arxiv.tex; May 10, 2014

On the extreme points of moments sets

Iosif Pinelis*

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract: Necessary and sufficient conditions for a measure to be an extreme point of the set of measures (on an abstract measurable space) with prescribed generalized moments are given, as well as an application to extremal problems over such moment sets; these conditions are expressed in terms of atomic partitions of the measurable space. It is also shown that every such extreme measure can be adequately represented by a linear combination of k Dirac probability measures with nonnegative coefficients, where k is the number of restrictions on moments; moreover, when the measurable space has appropriate topological properties, the phrase "can be adequately represented by" here can be replaced simply by "is". The proofs are elementary and mainly self-contained.

AMS 2000 subject classifications: Primary 49K30, 49K27, 52A40; secondary 26D15, 46A55, 46N10, 46N30, 60B05, 90C05, 90C25, 90C26, 90C48.
Keywords and phrases: moment sets, measures, generalized moments, extreme points, optimization, extremal problems, atoms, convex sets.

Let S be an arbitrary set and let Σ be a σ-algebra over S, so that (S, Σ) is a measurable space. Let M denote the set of all (nonnegative) real-valued measures on Σ. For any $\mu \in \mathrm{M}$ and any $A \in \Sigma$, define the truncation μ_A of the measure μ by A via the formula

$$\mu_A(B) := \mu(A \cap B)$$

for all $B \in \Sigma$; it is then clear that $\mu_A \in \mathrm{M}$.

For any $\mu \in \mathrm{M}$ and any $A \in \Sigma$, let us say that A is a *μ-atom* if $\mu_A(B) \in \{0, \mu(A)\}$ for all $B \in \Sigma$.

For any $n \in \mathbb{N}$ and $\mu \in \mathrm{M}$, let us refer to an n-tuple $(A_1, \dots, A_n)$ of members of Σ as a *non-null (n, μ)-partition of S* if the sets $A_1, \dots, A_n$ are pairwise disjoint, $A_1 \cup \dots \cup A_n = S$, and $\mu(A_i) > 0$ for all $i \in \overline{1,n}$; let us say that such a partition $(A_1, \dots, A_n)$ is *atomic* if A_i is a μ-atom for each $i \in \overline{1,n}$; here and in what follows, for any m and n in $\mathbb{Z} \cup \{\infty\}$ we let $\overline{m,n} := \{j \in \mathbb{Z} \colon m \leqslant j \leqslant n\}$.

Let $\mathcal{F}$ stand for the set of all Σ-measurable real-valued functions on S.

Take any $k \in \overline{1,\infty}$. For each $j \in \overline{1,k}$, take any $f_j \in \mathcal{F}$, and let

$$\mathbf{f} := (f_1, \dots, f_k).$$

For any $\mu \in \mathrm{M}$, let us write $\mathbf{f} \in L^1(\mu)$ if $f_j \in L^1(S, \Sigma, \mu)$ for all $j \in \overline{1,k}$, and also let

$$\int_A \mathbf{f}\,\mathrm{d}\mu := \Big(\int_A f_1 \mathrm{d}\mu, \dots, \int_A f_k \mathrm{d}\mu \Big)$$

*Supported by NSA grant H98230-12-1-0237

for all $\mathbf{f} \in L^1(\mu)$ and $A \in \Sigma$.

For any $C \subseteq \mathbb{R}^k$, consider the moment set

$$\mathrm{M}_{\mathbf{f},C} := \Big\{\mu \in \mathrm{M}\colon \mathbf{f} \in L^1(\mu), \int_S \mathbf{f}\,\mathrm{d}\mu \in C\Big\}.$$

In the case when C is a singleton set of the form $\{c\}$ for some $c \in \mathbb{R}^k$, let us write $\mathrm{M}_{\mathbf{f},c}$ in place of $\mathrm{M}_{\mathbf{f},C}$.

Let Λ denote a subset of M, and then consider the corresponding narrower moment sets

$$\Lambda_{\mathbf{f},C} := \Lambda \cap \mathrm{M}_{\mathbf{f},C} \quad \text{and} \quad \Lambda_{\mathbf{f},c} := \Lambda \cap \mathrm{M}_{\mathbf{f},c}, \tag{1}$$

again for any $C \subseteq \mathbb{R}^k$ and $c \in \mathbb{R}^k$.

Consider also

$$\Pi := \{\mu \in \Lambda\colon \mu(S) = 1\},$$

the set of all probability measures in Λ.

For any (not necessarily convex) subset K of M, let $\operatorname{ex} \mathrm{K}$ denote the set of all extreme points of K. Thus, $\mu \in \operatorname{ex}\mathrm{K}$ if and only if $\mu \in \mathrm{K}$ and for any $(t, \nu_0, \nu_1) \in (0,1) \times \mathrm{K} \times \mathrm{K}$ such that $(1-t)\nu_0 + t\nu_1 = \mu$ one has $\nu_0 = \nu_1$.

The following theorem presents a necessary condition for a measure to be an extreme point of the set $\Lambda_{\mathbf{f},C}$.

Theorem 1. (Necessity). *Take any $C \subseteq \mathbb{R}^k$. Suppose that the following condition holds:*

$$\Lambda \textit{ is a convex cone such that } \mu_A \in \Lambda \textit{ for all } \mu \in \Lambda \textit{ and } A \in \Sigma. \tag{2}$$

Take any $\mu \in \operatorname{ex}\Lambda_{\mathbf{f},C} \setminus \{0\}$. Then for some $m \in \overline{1,k}$ there is an atomic non-null (m,μ)-partition $(A_1,\dots,A_m)$ of S such that the vectors $\int_{A_1}\mathbf{f}\,\mathrm{d}\mu,\dots,\int_{A_m}\mathbf{f}\,\mathrm{d}\mu$ are linearly independent.

As usual, let us say that a measure $\mu \in \mathrm{M}$ is a 0,1 measure if all its values are in the set $\{0,1\}$; so, any 0,1 measure in Λ is also in Π.

In the case when $k = 1$, $f_1 = 1$ on S, and $C = \{1\} \subset \mathbb{R}^1$, Theorem 1 turns into

Corollary 2. *Suppose that condition (2) holds. Then any measure in $\operatorname{ex}\Pi$ is a 0,1 measure.*

Proof of Theorem 1. Suppose that for some $n \in \overline{k+1,\infty}$ there is a non-null (n,μ)-partition $(A_1,\dots,A_n)$ of S, so that $\mu(A_i) > 0$ for all $i \in \overline{1,n}$. Since $n > k$, the n vectors $\int_{A_1}\mathbf{f}\,\mathrm{d}\mu,\dots,\int_{A_n}\mathbf{f}\,\mathrm{d}\mu$ in $\mathbb{R}^k$ are linearly dependent, so that $\varepsilon_1\int_{A_1}\mathbf{f}\,\mathrm{d}\mu + \dots + \varepsilon_n\int_{A_n}\mathbf{f}\,\mathrm{d}\mu = 0$ for some nonzero vector $\varepsilon = (\varepsilon_1,\dots,\varepsilon_n) \in (-1,1)^n$. For any such vector ε, let $\nu_\pm := (1\pm\varepsilon_1)\mu_{A_1} + \dots + (1\pm\varepsilon_n)\mu_{A_n}$. Then, by (2), $\nu_\pm \in \Lambda$. Moreover, $\int_S \mathbf{f}\,\mathrm{d}\nu_\pm = \sum_1^n (1\pm\varepsilon_i)\int_{A_i}\mathbf{f}\,\mathrm{d}\mu = \sum_1^n \int_{A_i}\mathbf{f}\,\mathrm{d}\mu = \int_S \mathbf{f}\,\mathrm{d}\mu \in C$, so that $\nu_\pm \in \Lambda_{\mathbf{f},C}$. Also, $\frac{1}{2}(\nu_+ + \nu_-) = \mu$ and $(\nu_+ - \nu_-)(A_i) = 2\varepsilon_i\mu(A_i) \neq 0$ for some $i \in \overline{1,n}$. So, $\mu \notin \operatorname{ex}\Lambda_{\mathbf{f},C}$, which is a contradiction.

Therefore, $N_\mu \subseteq \overline{1,k}$, where N_μ denotes the set of all $n \in \mathbb{N}$ for which there is a non-null (n,μ)-partition of S. On the other hand, the condition $\mu \neq 0$ implies that $1 \in N_\mu$, so that $N_\mu \neq \emptyset$. Thus, $m := \max N_\mu \in \overline{1,k}$. Since $m \in N_\mu$, there is a non-null (m,μ)-partition $(A_1,\dots,A_m)$ of S, and the reasoning in the previous paragraph shows that the vectors $\int_{A_1} \mathbf{f}\mathrm{d}\mu, \dots, \int_{A_m} \mathbf{f}\mathrm{d}\mu$ are necessarily linearly independent. It remains to show that the non-null (m,μ)-partition $(A_1,\dots,A_m)$ of S is atomic. Indeed, suppose the contrary, so that, without loss of generality, A_1 is not a μ-atom. Then for some $B \in \Sigma$ one has $B \subset A_1$ and $0 < \mu(B) < \mu(A_1)$, and so, $(B, A_1 \setminus B, A_2, \dots, A_m)$ is a non-null $(m+1,\mu)$-partition of S, which contradicts the definition $m := \max N_\mu$. Now the proof of Theorem 1 is complete. □

Assuming the conditions of Theorem 1 and using it, we shall show that every measure $\mu \in \operatorname{ex}\Lambda_{\mathbf{f},C}$ can be represented, in a certain sense sufficient for applications, by a linear combination of Dirac probability measures with nonnegative coefficients. Recall that the Dirac probability measure δ_s with the mass at a point $s \in S$ is defined by the formula $\delta_s(A) = \mathbf{I}\{s \in A\}$ for all $A \in \Sigma$, where $\mathbf{I}\{\cdot\}$ is the indicator function. Introduce indeed the following sets of (discrete) measures on Σ:

$$\Delta_{\mathbf{f}}^{(k)} := \Big\{ \sum_1^m a_i\delta_{s_i} \colon m \in \overline{0,k}, (a_1,\dots,a_m) \in (0,\infty)^m, (s_1,\dots,s_m) \in S^m, \\ \mathbf{f}(s_1),\dots,\mathbf{f}(s_m) \text{ are linearly independent} \Big\},$$

and

$$\Delta^{(k)} := \Big\{ \sum_1^k a_i\delta_{s_i} \colon (a_1,\dots,a_k) \in [0,\infty)^k, (s_1,\dots,s_k) \in S^k \Big\}.$$

In the above definition of $\Delta_{\mathbf{f}}^{(k)}$, in the case when $m = 0$ it is assumed that $\sum_1^m a_i\delta_{s_i}$ is (the) zero (measure) and that the condition following $m \in \overline{0,k}$ between the braces is trivially satisfied. In particular, $0 \in \Delta_{\mathbf{f}}^{(k)} \subseteq \Delta^{(k)} \subseteq \mathrm{M}$.

Note that for any $g \in \mathcal{F}$ and

$$\text{for any } \nu = \sum_1^k a_i\delta_{s_i} \in \Delta^{(k)}, \text{ one has } \int_S g\mathrm{d}\nu = \sum_1^k a_i g(s_i).$$

To obtain Corollary 4 below, we shall need

Lemma 3. *Take any $\mu \in \mathrm{M}$. If A is a μ-atom and $g \in \mathcal{F}$, then there is a real number α such that $g = \alpha$ μ-almost everywhere (a.e.) on A.*

Proof of Lemma 3. For all $\beta \in [-\infty,\infty]$, let $A_\beta := \{s \in A \colon g(s) < \beta\}$. Then $\mu(A_\beta)$ is nondecreasing in $\beta \in \mathbb{R}$ and takes values in the set $\{0, \mu(A)\}$. Let $\beta_* := \sup\{\beta \in \mathbb{R} \colon \mu(A_\beta) = 0\} \in [-\infty,\infty]$. Then $\mu(A_\beta) = 0$ for all $\beta \in [-\infty, \beta_*)$ and $\mu(A_\beta) = \mu(A)$ for all $\beta \in (\beta_*, \infty]$. If $\beta_* > -\infty$ then $\mu(A_{\beta_*}) = \lim_{\beta\uparrow\beta_*} \mu(A_\beta) = 0$; if $\beta_* = -\infty$ then $A_{\beta_*} = \emptyset$ and hence again $\mu(A_{\beta_*}) = 0$. Thus, in all cases $\mu(A_{\beta_*}) = 0$ or, equivalently, $g \geqslant \beta_*$ μ-a.e. on A. Similarly, if $\beta_* < \infty$ then

$\mu(A_{\beta_*+}) = \lim_{\beta\downarrow\beta_*}\mu(A_\beta) = \mu(A)$, where $A_{\beta+} := \{s \in A\colon g(s) \leqslant \beta\}$; if $\beta_* = \infty$ then $A_{\beta_*+} = A$ and hence again $\mu(A_{\beta_*}) = \mu(A)$. Thus, in all cases $g \leqslant \beta_*$ μ-a.e. on A. We conclude that $g = \beta_*$ μ-a.e. on A. In particular, it follows that, if $\mu(A) > 0$, then necessarily $\beta_* \in \mathbb{R}$, because g is real-valued; and if $\mu(A) = 0$ then $g = \alpha$ μ-a.e. on A for any given real number α. □

From now on, take any $C \subseteq \mathbb{R}^k$ and take g to be any function in $\mathcal{F}$ such that the integral $\int_S g\,\mathrm{d}\mu$ exists in $[-\infty, \infty]$ for each $\mu \in \Lambda_{\mathbf{f},C}$.

Corollary 4. *Suppose that condition (2) holds. Take any $\mu \in \operatorname{ex}\Lambda_{\mathbf{f},C}$. Then*

(i) there is a measure $\tilde\mu$ such that

$$\tilde\mu \in \Delta^{(k)}_{\mathbf{f}} \cap \mathrm{M}_{\mathbf{f},C}, \quad \int_S \mathbf{f}\,\mathrm{d}\tilde\mu = \int_S \mathbf{f}\,\mathrm{d}\mu, \quad \text{and} \quad \int_S g\,\mathrm{d}\tilde\mu = \int_S g\,\mathrm{d}\mu. \tag{3}$$

(ii) If, in addition,

$$\Delta^{(k)}_{\mathbf{f}} \subseteq \Lambda, \tag{4}$$

then it follows that $\tilde\mu \in \Delta^{(k)}_{\mathbf{f}} \cap \Lambda_{\mathbf{f},C}$.

In this corollary, one can replace $\Delta^{(k)}_{\mathbf{f}}$ by $\Delta^{(k)}$ throughout.

Proof of Corollary 4. If $\mu = 0$ then, as discussed above, $\mu \in \Delta^{(k)}_{\mathbf{f}}$, so that (3) will hold with $\tilde\mu = \mu$. It remains to consider the case $\mu \in \operatorname{ex}\Lambda_{\mathbf{f},C} \setminus \{0\}$. Then, by Theorem 1, for some $m \in \overline{1,k}$ there is an atomic non-null (m,μ)-partition $(A_1, \dots, A_m)$ of S such that the vectors $\int_{A_1}\mathbf{f}\,\mathrm{d}\mu, \dots, \int_{A_m}\mathbf{f}\,\mathrm{d}\mu$ are linearly independent.

Let now $f_0 := g$. By Lemma 3, for each pair $(i,j) \in \overline{1,m} \times \overline{0,k}$ there exist a subset $B_{i,j}$ of A_i and a real number $a_{i,j}$ such that $B_{i,j} \in \Sigma$, $\mu(B_{i,j}) = \mu(A_i)$, and $f_j(s) = a_{i,j}$ for all $s \in B_{i,j}$.

Fix, in this paragraph, any $i \in \overline{1,m}$, and let $B_i := \cap_{j=0}^k B_{i,j}$. Then $B_i \subseteq A_i$, $\mu(B_i) = \mu(A_i) > 0$ and $f_j(s) = a_{i,j}$ for all $s \in B_i$ and $j \in \overline{0,k}$. Since $\mu(B_i) > 0$, there is some $s_i \in B_i$. For any such s_i and each $j \in \overline{0,k}$,

$$\int_{A_i} f_j\,\mathrm{d}\mu = a_{i,j}\mu(A_i) = f_j(s_i)\mu(A_i) = \int_{A_i} f_j\,\mathrm{d}\tilde\mu,$$

where $\tilde\mu := \mu(A_1)\delta_{s_1} + \dots + \mu(A_m)\delta_{s_m}$. Now (3) immediately follows, and then part (ii) of Corollary 4 immediately follows by (1).

The last sentence of the corollary now follows immediately as well, because $\Delta^{(k)}_{\mathbf{f}} \subseteq \Delta^{(k)}$. □

Corollary 5. *Suppose that condition (2) holds and*

$$\sup\Big\{\int_S g\,\mathrm{d}\mu\colon \mu \in \Lambda_{\mathbf{f},C}\Big\} = \sup\Big\{\int_S g\,\mathrm{d}\mu\colon \mu \in \operatorname{ex}\Lambda_{\mathbf{f},C}\Big\}. \tag{5}$$

Then

$$\sup\Big\{\int_S g\,\mathrm{d}\mu\colon \mu \in \Lambda_{\mathbf{f},C}\Big\} \leqslant \sup\Big\{\int_S g\,\mathrm{d}\mu\colon \mu \in \Delta^{(k)}_{\mathbf{f}} \cap \mathrm{M}_{\mathbf{f},C}\Big\}. \tag{6}$$

If condition (4) *holds as well, then*

$$\sup\Big\{\int_S g\mathrm{d}\mu\colon\, \mu\in\Lambda_{\mathbf{f},C}\Big\}=\sup\Big\{\int_S g\mathrm{d}\mu\colon\, \mu\in\Delta_{\mathbf{f}}^{(k)}\cap\Lambda_{\mathbf{f},C}\Big\}. \tag{7}$$

In (6)*, one can replace* $\Delta_{\mathbf{f}}^{(k)}$ *by* $\Delta^{(k)}$*; one can do so in* (7) *too if such a replacement is done in condition* (4) *as well.*

Proof of Corollary 5. By part (i) of Corollary 4, the right-hand side of (5) is no greater than the right-hand side of (6). Hence, by (5), the left-hand side of (6) is no greater than its right-hand side.

Now, if (4) holds, then $\Delta_{\mathbf{f}}^{(k)}\cap\mathrm{M}_{\mathbf{f},C}=\Delta_{\mathbf{f}}^{(k)}\cap\Lambda_{\mathbf{f},C}$, and hence the right-hand side of (6) equals the right-hand side of (7), which in turn is obviously no greater than the left-hand side of (7). So, (7) follows.

The last sentence of Corollary 5 is proved quite similarly, using the last sentence of Corollary 4. □

Applications of equalities of the form (7) can be found e.g. in [10].

Let us now indicate a number of generic cases when condition (5) holds:

Proposition 6. *Condition* (5) *is satisfied in each of the following cases:*

(i) when there exists an extreme point of the set

$$\Lambda_{\max g;\,\mathbf{f},C}:=\Big\{\nu\in\Lambda_{\mathbf{f},C}\colon\,\int_S g\mathrm{d}\nu\geqslant\int_S g\mathrm{d}\mu \text{ for all } \mu\in\Lambda_{\mathbf{f},C}\Big\};$$

(ii) when $\Lambda_{\max g;\,\mathbf{f},C}$ *is a nonempty compact convex subset of a locally convex space;*

(iii) when $\Lambda_{\max g;\,\mathbf{f},C}$ *is a nonempty compact finite-dimensional set;*

(iv) when $\Lambda_{\max g;\,\mathbf{f},C}$ *is a singleton set* (*that is, when the maximum of* $\int_S g\mathrm{d}\mu$ *over all* $\mu\in\Lambda_{\mathbf{f},C}$ *is attained at a unique measure* $\mu\in\Lambda_{\mathbf{f},C}$)*;*

(v) when the set S *is endowed with the structure of a Hausdorff topological space, the* σ*-algebra* Σ *coincides with the corresponding Borel* σ*-algebra* $\mathcal{B}$*,* $\Lambda=\mathrm{M}$*, and* $C=\{1\}\times I_2\times\cdots\times I_k$*, where* $I_2,\dots,I_k$ *are arbitrary closed convex subsets of* $\mathbb{R}$ (*so that all measures in* $\Lambda_{\mathbf{f},C}$ *are probability measures*)*.*

Proof of Proposition 6.
(i): Suppose that condition (i) of Proposition 6 holds, so that there exists an extreme point of the set $\Lambda_{\max g;\,\mathbf{f},C}$. Then it is easy to see that any such point (say $\mu_{\max}$) is in $\operatorname{ex}\Lambda_{\mathbf{f},C}$. At that, $\int_S g\mathrm{d}\mu_{\max}$ equals the left-hand side of (5). Thus, (5) follows.
(ii, iii, iv): Suppose that condition (ii) of Proposition 6 holds. Then, by the Krein–Milman theorem (see e.g. [8]), condition (i) of the proposition holds as well. Thus, (5) follows. Note also that condition (iv) of Proposition 6 implies condition (iii), which in turn implies (ii).
(v): Suppose that condition (v) of Proposition 6 holds. Then, by the arguments in [15, Theorems 3.1 and 3.2 and Proposition 3.1] (which in turn rely mainly on [14]), (5) follows. □

Let us now supplement the results presented above by a few other ones, which are perhaps of lesser interest, in that they are not needed for applications such as Corollary 5.

The following theorem supplements Theorem 1, as it presents a sufficient condition for a measure to be extreme in the moment set. Note that condition (2) is not needed in Theorem 7; on the other hand, here C is taken to be a singleton set.

Theorem 7. (Sufficiency). *Take any $c \in \mathbb{R}^k$. Take any $\mu \in \Lambda_{\mathbf{f},c}$ such that for some $m \in \overline{1,k}$ there is an atomic non-null (m,μ)-partition $(A_1,\dots,A_m)$ of S and at that the vectors $\int_{A_1}\mathbf{f}\,\mathrm{d}\mu,\dots,\int_{A_m}\mathbf{f}\,\mathrm{d}\mu$ are linearly independent. Then $\mu \in \operatorname{ex}\Lambda_{\mathbf{f},c}$.*

In the case when $k=1$, $f_1=1$ on S, and $c=1$, Theorem 7 turns into

Corollary 8. *Any* 0,1 *measure in* Π *is in* $\operatorname{ex}\Pi$ (cf. Corollary 2).

Proof of Theorem 7. Take any $(t,\nu_+,\nu_-) \in (0,1)\times\Lambda_{\mathbf{f},c}\times\Lambda_{\mathbf{f},c}$ such that $(1-t)\nu_+ + t\nu_- = \mu$. We need to show that then $\nu_+=\nu_-$.

Step 1. Let us note here that $\nu_\pm(B)=0$ for all $B\in\Sigma$ such that $\mu(B)=0$; that is, the measures $\nu_\pm$ are absolutely continuous with respect to μ. This follows because $1-t>0$, $t>0$, and for all $B\in\Sigma$ one has $0\leqslant(1-t)\nu_+(B)\leqslant\mu(B)$ and $0\leqslant t\nu_-(B)\leqslant\mu(B)$.

Step 2. Here we note that, if A is a μ-atom and a measure $\nu\in\mathrm{M}$ is absolutely continuous with respect to μ, then A is a ν-atom as well and, moreover, $\nu_A = a\mu_A$ for some $a\in[0,\infty)$. Indeed, take any $B\in\Sigma$ such that $B\subseteq A$. If $\mu(B)=0$ then $\nu(B)=0$, by the absolute continuity. Otherwise, $\mu(B)>0$ and $\mu(A\setminus B)=0$, whence $\nu(A\setminus B)=0$; so, $\nu(B)=\nu(A)$ and $\mu(B)=\mu(A)>0$. Thus, A is a ν-atom and $\nu_A=a\mu_A$, where $a:=0$ if $\mu(A)=0$ and $a:=\nu(A)/\mu(A)$ otherwise.

Step 3. From Steps 1 and 2 it follows that $(\nu_\pm)_{A_i}=a_{\pm;i}\,\mu_{A_i}$ for every $i\in\overline{1,m}$ and some $a_{\pm;i}\in[0,\infty)$. Therefore and because $\nu_\pm\in\Lambda_{\mathbf{f},c}$, one has $0=c-c=\int_S\mathbf{f}\,\mathrm{d}(\nu_+-\nu_-)=\sum_1^m(a_{+;i}-a_{-;i})\int_{A_i}\mathbf{f}\,\mathrm{d}\mu$. Therefore and because the vectors $\int_{A_1}\mathbf{f}\,\mathrm{d}\mu,\dots,\int_{A_m}\mathbf{f}\,\mathrm{d}\mu$ are linearly independent, $a_{+;i}-a_{-;i}=0$ and hence $(\nu_+)_{A_i}=(\nu_-)_{A_i}$ for all $i\in\overline{1,m}$, which implies that indeed $\nu_+=\nu_-$. Now the proof of Theorem 7 is complete. □

When the measurable space (S,Σ) has to do with a topology, usually one can somewhat simplify the condition in Theorems 1 and 7 of the existence of an atomic partition. More specifically, recall that by Corollary 4 under condition (2) any measure $\mu\in\operatorname{ex}\Lambda_{\mathbf{f},C}$ can be represented by a discrete measure $\tilde\mu\in\Delta_{\mathbf{f}}^{(k)}$ in the sense of (3). That was enough for applications presented in Corollary 5. Yet, it may be of interest to know under what conditions any measure in the set $\operatorname{ex}\Lambda_{\mathbf{f},C}$ is (not just represented by but) equal to a discrete measure in $\Delta_{\mathbf{f}}^{(k)}$.

To address this matter, suppose from now on to the rest of the paper that the set S is endowed with a Hausdorff topology, and let then the σ-algebra Σ contain the corresponding Borel σ-algebra $\mathcal{B}$.

Proposition 9. *Take any $\mu \in \mathrm{M}$ and $A \in \Sigma$. Then the following statements hold.*

(i) *If A is a μ-atom then $\operatorname{card}\operatorname{supp}\mu_A \leqslant 1$.*
(ii) *If $\operatorname{card}\operatorname{supp}\mu_A \leqslant 1$, $A \neq \emptyset$, and the measure μ_A is support-concentrated $\big($that is, $\mu_A(\operatorname{supp}\mu_A) = \mu(A)\big)$, then $\mu_A = a\delta_s$ for some $s \in A$ and $a \in [0, \infty)$.*
(iii) *For the measure μ_A to be support-concentrated, it is enough that the measure μ be a Radon one: $\mu(E) = \sup\{\mu(K) \colon K \subseteq E, K$ is compact$\}$ for all $E \in \Sigma$.*
(iv) *In turn, for every measure in M to be a Radon one, it is enough that $\Sigma = \mathcal{B}$ and S be a Polish space (that is, a separable completely metrizable topological space); for instance, $\mathbb{R}^d$ or, more generally, any separable Banach space is Polish.*
(v) *If $\mu_A = a\delta_s$ for some $s \in A$ and $a \in [0, \infty)$, then A is a μ-atom.*

Here, as usual, card stands for the cardinality and supp for the support. Recall that the support, $\operatorname{supp}\mu$, of a measure $\mu \in \mathrm{M}$ is the set of all points $s \in S$ such that $\mu(G) > 0$ for all open subsets G of S such that $G \ni s$; equivalently, $\operatorname{supp}\mu$ is the intersection of all closed subsets of S of "full" measure $\mu(S)$.

An immediate corollary of parts (i)–(iii) of Proposition 9 is the well-known fact that every 0,1-valued regular Borel measure on a Hausdorff space is a Dirac measure; see e.g. [1, Corollary 2.4]. As seen from [7, Example 2.3], part (iii) of Proposition 9 would be false in general if the condition that μ be a Radon measure were relaxed to it being regular; that is, if closed sets were used in place of compact sets K. For a further study of properties of support sets, see e.g. [12].

Proof of Proposition 9. For brevity, let $S_A := \operatorname{supp}\mu_A$.
(i): Suppose that A is a μ-atom, whereas S_A contains two distinct points, say s_1 and s_2. Let G_1 and G_2 be open sets in S such that $s_1 \in G_1$, $s_2 \in G_2$, and $G_1 \cap G_2 = \emptyset$. Then $\mu_A(G_1) > 0$ and $\mu(A) - \mu_A(G_1) = \mu_A(S \setminus G_1) \geqslant \mu_A(G_2) > 0$, so that $\mu_A(G_1) \in \big(0, \mu(A)\big)$, which contradicts the condition that A is a μ-atom. Thus, part (i) of Proposition 9 is proved.
(ii): Suppose that indeed $\operatorname{card} S_A \leqslant 1$, $A \neq \emptyset$, and $\mu_A(S_A) = \mu(A)$. If at that $\mu(A) = 0$ then $\mu_A = a\delta_s$ for $a = 0$ and any $s \in A$. Suppose now that $\mu(A) > 0$. Then $0 < \mu(A) = \mu_A(S_A)$, whence $S_A \neq \emptyset$, $\operatorname{card} S_A = 1$, $S_A = \{s\}$ for some $s \in S$, $\mu_A(\{s\}) = \mu_A(S_A) = \mu(A) > 0$, and hence $s \in A$. Also, $\mu_A(S \setminus \{s\}) = \mu_A(S) - \mu_A(\{s\}) = \mu_A(S) - \mu_A(S_A) = 0$, whence $\mu_A(B) = 0$ for all $B \in \Sigma$ such that $s \notin B$. On the other hand, for all $B \in \Sigma$ such that $s \in B$ one has $\mu(A) \geqslant \mu_A(B) \geqslant \mu_A(\{s\}) = \mu_A(S_A) = \mu(A)$, whence $\mu_A(B) = \mu(A)$. Thus, for any $B \in \Sigma$ one has $\mu_A(B) = 0$ if $s \notin B$ and $\mu_A(B) = \mu(A)$ if $s \in B$. That is, $\mu_A = a\delta_s$ for $a := \mu(A)$, which proves part (ii) of Proposition 9.
(iii): Note that, if μ is a Radon measure then μ_A is so too: if $B \in \Sigma$ and K is a compact subset of B, then $0 \leqslant \mu_A(B) - \mu_A(K) = \mu_A(B \setminus K) \leqslant \mu(B \setminus K) = \mu(B) - \mu(K)$. So, in the case when $\Sigma = \mathcal{B}$, part (iii) of Proposition 9 follows from the well known fact that any Radon measure on $\mathcal{B}$ is support-concentrated;

see e.g. [7, page 222] (where the terminology "μ has a strong support" is used in place of "μ is support-concentrated"). For the readers' convenience, let us present here an easy proof of the latter fact, which works whenever $\Sigma \supseteq \mathcal{B}$. By the definition of the support of a measure, for any $s \in S \setminus S_A$ there is a set G_s such that G_s is open, $s \in G_s \subseteq S$, and $\mu_A(G_s) = 0$. Take now any compact $K \subseteq S \setminus S_A$. Since $\bigcup_{s\in K} G_s \supseteq K$, there is some finite subset F of K such that $\bigcup_{s\in F} G_s \supseteq K$. So, $\mu_A(K) \leqslant \sum_{s\in F} \mu_A(G_s) = 0$. Thus, $\mu_A(K) = 0$ for all compact $K \subseteq S \setminus S_A$, and so, since μ_A is a Radon measure, $\mu_A(S \setminus S_A) = 0$ or, equivalently, $\mu_A(S_A) = \mu(A)$. This proves part (iii) of Proposition 9.
(iv): For part (iv) of the proposition, see e.g. [13, P16, page XIII].
(v): Part (v) of the proposition is trivial. □

Proposition 9 immediately yields

Corollary 10. *Suppose that S is a Polish space and $\Sigma = \mathcal{B}$. Take any $\mu \in \mathrm{M}$ and $A \in \Sigma$. Then A is a μ-atom iff $\mu_A = a\delta_s$ for some $s \in A$ and $a \in [0, \infty)$.*

Now Theorems 1 and 7 immediately imply

Corollary 11. *Suppose that S is a Polish space, $\Sigma = \mathcal{B}$, and $\Lambda = \mathrm{M}$. Then $\operatorname{ex}\Lambda_{\mathbf{f},c} = \Delta^{(k)}_{\mathbf{f}} \cap \Lambda_{\mathbf{f},c}$ for all $c \in \mathbb{R}^k$.*

This latter result should be enough for most applications. Yet, one may want to compare it with equality (7) in Corollary 5, which holds without any topological assumptions.

In conclusion, let us briefly discuss existing literature. The present paper was mainly motivated by the work of Winkler [15], especially by the principal result there:

Theorem 12. **([15, Theorem 2.1]).** *Suppose that the set Π of all probability measures in Λ is a Choquet-simplex and $\operatorname{ex}\Pi \subseteq \Delta^{(1)}$. Then the following conclusions hold.*

(a) If $C = (-\infty, c_1] \times \cdots \times (-\infty, c_k]$ for some $(c_1, \dots, c_k) \in \mathbb{R}^k$, then $\operatorname{ex}(\Pi \cap \Lambda_{\mathbf{f},C}) \subseteq \Delta^{(1+k)}_{(1,\mathbf{f})}$, where $(1, \mathbf{f}) := (1, f_1, \dots, f_k)$.

(b) For any $c \in \mathbb{R}^k$, one has $\operatorname{ex}(\Pi \cap \Lambda_{\mathbf{f},c}) = \Pi \cap \Lambda_{\mathbf{f},c} \cap \Delta^{(1+k)}_{(1,\mathbf{f})}$.

By a remark in [8, Section 9], the meaning of the condition that Π is a Choquet-simplex can be expressed as follows: Π is a convex set of probability measures such that for the cone $\Gamma := [0, \infty)\,\Pi$ generated by Π and any μ_1 and μ_2 in Γ there is some $\nu \in \Gamma$ such that $\mu_1 - \nu$ and $\mu_2 - \nu$ are in Γ and for any $\tilde\nu \in \Gamma$ such that $\mu_1 - \tilde\nu$ and $\mu_2 - \tilde\nu$ are in Γ one has $\nu - \tilde\nu \in \Gamma$.

To an extent, our Theorems 1 and 7 (cf. also Proposition 9 and Corollaries 10 and 11) correspond to parts (a) and (b), respectively, of Theorem 12. One may note that the second condition, $\operatorname{ex}\Pi \subseteq \Delta^{(1)}$, in Theorem 12 is of the same form (corresponding to $k = 0$ affine restrictions on the measure in addition to the requirement that it be a probability measure) as the conclusion $\operatorname{ex}(\Pi \cap \Lambda_{\mathbf{f},C}) \subseteq \Delta^{(1+k)}_{(1,\mathbf{f})}$ in part (a) of the theorem, where k additional affine restrictions on the measure are present. That is in distinction with Theorems 1 and 7 of this paper (cf. also Corollaries 2 and 8). Moreover, the set C in Theorem 1 may be any subset of $\mathbb{R}^k$ and not necessarily of the orthant form assumed in part (a) of Theorem 12.

Also in distinction with Theorem 12, the measures in this paper are not required to be probability ones; for instance, the set Λ may coincide with the set M of all measures on Σ.

Condition (2) in Theorem 1 appears to be easier to check than the conditions in Theorem 12 that Π be a Choquet-simplex and $\operatorname{ex}\Pi \subseteq \Delta^{(1)}$. In particular, (2) is trivially satisfied in the just mentioned case when $\Lambda = \mathrm{M}$, which appears to be of main interest in applications; condition (2) holds as well in the examples (a), (b), (c) in [15, page 585] – if one replaces there the sets P of probability measures by the corresponding cones $[0, \infty)\, P$. At this point one may also recall that condition (2) (or, in fact, any other special condition) is not needed or used to establish Theorem 7; however, the latter theorem appears not nearly as useful as Theorem 1 in such applications as Corollary 5.

One might also want to compare condition (4) that Λ contain the set $\Delta^{(k)}_{\mathbf{f}}$ (or $\Delta^{(k)}$) of discrete measures with the condition in Theorem 12 that the set $\operatorname{ex}\Pi$ be contained in $\Delta^{(1)}$, even though these two conditions go in opposite directions. It appears that (4) is generally easier to satisfy and check. Note also that condition (4) is used in this paper only to obtain the equality (7).

One can construct an example when Π is a Choquet-simplex with $\operatorname{ex}\Pi \subseteq \Delta^{(1)}$ while condition (2) fails to hold for the corresponding cone $\Lambda = [0, \infty)\Pi$. For instance, one may let Π be the set of all mixtures of the discrete probability distributions on $\mathbb{R}$ and the absolutely continuous probability distributions on $\mathbb{R}$ with everywhere continuous densities.

It is also easy to give an example when conditions (2) and (4) both hold while $\operatorname{ex}\Pi \not\subseteq \Delta^{(1)}$. For instance, suppose that S is any uncountable set, Σ is the σ-algebra over S generated by all countable subsets of S, and $\Lambda = \mathrm{M}$, so that Π is the set of all probability measures on Σ. Then (2) and (4) are both trivially satisfied. On the other hand, consider the 0,1 measure (say π) on Σ that takes the value 0 precisely on all countable sets in Σ. Then $\pi \in \operatorname{ex}\Pi$, by Corollary 8. Yet, $S \setminus \{s\} \in \Sigma$ and $\pi(S \setminus \{s\}) = 1$ for any $s \in S$ and hence $\pi \notin \Delta^{(1)}$.

However, these two examples may seem rather artificial. It appears that the results given here will be about as effective as those in [15] in most applications.

The methods presented in this paper seem different from and more elementary than those of [15]. In particular, the present paper is self-contained, except for quoting [15] and [13] concerning part (v) of Proposition 6 and part (iv) of Proposition 9, respectively.

As pointed out in [15], the results there generalize ones in Richter [11] (for $S \subseteq \mathbb{R}$ and piecewise-continuous f_j's), Mulholland and Rogers [6] (for $S = \mathbb{R}$), and Karr [4] (for compact metric spaces S).

An equality similar to (7) was given by Hoeffding [2] for $S = \mathbb{R}$; in fact, the result there holds for product measures on $\mathbb{R}^n$. When S is an interval in $\mathbb{R}$ and the functions $f_1, \dots, f_k, g$ form a Tchebycheff system, such results can be considerably improved: in that case, the support of extremal measures consists of only about $k/2$, rather than k, points; see e.g. [3, 5, 9].

References

[1] W. Adamski. Complete spaces and zero-one measures. *Manuscripta Math.*, 18(4):343–352, 1976.

[2] W. Hoeffding. The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.*, 26:268–275, 1955.

[3] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics*. Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[4] A. F. Karr. Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.*, 8(1):74–85, 1983.

[5] M. G. Kreĭn and A. A. Nudel′man. *The Markov moment problem and extremal problems*. American Mathematical Society, Providence, R.I., 1977. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.

[6] H. P. Mulholland and C. A. Rogers. Representation theorems for distribution functions. *Proc. London Math. Soc. (3)*, 8:177–223, 1958.

[7] S. Okada. Supports of Borel measures. *J. Austral. Math. Soc. Ser. A*, 27(2):221–231, 1979.

[8] R. R. Phelps. *Lectures on Choquet's theorem*. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1966.

[9] I. Pinelis. Tchebycheff systems and extremal problems for generalized moments: a brief survey, preprint, http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1.

[10] I. F. Pinelis and S. A. Utev. Sharp exponential estimates for sums of independent random variables. *Theory Probab. Appl.*, 34(2):340–346, 1989.

[11] H. Richter. Parameterfreie Abschätzung und Realisierung von Erwartungswerten. *Bl. Deutsch. Ges. Versicherungsmath.*, 3:147–162, 1957.

[12] W. Seidel. Supports of Borel measures. *Fund. Math.*, 133(1):67–80, 1989.

[13] F. Topsøe. *Topology and measure*. Lecture Notes in Mathematics, Vol. 133. Springer-Verlag, Berlin, 1970.

[14] H. von Weizsäcker and G. Winkler. Integral representation in the set of solutions of a generalized moment problem. *Math. Ann.*, 246(1):23–32, 1979/80.

[15] G. Winkler. Extreme points of moment sets. *Math. Oper. Res.*, 13(4):581–587, 1988.