Open access • Journal Article • DOI:10.1142/S0218202500000513

Maximum entropy for reduced moment problems — Source link

Michael Junk

Institutions: Kaiserslautern University of Technology

Published on: 01 Oct 2000 - Mathematical Models and Methods in Applied Sciences (World Scientific Publishing Company)

Topics: Maximum entropy probability distribution, Maximum entropy thermodynamics, Joint quantum entropy, Min entropy and Principle of maximum entropy

Mathematical Models and Methods in Applied Sciences

MAXIMUM ENTROPY FOR REDUCED MOMENT PROBLEMS

MICHAEL JUNK
Fachbereich Mathematik, Universität Kaiserslautern
Postfach 3049, 67653 Kaiserslautern, Germany

The existence of maximum entropy solutions for a wide class of reduced moment problems on arbitrary open subsets of $\mathbb{R}^d$ is considered. In particular, new results for the case of unbounded domains are obtained. A precise condition is presented under which solvability of the moment problem implies existence of a maximum entropy solution.

Keywords: maximum entropy; moment realizability; general multidimensional moment problem

1. Introduction

A variety of technical and physical problems leads to so called *reduced moment problems* where one tries to reconstruct a function f from a finite number of weighted integral averages

$$\int_\Omega a_i(x) f(x)\, dx = \rho_i, \qquad i = 1, \ldots, N. \tag{1.1}$$

Apart from (1.1), typical additional constraints on f are positivity or boundedness.

Obviously, relations like (1.1) occur whenever a process involves measurements of integral transforms like the Radon transform in image reconstruction, or the Fourier transform in spectral measurements. Other applications include the reconstruction of probability densities from a certain number of moments. For example, the problem to determine a density $f \geq 0$ satisfying $\int_{\mathbb{R}} f(x)\, dx = 1 = \rho_0$, having mean $\int_{\mathbb{R}} x f(x)\, dx = \rho_1$ and variance $\int_{\mathbb{R}} (x-\rho_1)^2 f(x)\, dx = \rho_2$ falls into the above class of moment problems. Without further constraints, the reconstruction of f from ρ_0, ρ_1, ρ_2 is not unique and thus (1.1) can be viewed as an ill–posed inverse problem. A common way to regularize the problem is the *maximum entropy* approach in which a solution of (1.1) is singled out as maximizer of the strictly concave entropy functional

$$H(f) = -\int_\Omega f \ln f\, dx.$$

In the example of reconstructing a probability density, the corresponding regularized

solution is just the Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi\rho_2}} \exp\left(-\frac{(x-\rho_1)^2}{2\rho_2}\right).$$

It turns out, however, that if the given data ρ_0, ρ_1, ρ_2 is extended by two more moments which are related to the skewness and the excess of f

$$\rho_3 = \int_{\mathbb{R}} (x-\rho_1)^3 f(x)\,dx, \qquad \text{and} \qquad \rho_4 = \int_{\mathbb{R}} (x-\rho_1)^4 f(x)\,dx,$$

the maximum entropy regularization may fail to give a solution of (1.1).

In the present article it is shown that an important ingredient for this behavior is the unboundedness of the domain of integration. In fact, for a quite general class of moment problems the set of data ρ where the maximum entropy approach fails is characterized (see Section 3 for assumptions and Section 4 for results). The developed characterization is useful to decide about solvability of a particular moment problem before applying any numerical solution algorithm.

While Sections 5 to 9 contain proofs of the results, a more intuitive argument, explaining the basic problem with unbounded domains of integration, is presented in Section 2. The particular moment problem considered there is a generalization of our introductory example on the reconstruction of probability densities. Other specific examples are listed in Section 10, including maximum entropy problems arising in connection with Levermore's moment systems which are extensions to the system of Euler equations of gas dynamics. The investigation of the latter example was the main motivation for this work.

2. A classical moment problem

Before treating the most general case, we want to present the basic ideas of the maximum entropy approach as well as the complications which can arise from its application to moment problems on unbounded domains. To this end, we consider the so called *Hamburger moment problem* where one wants to find a nonnegative function $f : \mathbb{R} \mapsto \mathbb{R}$ which has $2n+1$ given numbers $\rho_0, \ldots, \rho_{2n}$ as moments

$$\int_{\mathbb{R}} x^i f(x)\,dx = \rho_i \qquad i = 0, \ldots, 2n.$$

More precisely, if we exclude the trivial case $\rho_0 = 0$ which implies $f = 0$, we want to find a function in

$$D := \left\{ f \geq 0 \,:\, (1+x^{2n}) f \in \mathbb{L}^1(\mathbb{R}),\, f \not\equiv 0 \right\} \tag{2.2}$$

such that $\mu(f) = \rho$ with

$$\mu(f) := \int_{\mathbb{R}} a(x) f(x)\,dx, \qquad a(x) = (1, \ldots, x^{2n})^T.$$

Obviously, the problem is solvable only under certain assumptions on the prescribed moment vector ρ. A straight forward requirement is the positivity of all even moments. The exact conditions for solvability depend, of course, on the underlying set ($\Omega = \mathbb{R}$) as well as the moment functions $1, x, \ldots, x^{2n}$ (see for example [1,2] or the appendix). At this point, we just require $\rho \in \mu(D)$ so that the moment problem is solvable by definition. However, the solution is not unique and one often wants to find a particular solution of the problem which is compatible with the given information ρ but which contains as little additional information as possible. Such a solution is naturally obtained in the maximum entropy framework.[9,10,11] Among all solutions f of the moment problem $\mu(f) = \rho$ we choose the one with maximal entropy

$$H(f) = -\int_{\mathbb{R}} f \ln f \, dx. \tag{2.3}$$

This special solution of the reduced Hamburger moment problem will be called the *maximum entropy solution*. The following formal investigation quickly leads to the structure of the optimal density. The necessary condition on the first variation at the maximum of the constrained optimization is

$$0 = \delta(H(f) + \beta \cdot (\mu(f) - \rho)) = -\ln f - 1 + \sum_{i=0}^{2n} \beta_i x^i$$

with Lagrange multipliers β. Solving for f yields

$$f(x) = \exp\left(\sum_{i=0}^{2n} \lambda_i x^i\right) = \exp_\lambda(x)$$

(the Lagrange multipliers have been renamed λ after combining the constant -1 with β_0). If it is possible to determine the free parameters λ from the constraint $\mu(f) = \rho$ then f is the maximum entropy solution (which is unique in D due to strict concavity of H). Therefore, the main concern of this article is to answer the question whether the moment constraints can be satisfied with exponential densities. The required integrability of $\exp_\lambda$ restricts the parameter vector λ to the set

$$\Lambda := \left\{ \lambda \in \mathbb{R}^{2n+1} \,:\, \exp_\lambda \in \mathbb{L}^1(\mathbb{R}) \right\}.$$

For $\lambda \in \Lambda$, on the other hand, moments of $\exp_\lambda$ to any order are well defined so that the collection of integrable exponential densities

$$E := \{ \exp_\lambda \,:\, \lambda \in \Lambda \}$$

is a subset of D defined in (2.2). We will show that $\rho \in \mu(E)$ is a necessary and sufficient condition for the existence of a maximum entropy solution. In general, however, $\mu(E)$ is strictly contained in $\mu(D)$. Consequently, there exist admissible moment vectors $\rho \in \mu(D)$ for which the moment problem is solvable but the maximum entropy problem has no solution. The reason for this behavior is essentially

the following (a detailed analysis will be given for more general cases): If $(f_m)_{m\in\mathbb{N}}$ is a maximizing sequence for the entropy functional (2.3) such that $\mu(f_m) = \rho$ for all m, then f_m converges in $\mathbb{L}^1(\mathbb{R})$ (due to an estimate of the $\mathbb{L}^1$ norm in terms of the entropy). The uniform $\mathbb{L}^1$ bound on $x^{2n} f_m(x)$ then gives convergence of all lower moments. For the highest moment $\int x^{2n} f_m(x)\, dx$, Fatou's Lemma guarantees boundedness, however its value might drop in the limit so that the constraint $\mu(\lim_{m\to\infty} f_m) = \rho$ is not satisfied.

To show that this situation can occur, we focus our attention on those integrable exponential densities $\exp_{\hat\lambda}$ which satisfy $\hat\lambda_{2n} = 0$ (and thus also $\hat\lambda_{2n-1} = 0$). Since the two highest Lagrange multiplier vanish, it is clear that $\exp_{\hat\lambda}$ is an optimal density already under a reduced number of constraints. More precisely, with

$$\hat\rho_i = \int_{\mathbb{R}} x^i \exp_{\hat\lambda}(x)\, dx,$$

$\exp_{\hat\lambda}$ is the maximizer of H under the constraints that the moments are $\hat\rho_i$ for $i = 0, \dots, 2n-2$. In particular, any function f which satisfies

$$\mu_i(f) = \hat\rho_i, \quad i = 0, \dots, 2n-1, \qquad \mu_{2n}(f) > \hat\rho_{2n}$$

has less entropy than $\exp_{\hat\lambda}$. The trick to show that certain maximum entropy problems are not solvable now relies on a perturbation argument: We slightly perturb $\exp_{\hat\lambda}$ by adding mass at large x. The amount of mass is chosen in such a way that it contributes essentially only to the highest moment (for $|x| \gg 1$, the amplification of the mass through x^{2n} is much stronger than for the lower moments). By suitable additional perturbations (at small $|x|$) it is possible to construct a function f which satisfies

$$\mu_i(f) = \hat\rho_i, \quad i = 0, \dots, 2n-1, \qquad \mu_{2n}(f) = \hat\rho_{2n} + c$$

where $c > 0$ is some given number. Repeating this process and moving the perturbation at large x to infinity, we end up with some sequence (f_m) which converges to $\exp_{\hat\lambda}$ and whose highest moment drops in the limit (by $c > 0$). For the entropy $H(f_m)$, small perturbations at large x have no considerable contribution so that also $\lim_{m\to\infty} H(f_m) = H(\exp_{\hat\lambda})$. Due to our general observation, however, $H(\exp_{\hat\lambda})$ is an upper bound for the entropies of all functions which have the same lower moments as $\exp_{\hat\lambda}$ but whose highest moment exceeds $\hat\rho_{2n}$. Consequently, (f_m) is really an entropy maximizing sequence whose highest moment drops in the limit.

Obviously, this behavior is closely related to the unboundedness of the underlying domain $\Omega = \mathbb{R}$. Only in an unbounded situation, a contribution to the highest moment can escape to infinity in the limit.

In the following, our detailed considerations will not be restricted to the case of Hamburger moment problems. We just require some basic properties (see Section 3) which are also satisfied in many other situations.

For some special cases, our results are already presented elsewhere. In particular, moment problems on bounded intervals in $\mathbb{R}$ have been considered.[16] In a series of papers[8,19,20] the Hamburger and Stieltjes moment problems are considered (the

latter differs from the Hamburger case presented above in the domain $\Omega = [0, \infty)$ and possibly in the highest moment function which need not be of even order). The authors use a technique which is different from the one presented here and which relies on special properties of the moment problems under consideration. A more general approach based on probability measures on arbitrary measurable spaces is also possible[5] (applying, for example, to general bounded sets). In the work by Lewis[15] the problem is analyzed on compact Hausdorff spaces with general entropy functionals and continuous moment functions.

Our main emphasis is therefore on the case of unbounded Ω. Nevertheless, bounded domains are reconsidered in our general setting since they serve as important tool for the unbounded situation. It turns out, however, that the criterion for solvability of the maximum entropy problem is not so much boundedness of the underlying domain. It is rather related to the geometry of the set Λ of possible Lagrange multipliers (see Section 4 for a list of results).

3. Assumptions

First, the domain of interest can be any non empty, open subset $\Omega \subset \mathbb{R}^d$ with $d \in \mathbb{N}$. The moment functions $a_0 \equiv 1, a_1, \ldots, a_N$ should be measurable on Ω and satisfy the condition

$$\mathrm{vol}\{\, x \in \Omega \,:\, \lambda \cdot a(x) = 0 \,\} = 0 \qquad \forall \lambda \in \mathbb{R}^{N+1} \backslash \{0\}.$$

This implies in particular, that $\{a_0, \ldots, a_N\}$ are linearly independent functions (such a system of functions is called *pseudo–Haar*[15]). In addition, we assume that all a_i are bounded on bounded subsets of Ω. In cases where Ω is unbounded, we need the additional requirement $a_N \geq 0$ and

$$\frac{|a_i(x)|}{1 + a_N(x)} \xrightarrow[|x| \to \infty]{} 0 \quad i = 0, \ldots, N-1.$$

In accordance with the initial example, we define the basic space

$$D := \left\{\, f \geq 0 \,:\, (1 + a_N) f \in \mathbb{L}^1(\Omega),\, f \not\equiv 0 \,\right\}.$$

On D, the calculation of a–moments is possible since each a_i can be estimated by $1 + a_N$. We denote

$$\mu(f) := \int_\Omega a(x) f(x)\, dx, \qquad a(x) = (1, a_1(x), \ldots, a_N(x))^T.$$

If the dependence on Ω is important, we also write $\mu(f; \Omega)$. (Similarly, we will use the obvious notation $D(\Omega)$). The elements of $\mu(D)$ will be called admissible moment vectors because for $\rho \in \mu(D)$ there exists, by definition, at least one solution of the problem $\mu(f) = \rho$ with $f \in D$. Again, the exponential densities

$$\exp_\lambda(x) := \exp\left(\sum_{i=0}^{N} \lambda_i a_i(x) \right) = \exp(\lambda \cdot a(x))$$

will play an important role. The admissible set of parameters is called

$$\Lambda := \left\{ \lambda \in \mathbb{R}^{N+1} \,:\, \exp_\lambda \in D(\Omega) \right\}$$

and the corresponding densities are collected in

$$E := \{ \exp_\lambda \,:\, \lambda \in \Lambda \}.$$

Sometimes, it will be convenient to work with normalized densities

$$f^*(x) := \frac{f(x)}{\int_\Omega f\, dy} = \frac{f(x)}{\mu_0(f)}, \qquad f \in D.$$

Obviously, $D^* \subset D$ and also $E^* \subset E$ since $a_0 \equiv 1$ such that

$$(\exp_\lambda)^* = \exp_{\lambda^*}, \qquad \text{with} \qquad \lambda^* = \lambda - \ln(\mu_0(\exp_\lambda))e_0.$$

Here, e_0 is the first unit vector in $\mathbb{R}^{N+1}$. The parameters λ^* of normalized exponential densities will be collected in $\Lambda^* \subset \Lambda$.

4. The result

Given an admissible moment vector $\rho \in \mu(D)$ (so that the moment problem has at least one solution $f \in D$), we want to answer the question whether we can find a special solution of the moment problem which is optimal with respect to the entropy

$$H(f) = -\int_\Omega f \ln f\, dx.$$

The solvability turns out to be closely related to the structure of the set

$$\Lambda = \left\{ \lambda \in \mathbb{R}^{N+1} \,:\, \exp_\lambda \in D \right\}.$$

- If $\Lambda = \emptyset$, the problem has no solution (Corollary 2).
- If $\Lambda \neq \emptyset$ but $\Lambda \cap \partial\Lambda = \emptyset$, the problem is always solvable (Theorem 2).
- If $\Lambda \cap \partial\Lambda \neq \emptyset$, there are always some admissible vectors for which the problem has no solution (Theorem 4).

Whenever $\Lambda \neq \emptyset$, entropy maximizing sequences always converge to exponential densities $\exp_\lambda$. In the case $\Lambda \cap \partial\Lambda \neq \emptyset$, however, it is possible that the limit density does not satisfy the moment constraint.

The moment constraints for which this can happen are easily characterized:

- Pick any $\lambda \in \Lambda \cap \partial\Lambda$.
- Calculate the moment vector $\rho_\lambda = \mu(\exp_\lambda)$.
- Add any positive number to the highest component $\rho := \rho_\lambda + \epsilon e_N$, $\epsilon > 0$.

Then, ρ is an admissible vector (i.e., there exists a positive density f such that $\mu(f) = \rho$) but the maximum entropy problem with constraint ρ has no solution.

In Section 5 we give some preparatory remarks for the proofs of the above results (Sections 6 to 9). Finally, in Section 10, we show how standard results easily follow from our general considerations.

5. General remarks

Instead of working with the entropy functional $H(f) = -\int_\Omega f \ln f \, dx$ directly, we consider the so called relative entropy

$$H(f, \hat{f}) := \int_\Omega f \ln \frac{f}{\hat{f}} \, dx.$$

In our setting, it is very convenient to choose a fixed normalized exponential density $\hat{f} = \exp_{\hat{\lambda}}$ with $\hat{\lambda} \in \Lambda^*$ as reference density. (Here we assume that $\Lambda \neq \emptyset$. Nevertheless, the case of empty Λ will also be treated.) Then

$$H(f, \hat{f}) = -\hat{\lambda} \cdot \mu(f) - H(f). \tag{5.4}$$

Under the constraint $\mu(f) = \rho$, maximizing $H(f)$ is therefore equivalent to minimizing $H(f, \hat{f})$. Our second remark concerns the effect of normalizing the density f. A simple calculation shows that

$$H(f, \hat{f}) = \mu_0(f) H(f^*, \hat{f}) + \mu_0(f) \ln \mu_0(f)$$

and it is easy to prove

Lemma 1 *f minimizes relative entropy under the constraint $\mu = \rho$ if and only if f^* is a minimizer under the condition $\mu = \rho/\rho_0$.*

Due to Lemma 1 it is possible to characterize those moment vectors in $\mu(D)$ for which a maximum entropy solutions exist by restricting the considerations to the case $\rho_0 = 1$ (i.e. to probability densities).

A simple but powerful argument concerns the splitting of the entropy which is induced by a splitting of the domain Ω. If Ω_1 is an open subset of Ω we denote

$$H(f, \hat{f}; \Omega_1) := \int_{\Omega_1} f \ln \frac{f}{\hat{f}} \, dx$$

so that $H(f, \hat{f}) = H(f, \hat{f}; \Omega_1) + H(f, \hat{f}; \Omega_1^c)$. Now, if f is an optimal density with respect to the constraint $\rho = \mu(f)$ then $f|_{\Omega_1}$ is necessarily an optimal density with respect to the constraint $\rho^{(1)} = \mu(f; \Omega_1)$. Otherwise, one could replace f on Ω_1 by another density with less relative entropy but the same moments $\rho^{(1)}$ thus improving $H(f, \hat{f})$ in contradiction to the optimality of f. This argument will be useful to extend results known for bounded domains Ω to unbounded ones.

For our arguments we will also need some general results on maximum entropy problems which can be found for example in the book of Ihara.[9] For convenience, we list these results here.

Lemma 2 *Let $f \in D$. Then*

- $f \ln \frac{f}{\hat{f}} \geq -\frac{1}{e}\hat{f}$,
- $H(f, \hat{f})$ *is strictly convex in* f,
- *if* $f \in D^*$ *then* $H(f, \hat{f}) \geq 0$ *with equality only in the case* $f = \hat{f}$,
- *if* $f_n \to f$ *in* $\mathbb{L}^1(\Omega)$ *and* $(f_n)_{n\in\mathbb{N}} \subset D^*$ *then* $H(f, \hat{f}) \leq \liminf_{n\to\infty} H(f_n, \hat{f})$,
- *if* $F \subset D^*$ *is convex,* $(f_n)_{n\in\mathbb{N}} \subset F$ *and* $H(f_n, \hat{f}) \to \inf_{f\in F} H(f, \hat{f}) < \infty$ *then* f_n *converges in* $\mathbb{L}^1(\Omega)$.

Proof. Using the estimate $x \ln x \geq -\frac{1}{e}$ we get immediately

$$f \ln \frac{f}{\hat{f}} = \hat{f}\left(\frac{f}{\hat{f}} \ln \frac{f}{\hat{f}}\right) \geq -\frac{1}{e}\hat{f}.$$

The strict convexity of $H(f, \hat{f})$ in the first argument follows directly from the corresponding property of $x \mapsto x \ln \frac{x}{y}$. The remaining results can be taken from reference[9] (Theorems 1.4.1, 1.5.5 and Theorem 3.1.1) requiring probability densities. □

Since $f \ln(f/\hat{f}) \geq -\frac{1}{e}\hat{f}$, the negative part of $f \ln(f/\hat{f})$ is integrable and thus relative entropy is well defined. However, $H(f, \hat{f}) = +\infty$ is still possible. Nevertheless, for any $\rho \in \mu(D)$ we can find a particular solution $f \in D$ of the moment problem $\mu(f) = \rho$ which is bounded and compactly supported (see Theorem A.1 in the appendix). Consequently, $-\frac{1}{e} \leq H(f, \hat{f}) < \infty$ and we can always assume the existence of minimizing sequences with finite relative entropy.

The second result we need for our argument concerns the solvability of the entropy optimization problem if the prescribed moment vector ρ already belongs to an exponential density.

Theorem 1 *If $\rho \in \mu(E)$ then the unique solution of the problem to minimize relative entropy $H(f, \hat{f})$ under the constraint $f \in D$ and $\mu(f) = \rho$ is given by* $\exp_\lambda$ *where λ is uniquely determined through the relation $\mu(\exp_\lambda) = \rho$.*

Proof. According to Lemma 1 we can restrict ourselves to the case $\rho_0 = 1$. Theorem 3.1.4 in reference[9] then proves the result. □

All the remaining investigations aim at the relation between $\mu(E)$ and $\mu(D)$. If $\mu(E) = \mu(D)$ then, according to Theorem 1, the maximum entropy problem admits a unique solution for any admissible moment vector. In some cases, $\mu(E)$ is a proper subset of $\mu(D)$. Then, the maximum entropy problem will not be solvable for $\rho \in \mu(D)\backslash\mu(E)$.

6. Structure of Λ

The solvability of the maximum entropy problem is closely related to the structure of the set Λ or more precisely to the structure of its boundary. Throughout this

section, we assume that $\Lambda \neq \emptyset$. The following Lemma clarifies the geometry of the interior $\mathrm{int}(\Lambda)$ which is just an open half space in $\mathbb{R}^{N+1}$.

Lemma 3 *Let $\lambda_N^b := \sup\{\, \lambda_N \,:\, \lambda \in \Lambda \,\}$. Then the interior of Λ has the form*

$$\mathrm{int}(\Lambda) = \{\, \lambda \in \mathbb{R}^{N+1} \,:\, \lambda_N < \lambda_N^b \,\}.$$

In particular, if Ω is bounded then $\lambda_N^b = \infty$ and thus $\Lambda = \mathbb{R}^{N+1}$. If $K \subset \mathrm{int}(\Lambda)$ is compact and $P : \mathbb{R} \mapsto \mathbb{R}$ is a polynomial then there exists $\bar{\lambda} \in \mathrm{int}(\Lambda)$ such that for all $\lambda \in K$

$$|P(a_N)| \exp(\lambda \cdot a) \leq \exp(\bar{\lambda} \cdot a).$$

Proof. First, let us consider the case of bounded Ω. Since, by assumption, all a_i are bounded, the same holds for $\exp(\lambda \cdot a)$, where λ is an arbitrary vector in $\mathbb{R}^{N+1}$. Thus, $(1 + a_N)\exp_\lambda \in \mathbb{L}^1(\Omega)$ for all $\lambda \in \mathbb{R}^{N+1}$. For the second statement, we just set $\bar{\lambda} = \bar{\lambda}_0 e_0$ with

$$\exp(\bar{\lambda}_0) := \max\{\, |P(a_N(x))| \exp(\lambda \cdot a(x)) \,:\, x \in \Omega,\ \lambda \in K \,\}.$$

If Ω is unbounded, we denote $I := \{\, \lambda \in \mathbb{R}^{N+1} \,:\, \lambda_N < \Lambda_N^b \,\}$ and first show $I \subset \Lambda$. To every $\lambda \in I$ there exists $\bar{\lambda} \in \Lambda$ such that $\lambda_N < \bar{\lambda}_N$. Since the growth of a_N dominates over all a_i, we get

$$\frac{\bar{\lambda} \cdot a(x) - \lambda \cdot a(x)}{1 + a_N(x)} \xrightarrow[|x| \to \infty]{} \bar{\lambda}_N - \lambda_N > 0.$$

In particular, there exists some $R > 0$ such that $\bar{\lambda} \cdot a(x) > \lambda \cdot a(x)$ for all $|x| > R$. The integrability of $(1 + a_N)\exp_\lambda$ then follows from the one of $(1 + a_N)\exp_{\bar{\lambda}}$ since on $|x| \leq R$ the function $(1 + a_N)\exp_\lambda$ is bounded. By definition of λ_N^b, every $\lambda \in \Lambda$ satisfies $\lambda_N \leq \lambda_N^b$ so that $\mathrm{cl}(\Lambda) \subset \mathrm{cl}(I)$ which completes the proof of the first statement. For a compact $K \subset \mathrm{int}(\Lambda)$ we define $\beta := \max\{\, \lambda_N \,:\, \lambda \in K\}$ so that the difference $4\epsilon := \lambda_N^b - \beta$ is positive. Then, for any $\lambda \in K$

$$\frac{(\beta + 2\epsilon)a_N(x) - \lambda \cdot a(x)}{1 + a_N(x)} \xrightarrow[|x| \to \infty]{} 2\epsilon + \beta - \lambda_N \geq 2\epsilon > 0$$

and hence $(\beta + 2\epsilon)a_N(x) > \lambda \cdot a(x)$ for $|x| > R$ and R large enough. Consequently,

$$|P(a_N)| \exp(\lambda \cdot a) \leq |P(a_N)| \exp(-\epsilon a_N) \exp((\beta + 3\epsilon)a_N), \qquad |x| > R.$$

Now, $|P(a_N)| \exp(-\epsilon a_N)$ is uniformly bounded and $\exp((\beta + 3\epsilon)a_N)$ is integrable because $\beta + 3\epsilon < \lambda_N^b$ by construction. Setting $\bar{\lambda} := \bar{\lambda}_0 e_0 + (\beta + 3\epsilon)e_N$ with $\bar{\lambda}_0$ large enough, we have proved the lemma. □

Obviously, $\mathrm{int}(\Lambda)$ is a convex set but convexity of Λ can also be shown directly using Hölder inequality. Indeed, for any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$ and $\lambda_1, \lambda_2 \in \Lambda$ we find

$$(1 + a_N)\exp_{\frac{1}{q}\lambda_1 + \frac{1}{p}\lambda_2} = \big((1 + a_N)\exp_{\lambda_1}\big)^{\frac{1}{q}} \big((1 + a_N)\exp_{\lambda_2}\big)^{\frac{1}{p}} \in \mathbb{L}^1(\Omega).$$

The second statement in Lemma 3 is needed to show differentiability properties of the mapping $\lambda \mapsto \mu_0(\exp_\lambda)$. Besides this, the following result contains other important properties.

Lemma 4 *The function $\lambda \mapsto \mu_0(\exp_\lambda)$ is strictly convex on the convex set $\Lambda \subset \mathbb{R}^{N+1}$. On $\mathrm{int}(\Lambda)$ it is infinitely often differentiable and on $\Lambda \cap \partial\Lambda$ directional derivatives can be defined for admissible directions (a direction $\xi \in S^N$ is called admissible in $\lambda \in \Lambda$ if for some $\epsilon > 0$ also $\lambda + \epsilon\xi \in \Lambda$). More precisely, the function $\theta \mapsto \mu_0(\exp_{\theta\lambda_2+(1-\theta)\lambda_1})$ is in $C^1([0,1])$ for any pair $\lambda_1, \lambda_2 \in \Lambda$ and the derivative $(\lambda_2 - \lambda_1) \cdot \mu(\exp_{\theta\lambda_2+(1-\theta)\lambda_1})$ is strictly increasing in θ. If $\Lambda \cap \partial\Lambda = \emptyset$ then $\xi \cdot \mu(\exp_{\lambda^{(n)}}) \to +\infty$ whenever $\xi \in \mathbb{R}^{N+1}$ satisfies $\xi_N > 0$ and $(\lambda^{(n)})_{n\in\mathbb{N}}$ is a sequence which converges to a point on $\partial\Lambda$.*

Proof. If $\lambda \in \mathrm{int}(\Lambda)$, Lemma 3 yields integrable majorants $\exp_{\bar\lambda} \in \mathbb{L}^1(\Omega)$ for derivatives of $\exp_\lambda$ which are locally uniform in λ. Thus $\lambda \mapsto \mu_0(\exp_\lambda) \in C^\infty(\mathrm{int}(\Lambda))$. Its first derivative is just

$$\nabla_\lambda \mu_0(\exp_\lambda) = \mu(\exp_\lambda)$$

and for the Hessian, we find

$$H_\lambda[\mu_0(\exp_\lambda)] = \int_\Omega a \otimes a \exp_\lambda \, dx.$$

This matrix is positive definite (and thus $\lambda \mapsto \mu_0(\exp_\lambda)$ is strictly convex) because for any vector $0 \neq \beta \in \mathbb{R}^{N+1}$

$$\beta^T H_\lambda[\mu_0(\exp_\lambda)]\beta = \int_\Omega (\beta \cdot a)^2 \exp_\lambda \, dx > 0.$$

Here we have used the fact that $\{a_0, \ldots, a_N\}$ is pseudo–Haar. Indeed, $\beta \neq 0$ implies $\mathrm{vol}\{\, x \in \Omega : (\beta \cdot a(x))^2 > 0 \,\} > 0$. On a line segment $\lambda(\theta) = \theta\lambda_2 + (1-\theta)\lambda_1$ where $\theta \in [0,1]$ and $\lambda_1, \lambda_2 \in \mathrm{int}(\Lambda)$ we consider the function $g(\theta;\Omega) := \mu_0(\exp_{\lambda(\theta)};\Omega)$. Then $g(\cdot;\Omega) \in C^\infty([0,1])$ with

$$g'(\theta;\Omega) = (\lambda_2 - \lambda_1) \cdot \mu(\exp_{\lambda(\theta)};\Omega),$$

$$g''(\theta;\Omega) = \int_\Omega ((\lambda_2 - \lambda_1) \cdot a(x))^2 \exp_{\lambda(s)}(x)\, dx.$$

For $\lambda_1, \lambda_2 \in \Lambda$ (i.e. possibly on the boundary), we extend these relations through a limiting process over bounded sets $\Omega_R := \{x \in \Omega : |x| < R\}$. To investigate the approximation properties we focus on the difference $\mu(\exp_{\lambda(\theta)}) - \mu(\exp_{\lambda(\theta)};\Omega_R) = \mu(\exp_{\lambda(\theta)};\Omega_R^c)$. Due to our assumptions on a_i we can find constants $C_i > 0$ such that $|a_i| \leq C_i(1 + a_N)$. Then, using Hölder inequality, we get

$$\begin{aligned}
|\mu_i(\exp_{\lambda(\theta)};\Omega_R^c)| &\leq C_i \int_{\Omega_R^c} \left((1+a_N)\exp_{\lambda_2}\right)^\theta \left((1+a_N)\exp_{\lambda_1}\right)^{1-\theta} dx \\
&\leq C_i \left(\int_{\Omega_R^c} (1+a_N)\exp_{\lambda_2} \, dx\right)^\theta \left(\int_{\Omega_R^c} (1+a_N)\exp_{\lambda_1} \, dx\right)^{1-\theta} \\
&\leq C_i \max_{j=1,2} \int_{\Omega_R^c} (1+a_N)\exp_{\lambda_j} \, dx \xrightarrow[R\to\infty]{} 0.
\end{aligned}$$

Consequently, $g(\cdot;\Omega_R)$ and its derivative converge uniformly to $g(\cdot;\Omega)$ respectively $g'(\cdot;\Omega)$ so that $g \in C^1([0,1])$. If $\theta_1 < \theta_2$, we get for the increase in g' with monotone convergence

$$g'(\theta_2;\Omega) - g'(\theta_1;\Omega) = \lim_{R\to\infty} \int_{\theta_1}^{\theta_2} g''(s;\Omega_R)\, ds = \int_{\theta_1}^{\theta_2} g''(s;\Omega)\, ds > 0.$$

We remark that $g''(s;\Omega)$ need not be continuous so that we cannot infer $g \in C^2([0,1])$. Nevertheless, a strictly increasing derivative $g'(\theta;\Omega)$ implies strict convexity of g and thus also of $\lambda \mapsto \mu_0(\exp_\lambda)$ throughout Λ. Finally, let us investigate the behavior of $\lambda \mapsto \xi \cdot \mu(\exp_\lambda)$ at the boundary if $\Lambda \cap \partial\Lambda = \emptyset$ and $\xi_N > 0$. Choosing a sequence $(\lambda^{(n)})_{n\in\mathbb{N}}$ which converges to a point $\bar{\lambda} \in \partial\Lambda$ we split $\mu(\exp_{\lambda^{(n)}}) = \mu(\exp_{\lambda^{(n)}};\Omega_R) + \mu(\exp_{\lambda^{(n)}};\Omega_R^c)$. The constant $R > 0$ we choose in such a way that

$$a_N(x) \geq 1 \quad \text{and} \quad \sum_{i=0}^{N-1} \frac{\xi_i a_i(x)}{\xi_N a_N(x)} \geq -\frac{1}{2}, \qquad |x| > R.$$

On the bounded domain Ω_R we get immediately $\mu(\exp_{\lambda^{(n)}};\Omega_R) \to \mu(\exp_{\bar{\lambda}};\Omega_R)$ for $n \to \infty$, which is a bounded contribution. On $|x| > R$ we find

$$\begin{aligned}\xi \cdot \mu(\exp_{\lambda^{(n)}};\Omega_R^c) &= \int_{\Omega_R^c} \xi_N a_N \left(1 + \sum_{i=0}^{N-1} \frac{\xi_i a_i(x)}{\xi_N a_N(x)}\right) \exp_{\lambda^{(n)}}\, dx \\ &\geq \frac{1}{2}\xi_N \int_{\Omega_R^c} a_N \exp_{\lambda^{(n)}}\, dx = \frac{1}{2}\xi_N \mu_N(\exp_{\lambda^{(n)}};\Omega_R^c).\end{aligned}$$

If $\mu_N(\exp_{\lambda^{(n)}};\Omega_R^c)$ had a bounded subsequence then one could show with the help of Fatou's lemma that also $\mu_N(\exp_{\bar{\lambda}};\Omega_R^c) < \infty$. Since $a_N \geq 1$ on Ω_R^c we would find

$$\int_\Omega (1+a_N)\exp_{\bar{\lambda}}\, dx = \int_{\Omega_R} (1+a_N)\exp_{\bar{\lambda}}\, dx + \int_{\Omega_R^c} (1+a_N)\exp_{\bar{\lambda}}\, dx < \infty$$

in contradiction to $\Lambda \cap \partial\Lambda = \emptyset$. □

Corollary 1 *The moment map $\lambda \mapsto \mu(\exp_\lambda)$ is a diffeomorphism from* $\mathrm{int}(\Lambda)$ *onto* $\mathrm{int}(\mu(E))$.

Proof. According to Lemma 4, the moment map (which is just $\nabla_\lambda \mu_0(\exp_\lambda)$) is infinitely smooth on $\mathrm{int}(\Lambda)$. Since $\lambda \mapsto \mu_0(\exp_\lambda)$ is strictly convex, the Jacobian of the moment map never vanishes. Consequently, it is an injective, open mapping with a smooth inverse (inverse function theorem). □

7. The case $\Lambda \neq \emptyset$ and $\Lambda \cap \partial\Lambda = \emptyset$

From Theorem 1 we already know that the maximum entropy problem is solvable if $\rho \in \mu(E)$. We will now prove that in the cases under consideration $\mu(E) = \mu(D)$ so

that the problem is, in fact, solvable for any admissible moment vector. The basic idea is to show that the function

$$z(\lambda;\rho) = \mu_0(\exp_\lambda) - \lambda\cdot\rho$$

attains its minimum on Λ for any $\rho \in \mu(D)$. Then, the necessary condition for an extremum in $\bar\lambda$ yields

$$0 = \nabla_\lambda z(\bar\lambda) = \int_\Omega a(x)\exp(\bar\lambda\cdot a(x))\,dx - \rho$$

which gives $\rho = \mu(\exp_{\bar\lambda}) \in \mu(E)$.

Theorem 2 *Assume $\Lambda \neq \emptyset$ and $\Lambda \cap \partial\Lambda = \emptyset$. Then $\mu(E) = \mu(D)$, i.e. for any $\rho \in \mu(D)$ there exists a unique $\lambda \in \Lambda$ such that $\mu(\exp_\lambda) = \rho$. In particular, the maximum entropy problem is uniquely solvable for any admissible moment vector.*

The proof of Theorem 2 splits up into two lemmas. First, for given $\rho \in \mu(D)$, we investigate the behavior of

$$z(\lambda;\rho) := \mu_0(\exp_\lambda) - \lambda\cdot\rho, \qquad \lambda\in\Lambda \tag{7.5}$$

at $\partial\Lambda$ respectively at $|\lambda|\to\infty$. To this end, we pick some $\bar\lambda\in\Lambda$ and consider (7.5) along rays with directions $\xi\in S^N$. Due to the structure of Λ (which equals $\mathrm{int}(\Lambda)$ in the case under consideration), a ray in direction ξ hits the boundary if $\xi_N > 0$ unless $\partial\Lambda = \emptyset$ as for bounded Ω. To treat all cases notationally in the same manner we introduce

$$s_b(\xi,\bar\lambda) := \sup\{\, s \,:\, \bar\lambda + s\xi \in \Lambda\,\}$$

which is $+\infty$ if the boundary $\partial\Lambda$ is not met in direction ξ.

Lemma 5 *Let $\Lambda\neq\emptyset$ and $\Lambda\cap\partial\Lambda = \emptyset$. For $\rho\in\mu(D)$, $\bar\lambda\in\Lambda$ and $\xi\in S^N$, the function*

$$s \mapsto z(\bar\lambda + s\xi;\rho)$$

attains its unique minimum in the open interval $I(\xi,\bar\lambda) := (-s_b(-\xi,\bar\lambda), s_b(\xi,\bar\lambda))$.

Proof. From Lemma 4 we know that $z(s) := z(\bar\lambda + s\xi;\rho)$ is strictly convex and contained in $C^1(I(\xi,\bar\lambda))$. Also, if $s_b(\xi,\bar\lambda) < \infty$, i.e. ξ points towards $\partial\Lambda \neq \emptyset$, then

$$z'(s) = \xi\cdot(\mu(\exp_{\bar\lambda+s\xi}) - \rho) \xrightarrow[s\to s_b(\xi,\bar\lambda)]{} \infty.$$

To treat the case $s_b(\xi,\bar\lambda) = \infty$ we introduce two complementary half spheres

$$L^+ := \{\xi\in S^N \,:\, \xi\cdot\rho \geq 0\}, \qquad L^- := \{\xi\in S^N \,:\, \xi\cdot\rho < 0\}.$$

For $\xi\in L^-$ we get $-s\xi\cdot\rho\to+\infty$. Since μ_0 is always positive we see that $z(s)\to\infty$ for $s\to\infty$. In the case $\xi\in L^+$ we first assume that $\{\, x\in\Omega \,:\, \xi\cdot a(x) > 0\}$ has positive measure. Then, there exists some $\epsilon > 0$ such that also $B := \{\, x \,:\, \xi\cdot a(x) > \epsilon\,\}$ has positive measure in Ω. We obtain

$$z(s) \geq \left(\int_B \exp_{\bar\lambda}\,dx\right)\exp(s\epsilon) - (\bar\lambda + s\xi)\cdot\rho$$

which certainly goes to $+\infty$ due to the exponential increase. It remains to exclude the case where $\xi \cdot a(x) \leq 0$ on Ω. To this end we pick $f \in D$ such that $\rho = \mu(f)$. Since, by definition, $\{x : f(x) > 0\}$ has positive measure and $\{x : \xi \cdot a(x) = 0\}$ has measure zero due to the pseudo–Haar property, we find that $U := \{x : \xi \cdot a(x) f(x) < 0\}$ has positive measure and hence

$$\xi \cdot \rho = \int_\Omega \xi \cdot a(x) f(x)\, dx \leq \int_U \xi \cdot a(x) f(x)\, dx < 0$$

in contradiction to the assumption $\xi \in L^+$. We conclude that on each endpoint of the interval $I(\xi, \bar{\lambda})$, $z(s)$ either diverges or the slope $z'(s)$ becomes $+\infty$. In any case, there exist points $s_1, s_2 \in I(\xi, \bar{\lambda})$ such that $z'(s_1) z'(s_2) < 0$. Applying mean value theorem, we find $s^* \in I(\xi, \bar{\lambda})$ such that $z'(s^*) = 0$ and due to strict convexity of z, s^* is the unique minimizer. □

A basic result on optimization problems now shows that z always has a unique minimizer $\bar{\lambda}$ which concludes the proof of Theorem 2. The notation is the same as in Lemma 5 and the elementary proof is omitted.

Lemma 6 *Assume $\emptyset \neq \Lambda \subset \mathbb{R}^{N+1}$ is open and convex. Further, let $z : \Lambda \mapsto \mathbb{R}$ be strictly convex. If for $\bar{\lambda} \in \Lambda$*

$$s \mapsto z(\bar{\lambda} + s\xi), \qquad s \in I(\xi, \bar{\lambda})$$

attains its minimum for all directions $\xi \in S^N$ then z has a unique minimizer $\lambda^* \in \Lambda$.

8. The case $\Lambda = \emptyset$

Let us now turn to the extreme case $\Lambda = \emptyset$. The non solvability follows from a Theorem which applies in the most general setting. We start with a preparatory result.

Lemma 7 *Let Ω_1 be an open, bounded subset of a possibly unbounded, open $\Omega \subset \mathbb{R}^d$ and let $g \in D(\Omega)$ be a maximum entropy solution. Then $g|_{\Omega_1} \in E(\Omega_1)$ is of exponential type.*

Proof. We denote $g_1 := g|_{\Omega_1}$. If we assume $g_1 \not\equiv 0$ then $g_1 \in D(\Omega_1)$ is the maximum entropy solution under the constraint $\rho = \mu(g; \Omega_1)$ because otherwise, the global entropy

$$H(g) = -\int_{\Omega_1} g \ln g\, dx - \int_{\Omega \setminus \Omega_1} g \ln g\, dx$$

could be increased by modifying g on Ω_1 without changing the moment vector $\mu(g; \Omega_1)$. Since Ω_1 is bounded, we know $\Lambda(\Omega_1) = \mathbb{R}^{N+1}$ and thus Theorem 2 shows that $g_1 = \exp_\lambda$ for some $\lambda \in \mathbb{R}^{N+1}$. The remaining case $g_1 \equiv 0$ cannot appear. Since $g \not\equiv 0$, we can find an open, bounded $\Omega_2 \subset \Omega$ with $\Omega_1 \subset \Omega_2$ and $g_2 := g|_{\Omega_2} \not\equiv 0$. Applying the considerations to g_2 yields $g_1 = g_2|_{\Omega_1} = \exp_\lambda|_{\Omega_1} \not\equiv 0$. □

With the help of Theorem 1 and the above Lemma we can now completely characterize the moment vectors $\rho \in \mu(D)$ for which the maximum entropy problem is solvable.

Theorem 3 *Let $\rho \in \mu(D)$. Then the maximum entropy problem is solvable if and only if $\rho \in \mu(E)$. The optimal density is always of exponential type.*

Proof. First, we assume that $\rho \in \mu(E)$. Using Theorem 1 we see that the maximum entropy problem has a unique solution of exponential type. Conversely, if there exists a solution $g \in D$, we exhaust Ω with an increasing sequence of bounded, open sets $\Omega_n \subset \Omega_{n+1}$ and Lemma 7 shows that $g_n := g|_{\Omega_n} = \exp_{\lambda^{(n)}}$ for certain $\lambda^{(n)} \in \mathbb{R}^{N+1}$. Applying Lemma 7 again to a pair $\Omega_n \subset \Omega_{n+1}$ we also see that $\exp_{\lambda^{(n)}} = \exp_{\lambda^{(n+1)}}$ on Ω_n. Taking logarithm and regrouping yields $(\lambda^{(n+1)} - \lambda^{(n)}) \cdot a \equiv 0$ on Ω_n so that with the pseudo–Haar property $\lambda^{(n+1)} = \lambda^{(n)} =: \lambda$. Since also $g|_{\Omega_n} \to g$ in $\mathbb{L}^1(\Omega)$ we find that $g = \exp_\lambda$ (at least a.e.). In particular, $\mu(g) \in \mu(E)$. □

Corollary 2 *If $\Lambda = \emptyset$ then for any $\rho \in \mu(D)$ the maximum entropy problem is not solvable.*

Proof. $\Lambda = \emptyset$ implies $E = \emptyset$ and thus also $\mu(E) = \emptyset$. □

9. The case $\Lambda \cap \partial\Lambda \neq \emptyset$

With Theorem 3, the moment vectors for which the maximum entropy problem is solvable are completely characterized as moments of exponential densities. What remains to be investigated is the question whether $\mu(E)$ is possibly a proper subset of $\mu(D)$. In this case, there are admissible moment vectors $\rho \in \mu(D)$ for which the maximum entropy problem is not solvable. Our aim is to characterize these vectors.

Since Ω is necessarily unbounded in the case under consideration, the moment function a_N is distinguished in so far as it dominates all other a_i for large $|x|$. This property reflects itself in a special behavior of the highest components of λ and ρ. To stress this fact and to simplify notation, we introduce the order relation

$$(u_0, \ldots, u_N)^T \geq (v_0, \ldots, v_N)^T \quad \Longleftrightarrow \quad u_0 = v_0, \ldots, u_{N-1} = v_{N-1}, u_N \geq v_N.$$

Then, we have the following dual result.

Lemma 8 *Assume $\bar\lambda \in \Lambda$ and $\bar\rho \in \mu(D)$. Then all $\lambda < \bar\lambda$ and all $\rho > \bar\rho$ also belong to Λ respectively $\mu(D)$.*

Proof. The result on λ has already been shown in Lemma 3. For $\bar\rho \in \mu(D)$, Theorem A.1 in the appendix tells us that $\beta \cdot \bar\rho < 0$ for all $0 \neq \beta \in \mathbb{R}^{N+1}$ which satisfy $\beta \cdot a \leq 0$ a.e. on Ω. If $\rho = \bar\rho + \delta e_N$ with $\delta > 0$ then $\beta \cdot \rho = \beta \cdot \bar\rho + \beta_N \delta$. Consequently, also $\rho \in \mu(D)$ if $\beta_N \leq 0$. That this is indeed the case follows easily by contradiction. Assuming $\beta_N > 0$ we choose $R > 0$ so large that

$$\sum_{i=0}^{N-1} \frac{\beta_i a_i(x)}{\beta_N a_N(x)} > -\frac{1}{2}, \qquad |x| > R.$$

Then, we find

$$\beta \cdot a(x) = \beta_N a_N(x) \left(1 + \sum_{i=0}^{N-1} \frac{\beta_i a_i(x)}{\beta_N a_N(x)} \right) > 0, \qquad |x| > R.$$

in contradiction to our assumption on β. □

We have already seen, that $\Lambda \cap \partial\Lambda = \emptyset$ implies solvability of the maximum entropy problem for any choice of $\rho \in \mu(D)$. Conversely, if $\Lambda \cap \partial\Lambda \neq \emptyset$, we can always find admissible moment vectors such that the problem is not solvable.

Lemma 9 *Let $\bar{\lambda} \in \Lambda \cap \partial\Lambda$ and $\rho > \bar{\rho} = \mu(\exp_{\bar{\lambda}})$. Then $\rho \in \mu(D)\backslash\mu(E)$.*

Proof. Assume $\rho = \mu(\exp_\lambda)$ for some $\lambda \in \Lambda$. Then $\lambda \mapsto z(\lambda;\rho) = \mu_0(\exp_\lambda) - \lambda \cdot \rho$ satisfies on the line segment $\lambda(\theta) = \theta\lambda + (1-\theta)\bar{\lambda}$ (see Lemma 4)

$$\frac{d}{d\theta} z(\lambda(\theta);\rho) = (\lambda - \bar{\lambda})(\mu(\exp_{\lambda(\theta)}) - \rho).$$

The derivative of $\mu_0(\exp_{\lambda(\theta)})$, and thus also of $z(\lambda(\theta);\rho)$, is strictly increasing so that

$$(\lambda - \bar{\lambda})(\bar{\rho} - \rho) < (\lambda - \bar{\lambda})(\rho - \rho) = 0.$$

Using $\rho = \bar{\rho} + \delta e_N$ with $\delta > 0$ and $\bar{\lambda}_N = \lambda_N^b$, we get $\lambda_N > \lambda_N^b$ which contradicts $\lambda \in \Lambda$ and hence $\rho \notin \mu(E)$. □

The next Lemma shows that the converse statement is also true. Its proof is the mathematically strict version of the introductory argument which explained the possible failure of an optimizing sequence to converge to a solution of the maximum entropy problem.

Lemma 10 *Let $\rho \in \mu(D)\backslash\mu(E)$. Then $\rho > \mu(\exp_\lambda)$ for some $\lambda \in \partial\Lambda$. Any relative entropy minimizing sequence $(f_n)_{n\in\mathbb{N}} \subset \mu(D)$ which satisfies $\mu(f_n) = \rho$ converges to* $\exp_\lambda$.

Proof. By normalization, we will restrict our considerations to the case $\rho \in \mu(D^*)\backslash\mu(E^*)$. Also, we choose the reference density $\hat{f} = \exp_{\hat{\lambda}}$ with $\hat{\lambda} \in \partial\Lambda$. In the non–empty convex set F defined by

$$F := \{ f \in D^* \,:\, \mu(f) = \rho \}$$

we select a sequence $(f_n)_{n\in\mathbb{N}} \subset F$ which satisfies

$$H(f_n, \hat{f}) \xrightarrow[n\to\infty]{} \inf\{\, H(f, \hat{f}) \,:\, f \in F \,\}.$$

Modifying f_n on subsets of Ω, a new sequence $(g_n)_{n\in\mathbb{N}}$ is constructed. The corresponding sets Ω_n should be open, bounded and satisfy

$$f_n|_{\Omega_n} \not\equiv 0, \qquad |\mu(f_n;\Omega_n) - \rho| < \frac{1}{n}, \qquad |H(f_n, \hat{f}; \Omega_n^c)| < \frac{1}{n}$$

and

$$\Omega_n \subset \Omega_{n+1}, \qquad \bigcup_{n\in\mathbb{N}} \Omega_n = \Omega, \qquad \min\{\,|x| \,:\, x \in \Omega_n^c\,\} \xrightarrow[n\to\infty]{} \infty.$$

We set

$$g_n(x) := \begin{cases} \exp_{\lambda^{(n)}} & x \in \Omega_n \\ f_n(x) & x \in \Omega_n^c \end{cases}$$

where $\lambda^{(n)}$ is uniquely defined through $\mu(\exp_{\lambda^{(n)}}; \Omega_n) = \mu(f_n; \Omega_n)$. Obviously, $\mu(f_n) = \mu(g_n) = \rho$ and the relative entropy does not increase by the modification on Ω_n so that $(g_n)_{n\in\mathbb{N}} \subset F$ is also a minimizing sequence. Using Lemma 2, we can infer the $\mathbb{L}^1$–convergence of $(g_n)_{n\in\mathbb{N}}$. The limit function g is even contained in D^*. To see this, we first notice that with Fatou's lemma

$$\int_\Omega (1 + a_N) g \, dx \le \liminf_{n\to\infty} \int_\Omega (1 + a_N) g_n \, dx = 1 + \rho_N. \tag{9.6}$$

so that $g \in D$. On the other hand, the uniform bound on the a_N–moment yields convergence of all other moments. To see this, we remark that by assumption

$$\frac{|a_i(x)|}{1 + a_N(x)} \le \gamma_m, \qquad \forall x \in \Omega_m^c$$

with $\gamma_m \to 0$ for $m \to \infty$. Consequently,

$$\int_{\Omega_m^c} |a_i| g \, dx \le \gamma_m \int_\Omega (1 + a_N) g \, dx = \gamma_m (1 + \rho_N).$$

The same estimate also holds for each g_n. Since a_i are locally bounded, we have

$$\int_{\Omega_m} a_i g \, dx = \lim_{n\to\infty} \int_{\Omega_m} a_i g_n \, dx$$

and thus

$$\left| \int_{\Omega_m} a_i g \, dx - \rho_i \right| = \left| \lim_{n\to\infty} \int_{\Omega_m^c} a_i g_n \, dx \right| \le \gamma_m (1 + \rho_N).$$

Altogether, the lower moments converge and the highest moment might drop (due to (9.6)). Hence, $\mu(g) \le \rho$. To show that g is of exponential type, we take some $i \in \mathbb{N}$ such that $g|_{\Omega_i} \not\equiv 0$. By construction, we have

$$g_n|_{\Omega_i} = \exp_{\lambda^{(n)}}, \qquad \forall n \ge i.$$

Since Ω_i is bounded, $\Lambda(\Omega_i)$ and $\mu(E(\Omega_i)) = \mu(D(\Omega_i))$ are diffeomorphic so that

$$\lambda^{(n)} = \phi(\mu(g_n; \Omega_i)) \xrightarrow[n\to\infty]{} \phi(\mu(g; \Omega_i)) =: \lambda$$

(ϕ denotes the inverse of the mapping $\lambda \mapsto \mu(\exp_\lambda; \Omega_i)$). In particular, $g_n|_{\Omega_i}$ converges to $\exp_\lambda|_{\Omega_i}$. Repeating the same argument for $i+1$ we find $g_n|_{\Omega_{i+1}} \to \exp_{\lambda'}|_{\Omega_i}$ and due to the pseudo–Haar property $\lambda' = \lambda$. Consequently $g = \exp_\lambda$ and

$\lambda \in \Lambda$ because $g \in D$. Since we have assumed $\rho \notin \mu(E)$, the case $\mu(\exp_\lambda) = \rho$ cannot appear and we really find a drop $\mu(\exp_\lambda) < \rho$ in the limit. It remains to show that $\lambda \in \partial\Lambda$. Due to our construction of $(\Omega_n)_{n\in\mathbb{N}}$ we get

$$\begin{aligned} H(g_n, \hat{f}) &= H(\exp_{\lambda^{(n)}}, \hat{f}; \Omega_n) + H(f_n, \hat{f}; \Omega_n^c) \\ &= (\lambda^{(n)} - \hat{\lambda})\mu(f_n; \Omega_n) + H(f_n, \hat{f}; \Omega_n^c) \xrightarrow[n\to\infty]{} (\lambda - \hat{\lambda})\rho. \end{aligned}$$

In addition, we know from Lemma 2 that $H(g, \hat{f}) \leq \liminf_{n\to\infty} H(g_n, \hat{f})$ so that altogether

$$(\lambda - \hat{\lambda})\mu(\exp_\lambda) \leq (\lambda - \hat{\lambda})\rho.$$

Since $\mu(\exp_\lambda) < \rho$ and $\hat{\lambda}_N = \lambda_N^b$ we find $(\lambda_N - \lambda_N^b)(\mu_N(\exp_\lambda) - \rho_N) \leq 0$ which implies $\lambda_N = \lambda_N^b$ so that $\lambda \in \partial\Lambda$. Finally, the mixed sequence

$$f_1, g_1, f_2, g_2, f_3, \ldots$$

is also a minimizing sequence in F and thus converges in $\mathbb{L}^1(\Omega)$. Since the subsequence $(g_n)_{n\in\mathbb{N}}$ converges to $\exp_\lambda$ the same holds also for the subsequence $(f_n)_{n\in\mathbb{N}}$ which shows that every minimizing sequence converges to the same limit. □

We combine Lemma 9 and Lemma 10 in a final theorem.

Theorem 4 *Assume $\Lambda \cap \partial\Lambda \neq \emptyset$. Then*

$$\mu(D)\backslash\mu(E) = \{\, \rho \,:\, \rho > \mu(\exp_\lambda),\ \lambda \in \Lambda \cap \partial\Lambda \,\}.$$

In particular, the maximum entropy problem is solvable if and only if $\rho \in \mu(D)$ satisfies $\rho \not> \mu(\exp_\lambda)$ for all $\lambda \in \Lambda \cap \partial\Lambda$.

10. Examples

10.1. *The Hausdorff moment problem*

In the classical reduced Hausdorff problem, the setting is $\Omega = [0, 1]$ and $a_i(x) = x^i$ for $i = 0, \ldots, N$. Since Ω is bounded, $\Lambda = \mathbb{R}^{N+1}$ so that the maximum entropy problem is always solvable (Theorem 2). For further investigations on this special case like solvability conditions for the moment problem (i.e. the structure of $\mu(D) = \mu(E)$) or convergence of the maximum entropy distributions for $N \to \infty$ we refer to the article by Mead and Papanicolaou.[16] The case of bounded $\Omega \subset \mathbb{R}^d$ is also called generalized Hausdorff moment problem and Theorem 2 shows solvability of the maximum entropy problem whenever the moment problem itself has a solution (if all moments are prescribed, general solvability conditions can be given[4]). For the case of general entropy functionals (which are applied to measures on a compact Hausdorff space) and constraints based on continuous moment functions we refer to the work by Lewis.[15]

10.2. *The Stieltjes moment problem*

For the reduced Stieltjes moment problem the underlying space is the positive half line $\Omega = \mathbb{R}^+$ and the moment functions are again monomials $a_i(x) = x^i$ with $i = 0, \ldots, N$. For $N = 2$, non trivial conditions for the existence of a maximum entropy solution are found.[6] If ρ_0 is normalized to one, the condition is $1 < \rho_2/\rho_1^2 \leq 2$. While the lower bound is related to the solvability of the moment problem, the upper bound restricts the applicability of the maximum entropy approach (we know from Lemma 8 that the highest moment ρ_2 is not restricted in $\mu(D)$). According to Theorem 4, problems arise whenever $\rho > \mu(\exp_\lambda)$ with $\lambda \in \partial\Lambda$. For $N = 2$, normalized densities $\exp_\lambda$ with $\lambda \in \partial\Lambda$ are of the form

$$\exp_\lambda(x) = |\lambda_1| \exp(\lambda_1 x), \qquad \lambda_1 < 0.$$

Computing the moments yields $\mu_1(\exp_\lambda) = 1/|\lambda_1|$ and $\mu_2(\exp_\lambda) = 2/\lambda_1^2$. Consequently, the solvability condition $\rho \in \mu(D)$ and $\rho_2/\rho_1^2 \leq 2$ obtained from Theorem 4 coincides with known results.[6] The case $N = 3$ has also been treated in the literature.[13,19] The resulting conditions require knowledge of the moments of

$$\exp_\lambda(x) = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2), \qquad \lambda_2 < 0 \text{ or } \lambda_2 = 0 \text{ and } \lambda_1 < 0.$$

While $\lambda_2 = 0$ leads to the simple restriction $\rho_3/\rho_1^3 \leq 6$ if $\rho_2/\rho_1^2 = 2$, the case $\lambda_2 < 0$ is more complicated and involves error functions. If we restrict to the case $\rho_0 = \rho_1 = 1$ ($\rho_1 = 1$ can always be achieved by scaling x) the exponential densities $\exp_\lambda$ with $\lambda \in \partial\Lambda$ depend only on a single parameter s. If we plot the second and third moment depending on s, all points located above this curve in the (ρ_2, ρ_3) plane correspond to moment vectors for which the maximum entropy problem is not solvable (see Figure 1). For $\rho_2 > 2$ there are no restrictions on the highest moment ρ_3. The lower bound for ρ_3 is again caused by the solvability condition for the moment problem. Using Schwartz inequality we find immediately for any $f \in \mu(D)$

$$\rho_2 = \int_0^\infty x^{\frac{3}{2}} x^{\frac{1}{2}} f(x)\, dx \leq \sqrt{\rho_3}\sqrt{\rho_1}$$

which amounts to $\rho_3 \geq \rho_2^2$ since ρ_1 has been scaled to one. We remark that there are restrictions on the solvability for any $N \geq 2$ (see for example [7]) in contrast to the result presented in articles by Tagliani and Frontini[8,19] which negates restrictions for $N \geq 4$.

10.3. *The Hamburger moment problem*

The Hamburger moment problem has already been presented in the introduction. Here, $\Omega = \mathbb{R}$, the moment functions are monomials and $N = 2n$ is even. For $N = 2$, the boundary of Λ is empty because $\exp(\lambda_0 + \lambda_1 x)$ is never integrable on $\mathbb{R}$. Hence, the maximum entropy solution always exists (Theorem 2), giving rise to the well known Maxwellian distributions. In the case $N = 4$, restrictions apply.[8,20] The boundary $\partial\Lambda$ consists exactly of the Maxwellians

$$\exp_\lambda(x) = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2), \qquad \lambda_2 < 0.$$

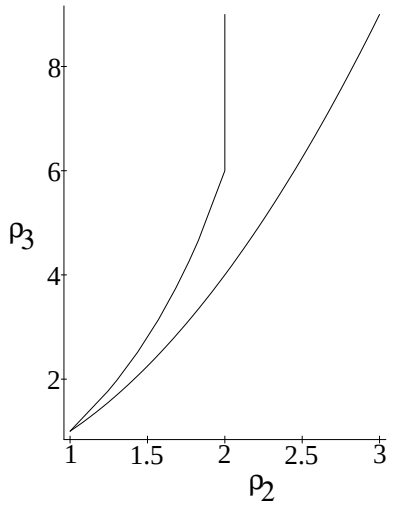

Fig. 1. The maximum entropy problem is solvable if the scaled moment constraint is located between the curves in the moment plane.

In terms of the centered moments

$$\hat{\mu}_i(f) := \int_{\mathbb{R}} (x-u)^i f(x)\,dx, \qquad u := \frac{\mu_1(f)}{\mu_0(f)},$$

the solvability conditions can easily be stated. If $\rho \in \mu(D)$ is given, we can calculate the corresponding centered moments $\hat{\rho}$. If $\hat{\rho}_3 \neq 0$ there can be no $\lambda \in \partial\Lambda$ such that $\rho > \mu(\exp_\lambda)$ because all Maxwellians satisfy $\hat{\mu}_3(\exp_\lambda) = 0$ so that $\mu_3(\exp_\lambda) \neq \rho_3$. On the other hand, if $\hat{\rho}_3 = 0$ then the relation $\hat{\rho}_0\hat{\rho}_4/\hat{\rho}_2^2 \leq 3$ is necessary for the existence of a maximum entropy solution. This is due to the fact that for Maxwellians $\hat{\mu}_0\hat{\mu}_4/\hat{\mu}_2^2 = 3$ holds. For $N = 6$, restrictions are reported only for the symmetric case where all odd centered moments vanish.[8] Otherwise, no restrictions are found. However, Theorem 4 shows that also for the non symmetric case such conditions exist. Indeed, every $\lambda = (\lambda_0, \ldots, \lambda_4, 0, 0)^T \in \Lambda$ with $\lambda_3 \neq 0$ gives rise to a half line consisting of non symmetric moments in $\mu(D)$ for which the maximum entropy problem is not solvable. Moreover, for any even $N \geq 4$ there are exceptional moment vectors in contrast to the statement in the articles by Tagliani and Frontini.[8,20]

10.4. *A multidimensional example*

Recently, the work by Levermore[14] has raised interest in maximum entropy solutions to moment problems on $\mathbb{R}^3$. In the Kinetic Theory of gases, it is well known[3] that the atomic velocities v of a gas in thermodynamical equilibrium are distributed

according to a special exponential density, the so called *Maxwellian*

$$\mathcal{M}(v) = \frac{\rho}{(2\pi T)^{\frac{3}{2}}} \exp\left(-\frac{|v-u|^2}{2T}\right), \qquad v \in \mathbb{R}^3. \tag{10.7}$$

Here, ρ denotes the density of the gas, u its average velocity and T the temperature. The time evolution of ρ, u, T is governed by the Euler equations of gas dynamics.

A crucial point in this context is that (10.7) maximizes the physical entropy

$$H(f) = -\int_{\mathbb{R}^3} f \ln f \, dv$$

under the constraint that the velocity moments based on

$$a_0(v) \equiv 1, \qquad a_1(v) = v_1,\ a_2(v) = v_2,\ a_3(v) = v_3, \qquad a_4(v) = \frac{1}{2}|v|^2$$

are prescribed. (The corresponding moments are algebraic combinations of ρ, u, T.) The basic idea in the work of Levermore is that for gases whose state is somewhat away from equilibrium, one can still describe the velocity distribution of the atoms with the help of a maximum entropy function.[14] Only the number of moment constraints is supposed to increase, in order to allow for more structure in the distributions. The simplest model in three dimensions, which contains more than quadratic moments, is characterized by the so called 14–moment system. The additional moment functions are

$$v_1^2, \quad v_2^2, \quad v_1 v_2, \quad v_1 v_3, \quad v_2 v_3,$$

as well as the cubic terms $|v|^2 v_i$, $i = 1, 2, 3$ and the quartic $a_{13}(v) = |v|^4$. The time evolution of the velocity moments is given in the form of a symmetric hyperbolic system of partial differential equations (the so called *14–moments system*) which is an extension of the Euler system.

Clearly, the assumptions of Section 3 are satisfied in this case. Moreover, the intersection of Λ with its boundary is not empty since $\lambda_{13} = 0$ does not rule out the integrability of $\exp_\lambda$ (all Maxwellians (10.7), for example, correspond to points on $\Lambda \cap \partial\Lambda$). Hence, Theorem 4 implies that there are moment vectors for which the maximum entropy problem has no solution. For the 14–moments system, this means that the domain of definition (the set of all possible moments $\mu(E)$) has a quite complicated structure. It is a convex set where half lines originating in moment vectors corresponding to Maxwellian distributions are excluded. In particular, the domain of definition of the 14–moments system is not convex in contrast to the one of the Euler system. For the case of one dimensional atomic velocities, the situation has been investigated in detail.[12]

10.5. *Extreme examples*

Finally, we want to give some examples where $\lambda_N^b \neq 0$ so that the boundary of Λ is different from $\{\, \lambda \in \Lambda \,:\, \lambda_N = 0 \,\}$. To obtain $\lambda_N^b > 0$, the domain Ω obviously has

to become very small at large x. Setting

$$\Omega := \bigcup_{n\in\mathbb{N}} \Omega_n, \qquad \Omega_n = \left[n, n + \ln\left(1 + \frac{\exp(-n)}{n^r}\right)\right], \qquad r > 1$$

and $a_1(x) = x$, we find

$$\int_\Omega \exp(x)\, dx = \sum_{n=1}^{\infty} \frac{1}{n^r} = \zeta(r)$$

where ζ is the Riemann Zeta function. Consequently, $x \mapsto \exp(x) \in \mathbb{L}^1(\Omega)$ for all $r > 1$. On the other hand, for $r = 2$ the decay of Ω_n is not strong enough to give also integrability of the first moment. We get

$$\int_{\Omega_n} x \exp(x)\, dx = \frac{1}{n} + R_n \tag{10.8}$$

where $\sum R_n$ is finite and $\sum 1/n$ diverges. In particular, $\exp((1+\epsilon)x) \notin \mathbb{L}^1(\Omega)$ for $\epsilon > 0$ and $x\exp((1-\epsilon)x) = x\exp(-\epsilon x)\exp(x) \in \mathbb{L}^1(\Omega)$ so that $\lambda_1^b = 1$. Since $x\exp(x) \notin \mathbb{L}^1(\Omega)$, however, this example falls into the class $\Lambda \cap \partial\Lambda = \emptyset$.

If we choose $r = 3$, we have

$$\int_{\Omega_n} x \exp(x)\, dx \approx \frac{1}{n^2}, \qquad \int_{\Omega_n} x^2 \exp(x)\, dx \approx \frac{1}{n}.$$

Using similar arguments, we find again $\lambda_1^b = 1$ but now $\lambda = (0,1)^T \in \Lambda \cap \partial\Lambda$.

In the last example we cover the case $\lambda_N^b < 0$. Obviously, this is only possible if the moment function a_N increases very slowly. Choosing $\Omega = \mathbb{R}$ and $a_1(x) = \ln(1 + |x|)$ we obtain

$$\int_\Omega \exp(\lambda_1 a_1(x))\, dx = \int_{\mathbb{R}} (1 + |x|)^{\lambda_1}\, dx$$

which is finite only if $\lambda_1 < -1$.

Appendix

The appendix is devoted to a characterization of those vectors $\rho \in \mathbb{R}^{N+1}$ which can be moments of a function in D. First, we want to derive a necessary condition so let us assume $\rho = \mu(f)$ with $f \in D$. Certainly, if there is a linear combination $\beta \cdot a(x)$ of the moment functions which is almost everywhere non positive on Ω we get

$$\beta \cdot \rho = \int_\Omega \beta \cdot a(x)\, f(x)\, dx \le 0.$$

Since $\{a_0, \ldots, a_N\}$ is pseudo–Haar, we know that $\mathrm{vol}(\{x \in \Omega : \beta \cdot a(x) = 0\}) = 0$ provided $\beta \neq 0$. Consequently, if $\beta \cdot a \le 0$ a.e. on Ω and $\beta \neq 0$ we even get $\beta \cdot a < 0$ a.e. on Ω. Since $f \in D$ we also have $\mathrm{vol}(\{x \in \Omega : f(x) > 0\}) > 0$ so that $\beta \cdot \rho < 0$.

Lemma A.1 *Let $\rho \in \mu(D)$. Then for all $0 \neq \beta \in \mathbb{R}^{N+1}$ which satisfy $\beta \cdot a \leq 0$ a.e. on Ω the relation $\beta \cdot \rho < 0$ holds.*

Geometrically, $\mu(D)$ forms a cone in $\mathbb{R}^{N+1}$ since D is a cone in $\mathbb{L}^1(\Omega)$ and μ is a linear map. Lemma (A.1) then relates $\mu(D)$ to polars of conic hulls

$$C_{\tilde{\Omega}} := \mathrm{ch}\left(\{\, a(x) \,:\, x \in \tilde{\Omega} \,\}\right), \qquad \tilde{\Omega} \subset \Omega, \quad \mathrm{vol}(\Omega \backslash \tilde{\Omega}) = 0.$$

Indeed, the polar of $C_{\tilde{\Omega}}$ is just[17]

$$\begin{aligned} C_{\tilde{\Omega}}^{\circ} :=& \{\, \beta \in \mathbb{R}^{N+1} \,:\, \beta \cdot \eta \leq 0 \ \forall \eta \in C_{\tilde{\Omega}} \,\} \\ =& \{\, \beta \in \mathbb{R}^{N+1} \,:\, \beta \cdot a(x) \leq 0 \ \forall x \in \tilde{\Omega} \,\} \end{aligned}$$

and thus Lemma A.1 implies $C_{\tilde{\Omega}}^{\circ} \subset \mu(D)^{\circ}$. More precisely, we find

$$\mu(D) \subset \mathrm{int}(C_{\tilde{\Omega}}) \tag{A.1}$$

as the following separation argument shows: assuming $\rho \in \mu(D)$ and $\rho \notin \mathrm{int}(C_{\tilde{\Omega}})$, we can find $0 \neq \beta \in C_{\tilde{\Omega}}^{\circ}$ such that $\beta \cdot \rho \geq 0$. However, Lemma A.1 yields $\beta \cdot \rho < 0$ for all $\beta \in C_{\tilde{\Omega}}^{\circ}$ which is a contradiction. The inverse inclusion to (A.1) can be shown if we select a special $\tilde{\Omega}$. We define Ω_L as the set of all Lebesgue points of the moment functions a_i, i.e. those $x \in \Omega$ for which

$$\lim_{r \to 0} \frac{1}{\mathrm{vol}(B_r)} \int_{B_r(x)} |a_i(y) - a_i(x)|\, dy = 0.$$

Since $\{a_0, \ldots, a_N\}$ are locally integrable, the non Lebesgue points form a set of measure zero.[18] Picking some $\rho \in \mathrm{int}(C_{\Omega_L})$ we can find $N+1$ linearly independent vectors $\eta_0, \ldots, \eta_N \in \mathrm{int}(C_{\Omega_L})$ on a small sphere around ρ such that ρ is in the interior of their convex hull, i.e.

$$\rho = \sum_{i=0}^{N} \alpha_i \eta_i, \qquad \sum_{i=0}^{N} \alpha_i = 1, \ \alpha_i > 0 \quad i = 0, \ldots, N.$$

By definition of C_{Ω_L}, each η_i can be written as a positive combination of vectors $a(x_j)$

$$\eta_i = \sum_{j=0}^{M} \beta_{ij} a(x_j), \qquad \beta_{ij} \geq 0, \ x_j \in \Omega_L.$$

Thus

$$\rho = \sum_{j=0}^{M} \gamma_j a(x_j), \qquad \gamma_j = \sum_{i=0}^{N} \alpha_i \beta_{ij} > 0, \ j = 0, \ldots, M.$$

We remark that $\{\, a(x_j) \,:\, j = 0, \ldots, M \,\}$ contains $N+1$ linearly independent vectors which are, without loss of generality, $a(x_0), \ldots, a(x_N)$ (otherwise span$\{\, a(x_j) \,:\, j = 0, \ldots, M \,\}$ could not contain $\eta_0, \ldots, \eta_N$). By construction, each x_j is a Lebesgue point of a so that we can approximate $a(x_j)$ by integral expressions

$$a^{(r)}(x_j) := \int_{\Omega} a(y) f_j^{(r)}(y)\, dy, \qquad f_j^{(r)}(y) = \frac{1}{\mathrm{vol}(B_r)} \mathcal{X}_{B_r(x_j)}(y)$$

where $\mathrm{cl}(B_r(x_j)) \subset \Omega$. Using implicit function theorem it is easy to see that, for r small enough, we can find positive parameters $\gamma_j^{(r)} > 0$ such that

$$\rho = \sum_{j=0}^{M} \gamma_j^{(r)} a^{(r)}(x_j) = \int_\Omega a(y) \left(\sum_{j=0}^{M} \gamma_j^{(r)} f_j^{(r)}(y) \right) dy \in \mu(D) \tag{A.2}$$

so that with (A.1) the equality $\mu(D) = \mathrm{int}(C_{\Omega_L})$ holds. To construct the parameters $\gamma_j^{(r)}$, we define the smooth mapping

$$\Phi(\nu, \zeta_0, \ldots, \zeta_M) := \sum_{j=0}^{M} \nu_j \zeta_j - \rho.$$

Then $\Phi(\gamma, a(x_0), \ldots, a(x_M)) = 0$ and the derivative with respect to $\nu_0, \ldots, \nu_N$ at this point has full rank

$$\det \left(\frac{\partial(\Phi_0 \cdots \Phi_N)}{\partial(\nu_0 \cdots \nu_N)} \right) = \det(a(x_0) \cdots a(x_N)) \neq 0$$

since $a(x_0), \ldots, a(x_N)$ are linearly independent. Thus, if $a(x_0), \ldots, a(x_M)$ are slightly varied (by going over to $a^{(r)}(x_0), \ldots, a^{(r)}(x_M)$), a corresponding slight change of $\gamma_0, \ldots \gamma_M$ can be found such that Φ stays zero. The implicit function theorem shows that the function $\hat{\nu}$ which relates changes in $a(x_i)$ to those in γ_j has the same smoothness as Φ and is, in particular, continuous so that positivity of γ_j is preserved under small deviations. (A.2) now follows with

$$\gamma_j^{(r)} := \hat{\nu}_j(a^{(r)}(x_0), \ldots, a^{(r)}(x_M)), \;\; 0 \leq j \leq N, \qquad \gamma_j^{(r)} = \gamma_j, \;\; j > N.$$

We conclude our considerations with

Theorem A.1 *A vector $\rho \in \mathbb{R}^{N+1}$ is contained in $\mu(D)$ if and only if for all $0 \neq \beta \in \mathbb{R}^{N+1}$ which satisfy $\beta \cdot a \leq 0$ a.e. on Ω the relation $\beta \cdot \rho < 0$ holds. Moreover, each $\rho \in \mu(D)$ is the moment vector of a bounded $f \in D$ which is compactly supported in Ω.*

Proof. One direction of the statement is just Lemma A.1. For the converse direction take $\rho \in \mathbb{R}^{N+1}$ such that $\beta \cdot \rho < 0$ for all $0 \neq \beta \in \mathbb{R}^{N+1}$ which satisfy $\beta \cdot a \leq 0$ a.e. on Ω. Using the definition of the polar $C^\circ_{\Omega_L}$ this shows that $\beta \cdot \rho < 0$ for all $\beta \in C^\circ_{\Omega_L}$ which implies $\rho \in \mathrm{int}(C_{\Omega_L}) = \mu(D)$ again by a separation argument. Finally, from (A.2) we deduce that $\rho \in \mu(D)$ is the moment vector of some bounded and compactly supported $f \in D$. □

References

1. N. I. Akhiezer, M. Krein, *Some Questions in the Theory of Moments*, Transl. of Math. Monographs (American Math. Soc., 1962), Vol. 2.

2. N. I. Akhiezer, **The classical moment problem and some related questions in analysis**, (Oliver & Boyd, 1965).
3. C. Cercignani, **The Boltzmann Equation And Its Applications**, (Springer, 1988).
4. J. D. Chandler, Jr., *Moment problems for compact sets*, *Proc. Amer. Math. Soc.*, **104** (1988) 1134–1140.
5. I. Csiszár, *I-Divergence geometry of probability distributions and minimization problems*, *Ann. of Prob.*, **3** (1975) 146–158.
6. D. C. Dowson, A. Wragg, *Fitting Continuous Probability Density Functions Over* $[0, \infty)$ *Using Information Theory Ideas*, *IEEE Trans. Inf. Theor*, **IT-16** (1973) 226–230.
7. J. M. Einbu, *On the Existence of a Class of Maximum–Entropy Probability Density functions*, *IEEE Trans. Inf. Theory*, **23** (1977) 772–775.
8. M. Frontini, A. Tagliani, *Maximum entropy in the finite Stieltjes and Hamburger moment problem*, *J. Math. Phys.*, **35** (1994) 6748–6756.
9. S. Ihara, **Information Theory for Continuous Systems**, (World Scientific, 1993).
10. E. T. Jaynes, *Information theory and statistical mechanics. I*, *Phys. Review*, **106** (1957) 620–630.
11. E. T. Jaynes, *Information theory and statistical mechanics. II*, *Phys. Review*, **108** (1957) 171–190.
12. M. Junk, *Domain of Definition of Levermore's Five-Moment System*, *J. Stat. Phys.*, **93** (1998) 1998.
13. A. Kociszewski, *The existence conditions for maximum entropy distributions, having prescribed the first three moments*, *J. Phys. A: Math. Gen.*, **19** (1986) L823–L827.
14. C. D. Levermore, *Moment Closure Hierarchies for Kinetic Theories*, *J. Stat. Phys.*, **83** (1996) 1021–1065.
15. A. S. Lewis, *Consistency of moment systems*, *Can. J. Math.*, **47** (1995) 995–1006.
16. L. R. Mead, N. Papanicolaou, *Maximum entropy in the problem of moments*, *J. Math. Phys.*, **25** (1984) 2404–2417.
17. R. T. Rockafellar, **Convex Analysis**, (Princeton University Press, New Jersey, 1970).
18. W. Rudin, **Real and complex analysis**, (Mc Graw Hill, 1987).
19. A. Tagliani, *On the application of maximum entropy to the moments problem*, *J. Math. Phys.*, **34** (1993) 326–337.
20. A. Tagliani, *Maximum entropy in the Hamburger moments problem*, *J. Math. Phys.*, **35** (1994) 5087–5096.