y; w, \lambda) \leq L_2(x, y; w, \lambda)$$

for all $x \in \mathscr{A}(T)$, $y \in \mathscr{L}_1(T)$, $w \in \mathfrak{B}\mathscr{V}(T)$, $\lambda \in \mathbb{R}^N$. However, by the argument given in Theorem 3.3, the values

$$\inf_{x \in \mathscr{A}(T),\, y \in \mathscr{L}_1(T)} \tilde{L}_2(x, y; w, \lambda) = \inf_{x \in \mathscr{A}(T),\, y \in \mathscr{L}_1(T)} L_2(x, y; w, \lambda)$$

coincide, and hence the primal and dual programs

$$(\tilde{P}) \qquad \text{minimize} \quad \sup_{w,\lambda} \tilde{L}_2(x, y; w, \lambda) \qquad \text{subject to} \quad x \in \mathscr{A}(T),\, y \in \mathscr{L}_1(T)$$

and

$$(\tilde{P}_2^*) \qquad \text{maximize} \quad \inf_{x,y} \tilde{L}_2(x, y; w, \lambda) \qquad \text{subject to} \quad w \in \mathfrak{B}\mathscr{V}(T),\, \lambda \in \mathbb{R}^N$$

have the same value as the programs (P) resp. (P_2^*). Also $\bar{x}$ resp. $(\bar{w}, \bar{\lambda})$ are again optimal solutions for $(\tilde{P})$ resp. $(\tilde{P}_2^*)$.

(b) We next observe that $\tilde{L}_2(x, y; w, \lambda)$ may be considered as a function of $x \in \mathscr{L}_\infty(T)$, $y \in \mathscr{L}_1(T)$, and we use the notation $\hat{L}_2(x, y; w, \lambda)$ for this extension We now claim that the values

$$\text{(4.5)} \qquad \inf_{x \in \mathscr{L}_\infty(T),\, y \in \mathscr{L}_1(T)} \hat{L}_2(x, y; w, \lambda) = \inf_{x \in \mathscr{A}(T),\, y \in \mathscr{L}_1(T)} \tilde{L}_2(x, y; w, \lambda)$$

again coincide. This is a consequence of the fact that given any $x \in \mathscr{L}_\infty(T)$, $y \in \mathscr{L}_1(T)$ having $(x, y) \in \operatorname{dom} J_\phi$, there exist $x_n \in \mathscr{A}(T)$, $y_n \in \mathscr{L}_1(T)$ such that $x_n \to x$, $y_n \to y$ both in $\mathscr{L}_1$ norm and with $J_\phi(x_n, y_n) \to J_\phi(x, y)$. The latter is proved by an argument in the spirit of, but slightly more elaborate than, the one given in Lemma 3.4.

(c) Consequently, the values of the primal and dual programs $(\hat{P}), (\hat{P}_2^*)$ arising from the Lagrangian $\hat{L}_2(x, y; w, \lambda)$ have the same values as (P) resp. (P_2^*), so primal resp. dual optimal solutions for $(P), (P_2^*)$ are again optimal for $(\hat{P}), (\hat{P}_2^*)$. Therefore, our optimal $\bar{x}$ and $(\bar{w}, \bar{\lambda})$ give rise to a saddle point of $\hat{L}_2(x, y; w, \lambda)$.

Now the conditions for a saddle point $((\bar{x}, \bar{x}'), (\bar{w}, \bar{\lambda}))$ of $\hat{L}_2$ imply

$$(0,0) \in \partial_{x,y} \hat{L}_2(\bar{x}, \bar{x}'; \bar{w}, \bar{\lambda})$$

or what is the same

$$\text{(4.6)} \qquad (\bar{x}, \bar{x}') \in \partial_{r,s} \hat{L}_2^*(0, 0; \bar{w}, \bar{\lambda}),$$

where the conjugate $\hat{L}_2^*(r, s; \bar{w}, \bar{\lambda})$ is now calculated with respect to the incomplete dual pairing $\langle \mathscr{L}_\infty(T) \times \mathscr{L}_1(T), \mathscr{L}_1(T) \times \mathscr{L}_\infty(T) \rangle$. Indeed, the following may be derived from Rockafellar's results (Rockafellar (1968)). (A direct argument similar to but easier than the one in Lemma 3.4 can be given.)

$$\hat{L}_2^*(r, s; w, \lambda) = J_{\phi^*}(r + v' - A^t\lambda, s + w) + \langle \lambda, b \rangle + Mu^+(1)$$
$$- mu^-(1) - Pw(1) + Qw(0), \qquad (u = w_s, v = w_a),$$

so (4.6) is equivalent to

$$\text{(4.7)} \qquad \langle \bar{x}, r \rangle + \langle \bar{x}', s \rangle \le J_{\phi^*}\big(r + \bar{w}_a' - A^t\bar{\lambda}, s + \bar{w}\big) - J_{\phi^*}\big(\bar{w}_a' - A^T\bar{\lambda}, \bar{w}\big)$$

for all $r \in \mathscr{L}_1(T)$, $s \in \mathscr{L}_x(T)$ having $(r + \overline{w}_a' - A^t\overline{\lambda}, s + \overline{w}) \in \operatorname{dom} J_{\phi^*}$. This proves the statement of the theorem. □

As an implicit consequence of the proof we obtain the following

COROLLARY 4.4. *$(\bar{x}, \bar{x}'; \overline{w}, \overline{\lambda})$ is a saddle point of the Lagrangian $\hat{L}_2(x, y; w, \lambda)$ if and only if $\bar{x}$ is optimal for (P) and $(\overline{w}, \overline{\lambda})$ is optimal for (P_2^*).*

EXAMPLE 4.1. Let us now exhibit the consequences of our main results for the Fisher information, or more generally, for the class of Csiszar distances (2.1).

Let $\phi(x, v) = x\psi(v/x)$ be an integrand of this type, and assume the conditions of Theorem 2.1 resp. Corollary 4.8 below are met, so that program (P) admits an optimal solution $\bar{x}$. Suppose the dual program (P_2^*) has an optimal solution $(\overline{w}, \overline{\lambda})$, guaranteed by the constraint qualification (CQ_F). Furthermore, assume that $\bar{x} > 0$ as a consequence of Theorem 2.2 resp. its analogue for the class of objectives (2.1). By Propositions 4.1 and 4.2, and since $m = 0, M = +\infty$, the singular part of the dual optimal $\overline{w}$ vanishes. Now a direct calculation gives

$$\phi^*(r, s) = \begin{cases} 0 & \text{if } r + \psi^*(s) \le 0, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.8}$$

so (4.7) becomes

$$\langle \bar{x}, r \rangle + \langle \bar{x}', s \rangle \le 0 \quad \text{for } r + \overline{w}' - A^t\overline{\lambda} + \psi^*(s + \overline{w}) \le 0 \text{ a.e.} \tag{4.9}$$

We claim that

$$\overline{w}' - A^t\overline{\lambda} + \psi^*(\overline{w}) = 0 \quad \text{a.e.}$$

Suppose not. Then $\overline{w}' - A^t\overline{\lambda} + \psi^*(w) \le -\xi$ on a set Ω of positive measure and for some $\xi > 0$. Define $s \equiv 0, r = \xi \cdot \chi_\Omega$, then the pair (r, s) is admitted in (4.9), so $\int_\Omega \bar{x} \cdot \xi \le 0$, which is impossible since $\bar{x} > 0$ on T.

Consequently, we may recast the statement in (4.9) as follows:

$$\langle \bar{x}, r \rangle + \langle \bar{x}', s \rangle \le 0 \quad \text{whenever } r + \psi^*(s + \overline{w}) - \psi^*(\overline{w}) \le 0.$$

Setting $r = \psi^*(\overline{w}) - \psi^*(s + \overline{w})$, we get

$$\langle \bar{x}, \psi^*(\overline{w}) - \psi^*(s + \overline{w}) \rangle + \langle \bar{x}', s \rangle \le 0$$

for all $s \in \mathscr{L}_x(T)$, which can only hold true if

$$-\bar{x}(t) \cdot \psi^{*\prime}(\overline{w}(t)) + \bar{x}'(t) = 0$$

for almost all $t \in T$. Therefore, we get the relation

$$\frac{\bar{x}'(t)}{\bar{x}(t)} = \psi^{*\prime}(\overline{w}(t))$$

or rather

$$\bar{x}(t) = C \cdot \exp\left\{ \int_0^t \psi^{*\prime}(\overline{w}(s))\, ds \right\}, \tag{4.10}$$

with the constant $C > 0$ determined by the constraints $A\bar{x} = b$. In particular, for the case of the Fisher information, we get the following result, which we state explicitly due to its relevance for our subsequent paper (Borwein, Lewis, Limber and Noll (1995)):

PROPOSITION 4.5. *Suppose the* $\{a_i\}$ *are pseudo-Haar with* $1 \in \text{lin}\{a_1, \dots, a_N\}$. *Let* (CQ_F) *be satisfied. Then for the Fisher information* I_F, *the dual program* (P_2^*) *may be stated in the form*

$$(P_2^*) \qquad \text{minimize} \quad \sum_{i=1}^{N} \lambda_i b_i$$

$$\text{subject to} \quad w' + \frac{1}{4}w^2 = \sum_{i=1}^{N} \lambda_i a_i,$$

$$w(0) = w(1) = 0, \qquad w \in \mathscr{A}(T).$$

The primal optimal $\bar{x}$ *may be recovered from the solution* $(\bar{w}, \bar{\lambda})$ *of* (P_2^*) *by means of the identity*

$$\bar{x}(t) = C \cdot \exp\left\{\frac{1}{2}\int_0^t \bar{w}(s)\, ds\right\}, \tag{4.11}$$

where $C > 0$ *is determined by the constraint* $A\bar{x} = b$. □

REMARK. As a consequence of (4.10), we derive that for analytic data a_i, $\psi^{*\prime}$, the solution $\bar{x}$ is of the same type. In particular, in the Fisher case, and for algebraic or trigonometric moments, $\bar{x}$ will be an entire function. Even for continuous a_i, $\bar{x}$ will at least be of class $\mathscr{C}^2$.

Let us end this section with a pleasant application of our duality theory. Suppose we do not know whether the primal program (P) admits an optimal solution. Starting with the dual program (P_2^*), it seems natural to consider a bidual program (P^{**}), which under some constraint qualification on (P_2^*) will have an optimal solution. It turns out that the solution of (P^{**}) may be viewed as a *generalized solution* for (P), and even more, in some cases, it is in fact a *solution* of (P). Let us notice that this bidual relaxation scheme fits a general pattern which has been used by various authors in different contexts; see, e.g., Rockafellar (1971, 1981), Ekeland and Teman (1989).

Let us consider the restricted dual program

$$(Q) \qquad \text{maximize} \quad -J_{\phi^*}(v' - A^t\lambda, v) - \langle \lambda, b\rangle$$

$$\text{subject to} \quad v \in \mathscr{A}(T), \qquad v(0) = v(1) = 0, \lambda \in \mathbb{R}^N.$$

We expect the values of (P_2^*) and (Q) to be identical. Instead of proving this directly, we consider the duality associated with (Q). This requires a constraint qualification for (Q):

$$(CQ)^* \qquad 0 \in \text{core}\left\{u - v' + A^t\lambda : (u, v) \in \text{dom}\, J_{\phi^*}, \lambda \in \mathbb{R}^N\right\}$$

where the core means the algebraic interior of a set, and the underlying space is $\mathscr{L}_1(T)$, and where dom J_{ϕ^*} is a subset of $\mathscr{L}_1(T) \times \mathscr{A}(T)$. The Lagrangian associated with program (Q) is now

$$L(u, v, \lambda; x) = J_{\phi^*}(u, v) + \langle \lambda, b \rangle - \langle u - v' + A^t\lambda, x \rangle, \tag{4.12}$$

defined for $u \in \mathscr{L}_1(T)$, $v \in \mathscr{A}(T)$, $x \in \mathscr{L}_1(T)^* = \mathscr{L}_\infty(T)$, $\lambda \in \mathbb{R}^N$. Then (Q) and (Q^*) may as usual be stated in terms of L. Our duality theory enables us to prove the following

PROPOSITION 4.6. *Suppose $(CQ)^*$ is satisfied. Then the dual program (Q^*) admits an optimal solution $\bar{x} \in \mathfrak{B}\mathscr{V}(T)$. The values of (Q) and (Q^*) coincide, and (Q^*) may be represented in the form*

$$(Q^*) \qquad \textit{minimize} \quad J_\phi(x, y') + M^* \int_0^1 dz^+ - m^* \int_0^1 dz^-$$

$$\textit{subject to} \quad Ax = b,$$

where $x = y + z$ is the Lebesgue decomposition of x, that is $y = x_a$, $z = x_s$ (Definition 3.1), *and with*

$$M^* = \sup\{\eta : \exists \xi \ \phi^*(\xi, \eta) < +\infty\}, \qquad m^* = \inf\{\eta : \exists \xi \ \phi^*(\xi, \eta) < +\infty\}.$$

PROOF. The existence of $\bar{x}$ follows from $(CQ)^*$ and Proposition 3.1, while $\bar{x} \in \mathfrak{B}\mathscr{V}(T)$ is a consequence of Proposition 3.2. Theorem 3.3 gives the above form of the dual program (Q^*). □

We are now ready to extend Theorem 2.1 to a more general class of objectives. The reader might compare this with the central existence result in Rockafellar (1971); see also Rockafellar (1972).

THEOREM 4.7. *Suppose J_{ϕ^*} satisfies $(CQ)^*$, and there exists $(\xi_0, \eta_0) \in \operatorname{dom} \phi$ such that $t \to \phi(\xi_0, \eta_0 + t)$ is coercive. Then program (P) admits an optimal solution.*

PROOF. As a consequence of $(CQ)^*$, we know that the dual program (Q^*) admits an optimal solution $\bar{x} \in \mathfrak{B}\mathscr{V}(T)$. If we can show $m^* = -\infty$ and $M^* = +\infty$, the singular part $\bar{x}_s$ of $\bar{x}$ will vanish as a consequence of Proposition 4.1, and then $\bar{x}$ will be an optimal solution of the original program (P).

In order to prove $m^* = -\infty$, $M^* = +\infty$, observe that $\sigma_{\operatorname{dom} \phi^*} = 0^+\phi$, where $\sigma_{\operatorname{dom} \phi^*}$ denotes the support function of $\operatorname{dom} \phi^*$, and where $0^+\phi$ denotes the recession function of ϕ. But now the coercivity of $t \to \phi(\xi_0, \eta_0 + t)$ implies

$$0^+\phi(0, 1) = 0^+\phi(0, -1) = +\infty,$$

hence $\sigma_{\operatorname{dom} \phi^*}(0, \pm 1) = +\infty$. This implies $M^* = +\infty$, $m^* = -\infty$, as desired. □

REMARK. We may split the statement. If $0^+\phi(0, 1) = +\infty$, then $M^* = +\infty$, while $0^+\phi(0, -1) = +\infty$ implies $m^* = -\infty$. The converse is also true in either case.

COROLLARY 4.8. *Let $\phi(x, v) = x\psi(v/x)$ be a Csiszar distance* (2.1). *Suppose ψ is coercive, and let $1 \in \operatorname{lin}\{a_1, \ldots, a_N\}$. Then the corresponding program (P) admits a unique optimal solution.*

PROOF. The coercivity assumption in Theorem 4.7 is satisfied. Hence we have to check that $(CQ)^*$ is satisfied.

Due to the special form (4.8) of the conjugate, it suffices to fix $u \equiv 1$ and $A^t\lambda = -1$. Then every $v \in \mathscr{A}(T)$ having $v \le -\psi^*(1)$ will give rise to a pair $(u, v) \in \operatorname{dom} J_{\phi^*}$. Now any $v' \in \mathscr{L}_1(T)$ having $\|v'\|_1 \le 1$ will provide such a v by setting

$v(t) = -1 - \psi^*(1) + \int_0^t v'(s)\,ds$. But then $\{u - v' + A^t\lambda\colon (u,v) \in \operatorname{dom} J_{\phi^*}, \lambda \in \mathbb{R}^N\}$ contains the $\mathscr{L}_1$ unit ball, which proves the statement. □

REMARK. Assuming conditions (1) and (2) for the Csiszar distance $\phi(x,v) = x\psi(v/x)$ as well as coercivity of ψ, we may now state the dual program (P_2^*) as follows:

$$(P_2^*) \qquad \text{maximize} \quad -\sum_{i=1}^{N} \lambda_i b_i$$

$$\text{subject to} \quad \psi^*(w) + w' = \sum_{i=1}^{N} \lambda_i a_i,$$

$$w(0) = w(1) = 0, \qquad w \in \mathscr{A}(T).$$

The formula for recovering the primal optimal solution $\bar{x}$ for (P) from the dual optimal $\bar{w}$ is (4.10). □

Conclusion. We have presented a maximum entropy type model for the spectral density estimation problem (1.1) which is based on entropy functions that control derivative values. A fairly general duality theory for this model was obtained in §§3 and 4. The general form of the dual program was derived in Theorem 3.3. More special but typical cases were presented in §4. A numerically tractable dual program formulation was obtained (Proposition 4.5 and the Remark following Corollary 4.8). A general existence result is Theorem 4.7, which was obtained by duality techniques. Special emphasis was given to the Fisher information model (§2, Example 3.1, Example 4.1, Proposition 4.5).

A presentation of explicit numerical results for the Fisher information model (1.1), (1.4) is given in the second part of our paper (Borwein, Lewis, Limber and Noll 1995).

5. Examples. In this section we present some examples which among others indicate that the constraint model (1.1) we used throughout the §§ 2 to 4 could be replaced by more general models—essentially without affecting the arguments presented. We indicate the straightforward changes as we go. We start, however, with an example that fits our model (1.1) verbally, and which is included in particular to show that the singular part of the dual variable may vanish for reasons which are different in nature from the ones which applied for the Fisher information.

Example 5.1. *Moment matching with minimum slope.* This example is included in particular to show that the singular part of the dual optimal variable may vanish even when the primal optimal solution hits the boundary.

Consider the program

$$(P) \qquad \text{minimize} \quad \frac{1}{2}\|x'\|_2^2$$

$$\text{subject to} \quad x \geq 0, \qquad \int_0^1 a_i(t)x(t)\,dt = b_i, \qquad i = 1, \ldots, N.$$

This is a problem in accordance with the model (1.1). The integrand (1.4) and its conjugate are given as

$$\phi(x,v) = \begin{cases} \frac{1}{2}v^2 & \text{for } x \geq 0, \\ +\infty & \text{for } x < 0, \end{cases} \qquad \phi^*(y,u) = \begin{cases} \frac{1}{2}u^2, & \text{for } y \leq 0, \\ +\infty & \text{for } y > 0. \end{cases}$$

Observe that (P) has a unique optimal solution $\bar{x} \in \mathscr{A}_2(T)$.

Let us consider the duality based on using the first Lagrangian. The conditions for a saddle point $(\bar{x}, \bar{\lambda})$ imply $A\bar{x} = b$ and

$$0 \leq \frac{\frac{1}{2}\|\bar{x}' + \tau h'\|_2^2 - \frac{1}{2}\|\bar{x}'\|_2^2}{\tau} + \int_0^1 \sum_{i=1}^N \bar{\lambda}_i a_i \cdot h < +\infty$$

for every $h \in \mathscr{A}_2(T)$ having $\bar{x} + \tau h \geq 0$ for small $\tau > 0$. Therefore the Dubois-Reymond Lemma only applies on the set $\{\bar{x} > 0\}$, where it provides the equation

$$\bar{x}''(s) = A^t\bar{\lambda}(s) = \sum_{i=1}^N \bar{\lambda}_i a_i(s).$$

This equation, however, need not hold on the interior of the set $\{\bar{x} = 0\}$. We may nevertheless deduce at this stage that $\bar{x}$ is of class C^1, and that $\bar{x}' \in \mathscr{A}(T)$.

Duality based on using the second Lagrangian provides more information. Notice that $m = 0$ and $M = +\infty$ in this case, so the dual program (P_2^*) is

$$(P)^* \qquad \text{maximize} \quad -\frac{1}{2}\|w\|_2^2 - \langle \lambda, b\rangle$$

$$\text{subject to} \quad w' - A^t\lambda \leq 0,$$

$$w(0) \leq 0, \qquad w(1) \geq 0, \qquad w \in \mathfrak{B}\mathscr{V}(T), \qquad \lambda \in \mathbb{R}^N.$$

We know that the differential inequality is in fact an equality on the set $\{\bar{x} > 0\}$, but might be a strict inequality in the interior of $\{\bar{x} = 0\}$. So we may not deduce $w \in \mathscr{A}(T)$ directly. However, Theorem 4.3 helps. Indeed, we have

$$\partial\phi^*(r,s) = \begin{cases} \{(0,s)\} & \text{for } r < 0, \\ \mathbb{R}_+ \times \{s\} & \text{for } r = 0, \\ \varnothing & \text{for } r > 0. \end{cases}$$

Then the second coordinate in formula (5.5) implies $\bar{x}'(t) = \bar{w}(t)$ for almost all t. Hence $\bar{w}$ must be absolutely continuous from what we have seen before.

We may not replace the differential inequality in the dual program by an equality, for this would mean solving the unrestricted program without the side condition $x \geq 0$. Therefore, whenever the optimal solution of the unrestricted program fails to be feasible for (P), the inequality will certainly be strict on some interval.

Example 5.2. *Interpolation on a strip.* We exhibit an example involving higher-order derivatives and slightly different constraints. The pattern for a duality as expounded in §§ 3 and 4 remains essentially the same.

Consider the problem of interpolation on a strip, which was discussed in Dontchev (1987) (see also Dontchev and Kalchev (1989)). For a partition $0 = t_1 < t_2 < \cdots < t_n = 1$ of $[0,1]$ we pose the interpolation problem

$$(IP) \qquad \text{minimize} \quad \|x''\|_2$$
$$\text{subject to} \quad x(t_i) = y_i, \qquad i = 1, \dots, n,$$
$$\alpha(t) \le x(t) \le \beta(t) \quad \text{for all } t \in [0,1].$$

We assume that the problem is feasible, that is $\alpha(t_i) < y_i < \beta(t_i)$, and that $\alpha(\cdot)$, $\beta(\cdot)$ are continuous and piecewise C^2 functions satisfying $\alpha(t) < \beta(t)$ for every t. It is well known that in the absence of the constraint $\alpha \le x \le \beta$, the solution is a cubic spline interpolating the data (t_i, y_i). We will show that, under reasonable conditions on α, β, the solution of (IP) is again a cubic spline, with a finite number of extra knots. In the case where α, β are piecewise linear on $[t_i, t_{i+1}]$, this has been demonstrated in Dontchev (1987).

It is well known that the problem may be reformulated as

$$(P) \qquad \text{minimize} \quad \frac{1}{2}\|x''\|_2^2$$
$$\text{subject to} \quad \int_0^1 a_i(s) x''(s)\, ds = b_i, \qquad i = 1, \dots, n-2, x(0) = y_0,$$
$$\alpha(t) \le x(t) \le \beta(t), \qquad t \in [0,1].$$

Here the a_i denote the second order B-splines, and the b_i are the ith divided differences of order 2 associated with the data (t_i, y_i); see Dontchev and Kalchev (1989), Dontchev (1987) or Borwein and Lewis (1991).

Let

$$(5.13) \qquad \phi(t,x,y,z) = \begin{cases} \frac{1}{2}z^2 & \text{if } \alpha(t) \le x \le \beta(t), \\ +\infty & \text{otherwise,} \end{cases}$$

$$\phi^*(t,u,v,w) = \begin{cases} \beta(t)u + \frac{1}{2}w^2 & \text{if } v = 0, u \ge 0, \\ \alpha(t)u + \frac{1}{2}w^2 & \text{if } v = 0, u \le 0, \\ +\infty & \text{if } v \ne 0. \end{cases}$$

Writing problem (P) in the form

$$\text{minimize} \quad J_\phi(x, x', x'') = \int_0^1 \phi(t, x(t), x'(t), x''(t))\, dt$$
$$\text{subject to} \quad Ax'' = b, \qquad x(0) = y_0,$$

we see that the correct choice for a Lagrangian of the second type is

$$L(x,y,z;\theta,w,\lambda,\rho) = J_\phi(x,y,z) + \langle \theta, x' - y\rangle + \langle w, y' - z\rangle + \langle \lambda, Az - b\rangle + \rho(x(0) - y_0),$$

with $x \in \mathscr{A}_{2,2}(T) = \{x \in \mathscr{L}_1 : x'' \in \mathscr{L}_2\}$, $y \in \mathscr{A}_2(T)$, $z \in \mathscr{L}_2$, $\theta \in \mathscr{A}_2(T)^*$, $w \in \mathscr{L}_2$. First we represent $\langle \theta, z \rangle = \langle v, z' \rangle + r \cdot z(0)$ for some $v \in \mathscr{L}_2$, $r \in \mathbb{R}$, hence $\langle \theta, x' - y \rangle = \langle v, x'' - y' \rangle + r(x'(0) - y(0))$. Then we establish duality in the spirit of Proposition 3.1. This requires a constraint qualification which in this case is satisfied as a consequence of the feasibility of (IP) resp. (P). We obtain a dual optimal $(\theta, w, \lambda, \rho)$. Now playing with the variable y, which is allowed to take on every value, we first show that $w \in \mathfrak{B}\mathscr{V}$. Playing with x', say, shows $v \in \mathfrak{B}V$, so we may integrate by parts. Using again the fact that the variable y is free, we can show successively that

$$v(1) = 0, \quad r - v(0) = 0, \quad w(0) = w(1) = 0 \quad \text{and} \quad dv = dw.$$

This implies $v = w$, and hence $r = 0$. Therefore the Lagrangian takes on the simplified form

$$L(x, y, z; v, \lambda, \rho) = J_\phi(x, y, z) - \langle dv, x' \rangle - \langle v, z \rangle + \langle \lambda, Az - b \rangle + \rho\big(x(0) - y_0\big).$$

Now let us fix $\hat{x}$ such that $\alpha(t) < \hat{x}(t) < \beta(t)$ and $\hat{x}(t_\iota) = y_\iota$. We may play with x near $\hat{x}$, producing arbitrarily high derivatives x on small measure sets. This allows us to prove that the singular part of the measure dv vanishes, proving $v \in \mathscr{A}(T)$. Notice here that we have subsumed the boundary values coming from the integration by parts into the singular measures, so we have the representation $\langle dv, x' \rangle = \langle v', x' \rangle$, and $v(0) = v(1) = 0$. Playing again with x', Proposition 3.2 shows $v' \in \mathfrak{B}\mathscr{V}(T)$, so we may again integrate by parts, which gives us

$$L(x, y, z; v, \lambda, \rho) = J_\phi(x, y, z) + \langle x, dv' \rangle - \langle v - A^t\lambda, z \rangle - \langle \lambda, b \rangle.$$

Here we used $\rho + v'(0) = 0$ and $v'(1) = 0$, which follows since problem (P) has an optimal solution.

We decompose according to Lebesgue: $v' = v_a' + v_s'$. By a result analogous to Theorem 3.3, the dual objective function is

$$\inf_{x, y, z} L = -J_{\phi^*}\big(-(v_a')', 0, v - A^t\lambda\big) - \langle \lambda, b \rangle + \int_0^1 \alpha(t)\, dv_s'^{+}(t) - \int_0^1 \beta(t)\, dv_s'^{-}(t).$$

Our next step is to evaluate the conditions for a saddle point of the Lagrangian. Similarly to the procedure as presented in §4, this leads to the inequality

$$\langle x, r \rangle + \langle x'', s \rangle \le J_{\phi^*}\big(r - (v_a')', 0, s + v - A^t\lambda\big) - J_{\phi^*}\big(-(v_a')', 0, v - A^t\lambda\big)$$

for all $r, s \in \mathscr{L}_\infty$. Letting $r = 0$ and s arbitrary, we first get

$$x''(t) = v(t) - \sum_{\iota=1}^{n-2} \lambda_\iota a_\iota(t) \quad \text{for a.a. } t. \tag{5.14}$$

Since the functions a_ι are piecewise linear and $v \in \mathscr{A}(T)$, we derive that x is of class C^2 and is three times differentiable. Secondly, suppose $(v_a')' < 0$ on a set M of positive measure. Setting $s = 0$ implies

$$\langle x, r \rangle \le \langle \beta, r \rangle$$

for all r having $r - (v_a')' \geq 0$, which implies $x = \beta$ on M. Similarly, $(v_a')' > 0$ on a set N of positive measure implies $x = \alpha$ on N. In particular, v_a' is constant whenever x stays strictly inside the strip.

An analogue of Proposition 4.2 tells us that the singular measure $dv_s'^{+}$ is concentrated on the set $\{t\colon\ x(t) = \alpha(t)\}$, while $dv_s'^{-}$ is concentrated on $\{t\colon\ x(t) = \beta(t)\}$. Therefore v is affine on any part of an interval $[t_i, t_{i+1}]$ on which x stays strictly inside the strip. Since the a_i are piecewise linear, so is x'' by (5.14), which means that x is a cubic spline as long as it stays strictly inside the strip. On the other hand, for general α and β it is possible that the solution x goes along the boundary for some time.

Let us consider the case where α and β are piecewise cubic on the $[t_i, t_{i+1}]$. Then we deduce that x is a cubic spline with a finite number of extra knots of the form $(t, \alpha(t))$ resp. $(t, \beta(t))$ in addition to the knots (t_i, y_i). In the case where α, β are piecewise linear, x may touch the upper resp. lower boundaries each at most once in a given interval $[t_i, t_{i+1}]$, so here the maximum number of extra knots is 2 per interval $[t_i, t_{i+1}]$. For piecewise quadratic or cubic α, β, the solution may follow the boundary on certain subintervals of the partition.

Example 5.3. *Fisher moment matching with tolerance.* It often happens in practical problems that the data b_i are noisy. One way of dealing with this phenomenon is to allow for tolerances in the moment matching model (1.1). In the case of the Fisher information, the relaxed model may be stated as

$$(P_\epsilon) \qquad \text{minimize} \quad I_F(x) = \int_0^1 \frac{x'^2(t)}{x(t)}\,dt$$

$$\text{subject to} \quad x \geq 0, \qquad \|Ax - b\| \leq \epsilon.$$

Here $\|\cdot\|$ is any fixed norm on $\mathbb{R}^N$. Introducing a new variable $e = (e_i)_{i=1}^N \in \mathbb{R}^N$, and separating x and x' we may recast the program as follows:

$$\text{minimize} \quad I_F(x) \qquad \text{subject to} \quad Ax = b + e \text{ and } \epsilon \geq \|e\|.$$

This leads us to consider the following Lagrangian of the second kind:

$$L(x, y, e; w, \lambda, \rho) = J_F(x, y) + \langle w, x' - y\rangle + \langle \lambda, Ax - b - e\rangle + \rho(\epsilon - \|e\|),$$

for $\|e\| \leq \epsilon$ and $L = +\infty$ otherwise. The duality as presented in §§ 3 and 4 will lead to the dual program

$$(P_\epsilon)^* \qquad \text{maximize} \quad -J_{\phi^*}(w' - A^t\lambda, w) - \langle \lambda, b\rangle - \epsilon\|\lambda\|_*$$

$$\text{subject to} \quad w \in \mathscr{A}(T), \quad \lambda \in \mathbb{R}^N.$$

Here $\|\cdot\|_*$ denotes the norm dual to $\|\cdot\|$. By the representation of the Fisher conjugate, this gives the dual program

$$(P_\epsilon)^* \qquad \text{maximize} \quad -\sum_{i=1}^{N} \lambda_i b_i - \epsilon \|\lambda\|_*$$

$$\text{subject to} \quad 4w' + w^2 = 4\sum_{i=1}^{N} \lambda_i a_i$$

$$w(0) = w(1) = 0.$$

It follows that $\langle \lambda, e \rangle = \epsilon \|\lambda\|_*$ with $\|e\| = \epsilon$, so the tolerance is fully used. For example, if $\|\cdot\|$ is the Euclidean norm, the solution $\bar{x}_\epsilon$ of the perturbed program (P_ϵ) coincides with the solution of the original moment matching problem (1.1) with the moments b_i replaced by either $b_i + \epsilon_i$ or $b_i - \epsilon_i$, where $\Sigma \epsilon_i^2 = \epsilon^2$. Numerical experiments using the Boltzmann-Shannon entropy suggest that in the case where $\|\cdot\|$ is the supremum norm, the perturbed program (P_ϵ) will be the original program with b_i replaced by either $b_i + \epsilon$ or $b_i - \epsilon$.

Appendix I. Let us briefly recall the origin of the Fisher information measure and its relation to what we call the averaged Fisher information.

Fisher information was introduced by R. A. Fisher (1930) in the context of maximum likelihood estimation. Let $f(x; p)$, $p \in T \subset \mathbb{R}$ be a parametrized family of probability densities on $\mathbb{R}^r$. For an independent sample $x_1, \dots, x_n$ the maximum likelihood estimate for the true parameter p_* is defined as the parameter value $p_n = p_n(x_1, \dots, x_n)$ where the log likelihood function

$$L(x_1, \dots, x_n; p) = \log \prod_{i=1}^{n} f(x_i; p) = \sum_{i=1}^{n} \log f(x_i; p),$$

attains its maximum (if any). It was known to R. A. Fisher (1930) and proved rigorously by J. L. Doob (1934) that, under reasonable conditions, and for large n, $\sqrt{n}(p_n - p_*)$ is asymptotically normally distributed with mean 0 and variance σ^2, where

$$\frac{1}{\sigma^2} = -\int \cdots \int f(x; p_*) \frac{\partial^2 \log f(x; p_*)}{\partial p^2} dx_1 \dots dx_r$$

$$= \int \cdots \int \left(\frac{\partial \log f(x; p_*)}{\partial p} \right)^2 f(x; p_*)\, dx_1 \dots dx_r.$$

The term $1/\sigma^2$ is known as the Fisher information of $f(\cdot; p_*)$, and it measures the expected negative curvature of the log likelihood function with regard to the distribution $f(x; p_*)\,dx$. In particular, the higher the negative curvature of the log likelihood function in a neighborhood of the true value p_*, the more accurate the maximum likelihood estimate p_n.

In order to measure the information of the parametrized family $f(\cdot, p)$, $p \in T = [t_0, t_1] \subset \mathbb{R}$, we assume that a priori all $p \in T$ are equally likely to occur. It is then convenient to consider the averaged information

$$(5.15) \qquad -\int_a^b f(x,p) \frac{\partial^2 \log f(x,p)}{\partial p^2} dp = f'(x,a) - f'(x,b) + \int_a^b \left(\frac{\partial f(x;p)}{\partial p} \right)^2 \frac{1}{f(x;p)} dp,$$

and the second term on the right-hand side of this expression, that is

$$(5.16) \qquad I_F(f) = \int_a^b \frac{f'^2(p)}{f(p)} dp,$$

is what we call the *averaged Fisher information* of $f(x, p)$ considered as a function of the parameter p.

It is intuitively clear from (5.15) that a sharp spike of the function f at some p, giving rise to a fairly negative curvature of $\log f$, will make a sizable contribution to the information $I_F(f)$. It is therefore heuristically clear that minimizing the averaged Fisher information will have the effect of "smoothing" the data as proposed in Silver (1992).

Appendix II

PROOF OF LEMMA 3.4. Let $v \in \mathscr{L}_1(T)$, $w \in \mathscr{L}_\infty(T)$. We have to show that

$$J_{\phi^*}(v,w) = \sup_{x \in \mathscr{A}(T),\, y \in \mathscr{L}_1(T)} \langle v, x\rangle + \langle w, y\rangle - J_\phi(x,y). \tag{5.17}$$

Assume first that the right-hand side is finite. We show that $J_{\phi^*}(v,w) < +\infty$. Our first step is to show that $t \to \phi^*(v(t), w(t))$ must be finite a.e.

Before starting, let us recall the notion of Lebesgue points. For a fixed realization f of an element of $\mathscr{L}_1(T)$, the set of Lebesgue points of f consists of those $t \in T$ for which

$$\lim_{\delta \to 0} \frac{1}{2\delta} \int_{t-\delta}^{t+\delta} f(s)\, ds = f(t).$$

It is known that the set of Lebesgue points has full measure (see Rudin (1973) or Hewitt and Stromberg (1969)).

Suppose now there exists a set Ω of positive measure such that $\phi^*(v(t), w(t)) = +\infty$ for $t \in \Omega$. Since the set E of common Lebesgue points of v and w has full measure, we may assume that Ω consists of such points. Now for $t \in \Omega$ and $n \in \mathbf{N}$, $\epsilon > 0$ fixed, select $(\xi_t, \eta_t) \in \operatorname{dom} \phi$ such that

$$\xi_t v(t) + \eta_t w(t) - \phi(\xi_t, \eta_t) > n. \tag{5.18}$$

Find $\delta(t) > 0$ such that

$$\frac{1}{2\delta} \int_{t-\delta}^{t+\delta} (\xi_t v(s) + \eta_t w(s))\, ds = \xi_t v(t) + \eta_t w(t) + o(1),$$

with $|o(1)| \le \epsilon$, say, for all $0 < \delta \le \delta(t)$. As the set of intervals $[t-\delta, t+\delta]$, $t \in \Omega$, $0 < \delta \le \delta(t)$ forms a Vitali covering of Ω, we find finitely many disjoint intervals $I_j = [t_j - \delta_j, t_j + \delta_j]$, $j = 1, \dots, r$ covering Ω up to a set of measure $< \epsilon$, where $t_j \in \Omega$, $\delta_j \le \delta(t_j)$.

Now let $\rho > 0$ be such that the intervals $I_{j\rho} = [t_j - \delta_j - \rho, t_j + \delta_j + \rho]$ are still disjoint. Define piecewise linear functions $x_{\epsilon\rho}$, $y_{\epsilon\rho}$ by

$$x_{\epsilon\rho} \equiv \xi_{t_j}, \qquad y_{\epsilon\rho} \equiv \eta_{t_j} \quad \text{on } I_j, \tag{5.19}$$

$$x_{\epsilon\rho} \equiv \xi, \qquad y_{\epsilon\rho} \equiv \eta \quad \text{outside } \bigcup_{j=1}^{r} I_{j\rho},$$

where $(\xi, \eta) \in \operatorname{dom} \phi$ is a fixed point. By convexity, $(x_{\epsilon\rho}(t), y_{\epsilon\rho}(t)) \in \operatorname{dom} \phi$ for all t, and moreover,

$$\begin{aligned} |\phi(x_{\epsilon\rho}(t), y_{\epsilon\rho}(t))| &= |\phi\big(\lambda(t)(\xi, \eta) + (1-\lambda(t))(\xi_{t_j}, \eta_{t_j})\big)| \\ &\le |\phi(\xi, \eta)| + \max_{j=1,\dots,r} |\phi(\xi_{t_j}, \eta_{t_j})| \end{aligned}$$

shows $\phi(x_{\epsilon\rho}, y_{\epsilon\rho})$ is integrable. Letting $\rho \to 0$, $\epsilon \to 0$, we derive that

$$\begin{aligned} J_\phi(x_{\epsilon\rho}, y_{\epsilon\rho}) - \langle v, x_{\epsilon\rho}\rangle - \langle w, y_{\epsilon\rho}\rangle &= o(1) + \sum_{j=1}^{r} \int_{I_j} \xi_{t_j} v(s) + \eta_{t_j} w(s) - \phi(\xi_{t_j}, \eta_{t_j})\, ds \\ &= o(1) + \sum_{j=1}^{r} 2\delta_j \big(\xi_{t_j} v(t_j) + \eta_{t_j} w(t_j) - \phi(\xi_{t_j}, \eta_{t_j})\big) + o(\delta_j) \\ &\ge o(1) + n \cdot \sum_{j=1}^{r} 2\delta_j = o(1) + n \cdot \operatorname{meas}(\Omega) \to +\infty, \end{aligned}$$

as $n \to +\infty$, contradicting the finiteness of the right-hand side in (5.17). Similar reasoning now shows that $\phi^*(v(\cdot), w(\cdot))$ must in fact be integrable.

Let us next check that the right-hand side in (5.17) equals $J_{\phi^*}(v,w)$. Fix $\epsilon > 0$ and let E be the set of common Lebesgue points of v, w and $\phi^*(v(\cdot), w(\cdot))$. For every $t \in E$ we may select $(\xi_t, \eta_t) \in \operatorname{dom} \phi$ such that

$$\phi^*(v(t), w(t)) \le \xi_t v(t) + \eta_t w(t) - \phi(\xi_t, \eta_t) + \epsilon. \tag{5.20}$$

As before, let $\delta(t) > 0$ be such that

$$(5.21)\qquad \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}(\xi_t v(s) + \eta_t w(s))\,ds = \xi_t v(t) + \eta_t w(t) + o(1),$$

$$\frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\phi^*(v(s), w(s))\,ds = \phi^*(v(t), w(t)) + o(1),$$

with the corresponding $|o(1)| \le \epsilon$ whenever $0 < \delta \le \delta(t)$.

Let $I_1, \dots, I_r$ be disjoint intervals $I_j = [t_j - \delta_j, t_j + \delta_j]$ having $\delta_j \le \delta(t_j)$, $t_i \in E$, and covering $T = [0,1]$ up to a set of measure $< \epsilon$. Let $\rho > 0$, and define the $x_{\epsilon\rho}, y_{\epsilon\rho}$ as above. Then the right-hand side in (5.17) is greater or equal to

$$J_\phi(x_{\epsilon\rho}, y_{\epsilon\rho}) - \langle v, x_{\epsilon\rho}\rangle - \langle w, x_{\epsilon\rho}\rangle = o(1) + \sum_{j=1}^{r}\int_{I_j}\xi_{t_j}v(s) + \eta_{t_j}w(s) - \phi\left(\xi_{t_j}, \eta_{t_j}\right)ds$$

$$= o(1) + \sum_{j=1}^{r} 2\delta_j\left(\xi_{t_j}v(t_j) + \eta_{t_j}w(t_j) - \phi\left(\xi_{t_j}, \eta_{t_j}\right)\right)$$

$$\ge o(1) + \sum_{j=1}^{r} 2\delta_j\phi^*\left(v(t_j), w(t_j)\right) - \epsilon.$$

By the second equation in (5.22), the last term has the same limit as

$$\sum_{j=1}^{r}\int_{I_j}\phi^*(v(s), w(s))\,ds = J_{\phi^*}(v, w) + o(1).$$

This proves that $J_{\phi^*}(v, w)$ is majorized by the right-hand side of (5.18). Since the reverse inequality is obvious, this proves the statement of the lemma. □

PROOF OF PROPOSITION 4.2. Suppose contrary to the statement that there exists a Lebesgue null set Ω such that $\bar{x}(t) > m$ on Ω, and $d\bar{u}^-(\Omega) > 0$. Since the measures $d\bar{u}^+$ and $d\bar{u}^-$ are mutually singular, we may assume that $d\bar{u}^+(\Omega) = 0$. Fix $\epsilon > 0$, and let $I_1, \dots, I_k$ disjoint intervals covering Ω up to a set of $d\bar{u}^-$-measure $< \epsilon$. Now let $J_i = I_i + [-\rho, \rho]$ be larger disjoint intervals such that still $(dx + d\bar{u}^+)(J_1 \cup \cdots \cup J_k) < \epsilon$ (dx = Lebesgue measure). Define a continuous function $\lambda: T \to [0,1]$ such that $\lambda = 1$ on each I_i, and $\lambda = 0$ outside $J_1 \cup \cdots \cup J_k$.

By the definition of m, and for ϵ small enough, there exists an m' having $(m + \epsilon, m') \in \operatorname{dom}\phi$. Now define an arc $x \in \mathscr{A}(T)$ and $y \in \mathscr{L}_1(T)$ by setting

$$x(t) = (1 - \lambda(t))\bar{x}(t) + \lambda(t)(m + \epsilon),$$

$$y(t) = (1 - \lambda(t))\bar{x}'(t) + \lambda(t)m'.$$

Notice that $(x, y) \in \operatorname{dom} I_\phi$, since by the convexity of ϕ we have on each J_i:

$$|\phi(x(t), y(t))| \le |\phi(\bar{x}(t), \bar{x}'(t))| + |\phi(m + \epsilon, m')|,$$

while $\phi(x, y) = \phi(\bar{x}, \bar{x}')$ outside the J_i. This estimate also shows $J_\phi(x, y) \to J_\phi(\bar{x}, \bar{x}')$, as $\epsilon, \rho \to 0$. The construction now implies

$$L_2(x, y; w, \lambda) \to L_2(\bar{x}, \bar{x}'; w, \lambda) - \langle d\bar{u}^-, \bar{x} - m\rangle, \qquad (\epsilon, \rho \to 0^+).$$

But $\bar{x}$ is optimal, so $\langle d\bar{u}^-, \bar{x} - m\rangle \le 0$, and since $x \ge m$ by definition, this implies that $d\bar{u}^- = 0$ on the set $\{t: \bar{x}(t) > m\}$ □

Acknowledgements. Research of the first and second authors was supported by Natural Sciences and Engineering Research Council of Canada (NSERC). Research of the third author was supported by Deutsche Forschungsgemeinschaft (DFG) and NSERC.

References

Borwein, J. M. (1983). Adjoint process duality. *Math. Oper. Res.* **8** 403–434.

———, A. S. Lewis (1991a). Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.* **29** 325–338.

———, ——— (1991b). Convergence of best entropy estimates. *SIAM J. Optim.* **1** 191–205.

———, ——— (1992). Partially finite convex programming I, II. *Math. Programming* **57** 15–48, 49–84.

———, ———, M. N. Limber, D. Noll (1995). Maximum entropy reconstruction using derivative information, Part 2: Computational results. *Numerische Mathematik* **69** 243–256.

———, M. N. Limber, D. Noll (1996) Fast heuristic methods for function reconstruction using derivative information. To appear in *Applicable Analysis*.

Clarke, F. H. (1990). Optimization and nonsmooth analysis. 2nd ed. Classics in Appl. Math. 5. Society for Industrial and Applied Mathematics (SIAM).

———, Ph.D. Loewen (1989). An intermediate existence theory in the calculus of variations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **16**(4).

Dacunha-Castelle, D., F. Gamboa (1990). Maximum d'entropie et problèmes des moments. *Ann. Inst. Henri Poincaré* **26** 567–596.

Decarreau, A., D. Hilhorst, C. Lemaréchal, J. Navaza (1992). Dual methods in entropy maximization: Applications to some problems in crystallography. *SIAM J. Optim.* **2** 173–197.

Dontchev, A. L. (1987). Duality methods for constrained interpolation. *Math. Balkanica* **1** 17–33.

———, B. D. Kalchev (1989). Duality and well-posedness in convex interpolation. *Numer. Funct. Analy. Optim.* **10** 673–688.

Doob, J. L. (1934). Probability and statistics. *Trans. Amer. Math. Soc.* **36** 759–775.

Dunford, N., T. Schwartz (1963). *Linear Operators*, Interscience, New York.

Ekeland, I., R. Teman (1989). *Convex Analysis and Variational Problems*. North Holland, Amsterdam.

Erickson, G. J., C. R. Smith (Ed.) (1988). *Maximum-entropy and Bayesian Methods in Science and Engineering*. Kluwer, Dordrecht, The Netherlands.

Fisher, R. A. (1930). Theory of statistical estimation. *Philos. Trans. Roy. Soc. London Ser* A **222** 309–368.

Hager, W. W., S. K. Mitter (1976). Lagrange duality theory for convex control problems. *SIAM J. Control Optim.* **14** 843–856.

Hewitt, E., K. Stromberg (1969). *Real and Abstract Analysis*. Springer Verlag, New York.

Jaynes, E. T. (1982). On the rationale of maximum entropy methods. *Proc. IEEE, Special issue on spectral estimation* **70** 939–952.

Lin, D. M., E. K. Wong (1990). A survey on the maximum entropy method and parameter spectral estimation. *Phys. Rep.* **193** 41–135.

Loewen, Ph.D. (1993). Optimal control via nonsmooth analysis. *CRM Proceedings and Lecture Notes*, Vol. **2**, Amer. Math. Soc., Providence, Rhode Island.

Navaza, J. (1986). The use of nonlocal constraints in maximum-entropy electron density reconstruction. *Acta Cryst. Sec.* A. **42** 212–223.

Noll, D. (1991). A unified approach to duality in convex programming. *J. Math. Anal. Appl.* **161** 508–521.

——— (1995). Rates of convergence for best entropy estimates. *Statistics and Decisions* **13** (1995), 141–165.

Rockafellar, R. T. (1968). Integrals which are convex functionals I, II. *Pacific J. Math.* **24** 525–539, **39** 439–469.

——— (1970). *Convex Analysis*. Princeton Univ. Press, Princeton, New Jersey.

——— (1971). Existence and duality theorems for convex problems of Bolza. *Trans. Amer. Math. Soc.* **159** 1–40.

——— (1972). State constraints in convex control problems of Bolza. *SIAM J. Control Optim.* **10** 691–715.

——— (1974). Conjugate duality and optimization. *SIAM*.

——— (1981). Optimality conditions for convex control problems with nonnegative states and the possibility of jumps. O. Moeschlin, ed., *Game Theory and Math. Economics*, North Holland, Amsterdam, 339–349.

Rudin, W. (1973). *Real and Complex Analysis*. McGraw-Hill, New York.

Silver, R. N. (1992). Quantum statistical inference. *IEEE Report*, Kluwer, Dordrecht, The Netherlands.

Skilling, J. (Ed.) (1989). *Maximum-entropy and Bayesian methods*. Kluwer, Dordrecht, The Netherlands.

J. M. Borwein: Centre for Experimental and Constructive Mathematics (CECM), Simon Fraser University, Burnaby, British Columbia, Canada V5A 156; e-mail: jborwein@cecm.sfu.ca

A. S. Lewis: Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; e-mail: aslewis@orion.uwaterloo.ca

D. Noll: Université Paul Sabatier, Départment de Mathématiques, 118, route de Narbonne, 31062 Toulouse, France; e-mail: noll@dumbo.ups-tlse.fr