JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **197**, 774–780 (1996)
ARTICLE NO. 0052

Superresolution in the Markov Moment Problem

A. S. Lewis

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

Submitted by E. W. Cheney

Received March 31, 1994

Suppose that $\bar{x}$ is a characteristic function and $a_1, a_2, \ldots, a_n$ are weight functions on a finite measure space. Recent work of Gamboa and Gassiat observed conditions guaranteeing that $\|x - \bar{x}\|_1$ is small whenever $0 \le x \le 1$ *a.e.* and the moment errors $\int (x - \bar{x}) a_i$ are small ($i = 1, 2, \ldots, n$). Using concise and elementary techniques we obtain similar results, under very mild assumptions. We also provide precise error bounds. © 1996 Academic Press, Inc.

1. INTRODUCTION

The Markov moment problem studies the relationship between a density $\bar{x}$ on a finite measure space S with $0 \le \bar{x} \le 1$ *a.e.* and its moments $\int a_i \bar{x}$ with respect to given integrable weight functions $a_1, a_2, \ldots, a_n$ (see [12]). In physical applications we seek to estimate $\bar{x}$ on the basis of this moment information, sometimes using a maximum entropy technique (see for example [6, 8]). A mathematical survey of such techniques may be found in [3].

Heuristically it has been observed that for some densities $\bar{x}$, any estimate $0 \le x \le 1$ *a.e.* with moments $\int a_i x$ close to the given moments $\int a_i \bar{x}$ must necessarily be close to $\bar{x}$ (in L_1 norm). This phenomenon is called "superresolution" (see [11] and [7]). An interesting explanation was presented in [9] and [10], based on some sophisticated probabilistic maximum entropy techniques introduced in [5]. The aim of this note is to give a concise, elementary, measure-theoretic approach to this problem. In this straightforward framework we derive results analogous to some of those in [9] and [10] under very simple assumptions and with an explicit error bound.

Let $[0, 1]_\infty = \{x \in L_\infty(S) | 0 \le x \le 1 \ a.e.\}$, and define a continuous linear map $A: L_\infty(S) \to \mathbb{R}^n$ by $Ax = \int ax$ (where a has components $a_1, a_2, \ldots, a_n$). Since superresolution requires that $\|\bar{x} - x\|_1$ is small whenever $0 \le x \le 1$

0022-247X/96 $18.00

$a.e.$ and $\int(x-\bar{x})a$ is small, an obvious prerequisite is that $x=\bar{x}$ is the unique solution in $[0,1]_\infty$ of $\int ax = \int a\bar{x}$. For this to be the case, except in trivial, finite-dimensional cases where the null space $N(A)=\{0\}$, it follows that $A\bar{x}$ cannot be in the relative interior of $A[0,1]_\infty$, $\mathrm{ri}(A[0,1]_\infty)$, since an easy standard argument shows that $A(\mathrm{int}[0,1]_\infty) = \mathrm{ri}(A[0,1]_\infty)$ (see for example Proposition 2.10 in [1]). Hence there exists a supporting hyperplane for $A[0,1]_\infty$ at $A\bar{x}$: for some nonzero λ in $\mathbb{R}^n$, $\int(x-\bar{x})\lambda^T a \le 0$ whenever $0 \le x \le 1$ $a.e.$ Thus providing that $\lambda^T a$ is nonzero a.e. we have that $\bar{x} = \chi_{S_{\lambda^T a}}$, the characteristic function of the set $S_{\lambda^T a} = \{s \in S \mid \lambda^T a(s) > 0\}$.

The above argument shows that we may often restrict our study of superresolution to cases where $\bar{x} = \chi_{S_{\lambda^T a}}$ for some nonzero λ in $\mathbb{R}^n$. This is our assumption for the main result, Theorem 2.2, which gives conditions guaranteeing that, for $0 \le x \le 1$ $a.e.$,

$$\|\bar{x}-x\|_1 = O\left(\left\|\int(\bar{x}-x)a\right\|^{1/2}\right).$$

The following example shows that the order of growth is best possible.

Example. Let $S=[0,1]$ with Lebesgue measure, $\bar{x} = \chi_{[0,1/2]}$, and $x_\varepsilon = \chi_{[0,1/2-\varepsilon]\cup[1/2,1/2+\varepsilon]}$ for $0 \le \varepsilon \le \frac{1}{2}$. Then with $a(s) = (1,s)^T$ we obtain $\int(\bar{x}-x_\varepsilon)a = (0,-\varepsilon^2)^T$ and $\|\bar{x}-x_\varepsilon\|_1 = 2\varepsilon$.

We have seen that the superresolution phenomenon is confined to cases where the underlying density $\bar{x}$ is the characteristic function of a set of the form $S_{\lambda^T a}$. It is therefore natural to ask how one recognizes such sets. In one case, familiar from approximation theory, this is extremely easy.

Continuous functions $a_1, a_2, \dots, a_n$ on a real interval I are said to satisfy the *Haar condition* if $\lambda^T a$ has at most $n-1$ zeroes for any nonzero λ in $\mathbb{R}^n$ (see for example [4]). The standard example is $a_i(s) = s^{i-1}$, for $i = 1,2,\dots,n$. Suppose that $S=[\mu,\nu]$ is contained in the interior of I (endowed with Lebesgue measure). Then any set $S_{\lambda^T a}$ clearly has the form (up to measure zero)

$$[s_0,s_1]\cup[s_2,s_3]\cup\cdots \qquad \text{or} \qquad [s_1,s_2]\cup[s_3,s_4]\cup\cdots,$$

where $\mu = s_0 < s_1 < \cdots < s_k < s_{k+1} = \nu$ and $k < n$.

In fact, the converse is also true: any set of this form can be written $S_{\lambda^T a}$ for some λ in $\mathbb{R}^n$. To see this, we simply choose λ so that $\lambda^T a$ has zeroes $s_1, s_2, \dots, s_k$ in the interior of S and any remaining zeroes outside S (and if necessary replace λ by $-\lambda$). Clearly a similar technique applies to the case where $S=[-\pi,\pi]$ and $a(s) = (1, \cos s, \sin s, \cos 2s, \sin 2s, \dots)^T$.

2. THE MAIN RESULT

We suppose that (S, ρ) is a fixed finite measure space. Our quantification of the superresolution phenomenon revolves around the following idea. We define, for any nonnegative function f in $L_1(S)$, the constant

$$\beta_f = \limsup_{\delta \downarrow 0} \delta^{-1} \rho\{s \in S | f(s) \leq \delta\}. \tag{2.1}$$

The constant β_f is finite exactly when the measure of the set $\{s | f(s) \leq \delta\}$ does not grow faster than linearly for small positive δ. For example, if the set S is a compact, nonsingleton, real interval with Lebesgue measure, then $\beta_{|g|}$ will be finite for any continuously differentiable function g on S with only a finite number of zeroes, all simple.

Our aim in this section is to prove the following result.

THEOREM 2.2. *Suppose that the functions $a_1, a_2, \ldots, a_n$ lie in $L_1(S)$, and that $\bar{x} = \chi_{S_{\lambda^T a}}$ for some λ in $\mathbb{R}^n$ with $\lambda^T a \neq 0$ a.e. Then for any sequence of measurable functions $x_r \colon S \to [0, 1]$, if $\int a x_r \to \int a \bar{x}$ it follows that $\|x_r - \bar{x}\|_1 \to 0$.*

Suppose further that the constant $\beta_{|\lambda^T a|}$ defined by (2.1) *is finite. Then for any norm $\|\cdot\|$ on $\mathbb{R}^n$ the following error estimate holds*:

$$\|\bar{x} - x_r\|_1 \leq K(r) \left\| \int (\bar{x} - x_r) a \right\|^{1/2}, \tag{2.3}$$

where the function K satisfies

$$K(r) \sim \left(2 \beta_{|\lambda^T a|} \|\lambda\|_*\right)^{1/2} \qquad \text{as } r \to \infty$$

(*and $\|\cdot\|_*$ is the dual norm*).

The proof will depend on a sequence of lemmas. For any function f in $L_1(S)$ we define a function $\Lambda_f \colon \mathbb{R} \to (-\infty, +\infty]$ by

$$\Lambda_f(\varepsilon) = \inf\left\{ \int f y \,\middle|\, 0 \leq y \leq 1 \text{ a.e.}, \int y \geq \varepsilon \right\}. \tag{2.4}$$

For u in $\mathbb{R}$ we write $u^+ = \max\{u, 0\}$.

LEMMA 2.5. *The function Λ_f is convex, nondecreasing, and continuous on its domain, $(-\infty, \rho(S)]$. The infimum in* (2.4) *is attained for any ε in $(-\infty, \rho(S)]$. For any ε in $\mathbb{R}$,*

$$\Lambda_f(\varepsilon) \geq \sup\left\{ \gamma \varepsilon - \int (\gamma - f)^+ \,\middle|\, 0 \leq \gamma \in \mathbb{R} \right\}. \tag{2.6}$$

Proof. The function Λ_f is just the value function of the convex program on the right-hand side of (2.4). Convexity is easy and standard to check, while attainment and lower semicontinuity are consequences of the weak-star compactness of $\{y \in L_\infty | 0 \leq y \leq 1 \ a.e.\}$. The inequality (2.6) is simply the weak duality inequality for problem (2.4), and is easily checked directly. ∎

In fact, a standard Fenchel duality result applied to the right-hand side of (2.6) shows that (2.6) holds with equality for all ε in $\mathbb{R}$ (see for example [13]). Furthermore, the supremum in (2.6) is attained, at least for $\varepsilon < \rho(S)$, by the results in [2]. We shall not need these stronger duality results.

The next result is the key tool in our general convergence analysis.

LEMMA 2.7. *Suppose that $f > 0$ a.e. Then $\Lambda_f(\varepsilon) > 0$ for all $\varepsilon > 0$, and if $\Lambda_f(\varepsilon_r) \to 0$ for some sequence $\varepsilon_r > 0$ then $\varepsilon_r \to 0$.*

Proof. The first statement is a consequence of the attainment in (2.4). Now since $\varepsilon_r \in [0, \rho(S)]$ for all large r, if $\varepsilon_r \nrightarrow 0$ then some subsequence $\varepsilon_{r'}$ has limit $\varepsilon > 0$, whence by continuity $\Lambda_f(\varepsilon_{r'}) \to \Lambda_f(\varepsilon) > 0$. This is a contradiction. ∎

This result can sometimes be quantified.

LEMMA 2.8. *If $f > 0$ a.e. then*

$$\limsup_{\delta \downarrow 0} \delta^{-2} \int (\delta - f)^+ \leq \beta_f / 2.$$

Proof. By Fubini's theorem,

$$\int (\delta - f)^+ = \int_{s \in S} \int_{t \in \mathbb{R}} \chi_{\{0 < f(s) \leq \delta,\, 0 \leq t \leq \delta - f(s)\}} \, dt \, d\rho$$

$$= \int_{t=0}^{\delta} \rho\{s \in S | f(s) \leq \delta - t\} \, dt$$

$$= \int_0^{\delta} \rho\{s \in S | f(s) \leq r\} \, dr.$$

For any $\varepsilon > 0$,

$$r^{-1} \rho\{s \in S | f(s) \leq r\} \leq \beta_f + \varepsilon, \qquad \text{for all small } r > 0,$$

so for small $\delta > 0$,

$$\delta^{-2} \int (\delta - f)^+ \leq \delta^{-2} \int_0^{\delta} (\beta_f + \varepsilon) r \, dr = (\beta_f + \varepsilon)/2.$$

The result follows. ∎

LEMMA 2.9. *Suppose that the function* f *lies in* $L_1(S)$ *with* $f > 0$ *a.e., and* β_f *defined by* (2.1) *finite. Then*

$$\liminf_{\varepsilon \downarrow 0} \varepsilon^{-2} \Lambda_f(\varepsilon) \geq (2\beta_f)^{-1}.$$

Proof. For any $k > 0$, setting $\gamma = k\varepsilon$ in (2.6) shows that

$$\begin{aligned}\liminf_{\varepsilon \downarrow 0} \varepsilon^{-2} \Lambda_f(\varepsilon) &\geq \liminf_{\varepsilon \downarrow 0} \varepsilon^{-2} \left\{ k\varepsilon^2 - \int (k\varepsilon - f)^+ \right\} \\ &= k - k^2 \limsup_{\delta \downarrow 0} \delta^{-2} \int (\delta - f)^+ \\ &\geq k - \beta_f k^2/2,\end{aligned}$$

by the previous lemma. Setting $k = \beta_f^{-1}$ gives the result. ∎

For any function g in $L_1(S)$, define $S_g = \{s \in S | g(s) > 0\}$.

LEMMA 2.10. *For any measurable functions* g *in* $L_1(S)$ *and* $x\colon S \to [0,1]$,

$$\int (\chi_{S_g} - x) g \geq \Lambda_{|g|}(\| \chi_{S_g} - x \|_1).$$

Proof. Choose $y = |\chi_{S_g} - x|$ in (2.4). Then $0 \leq y \leq 1$ *a.e.* and $\int y = \| \chi_{S_g} - x \|_1$. Furthermore,

$$\begin{aligned}\int |g| y &= \int |g| |\chi_{S_g} - x| \\ &= \int_{g>0} |g| |1 - x| + \int_{g<0} |g| |-x| \\ &= \int_{g>0} (1 - x) g + \int_{g<0} x(-g) \\ &= \int (\chi_{S_g} - x) g.\end{aligned}$$

The result follows. ∎

We can now prove the main result.

Proof of Theorem 2.2. By the previous lemma,

$$0 \le \Lambda_{|\lambda^T a|}(\|\bar{x} - x_r\|_1) \le \int (\bar{x} - x_r)\lambda^T a \to 0, \tag{2.11}$$

so $\|x - x_r\|_1 \to 0$ by Lemma 2.4. To see the second part, observe that

$$\|\lambda\|_* \left\| \int (\bar{x} - x_r) a \right\| \ge \int (\bar{x} - x_r)\lambda^T a \ge \Lambda_{|\lambda^T a|}(\|\bar{x} - x_r\|_1).$$

Without loss of generality, for all r, $x_r \ne \bar{x}$, and hence $\int(\bar{x} - x_r)a \ne 0$, by Lemma 2.4. Then by Lemma 2.9,

$$\liminf_{r\to\infty} \|\bar{x} - x_r\|_1^{-2} \|\lambda\|_* \left\| \int (\bar{x} - x_r) a \right\| \ge (2\beta_{|\lambda^T a|})^{-1},$$

and the result follows. ∎

Note that the proof of the first part of the theorem in fact only needs the assumption that $\int\lambda^T a x_r \to \int\lambda^T a\bar{x}$. An equivalent way to state this part of the result is then the following: for any function g in $L^1(S)$ with $g \ne 0$ a.e., and any sequence of measurable functions $x_r\colon S \to [0,1]$, if $\int g x_r \to \int g_+$ then it follows that $\|x_r - \chi_{S_g}\|_1 \to 0$. This can also be seen by a more direct argument.

EXAMPLE. Suppose in Theorem 2.2 that S is a compact interval of $\mathbb{R}$ (not a singleton) with Lebesgue measure, and that $\lambda^T a$ is continuously differentiable on S with zeroes $s_1, s_2, \dots, s_m$ $(m > 0)$, all simple. A straightforward calculation shows that

$$\beta_{|\lambda^T a|} = \sum_{i=1}^{m} \alpha_i \left|(\lambda^T a)'(s_i)\right|^{-1},$$

where $\alpha_i = 2$ if s_i lies in the interior of S and $\alpha_i = 1$ otherwise. Hence we obtain the error estimate

$$\limsup_{r\to\infty} \|\bar{x} - x_r\|_1 \left\| \int (\bar{x} - x_r) a \right\|^{-1/2} \le \left(2\|\lambda\|_* \sum_{i=1}^{m} \alpha_i \left|(\lambda^T a)'(s_i)\right|^{-1} \right)^{1/2}.$$

We can compare this inequality with the example at the end of Section 1, by setting $\lambda = (1, -2)^T$, so $s_1 = \frac{1}{2}$ and

$$\|\bar{x} - x_\varepsilon\|_1 \left\| \int (\bar{x} - x_\varepsilon) a \right\|_1^{-1/2} = 2 = \left(2\|\lambda\|_\infty \alpha_1 \left|(\lambda^T a)'(s_1)\right|^{-1}\right)^{1/2},$$

whence the error bound is tight.

REFERENCES

1. J. M. Borwein and A. S. Lewis, Partially finite convex programming. I. Duality theory, *Math. Programming B* **57** (1992), 15–48.
2. J. M. Borwein and A. S. Lewis, Partially finite convex programming. II. Explicit lattice models, *Math. Programming B* **57** (1992), 49–84.
3. J. M. Borwein and A. S. Lewis, Partially-finite programming in L_1 and the existence of maximum entropy estimates, *SIAM J. Optim.* **2** (1993), 248–267.
4. E. W. Cheney, "Introduction to Approximation Theory," McGraw–Hill, New York, 1966.
5. D. Dacunha-Castelle and F. Gamboa, Maximum d'entropie et problème des moments, *Ann. Inst. H. Poincaré* **26** (1990), 567–596.
6. A. Decarreau, D. Hilhorst, C. Lemaréchal, and J. Navaza, Dual methods in entropy maximization: Application to some problems in crystallography, *SIAM J. Optim.* **2** (1992), 173–197.
7. D. L. Donoho, Superresolution via sparsity constraints, *SIAM J. Math. Anal.* **23** (1992), 1309–1331.
8. D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, Maximum entropy and the nearly black object, *J. Roy. Statist. Soc. Ser. B* **54** (1992), 41–81.
9. F. Gamboa and E. Gassiat, M.E.M. techniques for solving moment problems, preprint, Statistics Laboratory, University of Orsay, Orsay, 1991.
10. F. Gamboa and E. Gassiat, Set of superresolution and the maximum entropy method on the mean, preprint, Statistics Laboratory, University of Orsay, Orsay, 1992.
11. P. A. Jansson (Ed.), "Deconvolution, with Applications in Spectroscopy," Academic Press, New York, 1984.
12. M. G. Krein and A. A. Nudel'man, "The Markov Moment Problem and Extremal Problems," American Math. Society, Providence, 1977.
13. R. T. Rockafellar, "Conjugate Duality and Optimization," SIAM, Philadelphia, 1974.