KYBERNETIKA — VOLUME *48* (2012), NUMBER 4, PAGES 637–689

GENERALIZED MINIMIZERS OF CONVEX INTEGRAL FUNCTIONALS, BREGMAN DISTANCE, PYTHAGOREAN IDENTITIES

IMRE CSISZÁR AND FRANTIŠEK MATÚŠ

This paper is dedicated to the memory of Igor Vajda (1942–2010)

Integral functionals based on convex normal integrands are minimized subject to finitely many moment constraints. The integrands are finite on the positive and infinite on the negative numbers, strictly convex but not necessarily differentiable. The minimization is viewed as a primal problem and studied together with a dual one in the framework of convex duality. The effective domain of the value function is described by a conic core, a modification of the earlier concept of convex core. Minimizers and generalized minimizers are explicitly constructed from solutions of modified dual problems, not assuming the primal constraint qualification. A generalized Pythagorean identity is presented using Bregman distance and a correction term for lack of essential smoothness in integrands. Results are applied to minimization of Bregman distances. Existence of a generalized dual solution is established whenever the dual value is finite, assuming the dual constraint qualification. Examples of 'irregular' situations are included, pointing to the limitations of generality of certain key results.

Keywords: maximum entropy, moment constraint, generalized primal/dual solutions, normal integrand, minimizing sequence, convex duality, Bregman projection, conic core, generalized exponential family, inference principles

Classification: 94A17, 49J53, 49K30, 62B10, 65K10, 90C46

CONTENTS

1. INTRODUCTION

1.A. Let μ be a σ-finite measure on a measurable space $(Z, \mathcal{Z})$ and $\varphi\colon Z \to \mathbb{R}^d$ a $\mathcal{Z}$-measurable vector-valued function referred to as the *moment mapping*. The linear space of the $\mathcal{Z}$-measurable functions $g\colon Z \to \mathbb{R}$ with μ-integrable φg is denoted by $\mathcal{G}$, and

$$\mathcal{G}_a \triangleq \left\{ g \in \mathcal{G}\colon \textstyle\int_Z \varphi g \, \mathrm{d}\mu = a \right\}, \qquad a \in \mathbb{R}^d .$$

Here, a is the *moment vector* of $g \in \mathcal{G}_a$ while the functions $g \notin \mathcal{G}$ have no moment vectors. The set of nonnegative functions in $\mathcal{G}/\mathcal{G}_a$ is denoted by $\mathcal{G}^+/\mathcal{G}_a^+$.

This work studies the minimization of *integral functionals* of the form

$$H_\beta(g) \triangleq \textstyle\int_Z \beta(z, g(z)) \, \mu(\mathrm{d}z) \tag{1}$$

subject to $g \in \mathcal{G}_a$. Here, $\beta\colon Z \times \mathbb{R} \to (-\infty, +\infty]$ is a *normal convex integrand* [50, Chapter 14] such that for $z \in Z$ the function $t \mapsto \beta(z,t)$ is finite and strictly convex when $t > 0$ and equals $+\infty$ when $t < 0$. The positive and negative parts of the integral in (1) may be both infinite, in which case the integral is taken to be $+\infty$ by convention. A necessary condition for $H_\beta(g) < +\infty$ is the nonnegativity of g, thus the minimization of H_β is actually over the family $\mathcal{G}_a^+$.

1.B. Minimization problems of this kind emerge across various scientific disciplines, notably in *inference*. When g is an unknown nonnegative function on Z whose moment vector $\int_Z \varphi g \, \mathrm{d}\mu$ can be measured in an experiment providing a vector a, typical *inference principles* call for adopting, as 'best guess' of g, a minimizer of H_β over $\mathcal{G}_a^+$, for a specific choice of β. The unknown g may be a probability density, or its integral may be known otherwise, in which case one coordinate function of the moment mapping φ is taken to be identically 1. Most often *autonomous* integrands are used, which means that β does not depend on the first coordinate $z \in Z$. Typical choices are $t \ln t$ or $-\ln t$ or t^2 giving $H_\beta(g)$ equal to the negative Shannon or Burg entropy[1] or squared L^2-norm of $g \geqslant 0$.

If a 'prior guess' h for g is available, that would be adopted before the measurement, related inference principles suggest to take as 'best guess' after the measurement the minimizer of some 'distance' of g from h subject to $g \in \mathcal{G}_a^+$. Two kinds of non-metric distance often used in this context are *Bregman distances*, see eq. (8), and γ-*divergences*, see eq. (44). The most familiar is the information (I-) divergence, also called Kullback–Leibler distance or relative entropy, that belongs to both families. For h fixed, both

[1] Here, 'entropy' is understood in a wide sense. Shannon entropy in the strict sense refers to the case when μ is the counting measure on a finite or countable set and g is a probability mass function.

kinds of distance are nonnegative integral functionals in g of the form (1), with non-autonomous integrands. The minimization of γ-divergences can be easily reduced to that of integral functionals with autonomous integrands, see Appendix C, but this is not possible for Bregman distances except in special cases. It will become evident below that Bregman distances inevitably enter the minimization of H_β over $\mathcal{G}_a^+$, even in the autonomous case.

Another common approach to inference problems as above is to specify a priori a family of functions f_ϑ parameterized by some ϑ and search in that family a function whose moment vector equals the experimentally measured vector a, thus solve the equation $\int_Z \varphi f_\vartheta \,\mathrm{d}\mu = a$ in the parameter ϑ. There is a close relationship between this approach and the one based on the minimization of H_β given moment constraints. Indeed, the latter will suggest to use the parametric family defined after eq. (4). If some function in that family has moment vector equal to a then this function minimizes H_β on $\mathcal{G}_a^+$. Even if no such function exists, it is usually possible to specify a 'best' function in the family, which is also a 'generalized solution' of the minimization problem.

1.C. The minimization of H_β over $\mathcal{G}_a$, or equivalently over $\mathcal{G}_a^+$, is approached here by convex duality theory, as in [8, 10]. A strategy is to introduce the *value function* J_β by

$$J_\beta(a) \triangleq \inf\nolimits_{g\in\mathcal{G}_a^+} H_\beta(g)\,, \qquad a\in\mathbb{R}^d\,, \tag{2}$$

and to study its conjugate and biconjugate. The value function ranges in $[-\infty,+\infty]$ and is convex. The case when it is identically $+\infty$ is often excluded, writing $J_\beta \not\equiv +\infty$, but it is sometimes not straightforward to recognize. Usually, the value function is *proper*, thus not identically $+\infty$ and never equal to $-\infty$. No general description of the effective domain $\mathit{dom}(J_\beta)$ of the value function, thus the set of $a\in\mathbb{R}^d$ with $J_\beta(a)<+\infty$, seems to be available in literature. This domain is contained in the set of the moment vectors $\int_Z \varphi g\,\mathrm{d}\mu$ of the functions $g\in\mathcal{G}^+$, that is called here the *φ-cone $cn_\varphi(\mu)$ of μ*. Theorem 6.8 describes $\mathit{dom}(J_\beta)$ in terms of faces of $cn_\varphi(\mu)$. A crucial point is to represent the φ-cone via a new concept of *conic core* for Borel measures on $\mathbb{R}^d$, introduced in Section 5 similarly to the convex cores in [25].

The minimization in (2) is the *primal problem* and the infimum $J_\beta(a)$ is the *primal value* for a. The value is attained if a minimizer exists. Since β is strictly convex, if $J_\beta(a)$ is finite then such a minimizer is unique[2] and it is referred to as the *primal solution* g_a for a. A first goal is to recognize whether the primal value is finite, then whether it is attained in which case a construction of the primal solution is desirable. A second goal is to understand the behavior of minimizing sequences g_n in $\mathcal{G}_a^+$ for which $H_\beta(g_n)$ converges to the primal value $J_\beta(a)$. When all minimizing sequences converge to a common limit locally in measure then the limit function will be called the *generalized primal solution* and denoted by $\hat{g}_a$. This convergence, denoted by $g_n \rightsquigarrow \hat{g}_a$, means that $\mu(Y\cap\{|g_n-\hat{g}_a|>\varepsilon\})\to 0$ for every $Y\in\mathcal{Z}$ of finite μ-measure and every $\varepsilon>0$. The fact justifying the terminology that each primal solution is also a generalized primal solution is discussed in Subsection 1.E. after eq. (5), see also Corollary 7.7.

[2] in the sense that any two minimizers are μ-a.e. equal. As a rule, equality of functions is understood μ-a.e., unless $z\in Z$ is included in the notation.

The *convex conjugate* J_β^* of the value function is defined by

$$J_\beta^*(\vartheta) \triangleq \sup_{a\in\mathbb{R}^d} \big[\langle\vartheta, a\rangle - J_\beta(a)\big]\,, \qquad \vartheta\in\mathbb{R}^d\,,$$

where $\langle\cdot,\cdot\rangle$ is the scalar product on $\mathbb{R}^d$. The conjugate β^* of β,

$$\beta^*(z,r) \triangleq \sup_{t\in\mathbb{R}} \big[\,rt - \beta(z,t)\,\big]\,, \qquad z\in Z\,,\ r\in\mathbb{R}\,,$$

is a convex normal integrand, giving rise to the integral functional H_{β^*} and the convex function K_β given by

$$K_\beta(\vartheta) \triangleq \int_Z \beta^*\big(z,\langle\vartheta,\varphi(z)\rangle\big)\,\mu(\mathrm{d}z) = H_{\beta^*}(\langle\vartheta,\varphi\rangle)\,, \qquad \vartheta\in\mathbb{R}^d\,.$$

The following key fact is referred to as the *integral representation* of J_β^*. Its proof, building on [47, 48, 50], is presented in Appendix A.

Theorem 1.1. *If $J_\beta \not\equiv +\infty$ then $J_\beta^* = K_\beta$.*

The convex conjugate J_β^* is proper if and only if J_β is proper [49, Theorem 12.2], which takes place if and only if $\mathit{dom}(J_\beta)$ and $\mathit{dom}(J_\beta^*)$ are both nonempty. When $\mathit{dom}(J_\beta)=\emptyset$, thus Theorem 1.1 does not apply, K_β may differ from $J_\beta^*\equiv -\infty$ and may be a proper convex function, see Example 10.1. A sufficient condition for J_β to be proper is the finiteness of K_β on an open set, see Corollary 3.11, a new result below.

The *biconjugate* of J_β is obtained by conjugating J_β^*,

$$J_\beta^{**}(a) \triangleq \sup_{\vartheta\in\mathbb{R}^d} \big[\langle\vartheta, a\rangle - J_\beta^*(\vartheta)\big]\,, \qquad a\in\mathbb{R}^d\,.$$

If $J_\beta \not\equiv +\infty$ then $J_\beta^{**} = K_\beta^*$, by Theorem 1.1.

The maximization in the conjugation of K_β

$$K_\beta^*(a) = \sup_{\vartheta\in\mathbb{R}^d} \Big[\langle\vartheta, a\rangle - \int_Z \beta^*\big(z,\langle\vartheta,\varphi(z)\rangle\big)\,\mu(\mathrm{d}z)\Big]\,, \qquad a\in\mathbb{R}^d\,, \tag{3}$$

is called the *dual problem* for a, also when $J_\beta\equiv+\infty$, thus when Theorem 1.1 does not apply. The supremum $K_\beta^*(a)$ in (3) is a *dual value.* If it is finite and attained, each maximizer is a *dual solution.* The latter situation is also referred to as existence of Lagrange multipliers, see [10]. There is an intimate relationship between the primal and dual problems discussed in detail below. The primal value always dominates the dual one, see Lemma 4.1. Their distance is the *duality gap.* If the gap is zero, thus the primal and dual values coincide, the dual problem provides valuable information on the primal one. What makes the strategy effective is that the dual problem is finite dimensional and unconstrained.

1.D. Standard results are typically proved under the pair of conditions

$$J_\beta \text{ is proper and } a\in \mathit{ri}(\mathit{dom}(J_\beta)) \tag{PCQ}$$

referred to jointly as the *primal constraint qualification.* Here, ri stands for the relative interior. A convex function that takes the value $-\infty$ somewhere, does so everywhere in the relative interior of its effective domain, and thus the PCQ can be equivalently stated replacing the first condition by $J_\beta(a) > -\infty$. By Remark 6.7, the second condition in PCQ is equivalent to the existence of a positive function g in $\mathcal{G}_a$. Under the PCQ for a, the duality gap is zero, $J_\beta(a) = K_\beta^*(a)$ [49, Theorems 7.4 and 12.2].

A special role will be played by the set Θ_β of those $\vartheta \in dom(K_\beta)$ for which the function $r \mapsto \beta^*(z,r)$ is finite in a neighborhood of $\langle\vartheta, \varphi(z)\rangle$ for μ-a.a. $z \in Z$. This set is convex but possibly empty. The assumption

$$\Theta_\beta \text{ is nonempty} \tag{DCQ}$$

is referred to as the *dual constraint qualification* (DCQ). For sufficient conditions of its validity see Remark 3.3. If the DCQ holds, maximization in the dual problem (3) can be restricted to Θ_β without changing the dual value or loosing a dual solution, see Lemmas 3.4 and 4.7.

Computation of directional derivatives of K_β features the following functions f_ϑ of $z \in Z$,

$$f_\vartheta(z) \triangleq \begin{cases} (\beta^*)'(z, \langle\vartheta, \varphi(z)\rangle)\,, & \text{if } \beta^*(z,\cdot) \text{ is differentiable at } \langle\vartheta, \varphi(z)\rangle, \\ 0\,, & \text{otherwise.} \end{cases} \tag{4}$$

The family $\mathcal{F}_\beta \triangleq \{f_\vartheta \colon \vartheta \in \Theta_\beta\}$ will play a similar role as exponential families do in the case of the negative Shannon entropy functional [4, 17].

1.E. Let the PCQ hold for $a \in \mathbb{R}^d$. Then, the primal and dual values for a are finite, coincide, $J_\beta(a) = K_\beta^*(a)$, and a dual solution $\vartheta \in dom(K_\beta)$ exists, by Lemma 4.2. If the DCQ fails then no primal solution exists and the generalized primal solution does not exist either, see Lemma 4.10 and Theorem 4.17. Otherwise, if $\Theta_\beta \neq \emptyset$, each dual solution ϑ belongs to Θ_β and gives rise to the same function f_ϑ, by Corollary 4.8. This unique function from $\mathcal{F}_\beta$ is called here the *effective dual solution* for a and is denoted by g_a^*, see Remark 4.9. The primal solution g_a exists if and only if $\int_Z \varphi g_a^* \,\mathrm{d}\mu$ exists and equals a, in which case $g_a = g_a^*$, see Lemma 4.10. Alternatively, by the same lemma, the primal solution g_a exists if and only if $\mathcal{F}_\beta$ intersects $\mathcal{G}_a$, thus the equation $\int_Z \varphi f_\vartheta \,\mathrm{d}\mu = a$ has a solution $\vartheta \in \Theta_\beta$. In this case, $\mathcal{F}_\beta \cap \mathcal{G}_a = \{g_a\}$. Subject to the PCQ and DCQ, these conditions are always satisfied if K_β is essentially smooth, then g_a exists and equals g_a^*, see Corollary 4.12. Otherwise, g_a may not exist, for g_a^* need not have moment vector, or its moment vector may differ from a, see Examples 10.3 and 10.6. Under the PCQ for a, however, the DCQ is necessary and sufficient for the existence of the generalized primal solution $\hat{g}_a$, which then coincides with the effective dual solution g_a^*, see Theorem 4.17, a new result.

The main results of this paper include extensions of the above assertions to the cases when the PCQ is relaxed to the finiteness of $J_\beta(a)$, see Section 7. These are relevant when the effective domain of J_β includes a nontrivial relative boundary. Depending on the position of a in the convex set $dom(J_\beta)$, the dual problem is modified, restricting the integration to a subset of Z that corresponds to a face of the convex cone $cn_\varphi(\mu)$. In the modified problem a solution exists and the above assertions have appropriate reformulations, see Theorem 7.6. In particular, a primal solution exists if and only if an extension of the family $\mathcal{F}_\beta$ intersects $\mathcal{G}_a$, see Corollary 7.9. This resolves existence of the primal and generalized primal solutions for $a \in \mathbb{R}^d$ and their construction without the PCQ, whenever $J_\beta(a)$ is finite, even if $J_\beta(b) = -\infty$ for some $b \neq a$.

Another main result is the *generalized Pythagorean identity*, see Theorem 7.10, asserting that for any $a \in \mathbb{R}^d$ with $J_\beta(a)$ finite there exists a unique function $\tilde{g}_a$ such that

$$H_\beta(g) = J_\beta(a) + B_\beta(g, \tilde{g}_a) + C_\beta(g)\,, \qquad g \in \mathcal{G}_a^+\,, \tag{5}$$

under a condition not stronger than the DCQ. Under the PCQ for a and DCQ, the function $\tilde{g}_a$ equals g_a^* while in general $\tilde{g}_a$ is constructed as the effective dual solution of a modified dual problem. In (5), B_β denotes Bregman distance defined by eq. (8) and C_β is a nonnegative correction functional, defined in special cases by eq. (18) and in general by eq. (22). The idea to involve a correction is new even under the PCQ and DCQ. When β is essentially smooth then C_β is identically zero and (5) without the correction becomes a *Pythagorean identity*. In general, omitting C_β in (5) a *Pythagorean inequality* arises. The inequality allows for the conclusions that the generalized primal solution $\hat{g}_a$ exists and equals $\tilde{g}_a$, and that if the primal solution g_a exists then $g_a = \tilde{g}_a = \hat{g}_a$, using Corollary 2.14.

In absence of the PCQ, generalized solutions are introduced also for the dual problem, and their existence is proved under general conditions, see Theorem 9.5. The generalized primal and dual solutions coincide if the duality gap is zero. In general, their Bregman distance is not larger that the duality gap, see Remark 9.10.

1.F. This work is organized as follows. Section 2 collects definitions, technicalities, auxiliary lemmas, and presents general results on the normal integrands and Bregman distances. In Section 3, the function K_β is studied, its directional derivatives computed, and a new sufficient condition for $J_\beta \not\equiv +\infty$ is presented in terms of this function.

Section 4 summarizes results about the primal a dual problems, mostly familiar in the case of autonomous and essentially smooth integrands. These results cover the case when the PCQ holds, but some describe also the more general situation when the primal and dual values coincide and a dual solution exists. The concepts of effective dual solutions and of the correction functional are introduced and a first restricted version of the generalized Pythagorean identity is elaborated, which appears new already in this form. Another new result relates the existence of generalized primal solutions to the DCQ.

Conic cores are introduced and studied in Section 5. In Section 6 a geometric description of the effective domain of the value function is given via the φ-cone of μ. Section 7 formulates the main results on the primal problem without the PCQ, including the generalized Pythagorean identity. The main results are specialized to the problem of Bregman projections in Section 8, and general Pythagorean identities are also treated there. Section 9 is devoted to the dual problem, its main result is a theorem on existence of generalized dual solutions. All examples are collected in Section 10. The relations of this work to previous ones are discussed in Section 11.

Appendix A presents a proof of Theorem 1.1, extending a standard result about the interchange of integration and minimization. Appendix B describes how the usual approach to Shannon entropy maximization is embedded into the framework. Appendix C addresses γ-divergences, and presents a lemma that would admit to restrict attention to finite measures μ throughout this paper.

2. PRELIMINARIES

The terminology and notation of [49] are mostly adopted. If $C \subseteq \mathbb{R}^d$ then $cl(C)$ is the closure and $ri(C)$ the relative interior of C, thus the interior in the topology of the affine hull of C. Subsets of Z on which certain relations hold are denoted briefly by these relations in the curly brackets. For example, the level set $\{z \in Z \colon g(z) > t\}$ of a function $g \colon Z \to \mathbb{R}$ is denoted by $\{g > t\}$. Shorthand notations for μ-almost everywhere are μ-a.e. or $[\mu]$. The function $sgn \colon \mathbb{R} \mapsto \{+, -\}$ assigns $+$ to the *nonnegative* and $-$ to the negative numbers.

2.A. Let Γ denote the family of functions $\gamma \colon \mathbb{R} \to (-\infty, +\infty]$ that are finite and strictly convex for $t > 0$, equal to $+\infty$ for $t < 0$, and satisfy $\gamma(0) = \lim_{t\downarrow 0} \gamma(t)$. In terms of [49], γ is proper and closed, thus lower semicontinuous (lsc). The effective domain $dom(\gamma)$ equals $(0, +\infty)$ or $[0, +\infty)$. The left/right derivatives of γ at $t > 0$ are finite, $\gamma'_-(t) \leqslant \gamma'_+(t)$, and both γ'_- and γ'_+ increase. Let $\gamma'_-(0) \triangleq -\infty$ and $\gamma'_+(0) \triangleq \lim_{t\downarrow 0} \gamma'_+(t)$, which is the standard right derivative at 0 if $\gamma(0) < +\infty$. Further, let $\gamma(+\infty) \triangleq \lim_{t\uparrow+\infty} \gamma(t)$ and $\gamma'(+\infty) \triangleq \lim_{t\uparrow+\infty} \gamma'_+(t)$. If $\gamma'(+\infty) = +\infty$ then γ is called *cofinite*. Otherwise, the function $t \mapsto t\gamma'(+\infty) - \gamma(t)$ is increasing. If it has a finite limit as $t \uparrow +\infty$ then γ is called *asymptotically linear*.

The convex conjugate γ^* of $\gamma \in \Gamma$ is given by $\gamma^*(r) = \sup_{t>0} [r\,t - \gamma(t)]$, $r \in \mathbb{R}$. It is finite and nondecreasing in the interval $(-\infty, \gamma'(+\infty))$, and $\gamma^*(r) = +\infty$ for $r > \gamma'(+\infty)$. When γ is not cofinite then $\gamma^*(\gamma'(+\infty)) = \lim_{t\uparrow+\infty} [t\gamma'(+\infty) - \gamma(t)]$, thus γ^* is finite at $\gamma'(+\infty)$ if and only if γ is asymptotically linear. If $r \downarrow -\infty$ then $\gamma^*(r) \downarrow -\gamma(0)$ where the limit is denoted also by $\gamma^*(-\infty)$. If $\gamma(0)$ is finite then $\gamma^*(r) = -\gamma(0)$ for $r \leqslant \gamma'_+(0)$. The strict convexity of γ implies that γ^* is *essentially smooth* [49, Theorem 26.3], thus γ^* is differentiable in $(-\infty, \gamma'(+\infty))$ and if γ is not cofinite then $(\gamma^*)'(r) \uparrow +\infty$ as $r \uparrow \gamma'(+\infty)$. The latter holds also when γ is cofinite.

Let u denote the function defined for $r < \gamma'(+\infty)$ by $u(r) = (\gamma^*)'(r)$. The following lemma contains an elementary reformulation of the fact that the subgradient mappings of γ and γ^* are mutually inverse [49, Corollary 23.5.1].

Lemma 2.1. *Let $\gamma \in \Gamma$. For $t \geqslant 0$ and $r \in \mathbb{R}$, if $\gamma'_-(t) \leqslant r \leqslant \gamma'_+(t)$ then $r < \gamma'(+\infty)$, $u(r) = t$ and $\gamma^*(r) = tr - \gamma(t)$.*

The function u is nondecreasing on $(-\infty, \gamma'(+\infty))$. It is strictly increasing if and only if γ^* is strictly convex which is equivalent to the essential smoothness of γ, thus the differentiability of γ in $(0, +\infty)$ together with $\gamma'_+(0) = -\infty$. Further, $u(r) \downarrow 0$ as $r \downarrow -\infty$, where the limit 0 is interpreted as $(\gamma^*)'(-\infty)$, and $u(r) \uparrow +\infty$ as $r \uparrow \gamma'(+\infty)$. The function u vanishes on the interval $(-\infty, \gamma'_+(0)]$.

Another reformulation of Lemma 2.1 is convenient for future references.

Lemma 2.2. *For $\gamma \in \Gamma$, u defined as above and $r < \gamma'(+\infty)$*

(i) $\gamma'_-(u(r)) \leqslant r \leqslant \gamma'_+(u(r))$,

(ii) $\gamma(u(r)) = ru(r) - \gamma^*(r)$.

2.B. Bregman distances will be defined by means of the following functions of two variables. Given $\gamma \in \varGamma$, for $s, t \geqslant 0$ let

$$\varDelta_\gamma(s,t) \triangleq \gamma(s) - \gamma(t) - \gamma'_{sgn(s-t)}(t)[s-t] \quad \text{if } \gamma'_+(t) \text{ is finite,} \tag{6}$$

and $\varDelta_\gamma(s,0) \triangleq s \cdot (+\infty)$ otherwise. This definition of $\varDelta_\gamma$ is extended to all $(s,t) \in \mathbb{R}^2$, letting $\varDelta_\gamma(s,t) \triangleq +\infty$ if $s < 0$ or $t < 0$. Beyond these cases, $\varDelta_\gamma(s,t)$ equals $+\infty$ if and only if $0 = s < t$ and $\gamma(0) = +\infty$, or $s > t = 0$ and $\gamma'_+(0) = +\infty$. The strict convexity of γ implies that $\varDelta_\gamma(s,t) \geqslant 0$ with the equality if and only if $s = t \geqslant 0$.

Lemma 2.3. *For $\gamma \in \varGamma$ the function $\varDelta_\gamma \colon \mathbb{R}^2 \to [0, +\infty]$ is lower semicontinuous.*

Proof. By definition, $\varDelta_\gamma$ is lsc at (s,t) if $s < 0$ or $t < 0$. By nonnegativity, $\varDelta_\gamma$ is lsc at (s,t) if $s = t \geqslant 0$. Otherwise, for $s, t \geqslant 0$ different let $s_n \to s$ and $t_n \to t$ such that the sequence $\varDelta_\gamma(s_n, t_n)$ has a finite limit r. Thus, s_n and t_n are eventually nonnegative. If $t > 0$ then the sequence $\gamma'_{sgn(s_n - t_n)}(t_n)(s_n - t_n)$ has at most two accumulation points $\gamma'_\pm(t)(s-t)$. Hence, $\varDelta_\gamma$ is lsc at (s,t) because γ is lsc at s and continuous at t. If $t = 0$ then $\gamma'_+(0) < +\infty$ since r is finite. Therefore, $\varDelta_\gamma(s_n, t_n)$ converges to $\varDelta_\gamma(s,0) = \gamma(s) - \gamma(0) - \gamma'_+(0) \cdot s$ whence $\varDelta_\gamma$ is lsc at (s,t). □

Lemma 2.4. *If $s, t \geqslant 0$ then there exist sequences s_n and t_n of positive rational numbers such that $s_n \to s$, $t_n \to t$ and $\varDelta_\gamma(s_n, t_n) \to \varDelta_\gamma(s,t)$.*

Proof. The assertion is trivial for $s = t$, taking $s_n = t_n$ rational. Otherwise, if $0 \leqslant s < t$ then limiting along sequences $s_n \downarrow s$, $t_n \uparrow t$ works by the continuity of γ'_- from the left. Analogously, if $0 \leqslant t < s$ then $t_n \downarrow t$, $s_n \uparrow s$ works. □

Lemma 2.5. *If $K > 0$ and $\varepsilon > 0$ then*

$$m_\gamma^{K,\varepsilon} \triangleq \min\big\{ \min_{s \leqslant K} \varDelta_\gamma(s, s+\varepsilon), \min_{t \leqslant K} \varDelta_\gamma(t+\varepsilon, t) \big\}$$

is a positive lower bound on $\varDelta_\gamma(s,t)$ whenever $0 \leqslant \min\{s,t\} \leqslant K$ and $|s-t| \geqslant \varepsilon$.

Proof. The two minima are finite and attained since $\varDelta_\gamma$ is nonnegative and lsc, by Lemma 2.3. Then, they cannot vanish whence $m_\gamma^{K,\varepsilon} > 0$. For any $s \geqslant 0$ the function $t \mapsto \varDelta_\gamma(s,t)$ is non-decreasing in $[s, +\infty)$, and thus lower bounded by $\varDelta_\gamma(s, s+\varepsilon)$ for $t \geqslant s + \varepsilon$. For any $t \geqslant 0$ the function $s \mapsto \varDelta_\gamma(s,t)$ is non-decreasing in $[t, +\infty)$, and thus lower bounded by $\varDelta_\gamma(t+\varepsilon, t)$ when $s \geqslant t + \varepsilon$. It follows for $s, t \geqslant 0$ that if $|s-t| \geqslant \varepsilon$ then $\varDelta_\gamma(s,t)$ is lower bounded by $\varDelta_\gamma(s, s+\varepsilon)$ or $\varDelta_\gamma(t+\varepsilon, t)$ which implies the assertion. □

2.C. For fixed s with $\gamma(s)$ finite, the function $t \mapsto \varDelta_\gamma(s,t)$ need not be continuous on $(0, +\infty)$, and it need not be convex even if γ is differentiable. For fixed $t \geqslant 0$, with $t = 0$ allowed only when $\gamma'_+(0)$ is finite, the function

$$[\gamma t] \colon s \mapsto \varDelta_\gamma(s,t)$$

differs from γ on $(0, +\infty)$ by a piecewise linear function. This notation is introduced for the purposes of Section 8.

Lemma 2.6. *Let $\gamma \in \varGamma$ and $t > 0$, or $t = 0$ if $\gamma'_+(0)$ is finite. Then, $[\gamma t] \in \varGamma$, $[\gamma t]'(t) = 0$ for $t > 0$, $[\gamma t]'_+(t) = 0$ for $t = 0$,*

$$\begin{aligned}
[\gamma t]'_\pm(s) &= \gamma'_\pm(s) - \gamma'_{sgn(s-t)}(t)\,, && s \geqslant 0\,,\, t \neq s\,,\\
[\gamma t]'(+\infty) &= \gamma'(+\infty) - \gamma'_+(t), &&\\
[\gamma t]^*(r) &= \gamma^*(r + \gamma'_{sgn(r)}(t)) - \gamma^*(\gamma'_{sgn(r)}(t))\,, && r \in \mathbb{R}\,,\\
([\gamma t]^*)'(r) &= (\gamma^*)'(r + \gamma'_{sgn(r)}(t))\,, && r < [\gamma t]'(+\infty)\,.
\end{aligned}$$

If $t = 0$ and $r < 0$ then $\gamma'_{sgn(r)}(t) = \gamma'_-(0) = -\infty$, and thus the values $\gamma^*(\gamma'_{sgn(r)}(t))$ and $\gamma^*(-\infty)$ are equal to $-\gamma(0)$, see Subsection 2.A. Hence, the right-hand sides of the last two equations in Lemma 2.6 equal 0 when $t = 0$ and $r < 0$.

Proof. The assertions on derivatives of $[\gamma t]$ follow by differentiation in (6) and limiting. Further, $[\gamma t]^*(0) = 0$ because $[\gamma t]$ has the global minimum 0 attained at t. When computing the conjugate

$$[\gamma t]^*(r) = \sup_{s>0} \left[rs - \gamma(s) + \gamma(t) + \gamma'_{sgn(s-t)}(t)[s-t]\right]$$

of $[\gamma t]$ at $r > 0$, the supremum can be restricted to $s > t$, thus

$$[\gamma t]^*(r) = [\gamma(t) - \gamma'_+(t)t] + \sup_{s>t} \left[s[r + \gamma'_+(t)] - \gamma(s)\right].$$

By Lemma 2.1, the first term is equal to $-\gamma^*(\gamma'_+(t))$, and the second one to $\gamma^*(r+\gamma'_+(t))$ since $r > 0$. The conjugate is computed similarly at $r < 0$ when $t > 0$, in which case the supremum can be restricted to $s < t$. If $r < 0$ and $t = 0$ then $[\gamma t]^*(r)$ is equal to $\gamma(0) + \gamma^*(r + \gamma'_+(0)) = 0$. The last assertion follows from the last but one by differentiation. □

Lemma 2.7. *Let $\gamma \in \varGamma$, $s \geqslant 0$ and $t > 0$, or $t = 0$ if $\gamma'_+(0)$ is finite. Then, $\varDelta_{[\gamma t]}(s,t)$ equals $\varDelta_\gamma(s,t)$ and for $r \geqslant 0$ different from t*

$$\varDelta_\gamma(s,r) = \varDelta_{[\gamma t]}(s,r) + [\gamma'_{sgn(s-t)}(t) - \gamma'_{sgn(r-t)}(t)][s-t]\,.$$

Proof. Excluding the case $r = 0$, $\gamma'_+(0) = -\infty$, Lemma 2.6 implies that $[\gamma t]'_+(r)$ is finite and

$$\varDelta_{[\gamma t]}(s,r) = \varDelta_\gamma(s,t) - \varDelta_\gamma(r,t) - [\gamma'_{sgn(s-r)}(r) - \gamma'_{sgn(r-t)}(t)][s-r]$$

when $r \neq t$. The assertion follows by a simple calculation. In the excluded case both sides are 0 or $+\infty$ according as $s = 0$ or $s > 0$. □

2.D. The correction term in (5) will be constructed by means of the function

$$\varUpsilon_\gamma(s,r) \triangleq [\gamma'_{sgn(s-u(r))}(u(r)) - r][s - u(r)]\,, \qquad s \geqslant 0\,,\, r < \gamma'(+\infty)\,, \tag{7}$$

where $\gamma \in \varGamma$ and $u(r) = (\gamma^*)'(r)$ as in Subsection 2.A. By Lemma 2.2*(i)*, $\varUpsilon_\gamma(s,r) \geqslant 0$. This quantity is identically zero if γ is essentially smooth, in which case $u(r) > 0$ and $\gamma'(u(r)) = r$ for $r < \gamma'(+\infty)$. If γ is differentiable on $(0,+\infty)$ then $\varUpsilon_\gamma(s,r)$ equals $|\gamma'_+(0) - r|_+ \cdot s$.

Lemma 2.8. *For $\gamma \in \Gamma$, $s \geqslant 0$ and $r < \gamma'(+\infty)$*

$$\gamma(s) + \gamma^*(r) = rs + \Delta_\gamma(s, u(r)) + \Upsilon_\gamma(s, r)\,.$$

Proof. The assumptions and Lemma 2.2*(i)* imply that $\gamma'_+(u(r))$ is finite. Then, by the definitions of Δ_γ and Υ_γ, the right-hand side equals

$$rs + \gamma(s) - \gamma(u(r)) - r[s - u(r)] = \gamma(s) - \gamma(u(r)) + ru(r)\,.$$

Hence, the assertion follows by Lemma 2.2*(ii)*. □

In Section 9, the following analogue of Δ_γ

$$\Delta_{\gamma^*}(r_2, r_1) \triangleq \gamma^*(r_2) - \gamma^*(r_1) - u(r_1)[r_2 - r_1]\,, \qquad r_1, r_2 < \gamma'(+\infty)\,,$$

with γ^* replacing γ is needed.

Lemma 2.9. *For $\gamma \in \Gamma$ and $r_1, r_2 < \gamma'(+\infty)$*

$$\Delta_{\gamma^*}(r_2, r_1) = \Delta_\gamma(u(r_1), u(r_2)) + \Upsilon_\gamma(u(r_1), r_2)\,.$$

Proof. By definition and Lemma 2.2*(ii)*, $\Delta_{\gamma^*}(r_2, r_1) = \gamma^*(r_2) + \gamma(u(r_1)) - r_2 u(r_1)$. Hence the assertion follows by Lemma 2.8. □

2.E. Following [50, Chapter 14D], a function $f\colon Z \times \mathbb{R}^k \to [-\infty, +\infty]$ is an *integrand* if for every $x \in \mathbb{R}^k$ the function $z \mapsto f(z, x)$ on Z, denoted also by $f(\cdot, x)$, is $\mathcal{Z}$-measurable. An integrand as above is convex/lsc if $f(z, \cdot)$ is convex/lsc for every $z \in Z$, and this convention extends naturally. A function $f\colon Z \times \mathbb{R}^k \to [-\infty, +\infty]$ is a *normal integrand* if $f(z, \cdot)$ is lsc for each $z \in Z$ and $z \mapsto \inf_{x \in D} f(z, x)$ is $\mathcal{Z}$-measurable for each open ball $D \subset \mathbb{R}^k$. This is not the original definition of [50, Chapter 14D] but is equivalent to it by [50, Proposition 14.40 and Proposition 28]. In particular, a normal integrand is an integrand.

The class of integrands $\beta\colon Z \times \mathbb{R} \to (-\infty, +\infty]$ such that $\beta(z, \cdot) \in \Gamma$ for all $z \in Z$ is denoted by B. The assumption $\beta \in B$ is the only restriction on β adopted throughout this paper. By [50, Proposition 14.39], each $\beta \in B$ is a normal convex integrand. The conjugate β^* of $\beta \in B$ inherits this property [50, Theorem 14.50]. Hence, the function $z \mapsto \beta(z, g(z))$ in the definition of H_β, and $z \mapsto \beta^*(z, \langle \vartheta, \varphi(z) \rangle)$ in the definition of K_β, are $\mathcal{Z}$-measurable [50, Proposition 14.28].

By [50, Proposition 14.56], for $t > 0$ the functions $\beta'_+(\cdot, t)$ and $\beta'_-(\cdot, t)$ are $\mathcal{Z}$-measurable and so are their monotone limits $\beta'_+(\cdot, 0)$ and $\beta'(\cdot, +\infty)$. Hence, the set $\{\beta'(\cdot, +\infty) = +\infty\}$ of $z \in Z$ for which $\beta(z, \cdot)$ is cofinite, denoted by $Z_{\beta,\mathrm{cf}}$, belongs to $\mathcal{Z}$. Let $Z_{\beta,\mathrm{al}}$ denote the set of those $z \in Z \setminus Z_{\beta,\mathrm{cf}}$ for which $\beta(z, \cdot)$ is asymptotically linear. Then, $Z_{\beta,\mathrm{al}} \in \mathcal{Z}$ since $z \mapsto \beta^*(z, \beta'(z, +\infty))$ is $\mathcal{Z}$-measurable on $Z \setminus Z_{\beta,\mathrm{cf}}$.

Recall the notation Δ_γ from Subsection 2.D. Assuming the integrand β is in B, the function $(z, s, t) \mapsto \Delta_\gamma(s, t)$, where $\gamma = \beta(z, \cdot)$ depends on z, is denoted by Δ_β.

Lemma 2.10. *If $\beta \in B$ then $\Delta_\beta\colon Z \times \mathbb{R}^2 \to [0, +\infty]$ is a normal integrand.*

P r o o f. By Lemma 2.3, for each $z \in Z$ the function $(s,t) \mapsto \Delta_\beta(z,s,t)$ is lsc. Hence, it suffices to prove that the function $z \mapsto \inf_{(s,t)\in D} \Delta_\beta(z,s,t)$ is $\mathcal{Z}$-measurable for each open ball $D \subset \mathbb{R}^2$. The infimum is identically $+\infty$ if D contains no points with positive coordinates. Otherwise, Lemma 2.4 implies that to each $z \in Z$ and $(s,t) \in D$ with $s,t \geqslant 0$ there exist sequences of positive rational numbers $s_n \to s$ and $t_n \to t$ such that $\Delta_\beta(z,s_n,t_n) \to \Delta_\beta(z,s,t)$. It follows that the above infimum does not change when admitting only $(s,t) \in D$ with positive rational coordinates. The measurability of this infimum over countably many pairs is implied by the measurability of the individual functions $z \mapsto \Delta_\beta(z,s,t)$, $s,t>0$, which in turn follows from the measurability of $\beta(\cdot,s)$, $\beta(\cdot,t)$, $\beta'_+(\cdot,t)$ resp. $\beta'_-(\cdot,t)$. $\square$

2.F. For $\beta \in B$ the *Bregman distance* of $\mathcal{Z}$-measurable functions g,h is defined as

$$B_\beta(g,h) \triangleq \int_Z \Delta_\beta(z,g(z),h(z))\,\mu(\mathrm{d}z) \tag{8}$$

where the integrated function is $\mathcal{Z}$-measurable by Lemma 2.10 and [50, Proposition 14.28]. The quantity $B_\beta(g,h)$ can be finite only for g and h nonnegative μ-a.e., by the definition of Δ_β. The Bregman distance is not a metric on nonnegative functions, nevertheless $B_\beta(g,h) \geqslant 0$ with the equality if and only if $g = h \geqslant 0\,[\mu]$.

The Bregman distances corresponding to the autonomous integrands $t\ln t$, $-\ln t$, t^2 mentioned in Subsection 1.B. are the Kullback–Leibler distance (I-divergence), Itakura-Saito distance, and squared L^2-distance.

Remark 2.11. In the literature the term Bregman distance frequently refers to non-metric distances on $\mathbb{R}^k$ associated with convex functions $\phi\colon \mathbb{R}^k \to (-\infty,+\infty]$. Typically, ϕ is assumed differentiable in the interior of its effective domain and the Bregman distance of x in $\mathit{dom}(\phi)$ from y in the interior is defined as $\phi(x) - \phi(y) - \langle \nabla\phi(y), x-y\rangle$. For $k=1$ and $\phi = \gamma \in \Gamma$ differentiable in $(0,+\infty)$ this reduces to $\Delta_\gamma(x,y)$ from Subsection 2.D. The special case of (8) for $Z = \{1,\dots,k\}$ gives the Bregman distance in this sense of the vector $x = (g(1),\dots,g(k))$ from $y = (h(1),\dots,h(k))$, associated with the convex function

$$\phi(x_1,\dots,x_k) = \textstyle\sum_{z\in Z} \mu(z)\,\beta(z,x_z)\,, \qquad (x_1,\dots,x_k) \in \mathbb{R}^k\,.$$

As this ϕ is a sum of convex functions of individual coordinates, the associated Bregman distance is *separable.*

Lemma 2.12. *For $\beta \in B$ the Bregman distance B_β is lsc for local convergence in measure, jointly in both coordinates.*

P r o o f. Assuming $g_n \rightsquigarrow g$, $h_n \rightsquigarrow h$ and $\liminf_{n\to+\infty} B_\beta(g_n,h_n) = r$, there exists an increasing sequence n_k such that the subsequence g_{n_k} converges to g and h_{n_k} to h, both μ-a.e. Then,

$$r \geqslant \int_Z \liminf_{k\to\infty}\ \Delta_\beta(z,g_{n_k}(z),h_{n_k}(z))\,\mu(\mathrm{d}z) \geqslant B_\beta(g,h)\,,$$

by Fatou lemma and the lower semicontinuity of Δ_γ, see Lemma 2.3. $\square$

Lemma 2.13. *Given $\beta \in B$, to any set $C \in \mathcal{Z}$ of finite μ-measure and positive numbers K, ξ and ε there exists $\delta > 0$ such that for $\mathcal{Z}$-measurable functions g and h either of $\mathsf{B}_\beta(g,h) \leqslant \delta$ or $\mathsf{B}_\beta(h,g) \leqslant \delta$ implies*

$$\mu(C \cap \{|g-h| > \varepsilon\}) < \xi + \mu(C \cap \{g > K\}) .$$

P r o o f. The function $m_\beta^{K,\varepsilon}$ given by

$$m_\beta^{K,\varepsilon}(z) \triangleq \min\big\{ \min_{s \leqslant K} \Delta_\beta(z,s,s+\varepsilon), \min_{t \leqslant K} \Delta_\beta(z,t+\varepsilon,t)\big\} , \qquad z \in Z ,$$

is positive by Lemma 2.5 and $\mathcal{Z}$-measurable by Lemma 2.10 and [50, Proposition 14.37]. This and $\mu(C) < +\infty$ imply that $\mu(C \cap \{m_\beta^{K,\varepsilon} < \eta\}) < \frac{1}{2}\xi$ for $\eta > 0$ sufficiently close to 0. Let δ be equal to $\frac{1}{2}\eta\xi$. By the definition of Bregman distance and Lemma 2.5, whenever $\delta \geqslant \min\{\mathsf{B}_\beta(g,h), \mathsf{B}_\beta(h,g)\}$

$$\begin{aligned} \tfrac{1}{2}\eta\xi &\geqslant \int_{\{g \leqslant K, |g-h| > \varepsilon\}} \min\big\{\Delta_\beta(z,g(z),h(z)), \Delta_\beta(z,h(z),g(z))\big\}\, \mu(\mathrm{d}z) \\ &\geqslant \int_{\{g \leqslant K, |g-h| > \varepsilon\}} m_\beta^{K,\varepsilon}\, \mathrm{d}\mu \geqslant \eta \cdot \mu\big(C \cap \{m_\beta^{K,\varepsilon} \geqslant \eta\} \cap \{g \leqslant K, |g-h| > \varepsilon\}\big) . \end{aligned}$$

Therefore,

$$\mu(C \cap \{|g-h| > \varepsilon\}) \leqslant \mu(C \cap \{m_\beta^{K,\varepsilon} < \eta\}) + \mu(C \cap \{g > K\}) + \tfrac{1}{2}\xi$$

and the assertion follows by the choice of η. □

Corollary 2.14. *If a sequence of $\mathcal{Z}$-measurable functions g_n converges to a $\mathcal{Z}$-measurable function g either in the sense $\mathsf{B}_\beta(g_n,g) \to 0$ or $\mathsf{B}_\beta(g,g_n) \to 0$, then $g_n \rightsquigarrow g$.*

Note that for certain integrands $\beta \in B$ the hypothesis of this corollary admits to conclude even L_1-convergence, see [23, Lemma 3].

3. PRELIMINARIES ON THE DUAL PROBLEM

This section collects auxiliary results on the function K_β and the subset Θ_β of its effective domain. Proposition 3.10 provides a new sufficient condition for $\mathsf{J}_\beta \not\equiv +\infty$.

Lemma 3.1. *If $\vartheta \in \mathsf{dom}(\mathsf{K}_\beta)$ then $\langle \vartheta, \varphi\rangle \leqslant \beta'(\cdot,+\infty)\,[\mu]$ with the strict inequality on $Z \setminus Z_{\beta,\,\mathrm{al}}$.*

P r o o f. The effective domain of K_β consists of those $\vartheta \in \mathbb{R}^d$ for which the positive part of the integral $\int_Z \beta^*(z,\langle\vartheta,\varphi(z)\rangle)\,\mu(\mathrm{d}z)$ is finite. For such ϑ the integrand is finite μ-a.e., and since $\beta \in B$, the assertion follows. □

Recall that the set Θ_β consists of those $\vartheta \in \mathsf{dom}(\mathsf{K}_\beta)$ for which $\beta^*(z,\cdot)$ is finite around $\langle\vartheta,\varphi(z)\rangle$ for μ-a.a. $z \in Z$. Since $\beta \in B$, by properties of γ^* discussed in Subsection 2.A,

$$\vartheta \in \mathsf{dom}(\mathsf{K}_\beta) \text{ belongs to } \Theta_\beta \text{ if and only if } \langle\vartheta,\varphi\rangle < \beta'(\cdot,+\infty)\,[\mu]. \tag{9}$$

The inequality in (9) is equivalent to existence of the derivative $(\beta^*)'(z,\cdot)$ at $\langle\vartheta,\varphi(z)\rangle$.

Corollary 3.2. $\Theta_\beta = \{\vartheta \in \mathsf{dom}(K_\beta)\colon \langle\vartheta,\varphi\rangle < \beta'(\cdot,+\infty)\ \mu\text{-a.e. on } Z_{\beta,\mathrm{al}}\}$.

Remark 3.3. The set Θ_β may be empty even if K_β is proper, see Examples 10.1 and 10.5. However, $\mu(Z_{\beta,\mathrm{al}}) = 0$ implies $\Theta_\beta = \mathsf{dom}(K_\beta)$, by Corollary 3.2. In case $\mathsf{dom}(K_\beta) \neq \emptyset$, another sufficient condition for $\Theta_\beta \neq \emptyset$ is the *moment assumption*

$$\langle\theta,\varphi\rangle > 0 \quad [\mu], \quad \text{for some } \theta \in \mathbb{R}^d. \tag{10}$$

In fact, (10) implies $\vartheta - t\theta \in \Theta_\beta$ for $\vartheta \in \mathsf{dom}(K_\beta)$ and $t > 0$, by the monotonicity of each $\beta^*(z,\cdot)$ and (9). The moment assumption holds for example when one coordinate of φ is a nonzero constant. If $\beta'(\cdot,+\infty) < 0$ $[\mu]$ then (10) holds for each $\theta \in \mathbb{R}^d$ with $-\theta \in \mathsf{dom}(K_\beta)$.

If $\beta'(\cdot,+\infty) > 0$ $[\mu]$ then a trivial sufficient condition for $\Theta_\beta \neq \emptyset$ is the μ-integrability of $\beta^*(z,0)$, for it implies that $\vartheta = \mathbf{0}$ belongs to Θ_β.

Lemma 3.4. *Under the* DCQ $\Theta_\beta \neq \emptyset$, *the sets* $\mathsf{dom}(K_\beta)$ *and* Θ_β *have the same relative interior, and dual values do not change when the maximization in* (3) *is restricted to* Θ_β.

Proof. Let $\vartheta \in \mathsf{dom}(K_\beta)$, $\theta \in \Theta_\beta$ and $0 < t < 1$. Then, $\tau_t = t\theta + (1-t)\vartheta$ belongs to $\mathsf{dom}(K_\beta)$. By Lemma 3.1, $\langle\vartheta,\varphi\rangle \leqslant \beta'(\cdot,+\infty)\,[\mu]$, and $\langle\theta,\varphi\rangle < \beta'(\cdot,+\infty)\,[\mu]$ follows from (9). Then, the latter inequality holds for τ_t instead of θ. By (9), τ_t belongs to Θ_β. Since $\tau_t \to \vartheta$ as $t \downarrow 0$, this proves that $\vartheta \in \mathsf{cl}(\Theta_\beta)$. Therefore, $\mathsf{dom}(K_\beta)$ is contained in $\mathsf{cl}(\Theta_\beta)$. As $\Theta_\beta \subseteq \mathsf{dom}(K_\beta)$ is convex, this proves the first assertion. The second assertion follows from the convexity of K_β. □

Each function $f_\vartheta \in \mathcal{F}_\beta$, $\vartheta \in \Theta_\beta$, defined in (4) is $\mathcal{Z}$-measurable by [50, Proposition 14.56]. It is nonnegative, vanishes on the set $\{\langle\vartheta,\varphi\rangle \leqslant \beta'_+(\cdot,0)\}$, and is μ-a.e. positive on the complement of this set. In particular, if β is essentially smooth then $f_\vartheta > 0$ $[\mu]$.

Remark 3.5. The parametrization of $\mathcal{F}_\beta$ in (4) is not bijective, in general. However, if β is essentially smooth then $(\beta^*)'(z,\cdot)$ is strictly increasing and maps $(-\infty,\beta'(z,+\infty))$ onto $(0,+\infty)$, hence the function $f_\vartheta \in \mathcal{F}_\beta$ does determine the function $\langle\vartheta,\varphi\rangle$. In this case, the parametrization is bijective when the component functions of φ are linearly independent under μ. This has not been assumed because in modified primal and dual problems of Section 7 restrictions of μ occur under which the independence is lost anyhow.

Lemma 3.6. *For* $\theta,\vartheta \in \mathbb{R}^d$ *with* $K_\beta(\theta)$ *and* $K_\beta(\vartheta)$ *finite, the directional derivative*

$$K'_\beta(\vartheta;\,\theta-\vartheta) \triangleq \lim_{t\downarrow 0} \tfrac{1}{t}[K_\beta(\vartheta + t(\theta-\vartheta)) - K_\beta(\vartheta)]$$

is equal to $-\infty$ *if the set* $\{\langle\vartheta,\varphi\rangle = \beta'(\cdot,+\infty)\,,\ \langle\theta,\varphi\rangle \neq \langle\vartheta,\varphi\rangle\}$ *has positive* μ*-measure. Otherwise, it equals*

$$\int_{\{\langle\vartheta,\varphi\rangle<\beta'(\cdot,+\infty)\}} \langle\theta-\vartheta,\varphi(z)\rangle\cdot(\beta^*)'(z,\langle\vartheta,\varphi(z)\rangle)\,\mu(\mathrm{d}z) \tag{11}$$

where the positive part of the integral is finite.

P r o o f. Since $K_\beta(\theta)$ and $K_\beta(\vartheta)$ are finite, $\beta^*(z, \langle\vartheta, \varphi(z)\rangle)$ and $\beta^*(z, \langle\theta, \varphi(z)\rangle)$ are finite for μ-a.a. $z \in Z$. For those z, the function ϕ_z given by

$$\phi_z(t) = \tfrac{1}{t}\left[\beta^*\big(z, \langle\vartheta, \varphi(z)\rangle + t\langle\theta - \vartheta, \varphi(z)\rangle\big) - \beta^*\big(z, \langle\vartheta, \varphi(z)\rangle\big)\right], \qquad t > 0\,,$$

is non-increasing as $t \downarrow 0$, by convexity. If $\langle\vartheta, \varphi(z)\rangle < \beta'(z, +\infty)$ then $\beta^*(z, \cdot)$ is differentiable at $\langle\vartheta, \varphi(z)\rangle$ and $\phi_z(t)$ tends to $\langle\theta - \vartheta, \varphi(z)\rangle \cdot (\beta^*)'(z, \langle\vartheta, \varphi(z)\rangle)$. If $\langle\vartheta, \varphi(z)\rangle = \beta'(z, +\infty)$, thus $z \in Z_{\beta,\,\mathrm{al}}$, then $\phi_z(t)$ tends to 0 or $-\infty$ according to $\langle\theta, \varphi(z)\rangle = \langle\vartheta, \varphi(z)\rangle$ or not. It follows that

$$K'_\beta(\vartheta;\, \theta - \vartheta) = \lim_{t\downarrow 0} \int_{\{\langle\theta,\varphi\rangle \neq \langle\vartheta,\varphi\rangle\}} \phi_z(t)\, \mu(\mathrm{d}z)$$

where the limiting and integration can be interchanged by monotone convergence. Then, the limit is $-\infty$ if the set $\{\langle\vartheta, \varphi\rangle = \beta'(\cdot, +\infty)\,,\, \langle\theta, \varphi\rangle \neq \langle\vartheta, \varphi\rangle\}$ is not μ-negligible. Otherwise, $\langle\vartheta, \varphi\rangle < \beta'(\cdot, +\infty)\,[\mu]$ on $\{\langle\theta, \varphi\rangle \neq \langle\vartheta, \varphi\rangle\}$ on account of Lemma 3.1, and the second assertion follows by the interchange. The integral (11) cannot be $+\infty$ by monotonicity. □

Remark 3.7. If $\vartheta \in \Theta_\beta$ then the integral (11) is equal to $\int_Z \langle\theta - \vartheta, \varphi\rangle f_\vartheta \,\mathrm{d}\mu$, see (9).

Corollary 3.8. *If K_β is finite in a neighborhood of ϑ then $\beta^*(z, \cdot)$ is differentiable at $\langle\vartheta, \varphi(z)\rangle$ for μ-a.a. $z \in Z$ with $\varphi(z) \neq \mathbf{0}$, K_β is differentiable at ϑ, and*

$$\nabla K_\beta(\vartheta) = \int_{\{\varphi\neq\mathbf{0}\}} \varphi(z) \cdot (\beta^*)'(z, \langle\vartheta, \varphi(z)\rangle)\, \mu(\mathrm{d}z)\,.$$

If additionally the set $\{\varphi = \mathbf{0}\}$ is μ-negligible then $\vartheta \in \Theta_\beta$.

P r o o f. Since K_β is convex, the hypothesis implies that all directional derivatives at ϑ are finite. Therefore, Lemma 3.6 implies for each θ sufficiently close to ϑ that $\langle\theta, \varphi\rangle \neq \langle\vartheta, \varphi\rangle$ $[\mu]$ on the set $\{\langle\vartheta, \varphi\rangle = \beta'(\cdot, +\infty)\}$. It follows that on this set $\varphi = \mathbf{0}\,[\mu]$. Hence, recalling Lemma 3.1, the integral (11) can be equivalently taken over $\{\varphi \neq \mathbf{0}\}$ and the assertion follows. If $\mu(\{\varphi = \mathbf{0}\}) = 0$ then $\vartheta \in \Theta_\beta$ by (9). □

Corollary 3.9. *Under finiteness of K_β on an open set, the* DCQ *holds if and only if the set where $\varphi = \mathbf{0}$ and $\lim_{t\uparrow+\infty} \beta(\cdot, t)$ is finite is μ-negligible.*

P r o o f. This follows from (9) and Corollary 3.8. □

Proposition 3.10. *If K_β is finite in a neighborhood of ϑ then $J_\beta(\nabla K_\beta(\vartheta)) < +\infty$.*

P r o o f. Since $K_\beta(\vartheta)$ is finite, both integrals in the sum

$$K_\beta(\vartheta) = \int_{\{\varphi\neq\mathbf{0}\}} \beta^*\big(z, \langle\vartheta, \varphi(z)\rangle\big)\, \mu(\mathrm{d}z) + \int_{\{\varphi=\mathbf{0}\}} \beta^*(\cdot, 0)\, \mathrm{d}\mu$$

are finite. Then, the function $\inf_t \beta(\cdot, t) = -\beta^*(\cdot, 0)$ is μ-integrable on the set $\{\varphi = \mathbf{0}\}$. By Lemma A.5, $\int_{\{\varphi=\mathbf{0}\}} \beta(z, h(z))\, \mu(\mathrm{d}z)$ is finite for some $\mathcal{Z}$-measurable function h. Let g be the function defined by $g(z) = h(z)$ if $\varphi(z) = \mathbf{0}$ and $g(z) = (\beta^*)'(z, \langle\vartheta, \varphi(z)\rangle)$

otherwise, $z \in Z$. Corollary 3.8 implies that this derivative exists μ-a.e. and $\int_Z \varphi g \, \mathrm{d}\mu$ equals $\nabla K_\beta(\vartheta)$. By Lemma 2.2*(ii)*,

$$\int_{\{\varphi \neq \mathbf{0}\}} \beta\big(z, g(z)\big)\, \mu(\mathrm{d}z) = \langle \vartheta, \nabla K_\beta(\vartheta) \rangle - \int_{\{\varphi \neq \mathbf{0}\}} \beta^*\big(z, \langle \vartheta, \varphi(z) \rangle\big)\, \mu(\mathrm{d}z)$$

where the right-hand side is finite. It follows that $H_\beta(g)$ is finite and the assertion follows. □

Corollary 3.11. *If K_β is finite on an open set then J_β is proper, and K_β is lsc.*

Proof. By Proposition 3.10, $J_\beta \not\equiv +\infty$. Therefore, $K_\beta = J_\beta^*$ by Theorem 1.1. The assumption implies that K_β is proper hence so is also J_β. As K_β is a convex conjugate, it is lsc. □

The hypothesis in Proposition 3.10 is equivalent to assuming that K_β is proper and its effective domain has nonempty interior. To conclude $J_\beta \not\equiv +\infty$, neither of the two assumptions can be omitted, see Examples 10.1 and 10.2.

The following lemma addresses, for later reference, essential smoothness of K_β.

Lemma 3.12. *The function K_β is essentially smooth if and only if it is finite on an open set and the subdifferential $\partial K_\beta(\vartheta)$ is empty for those ϑ in $dom(K_\beta)$ that are not in its interior. Here, the condition on emptiness is equivalent to $K'_\beta(\vartheta; \theta - \vartheta) = -\infty$ for each θ in the interior of $dom(K_\beta)$.*

Proof. If K_β is finite on an open set then it is lsc by Corollary 3.11, and differentiable in the interior of $dom(K_\beta)$ by Corollary 3.8. Hence, the assertion follows from [49, Theorem 26.1] and the proof of [49, Lemma 26.2]. □

4. THE CONSTRAINT QUALIFICATIONS

Most results of this section are well-known in more restrictive frameworks, typically for autonomous integrands which are essentially smooth or at least differentiable. The examples and Figure 1 in Section 10 illustrate several situations encountered below. Lemma 4.16 and Theorem 4.17 are new results.

4.A. Two simple lemmas are sent forward.

Lemma 4.1. $J_\beta \geqslant K_\beta^*$.

In particular, $J_\beta = K_\beta^* \equiv +\infty$ if K_β attains the value $-\infty$.

Proof. By Fenchel inequality, for any $\vartheta \in \mathbb{R}^d$ and function g on Z

$$\beta(z, g(z)) + \beta^*\big(z, \langle \vartheta, \varphi(z) \rangle\big) \geqslant \langle \vartheta, \varphi(z) \rangle g(z)\,, \quad z \in Z\,. \tag{12}$$

Integrating, for $a \in \mathbb{R}^d$

$$\int_Z \big[\beta(z, g(z)) + \beta^*\big(z, \langle \vartheta, \varphi(z) \rangle\big)\big]\, \mu(\mathrm{d}z) \geqslant \langle \vartheta, a \rangle\,, \qquad g \in \mathcal{G}_a^+\,. \tag{13}$$

If $K_\beta(\vartheta) = -\infty$ for some $\vartheta \in \mathbb{R}^d$ then this inequality implies $H_\beta \equiv +\infty$. Otherwise, K_β is finite on its effective domain and $H_\beta(g) \geqslant \langle \vartheta, a \rangle - K_\beta(\vartheta)$ holds for every $g \in \mathcal{G}_a$ and $\vartheta \in \mathbb{R}$. □

Lemma 4.2. *If the* PCQ *holds for* $a \in \mathbb{R}^d$ *then* $J_\beta(a) = K^*_\beta(a)$ *and a dual solution for* a *exists.*

P r o o f. By Theorem 1.1, $J^*_\beta = K_\beta$ whence the equality is a consequence of the equality $J_\beta(a) = J^{**}_\beta(a)$, valid in $ri(dom(J_\beta))$. The existence of a dual solution follows from [49, Theorems 23.4 and 23.5]. □

Remark 4.3. The PCQ is also necessary for the existence of a dual solution if $J_\beta \not\equiv +\infty$ and β is essentially smooth, see Corollary 9.3, but not in general, see Example 10.4.

In a 'regular' situation the families $\mathcal{G}_a$ and $\mathcal{F}_\beta$ intersect.

Lemma 4.4. *If* $a \in \mathbb{R}^d$ *and* $f_\theta \in \mathcal{G}_a$ *for some* $\theta \in \Theta_\beta$ *with* $K_\beta(\theta)$ *finite then*

$$J_\beta(a) = H_\beta(f_\theta) = \langle \theta, a\rangle - K_\beta(\theta) = K^*_\beta(a) = J^{**}_\beta(a)\,,$$

the primal solution g_a *exists,* $g_a = f_\theta$*,* θ *is a dual solution for* a*, and* $\mathcal{G}_a \cap \mathcal{F}_\beta = \{g_a\}$*.*

P r o o f. For $\vartheta \in \mathbb{R}^d$, $z \in Z$ and a function g, ineq. (12) is tight if and only if $g(z)$ is equal to the derivative of $\beta^*(z, .)$ at $\langle \vartheta, \varphi(z)\rangle$ [49, Theorem 23.5]. It follows that ineq. (13) is tight if and only if $\vartheta \in \Theta_\beta$ and $g = f_\vartheta$. This and finiteness of $K_\beta(\theta)$ imply $H_\beta(g) + K_\beta(\theta) \geqslant \langle \theta, a\rangle$, $g \in \mathcal{G}_a$, with the equality if and only if $g = f_\theta$. By assumption, $f_\theta \in \mathcal{G}_a$, and thus $H_\beta(f_\theta)$ is finite and equals $\langle \theta, a\rangle - K_\beta(\theta)$. Therefore, $H_\beta(g) \geqslant H_\beta(f_\theta)$, $g \in \mathcal{G}_a$. This proves that $J_\beta(a) = H_\beta(f_\theta)$, the primal solution g_a exists and equals f_θ. By Theorem 1.1, $J^*_\beta = K_\beta$, and thus $K^*_\beta(a) = J^{**}_\beta(a)$. Therefore, the two inequalities in the chain

$$J^{**}_\beta(a) \leqslant J_\beta(a) = \langle \theta, a\rangle - K_\beta(\theta) \leqslant K^*_\beta(a)$$

are tight and θ is a dual solution for a. □

Corollary 4.5. *A sufficient condition for* $J_\beta \not\equiv +\infty$ *is* $f_\vartheta \in \mathcal{G}$ *(integrability of* φf_ϑ*) for some* $\vartheta \in \Theta_\beta$ *with* $K_\beta(\vartheta)$ *finite.*

Remark 4.6. The hypotheses in Lemma 4.4 may hold also when the PCQ is not valid for a, see Example 10.4. If, however, $\beta'_+(\cdot, 0) \equiv -\infty$ and in particular if β is essentially smooth, then all the functions f_ϑ in $\mathcal{F}_\beta$ are positive μ-a.e., and the assumption $f_\vartheta \in \mathcal{G}_a$ does imply $a \in ri(dom(J_\beta))$, by Corollary 6.2 and Lemma 6.5. For $a \in \mathbb{R}^d$ satisfying the PCQ, the assertion of Lemma 4.4 admits a conversion, see Lemma 4.10. The assumption that $K_\beta(\vartheta)$ is finite is essential in Lemma 4.4 and Corollary 4.5, see Example 10.2. That assumption is automatically satisfied when K_β is proper, since $\vartheta \in \Theta_\beta \subseteq dom(K_\beta)$ implies that $K_\beta(\vartheta) < +\infty$.

4.B. This subsection introduces effective dual solutions in general.

Lemma 4.7. *Under the* DCQ*, if* $\vartheta \in \mathbb{R}^d$ *is a dual solution for* $a \in \mathbb{R}^d$ *then* $\vartheta \in \Theta_\beta$ *and*

$$\int_Z \langle \theta - \vartheta, \varphi\rangle f_\vartheta \,\mathrm{d}\mu \geqslant \langle \theta - \vartheta, a\rangle\,, \qquad \theta \in dom(K_\beta)\,, \tag{14}$$

where the integrals are finite.

P r o o f. By the assumption, $K^*_\beta(a)$ is finite and equals $\langle\vartheta, a\rangle - K_\beta(\vartheta)$ whence $K_\beta(\vartheta)$ is finite and K_β proper. If $\theta \in dom(K_\beta)$ and $0 < t < 1$ then

$$\langle\vartheta, a\rangle - K_\beta(\vartheta) = K^*_\beta(a) \geqslant \langle t\theta + (1-t)\vartheta, a\rangle - K_\beta(t\theta + (1-t)\vartheta)$$

by the definition of conjugation. This implies $\frac{1}{t}[K_\beta(\vartheta + t(\theta - \vartheta)) - K_\beta(\vartheta)] \geqslant \langle\theta - \vartheta, a\rangle$. Then, the limit of the left-hand side as $t \downarrow 0$ is at least $\langle\theta - \vartheta, a\rangle$. By Lemma 3.6, the limit is finite, equals the integral (11), and the set $\{\langle\vartheta, \varphi\rangle = \beta'(\cdot, +\infty)\,, \langle\theta, \varphi\rangle \neq \langle\vartheta, \varphi\rangle\}$ is μ-negligible. Therefore, if $\langle\theta, \varphi\rangle < \beta'(\cdot, +\infty)\,[\mu]$ then also $\langle\vartheta, \varphi\rangle < \beta'(\cdot, +\infty)\,[\mu]$, by Lemma 3.1. By (9) and the assumption $\Theta_\beta \neq \emptyset$, it follows that $\vartheta \in \Theta_\beta$. By Remark 3.7, the integral (11) rewrites to $\int_Z \langle\theta - \vartheta, \varphi\rangle f_\vartheta \,\mathrm{d}\mu$ and the assertion follows. ☐

Corollary 4.8. *Under the* DCQ, *if ϑ, θ are dual solutions for $a \in \mathbb{R}^d$ then $f_\vartheta = f_\theta$ (μ-a.e.) and*

$$\int_Z \langle\theta - \vartheta, \varphi\rangle f_\vartheta \,\mathrm{d}\mu = \langle\theta - \vartheta, a\rangle\,.$$

P r o o f. Summing ineq. (14) and its instance with ϑ and θ interchanged,

$$\int_Z \big[\langle\vartheta, \varphi\rangle - \langle\theta, \varphi\rangle\big]\big[f_\vartheta - f_\theta\big] \,\mathrm{d}\mu \leqslant 0\,.$$

The definition (4) of f_ϑ and monotonicity of the functions $(\beta^*)'(z, \cdot)$, $z \in Z$, imply that the product in the integral is nonnegative μ-a.e. Hence, the product vanishes μ-a.e. and $f_\vartheta = f_\theta$ holds by (4). In turn, ineq. (14) is tight. ☐

Remark 4.9. Assuming the DCQ and existence of dual solutions for $a \in \mathbb{R}^d$, Corollary 4.8 implies that each dual solution ϑ for a gives rise to the same function f_ϑ. The unique function defined thereby is denoted by g^*_a and referred to as the *effective dual solution* for a. In Subsection 1.E., g^*_a appeared assuming additionally to the DCQ also the PCQ for a, which is a sufficient but not always necessary condition for existence of dual solutions, see Remark 4.3. Corollary 4.8 goes beyond the situation $a \in ri(dom(J_\beta))$, see Example 10.4.

Lemma 4.10. *Assuming the* PCQ *holds for $a \in \mathbb{R}^d$, the primal solution g_a exists if and only if $\Theta_\beta \neq \emptyset$ and the moment vector of g^*_a exists and equals a. This takes place if and only if $\mathcal{G}_a$ intersects $\mathcal{F}_\beta$. In this case, $g_a = g^*_a$ and $\mathcal{G}_a \cap \mathcal{F}_\beta = \{g_a\}$.*

P r o o f. By Lemma 4.2, a dual solution ϑ for a exists and the primal value $J_\beta(a)$ coincides with the dual one $K^*_\beta(a) = \langle\vartheta, a\rangle - K_\beta(\vartheta)$. Hence, K_β is proper.

If the primal solution $g_a \in \mathcal{G}_a$ exists, thus $H_\beta(g_a) = J_\beta(a)$, then ineq. (13) applies to g_a and ϑ. It rewrites to $H_\beta(g_a) + K_\beta(\vartheta) \geqslant \langle\vartheta, g_a\rangle$. It follows that this inequality is tight whence ineq. (12) is tight for μ-a.a. $z \in Z$. In such a case, $g_a(z)$ is equal to the derivative of $\beta^*(z, .\,)$ at $\langle\vartheta, \varphi(z)\rangle$ [49, Theorem 23.5]. Therefore, $\vartheta \in \Theta_\beta$ and $g_a = f_\vartheta$. This implies that the moment vector of $g^*_a = f_\vartheta$ exists and equals a, and also that $\mathcal{G}_a$ and $\mathcal{F}_\beta$ intersect. Thus, both conditions for existence are necessary.

If $\Theta_\beta \neq \emptyset$ then the dual solution ϑ belongs to this set by Lemma 4.7. If also the moment vector of the effective dual solution $g^*_a = f_\vartheta$ exists and equals a then $\mathcal{G}_a$ intersects $\mathcal{F}_\beta$ in f_ϑ. More generally, if the intersection contains f_θ for some $\theta \in \Theta_\beta$ then $K_\beta(\theta)$

is finite because K_β is proper. Lemma 4.4 implies that the primal solution g_a exists and equals g_a^*, and $\mathcal{G}_a \cap \mathcal{F}_\beta = \{g_a\}$. Thus, both conditions for existence are sufficient and the last assertion holds. □

Proposition 4.11. *If the* DCQ *holds and* $\mathit{dom}(K_\beta)$ *has nonempty interior then each effective dual solution belongs to* $\mathcal{G}$ *(has a moment vector). If the* DCQ *holds and the dual problem for* $a \in \mathbb{R}^d$ *has a solution in the interior of* $\mathit{dom}(K_\beta)$ *then the primal solution* g_a *exists and equals* g_a^*.

Proof. If for some $a \in \mathbb{R}^d$ a dual solution ϑ exists then K_β is proper. The DCQ and Lemma 4.7 imply $\vartheta \in \Theta_\beta$. Since the interior is nonempty, the set of $\theta - \vartheta \in \mathbb{R}^d$ with $\theta \in \mathit{dom}(K_\beta)$ has full dimension d. This and finiteness of the integrals in (14) imply that φf_ϑ is integrable, thus the first assertion holds. If additionally the dual solution ϑ is in the interior then the inequalities in (14) turn into equalities whence $g_a^* = f_\vartheta \in \mathcal{G}_a$. By Lemma 4.4, the primal solution g_a exists and equals g_a^*. □

Corollary 4.12. *If the* DCQ *holds and* K_β *is essentially smooth then the primal solution* g_a *exists for* $a \in \mathbb{R}^d$ *whenever a dual solution does, and then* $g_a = g_a^*$. *In particular, the primal solution exists and equals the effective dual solution for each* $a \in \mathit{ri}(\mathit{dom}(J_\beta))$.

For essential smoothness of K_β, see Lemma 3.12, a trivial sufficient condition is that K_β is proper with its effective domain open. When K_β is essentially smooth, Corollary 3.9 gives a necessary and sufficient condition for the DCQ. Note that the essential smoothness of K_β is not related to that of the integrand β.

Proof. Assuming a dual solution ϑ for $a \in \mathbb{R}^d$ exists, f_ϑ has a moment vector by Proposition 4.11. Thus, for $\theta \in \Theta_\beta$ the directional derivative $K'_\beta(\vartheta; \theta - \vartheta)$ is finite, see Lemma 3.6. In particular, this holds for θ in the interior of $\mathit{dom}(K_\beta)$, hence the essential smoothness of K_β implies by Lemma 3.12 that $\vartheta \in \Theta_\beta$ is not on the boundary of $\mathit{dom}(K_\beta)$. Having ϑ in the interior, the second part of Proposition 4.11 gives that g_a exists and equals g_a^*. The last assertion follows by Lemma 4.2. □

4.C. This subsection introduces the correction term for the functions $g \in \mathcal{G}^+$ whose moment vector belongs to $\mathit{ri}(\mathit{dom}(J_\beta))$, provided the DCQ holds and $J_\beta > -\infty$.

Recall the function $\varUpsilon_\gamma$, $\gamma \in \Gamma$, given by (7) in Subsection 2.D. For $\vartheta \in \Theta_\beta$ let $\varUpsilon_\beta^\vartheta$ denote the function given for $s \geqslant 0$ and $z \in Z$ by

$$\varUpsilon_\beta^\vartheta(z,s) \triangleq \varUpsilon_{\beta(z,\cdot)}(s, \langle\vartheta, \varphi(z)\rangle) = \left[\beta'_{\mathit{sgn}(s-f_\vartheta(z))}(z, f_\vartheta(z)) - \langle\vartheta, \varphi(z)\rangle\right]\left[s - f_\vartheta(z)\right] \tag{15}$$

if $\langle\vartheta, \varphi(z)\rangle < \beta'(z, +\infty)$ and $\varUpsilon_\beta^\vartheta(z,s) \triangleq 0$ otherwise. Since $\vartheta \in \Theta_\beta$ the latter case is μ-negligible. The function $\varUpsilon_\beta^\vartheta$ is nonnegative. It vanishes if β is essentially smooth. If β is differentiable then $\varUpsilon_\beta^\vartheta(z,s) = |\beta'_+(z,0) - \langle\vartheta, \varphi(z)\rangle|_+ \cdot s$.

For a $\mathcal{Z}$-measurable function $g \geqslant 0$ let

$$D_\beta^\vartheta(g) \triangleq \int_Z \varUpsilon_\beta^\vartheta(z, g(z))\, \mu(\mathrm{d}z)\,, \qquad \vartheta \in \Theta_\beta\,. \tag{16}$$

Remark 4.13. The $\mathcal{Z}$-measurability of the nonnegative function in the integral (16) follows from the identity in the proof of Lemma 4.15. By (15), $D_\beta^\vartheta(f_\vartheta) = 0$, and if β is essentially smooth then $D_\beta^\vartheta \equiv 0$ on $\mathcal{G}^+$. If β is differentiable then for $g \geqslant 0$

$$D_\beta^\vartheta(g) = \int_Z |\beta'_+(\cdot,0) - \langle\vartheta,\varphi\rangle|_+ \cdot g \,\mathrm{d}\mu\,, \qquad \vartheta \in \Theta_\beta\,, \tag{17}$$

hence in this case $\langle\vartheta,\varphi\rangle \geqslant \beta'_+(\cdot,0)$ $[\mu]$ is a sufficient condition for $D_\beta^\vartheta(g) = 0$.

Lemma 4.14. *Assuming the* DCQ*, if ϑ, θ are dual solutions for $a \in \mathbb{R}^d$ then*

$$D_\beta^\vartheta(g) = D_\beta^\theta(g), \quad g \in \mathcal{G}_a^+\,.$$

Proof. By Lemma 4.7, ϑ, θ are in Θ_β. By Corollary 4.8, two such solutions ϑ, θ give $f_\vartheta = f_\theta$ $[\mu]$ and $\int_Z \langle\theta-\vartheta,\varphi\rangle f_\vartheta \,\mathrm{d}\mu = \langle\theta-\vartheta,a\rangle$. Eq. (15) implies that

$$\Upsilon_\beta^\vartheta(z,s) = \Upsilon_\beta^\theta(z,s) + \langle\theta-\vartheta,\varphi(z)\rangle\,[s - f_\vartheta(z)]\,, \qquad \text{for } s \geqslant 0 \text{ and } \mu\text{-a.a. } z \in Z\,.$$

For $g \in \mathcal{G}_a^+$ the assertion follows by substituting $s = g(z)$ and integrating. □

The *correction functional* C_β alluded to in (5) is defined, temporarily, for the functions $g \in \mathcal{G}^+$ whose moment vectors belong to $ri(dom(J_\beta))$, by

$$C_\beta(g) \triangleq D_\beta^\vartheta(g) \quad \text{where } \vartheta \in \Theta_\beta \text{ is any dual solution for } \int_Z \varphi g \,\mathrm{d}\mu\,, \tag{18}$$

provided that J_β is proper and the DCQ holds. Here, a dual solution exists by Lemma 4.2 and the definition does not depend on its choice by Lemma 4.14. This definition of the correction functional is further extended in Section 7.

4.D. The key lemma of this subsection is formulated as follows.

Lemma 4.15. *For $a \in \mathbb{R}^d$, $g \in \mathcal{G}_a^+$ and $\vartheta \in \Theta_\beta$*

$$H_\beta(g) = \langle\vartheta,a\rangle - K_\beta(\vartheta) + B_\beta(g,f_\vartheta) + D_\beta^\vartheta(g)\,.$$

Proof. For $z \in Z$, Lemma 2.8 is applied to $\beta(z,\cdot)$ in the role of γ, with $s = g(z)$ and $r = \langle\vartheta,\varphi(z)\rangle$. It follows that if $\langle\vartheta,\varphi(z)\rangle < \beta'(z,+\infty)$, which holds for μ-a.e. $z \in Z$ by (9), then

$$\beta(z,g(z)) + \beta^*(z,\langle\vartheta,\varphi(z)\rangle) = \langle\vartheta,\varphi(z)\rangle g(z) + \Delta_\beta(z,g(z),f_\vartheta(z)) + \Upsilon_\beta^\vartheta(z,g(z))\,.$$

The assertion is obtained by integration since $K_\beta(\vartheta) < +\infty$. □

Below, the generalized Pythagorean identity (5) is formulated, under restrictive assumptions alleviated later in Theorem 7.10.

Lemma 4.16. *Assuming the* PCQ *for $a \in \mathbb{R}^d$ and the* DCQ*,*

$$H_\beta(g) = J_\beta(a) + B_\beta(g,g_a^*) + C_\beta(g), \qquad g \in \mathcal{G}_a^+\,.$$

P r o o f. By Lemmas 4.2 and 4.7, a dual solution ϑ for a exists, it belongs to Θ_β and $\langle\vartheta, a\rangle - K_\beta(\vartheta) = K^*_\beta(a) = J_\beta(a)$. It suffices to invoke Lemma 4.15, using Remark 4.9 and the definition (18). $\square$

The main result of this section is based on the hypotheses that the duality gap between the primal and dual values is zero and the dual value is attained. By Lemma 4.2, the PCQ is a sufficient condition for this. However, it is not necessary, see Example 10.4. In general, it is difficult to recognize whether the gap is zero and this problem is not addressed here.

Theorem 4.17. *For $a \in \mathbb{R}^d$ let the duality gap be zero and the dual value be attained. Then, the generalized primal solution for a exists if and only if the* DCQ *holds, in which case $\hat{g}_a = g^*_a$.*

P r o o f. The hypotheses imply that $J_\beta(a) = K^*_\beta(a)$ which is equal to $\langle\vartheta, a\rangle - K_\beta(\vartheta)$ for some dual solution $\vartheta \in \mathbb{R}^d$. Let g_n be a sequence in $\mathcal{G}^+_a$ such that $H_\beta(g_n)$ converges to the finite primal value $J_\beta(a)$. Then the functions

$$h_n \colon z \mapsto \beta(z, g_n(z)) + \beta^*\big(z, \langle\vartheta, \varphi(z)\rangle\big) - \langle\vartheta, \varphi(z)\rangle g_n(z)$$

are nonnegative, $\mathcal{Z}$-measurable and their integrals $H_\beta(g_n) + K_\beta(\vartheta) - \langle\vartheta, a\rangle$ go to zero. Then, going to a subsequence if necessary, $h_n \to 0$, μ-a.e. If $\Theta_\beta = \emptyset$ then Corollary 3.2 implies that $\langle\vartheta, \varphi(z)\rangle = \beta'(z, +\infty)$ for z in a subset $Y \in \mathcal{Z}$ of $Z_{\beta,\mathrm{al}}$ of positive μ-measure, and thus

$$\big[\beta'(z, +\infty)\, g_n(z) - \beta(z, g_n(z))\big] \to \beta^*\big(z, \beta'(z, +\infty)\big)\,, \qquad \text{for } \mu\text{-a.e. } z \in Y \subseteq Z_{\beta,\mathrm{al}}\,.$$

Since $Y \subseteq Z_{\beta,\mathrm{al}}$ it follows that g_n goes to $+\infty$ μ-a.e. on Y. Therefore, the sequence g_n is not convergent locally in measure, and thus the generalized primal solution for a does not exist.

Assuming the DCQ holds, the dual solution ϑ belongs to Θ_β by Lemma 4.7. Thus, $f_\vartheta = g^*_a$ by Remark 4.9. Lemma 4.15 implies that $H_\beta(g) \geqslant J_\beta(a) + B_\beta(g, g^*_a)$, $g \in \mathcal{G}^+_a$. For any sequence g_n in $\mathcal{G}^+_a$ with $H_\beta(g_n) \to J_\beta(a)$ necessarily $B_\beta(g_n, g^*_a) \to 0$. By Corollary 2.14, $g_n \rightsquigarrow g^*_a$. This proves that the generalized primal solution $\hat{g}_a$ exists and equals g^*_a. $\square$

Let the Bregman closure of $\mathcal{G}_a$ be defined as the set of $\mathcal{Z}$-measurable functions h such that $B_\beta(g_n, h) \to 0$ for some sequence g_n in $\mathcal{G}_a$. In the 'irregular' situation when $\mathcal{G}_a$ and $\mathcal{F}_\beta$ are disjoint, $\mathcal{F}_\beta$ can still intersect the closure. For example, $g^*_a \in \mathcal{F}_\beta$ belongs to the closure if the duality gap is zero and the dual value for a is attained, by the last part of the above proof. The following assertion provides the converse under a regularity condition. Let Θ^+_β denote the set of those $\vartheta \in \Theta_\beta$ for which $\langle\vartheta, \varphi\rangle \geqslant \beta'_+(\cdot, 0)$ $[\mu]$. If β is essentially smooth then $\Theta^+_\beta = \Theta_\beta$. In general, Θ^+_β cannot be replaced by Θ_β in Proposition 4.18, see Example 10.12.

Proposition 4.18. *Let β be differentiable and $a \in \mathbb{R}^d$. If $\vartheta \in \Theta^+_\beta$ has $K_\beta(\vartheta)$ finite and f_ϑ belongs to the Bregman closure of $\mathcal{G}_a$ then $J_\beta(a) = K^*_\beta(a)$, ϑ is a dual solution for a and $f_\vartheta = g^*_a$.*

P r o o f. The hypotheses on β and ϑ imply that D_β^ϑ vanishes on $\mathcal{G}^+$, on account of (17) in Remark 4.13. By assumption, there exists a sequence g_n in $\mathcal{G}_a^+$ with $B_\beta(g_n, f_\vartheta) \to 0$. It follows from Lemma 4.15 that

$$J_\beta(a) \leqslant \lim_{n\to\infty} H_\beta(g_n) = \langle \vartheta, a\rangle - K_\beta(\vartheta) \leqslant K_\beta^*(a)\,.$$

Lemma 4.1 implies that the above inequalities are tight and the assertions follow. □

5. CONIC CORES

A set C in $\mathbb{R}^d$ is a *cone* if it contains the origin $\mathbf{0}$ and $tx \in C$ whenever $t > 0$ and $x \in C$. The convex/conic hull of C is denoted by $\mathsf{conv}(C)/\mathsf{cone}(C)$.

In this section Q typically denotes a σ-finite Borel measure on $\mathbb{R}^d$. A Borel subset of $\mathbb{R}^d$ is *Q-full* if its complement has Q-measure zero. The intersection of all closed Q-full sets in $\mathbb{R}^d$ is the *support* $\mathsf{s}(Q)$ of Q and the intersection of all convex, closed and Q-full sets is the *convex support* $\mathsf{cs}(Q)$ of Q.

The *convex core* $\mathsf{cc}(Q)$ of a probability measure (pm) Q was introduced in [25] as the intersection of all Q-full convex Borel sets in $\mathbb{R}^d$. The concept extends naturally to the σ-finite measures [28] since $\mathsf{cc}(Q)$ does not change when Q is replaced by a finite measure equivalent to Q. An equivalent definition involves the means of probability measures dominated by Q, namely by [25, Theorem 3],

$$\mathsf{cc}(Q) = \left\{ \textstyle\int_{\mathbb{R}^d} x\, P(\mathrm{d}x) \colon P \text{ is a pm with mean and } P \ll Q \right\}. \tag{19}$$

Definition 5.1. The *conic core* $\mathsf{cnc}(Q)$ of a σ-finite Borel measure Q on $\mathbb{R}^d$ is the intersection of the convex, Borel and Q-full cones. The *conic support* $\mathsf{cns}(Q)$ is the intersection of the convex, closed and Q-full cones.

Remark 5.2. The conic core is a convex cone, not necessarily Q-full. The conic support is a convex, closed and Q-full cone. Both are nonempty since they contain the origin; they are equal to the singleton $\{\mathbf{0}\}$ if and only if $\mu(\mathbb{R}^d \setminus \{\mathbf{0}\}) = 0$. The conic core and support do not depend on the weight assigned by Q to $\{\mathbf{0}\}$. Thus, in Definition 5.1, one can admit infinite Q-mass at $\mathbf{0}$, going slightly beyond σ-finiteness.

Some properties of conic cores can be derived also from known facts on the convex cores, but a direct self-contained approach is preferred in this section.

Lemma 5.3. $\mathsf{cl}(\mathsf{cnc}(Q)) = \mathsf{cns}(Q)$ *and* $\mathsf{ri}(\mathsf{cnc}(Q)) = \mathsf{ri}(\mathsf{cns}(Q)) = \mathsf{ri}(\mathsf{cone}(\mathsf{conv}(\mathsf{s}(Q))))$.

P r o o f. By definition, $\mathsf{cnc}(Q) \subseteq \mathsf{cns}(Q)$ whence $\mathsf{cl}(\mathsf{cnc}(Q)) \subseteq \mathsf{cns}(Q)$ using that $\mathsf{cns}(Q)$ is closed. There is no loss of generality in assuming that the dimension of $\{\mathbf{0}\} \cup \mathsf{s}(Q)$ is d. If K is a convex, Borel and Q-full cone then $\mathsf{cl}(K)$ is a convex, closed and Q-full cone, and thus $\mathsf{cns}(Q) \subseteq \mathsf{cl}(K)$. By the assumption on dimension, $\mathsf{ri}(\mathsf{cns}(Q)) \subseteq K$, and, in turn, $\mathsf{ri}(\mathsf{cns}(Q))$ is contained in $\mathsf{cnc}(Q)$. Hence, $\mathsf{cns}(Q)$ is contained in $\mathsf{cl}(\mathsf{cnc}(Q))$. The second assertion is a consequence of the first one and $\mathsf{cns}(Q) = \mathsf{cl}(\mathsf{cone}(\mathsf{conv}(\mathsf{s}(Q))))$. □

A *supporting hyperplane* to a convex cone K is a hyperplane H intersecting K such that one of the closed half-spaces bordered by H contains K. A *nontrivial* supporting hyperplane does not contain K. Any supporting hyperplane to a convex cone K contains the origin. Thus, there exists $\vartheta \in \mathbb{R}^d$ nonzero such that $H = \{x \colon \langle \vartheta, x\rangle = 0\}$ and $K \subseteq H \cup H_<$ where $H_< = \{x \colon \langle \vartheta, x\rangle < 0\}$.

The restriction of a σ-finite measure Q to a Borel set $A \subseteq \mathbb{R}^d$ is denoted by Q^A. It is given by $Q^A(B) = Q(A \cap B)$ for every $B \subseteq \mathbb{R}^d$ Borel.

Lemma 5.4. *If H is a supporting hyperplane of $\mathit{cns}(Q)$ then $\mathit{cnc}(Q) \cap H = \mathit{cnc}(Q^H)$ and $Q(H \setminus \mathit{cl}(\mathit{cnc}(Q) \cap H)) = 0$.*

Proof. The hyperplane H contains the origin and $\mathit{cns}(Q) \subseteq H \cup H_<$ as above. If K is any convex, Borel and Q^H-full cone then $K \cap H$ has the same properties. Then, $(K \cap H) \cup H_<$ is a convex, Borel cone which is Q-full by $\mathit{cns}(Q) \subseteq H \cup H_<$. Hence, $\mathit{cnc}(Q) \subseteq (K \cap H) \cup H_<$ and, intersecting with H, $\mathit{cnc}(Q) \cap H \subseteq K$. This implies that $\mathit{cnc}(Q) \cap H \subseteq \mathit{cnc}(Q^H)$. The opposite inclusion holds because $\mathit{cnc}(Q)$ and H contain $\mathit{cnc}(Q^H)$, by definitions. The first assertion and Lemma 5.3 imply that $\mathit{cl}(\mathit{cnc}(Q) \cap H)$ equals $\mathit{cns}(Q^H)$. Then, the second assertion follows since this set is Q^H-full. $\square$

A *face* of a convex set $C \subseteq \mathbb{R}^d$ is a nonempty convex subset $F \subseteq C$ such that every closed line segment in C with a relative interior point in F is contained in F. The face is proper if $F \neq C$. The relative interiors $\mathit{ri}(F)$ of the faces F partition the set C [49, Theorem 18.2]. A face of a convex cone is a convex cone. The *smallest face* of a convex cone (the intersection of all faces) is either the singleton $\{\mathbf{0}\}$ or a linear subspace of $\mathbb{R}^d$.

Lemma 5.5. *If F is a face of $\mathit{cnc}(Q)$ then $\mathit{cnc}(Q^{\mathit{cl}(F)}) = F$.*

Proof. Induction on the dimension of $\mathit{cnc}(Q)$ is employed. If $F = \mathit{cnc}(Q)$ then the assertion follows by Lemma 5.3 using that $\mathit{cns}(Q)$ is Q-full. Otherwise, F is a proper face and there exists a nontrivial supporting hyperplane H to $\mathit{cnc}(Q)$ containing F [49, Theorem 11.6]. Then, $\mathit{cnc}(Q) \cap H$ is a proper face of $\mathit{cnc}(Q)$ containing F. Lemma 5.4 implies that F is a face of $\mathit{cnc}(Q^H)$. As $\mathit{cnc}(Q^H)$ has smaller dimension than $\mathit{cnc}(Q)$, by induction, $\mathit{cnc}((Q^H)^{\mathit{cl}(F)}) = F$. $\square$

Corollary 5.6. *$Q(\mathit{cl}(F)) > 0$ for each face F of $\mathit{cnc}(Q)$, except perhaps for $F = \{\mathbf{0}\}$.*

Lemma 5.7. *If the integral $\int_{\mathbb{R}^d} x\, Q(\mathrm{d}x)$ exists then it belongs to $\mathit{ri}(\mathit{cnc}(Q))$.*

Proof. Let H be a supporting hyperplane of $\mathit{cns}(Q)$, thus $\mathit{cns}(Q) \subseteq H \cup H_<$ where H and $H_<$ are parameterized by ϑ as above. Denoting the integral by a, $\langle \vartheta, a\rangle$ equals $\int_{H_<} \langle \vartheta, x\rangle\, Q(\mathrm{d}x)$. If $Q(H_<) = 0$ then $\langle \vartheta, a\rangle = 0$ whence $a \in H$. Otherwise, the supporting hyperplane H is nontrivial and $\langle \vartheta, a\rangle < 0$, since $\langle \vartheta, x\rangle < 0$ for $x \in H_<$. Thus, $a \in H_<$. It follows that a belongs to the intersection of all closed halfspaces $H \cup H_<$ as above, which equals $\mathit{cns}(Q)$, but to none of the nontrivial supporting hyperplanes of $\mathit{cns}(Q)$. Therefore, $a \in \mathit{ri}(\mathit{cns}(Q))$ and the assertion follows by Lemma 5.3. $\square$

Corollary 5.8. *If $P \ll Q$ and the integral $\int_{\mathbb{R}^d} x\, P(\mathrm{d}x)$ exists then it belongs to $\mathit{cnc}(Q)$.*

Proof. The integral belongs to $\mathit{ri}(\mathit{cnc}(P)) \subseteq \mathit{cnc}(P) \subseteq \mathit{cnc}(Q)$, where the latter inclusion follows from $P \ll Q$, by the definition of the conic core. $\square$

Lemma 5.9. *Each $a \in \mathsf{ri}(\mathsf{cnc}(Q))$ can be represented as $\int_{\mathbb{R}^d} x\, P(\mathrm{d}x)$ where P is a finite measure that is dominated by Q, has compact support, and its Q-density takes a finite number of values.*

Proof. Let C_Q denote the set of points that can be represented as the above integral with P having the stated properties. By Corollary 5.8, C_Q is a convex subcone of $\mathsf{cnc}(Q)$. Then, it suffices to show that $\mathsf{ri}(\mathsf{cnc}(Q)) \subseteq \mathsf{cl}(C_Q)$, since this implies the assertion $\mathsf{ri}(\mathsf{cnc}(Q)) \subseteq C_Q$.

By Lemma 5.3, each $a \in \mathsf{ri}(\mathsf{cnc}(Q))$ can be represented as $\sum_{y \in Y} t_y\, y$ where Y is a finite subset of $\mathsf{s}(Q)$ and all t_y are positive. Since Q is σ-finite, for any $\varepsilon > 0$ and $y \in Y$ there exists a Borel subset $A_{\varepsilon,y}$ of the ε-ball $B_y(\varepsilon)$ around y such that $Q(A_{\varepsilon,y})$ is positive and finite. Let $y_\varepsilon = \int_{A_{\varepsilon,y}} x\, Q(\mathrm{d}x)/Q(A_{\varepsilon,y})$. Then, each y_ε belongs to C_Q and $\|y_\varepsilon - y\| \leqslant \varepsilon$ because y_ε is the mean of a pm concentrated on $B_y(\varepsilon)$. Therefore, the point $\sum_{y \in Y} t_y\, y_\varepsilon$ of C_Q is arbitrarily close to a if ε is sufficiently small. It follows that $a \in \mathsf{cl}(C_Q)$. □

Theorem 5.10. *The conic core $\mathsf{cnc}(Q)$ consists of the integrals $\int_{\mathbb{R}^d} x\, P(\mathrm{d}x)$ where P runs over all finite measures dominated by Q.*

Proof. One inclusion follows from Corollary 5.8. If $a \in \mathsf{cnc}(Q)$ then $a \in \mathsf{ri}(F)$ for a face F of $\mathsf{cnc}(Q)$. By Lemma 5.5, $a \in \mathsf{ri}(\mathsf{cnc}(Q^{\mathsf{cl}(F)}))$. By Lemma 5.9, $a = \int_{\mathbb{R}^d} x\, P(\mathrm{d}x)$ for a finite measure P dominated by $Q^{\mathsf{cl}(F)}$, and thus by Q. □

Remark 5.11. The measures P in Theorem 5.10 can be also restricted as in Lemma 5.9.

Corollary 5.12. $\mathsf{cnc}(Q) = \mathsf{cone}(\mathsf{cc}(Q))$.

Proof. The equality follows from (19), which is [25, Theorem 3], and Theorem 5.10. □

Remark 5.13. The faces of $\mathsf{cnc}(Q)$ and $\mathsf{cc}(Q)$ are not related to each other in general. However, if Q is concentrated on a hyperplane that does not contain the origin then there is a bijection between the families of faces of $\mathsf{cc}(Q)$ and $\mathsf{cnc}(Q)$, up to the face $\{\mathbf{0}\}$ of the latter: the faces of $\mathsf{cnc}(Q)$ are the conic hulls of the faces of $\mathsf{cc}(Q)$.

Remark 5.14. The number of faces of any convex core is at most countable [25, Theorem 3]. This remains true also for the conic cores. In fact, it suffices to prove that if Q is a pm then $\mathsf{cnc}(Q) = \mathsf{cc}(R)$ for $R = \sum_{n \geqslant 0} Q^{[n]} 2^{-n}$ where $Q^{[n]}$ is the image of Q under the scaling $x \mapsto nx$. If $a \in \mathsf{cnc}(Q)$ then $a = t \int_{\mathbb{R}^d} x\, P(\mathrm{d}x)$ for $t \geqslant 0$ and a pm $P \ll Q$, by Theorem 5.10. Then a is the mean of a convex combination of $P^{[0]}$ and $P^{[n]}$, $n \geqslant t$. Since R dominates these pm's, $a \in \mathsf{cc}(R)$ by (19). In the opposite direction, any convex, Borel and Q-full cone is also $Q^{[n]}$-full whence R-full. Therefore, $\mathsf{cnc}(Q) \supseteq \mathsf{cnc}(R) \supseteq \mathsf{cc}(R)$ by definitions.

6. THE EFFECTIVE DOMAIN OF THE VALUE FUNCTION

Recall that the φ-cone $\mathsf{cn}_\varphi(\mu)$ of μ consists of the moment vectors $\int_Z \varphi g \, \mathrm{d}\mu$ of the functions $g \in \mathcal{G}^+$. The φ-cone contains the effective domain of J_β for each $\beta \in B$. In this section, a geometric description of this domain is presented that relies upon results on conic cores from Section 5 and pays special attention to the relative boundary.

Let μ_φ denote the φ-image of μ. The intuitive meaning of the following lemma is that the φ-cone of μ is equal to the conic core of μ_φ. Conic cores, however, have been defined only for measures on $\mathbb{R}^d$ which are σ-finite on $\mathbb{R}^d \setminus \{\mathbf{0}\}$. As the measure μ_φ may fail to satisfy this condition, an auxiliary measure ν is invoked.

Lemma 6.1. *If ν is a measure equivalent to μ and the image ν_φ is σ-finite on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ then $\mathsf{cn}_\varphi(\mu) = \mathsf{cnc}(\nu_\varphi)$.*

Proof. To prove that $\mathsf{cnc}(\nu_\varphi) \subseteq \mathsf{cn}_\varphi(\mu)$, it can be assumed that ν is finite because measures which are σ-finite and equivalent on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ have the same conic core, see Remark 5.2. Let h be a positive μ-density of ν. By Theorem 5.10, any $a \in \mathsf{cnc}(\nu_\varphi)$ can be written as $\int_{\mathbb{R}^d} x f(x) \, \nu_\varphi(\mathrm{d}x)$ where $f \geqslant 0$ is Borel. If $g(z) = f(\varphi(z))h(z)$ then $\int_Z \varphi g \, \mathrm{d}\mu = \int_Z \varphi f(\varphi) \, \mathrm{d}\nu = a$ which implies $a \in \mathsf{cn}_\varphi(\mu)$.

In the opposite direction, suppose $a = \int_Z \varphi g \, \mathrm{d}\mu$ for $g \in \mathcal{G}^+$. There is no loss of generality in assuming that g vanishes on the set $\{\varphi = \mathbf{0}\}$. Denote by λ the measure with μ-density g. Then, $\{\varphi = 0\}$ is λ-negligible and $\{\varphi \neq \mathbf{0}\}$ partitions into at most countably many sets $A_n \in \mathcal{Z}$ with $\lambda(A_n)$ finite. Let Q_n be the φ-image of λ^{A_n} and Q denote the sum of the measures Q_n. By the assumption on g, $Q(\{\mathbf{0}\}) = 0$. Since

$$+\infty > \int_Z \|\varphi\| \, \mathrm{d}\lambda = \sum_n \int_Z \|\varphi\| \, \mathrm{d}\lambda^{A_n} = \sum_n \int_{\mathbb{R}^d} \|x\| \, Q_n(\mathrm{d}x) \, ,$$

the complement of any ball around the origin has finite Q-measure. Therefore, Q is σ-finite and $a = \int_{\mathbb{R}^d} x \, Q(\mathrm{d}x)$. Since $Q_n \ll \nu_\varphi$, it follows that $a \in \mathsf{cnc}(\nu_\varphi)$, using Corollary 5.8. □

Corollary 6.2. *The set $\{\varphi \notin \mathsf{cl}(\mathsf{cn}_\varphi(\mu))\}$ is μ-negligible.*

Proof. By Lemma 6.1, it ν is finite and equivalent to μ then $\mathsf{cl}(\mathsf{cn}_\varphi(\mu))$ is equal to $\mathsf{cl}(\mathsf{cnc}(\nu_\varphi))$ which is ν_φ-full by Lemma 5.3. This implies that $\varphi^{-1}(\mathsf{cl}(\mathsf{cnc}(\nu_\varphi)))$ is ν-full and the assertion follows. □

Remark 6.3. The φ-cone $\mathsf{cn}_\varphi(\mu)$ can be equivalently defined to consist of the moment vectors $\int_Z \varphi g \, \mathrm{d}\mu$ of the μ-integrable (rather than all) functions g from $\mathcal{G}^+$. This follows from Lemma 6.1 and the first part of its proof. In fact, f can be taken ν_φ-integrable by Theorem 5.10, and then $g = f(\varphi) \cdot h$ is μ-integrable.

Lemma 6.4. *If F is a face of $\mathsf{cn}_\varphi(\mu)$ then $\mathsf{cn}_\varphi(\mu^{\varphi^{-1}(\mathsf{cl}(F))}) = F$.*

Proof. This follows by Lemma 5.5 and Lemma 6.1. □

Lemma 6.5. *The moment vector $\int_Z \varphi g \, \mathrm{d}\mu$ of a function $g \in \mathcal{G}^+$ belongs to a face F of $\mathsf{cn}_\varphi(\mu)$ if and only if g vanishes μ-a.e. on $\{\varphi \notin \mathsf{cl}(F)\}$.*

P r o o f. Since any face contains the origin, there is no loss of generality in assuming that $g = 0$ on $\{\varphi = \mathbf{0}\}$. Let $a = \int_Z \varphi g \,\mathrm{d}\mu$, ν be a finite measure equivalent to μ, and λ denote the measure with μ-density g. Arguing as in the second part of the proof of Lemma 6.1, $a = \int_{\mathbb{R}^d} x\, Q(\mathrm{d}x)$ for a σ-finite measure $Q = \sum_n (\lambda^{A_n})_\varphi \ll \nu_\varphi$. Then, $\mathsf{cnc}(Q) \subseteq \mathsf{cnc}(\nu_\varphi) = \mathsf{cn}_\varphi(\mu)$ by Lemma 6.1, and $a \in \mathsf{ri}(\mathsf{cnc}(Q))$ by Lemma 5.7.

It follows that if $a \in F$ then $\mathsf{cnc}(Q) \subseteq F$. By Lemma 5.3, $\mathsf{cl}(F)$ is Q-full. Hence, $\varphi^{-1}(\mathsf{cl}(F))$ is λ^{A_n}-full for all n, and thus λ-full. This implies that g vanishes μ-a.e. on $\{\varphi \notin \mathsf{cl}(F)\}$. In the opposite direction, the vanishing of g implies that a belongs to $\mathsf{cn}_\varphi(\mu^{\varphi^{-1}(\mathsf{cl}(F))})$ which equals F, by Lemma 6.4. $\square$

Lemma 6.6. *If $\mathsf{dom}(J_\beta)$ is nonempty then $\mathsf{ri}(\mathsf{dom}(J_\beta)) = \mathsf{ri}(\mathsf{cn}_\varphi(\mu))$.*

P r o o f. As $\mathsf{dom}(J_\beta) \subseteq \mathsf{cn}_\varphi(\mu)$ are convex sets, the assertion is a consequence of the inclusion $\mathsf{cn}_\varphi(\mu) \subseteq \mathsf{cl}(\mathsf{dom}(J_\beta))$ which is proved in two steps as follows.

First, let $a \in \mathsf{cn}_\varphi(\mu)$ equal $\int_Z \varphi g \,\mathrm{d}\mu$ for a function $g \in \mathcal{G}$ that is everywhere positive. Since $\mathsf{dom}(J_\beta)$ is nonempty, $H_\beta(h) < +\infty$ for some $h \in \mathcal{G}^+$. Since μ is σ-finite, there exists a positive integrable function f on Z. Let Y_n denote the set of those $z \in Z$ that satisfy the inequality $\beta(z, g(z)) \leqslant \beta(z, h(z)) + nf(z)$. As g and f are positive and $\beta(z, h(z)) < +\infty$ for μ-a.a. $z \in Z$, the sequence $Y_n \in \mathcal{Z}$ increases to a μ-full set. Let g_n equal g on Y_n and h otherwise. It follows that $g_n \in \mathcal{G}$, the moments $a_n = \int_Z \varphi g_n \,\mathrm{d}\mu$ converge to a, and

$$H_\beta(g_n) = \int_{Y_n} \beta(z, g(z))\, \mu(\mathrm{d}z) + \int_{Z \setminus Y_n} \beta(z, h(z))\, \mu(\mathrm{d}z) \leqslant H_\beta(h) + n \int_{Y_n} f \,\mathrm{d}\mu < +\infty\,.$$

Hence, $a_n \in \mathsf{dom}(J_\beta)$, and in turn $a \in \mathsf{cl}(\mathsf{dom}(J_\beta))$.

Second, let a be the moment vector $\int_Z \varphi g \,\mathrm{d}\mu$ of some function $g \in \mathcal{G}^+$ that may vanish somewhere. The family $\mathcal{G}$ contains a positive function f. As the function $g + \frac{1}{n} f$ is positive and belongs to $\mathcal{G}$, its moment $b_n = \int_Z \varphi (g + \frac{1}{n} f) \,\mathrm{d}\mu$ belongs to the closure of $\mathsf{dom}(J_\beta)$ by the previous part of the proof. Since $b_n \to a$, this completes the proof. $\square$

Remark 6.7. The relative interior of $\mathsf{cn}_\varphi(\mu)$ is equal to the set $\mathsf{cn}^+_\varphi(\mu)$ of points that are representable as $\int_Z \varphi g \,\mathrm{d}\mu$ with strictly positive $g \in \mathcal{G}^+$. Indeed, $\mathsf{cn}^+_\varphi(\mu)$ is a convex subset of $\mathsf{ri}(\mathsf{cn}_\varphi(\mu))$ by Lemma 6.5. Arguing as in the second part of the proof of Lemma 6.6, the closure of $\mathsf{cn}^+_\varphi(\mu)$ contains $\mathsf{cn}_\varphi(\mu)$ whence $\mathsf{ri}(\mathsf{cn}_\varphi(\mu)) \subseteq \mathsf{cn}^+_\varphi(\mu)$.

For a face F of $\mathsf{cn}_\varphi(\mu)$, let

$$\omega_{F,\beta} \triangleq \int_{\{\varphi \notin \mathsf{cl}(F)\}} \beta(\cdot, 0) \,\mathrm{d}\mu\,,$$

and $\boldsymbol{F}_\beta$ denote the family of the faces F such that $\omega_{F,\beta} < +\infty$. By Corollary 6.2, $\mathsf{cn}_\varphi(\mu) \in \boldsymbol{F}_\beta$. If $F \subseteq G$ are faces of $\mathsf{cn}_\varphi(\mu)$ and F belongs to $\boldsymbol{F}_\beta$ then so does also G. In particular, $\boldsymbol{F}_\beta$ contains all faces of $\mathsf{cn}_\varphi(\mu)$ if and only if the smallest face of $\mathsf{cn}_\varphi(\mu)$ belongs to $\boldsymbol{F}_\beta$.

Theorem 6.8. *If $\mathsf{dom}(J_\beta)$ is nonempty then it is equal to $\bigcup_{F \in \boldsymbol{F}_\beta} \mathsf{ri}(F)$.*

P r o o f. In this proof the notation, Definition 7.1 and Lemma 7.3 from the beginning of Section 7 are employed. Supposing $\mathsf{dom}(J_\beta) \neq \emptyset$, there exists $g \in \mathcal{G}^+$ such that $H_\beta(g) < +\infty$. If F is a face of $\mathsf{cn}_\varphi(\mu)$ then $g \in \mathcal{G}^+_F$ and $H_{F,\beta}(g) < +\infty$. Denoting $\int_{\{\varphi \in \mathsf{cl}(F)\}} \varphi g \,\mathrm{d}\mu$ by a, the function g is in $\mathcal{G}^+_{F,a}$, and $J_{F,\beta}(a) < +\infty$, see Definition 7.1. Thus, $\mathsf{dom}(J_{F,\beta})$ is nonempty, and by Lemma 6.6 it contains $\mathsf{ri}(\mathsf{cn}_\varphi(\mu^{\{\varphi\in\mathsf{cl}(F)\}}))$ that equals $\mathsf{ri}(F)$ by Lemma 6.4. Hence $J_{F,\beta} < +\infty$ on $\mathsf{ri}(F)$. If $F \in \boldsymbol{F}_\beta$ then (21) from Lemma 7.3 implies that $J_\beta < +\infty$ on $\mathsf{ri}(F)$. This proves that $\mathsf{dom}(J_\beta)$ contains the union.

Conversely, if $\mathsf{dom}(J_\beta)$ intersects a face F of $\mathsf{cn}_\varphi(\mu)$ then $H_\beta(g) < +\infty$ for some $g \in \mathcal{G}$ with the moment vector $\int_Z \varphi g \,\mathrm{d}\mu \in F$. By Lemma 6.5, $g = 0\,[\mu]$ on $\{\varphi \notin \mathsf{cl}(F)\}$, and thus $H_\beta(g) < +\infty$ implies that $F \in \boldsymbol{F}_\beta$. Since $\mathsf{dom}(J_\beta)$ is a subset of $\mathsf{cn}_\varphi(\mu)$, it is contained in the union. □

Corollary 6.9. *The effective domain of J_β is closed to positive multiples.*

Corollary 6.10. *A sufficient condition for $\mathsf{dom}(J_\beta) = \mathsf{cn}_\varphi(\mu)$ is $\int_Z \beta(\cdot,0)\,\mathrm{d}\mu < +\infty$. If $\{\mathbf{0}\}$ is a face of $\mathsf{cn}_\varphi(\mu)$ and $\{\varphi = \mathbf{0}\}$ is μ-negligible then this condition is necessary, as well. If the integral equals $-\infty$ then $J_\beta = -\infty$ on $\mathsf{cn}_\varphi(\mu)$.*

P r o o f. The first assertion follows from Theorem 6.8, for the hypothesis implies that each face of $\mathsf{cn}_\varphi(\mu)$ belongs to $\boldsymbol{F}_\beta$ and J_β is not $+\infty$ at the origin $\mathbf{0}$. By Theorem 6.8, the equality $\mathsf{dom}(J_\beta) = \mathsf{cn}_\varphi(\mu)$ implies that the smallest face of $\mathsf{cn}_\varphi(\mu)$ belongs to $\boldsymbol{F}_\beta$. If this smallest face is the singleton $\{\mathbf{0}\}$ and $\mu(\{\varphi = \mathbf{0}\}) = 0$ then $\omega_{\{\mathbf{0}\},\beta} = \int_Z \beta(\cdot,0)\,\mathrm{d}\mu$, and the second assertion follows. If the integral equals $-\infty$ then $J_\beta(\mathbf{0}) = -\infty$, and the third assertion follows by convexity of J_β. □

Remark 6.11. The hypotheses of the second assertion of Corollary 6.10, guaranteeing the necessity, are valid under the moment assumption. In fact, if (10) holds with $\theta \in \mathbb{R}^d$ then $\mathsf{cn}_\varphi(\mu)$ is contained in the cone $\{\mathbf{0}\} \cup \{x \in \mathbb{R}^d \colon \langle\theta, x\rangle > 0\}$, due to Lemma 6.1. Hence, $\{\mathbf{0}\}$ is the smallest face of $\mathsf{cn}_\varphi(\mu)$. The second hypothesis $\mu(\{\varphi = \mathbf{0}\}) = 0$ follows directly from (10).

Corollary 6.12. *If $\int_{\{\varphi\notin H\}} \beta(\cdot,0)\,\mathrm{d}\mu = +\infty$ for each nontrivial supporting hyperplane H of the cone $\mathsf{cn}_\varphi(\mu)$ then $\mathsf{dom}(J_\beta)$ is either empty or equals $\mathsf{ri}(\mathsf{cn}_\varphi(\mu))$. Moreover, this condition is necessary for $\mathsf{dom}(J_\beta) = \mathsf{ri}(\mathsf{cn}_\varphi(\mu))$.*

P r o o f. Each proper face of $\mathsf{cn}_\varphi(\mu)$ is contained in a nontrivial supporting hyperplane, hence the hypothesis implies that no proper face of $\mathsf{cn}_\varphi(\mu)$ belongs to $\boldsymbol{F}_\beta$. By Theorem 6.8, the first assertion follows. Further, the intersection of $\mathsf{cn}_\varphi(\mu)$ with a nontrivial supporting hyperplane H is a proper face F. If $\mathsf{dom}(J_\beta) = \mathsf{ri}(\mathsf{cn}_\varphi(\mu))$, Theorem 6.8 implies that F does not belong belong to $\boldsymbol{F}_\beta$, thus $\omega_{F,\beta} = +\infty$. By Lemmas 5.4 and 6.1, this is equivalent to $\int_{\{\varphi\notin H\}} \beta(\cdot,0)\,\mathrm{d}\mu = +\infty$. □

Remark 6.13. If β is autonomous, $\beta(z,t) = \gamma(t)$ for $z \in Z$ and $t \in \mathbb{R}$, then Corollary 6.10 states that $\mathsf{dom}(J_\gamma)$ coincides with $\mathsf{cn}_\varphi(\mu)$ whenever $\gamma(0)$ and μ are finite, or $\gamma(0) \leqslant 0$. If $\gamma(0) = +\infty$ then Corollary 6.12 gives that $\mathsf{dom}(J_\gamma)$ is either empty or equals $\mathsf{ri}(\mathsf{cn}_\varphi(\mu))$, as observed in [29, Section 3].

7. DISPENSING WITH THE PCQ IN THE PRIMAL PROBLEM

In this section, the primal problem is studied for the vectors $a \in \mathbb{R}^d$ with a finite value $J_\beta(a)$. The PCQ is not assumed. Recall that the primal and dual problems are constructed from three objects: the measure μ, the moment mapping φ and the integrand $\beta \in B$. The notation has not made explicit the dependence on μ and φ. In this section, μ will be replaced by its restriction to $\{\varphi \in cl(F)\}$ where F is a face of the φ-cone $cn_\varphi(\mu)$. To indicate this restriction, the letter F is added to indices.

Correspondingly, for a face F of $cn_\varphi(\mu)$ let $\mathcal{G}_F$ denote the linear space of the $\mathcal{Z}$-measurable functions $g\colon Z \to \mathbb{R}$ such that φg is μ-integrable on $\{\varphi \in cl(F)\}$, and

$$\mathcal{G}_{F,a} \triangleq \left\{g \in \mathcal{G}_F : \int_{\{\varphi\in cl(F)\}} \varphi g \,\mathrm{d}\mu = a\right\}, \qquad a \in \mathbb{R}^d .$$

Let $\mathcal{G}_F^+/\mathcal{G}_{F,a}^+$ denote the set of nonnegative functions in $\mathcal{G}_F/\mathcal{G}_{F,a}$.

Definition 7.1. For a face F of $cn_\varphi(\mu)$ and $a \in \mathbb{R}^d$, the minimization in

$$J_{F,\beta}(a) \triangleq \inf_{g\in\mathcal{G}_{F,a}^+} H_{F,\beta}(g) \quad \text{where} \quad H_{F,\beta}(g) \triangleq \int_{\{\varphi\in cl(F)\}} \beta(z,g(z))\,\mu(\mathrm{d}z)$$

is the F*-primal problem* and the maximization in

$$K_{F,\beta}^*(a) \triangleq \sup_{\vartheta\in\mathbb{R}^d} [\langle\vartheta,a\rangle - K_{F,\beta}(\vartheta)] \;\text{ where }\; K_{F,\beta}(\vartheta) \triangleq \int_{\{\varphi\in cl(F)\}} \beta^*\big(z,\langle\vartheta,\varphi(z)\rangle\big)\,\mu(\mathrm{d}z)$$

is the F*-dual problem* for a. If $J_{F,\beta}(a)$ is finite and the infimum is attained then the minimizers can be assumed to vanish outside $\{\varphi \in cl(F)\}$. These minimizers define the μ-unique F*-primal solution* $g_{F,a}$ for a. The *generalized* F*-primal solution* $\hat{g}_{F,a}$ is defined likewise.

Remark 7.2. The F-primal/F-dual problem constructed from μ, φ and β is identical to the primal/dual problem constructed from $\mu^{\{\varphi\in cl(F)\}}$, φ and β. Note that if $F = cn_\varphi(\mu)$ then μ does not change when restricted to $\{\varphi \in cl(F)\}$, by Corollary 6.2.

Two lemmas are sent forward.

Lemma 7.3. *Let a be a point in a face F of $cn_\varphi(\mu)$. A function g belongs to $\mathcal{G}_a^+$ if and only if it belongs to $\mathcal{G}_{F,a}^+$ and vanishes μ-a.e. on $\{\varphi \notin cl(F)\}$. Assuming that a left-hand side is not $+\infty$ or that no term on a right-hand side is $+\infty$,*

$$H_\beta(g) = \omega_{F,\beta} + H_{F,\beta}(g)\,, \qquad g \in \mathcal{G}_a^+\,, \tag{20}$$

$$J_\beta(a) = \omega_{F,\beta} + J_{F,\beta}(a)\,. \tag{21}$$

Proof. The first assertion follows from Lemma 6.5. Then, for g in $\mathcal{G}_a^+$

$$\int_Z \beta(z,g(z))\,\mu(\mathrm{d}z) = \int_{\{\varphi\notin cl(F)\}} \beta(z,0)\,\mu(\mathrm{d}z) + \int_{\{\varphi\in cl(F)\}} \beta(z,g(z))\,\mu(\mathrm{d}z)\,,$$

if the integral on the left differs from $+\infty$ or if neither integral on the right equals $+\infty$. Hence, (20) holds, and the quantification there is equivalently over $g \in \mathcal{G}_{F,a}^+$ vanishing on $\{\varphi \notin cl(F)\}$. As $J_{F,\beta}(a)$ equals the infimum of $H_{F,\beta}(g)$ over such functions g, eq. (21) follows. □

Lemma 7.4. *If F is a face of $\mathit{cn}_\varphi(\mu)$, and $J_\beta(a) < +\infty$ for some $a \in \mathit{ri}(F)$, then $\mathit{ri}(\mathit{dom}(J_{F,\beta}))$ is equal to $\mathit{ri}(F)$.*

Proof. The assumptions and Theorem 6.8 imply that $\mathit{ri}(F)$ is contained in $\mathit{dom}(J_\beta)$. By eq. (21), $\mathit{ri}(F)$ is contained in $\mathit{dom}(J_{F,\beta})$. Since $\mathit{dom}(J_{F,\beta}) \subseteq \mathit{cn}_\varphi(\mu^{\varphi^{-1}(\mathit{cl}(F))}) = F$, using Lemma 6.4, the assertion follows. □

The set $\Theta_{F,\beta}$ consists of those $\vartheta \in \mathit{dom}(K_{F,\beta})$ for which the function $r \mapsto \beta^*(z,r)$ is finite around $r = \langle\vartheta, \varphi(z)\rangle$ when $\varphi(z) \in \mathit{cl}(F)$, for μ-a.a. $z \in Z$. If $\vartheta \in \Theta_{F,\beta}$ let

$$f_{F,\vartheta}(z) \triangleq \begin{cases} (\beta^*)'(z, \langle\vartheta, \varphi(z)\rangle)\,, & \text{if } \varphi(z) \in \mathit{cl}(F) \text{ and the derivative exists,} \\ 0\,, & \text{otherwise.} \end{cases}$$

Remark 7.5. The assumption $\Theta_{F,\beta} \neq \emptyset$ plays the role of DCQ in the F-dual problems and is implied by the DCQ for the original problem (3). Under this assumption and attainment in the F-dual problem for $a \in \mathbb{R}^d$, each F-dual solution ϑ belongs to $\Theta_{F,\beta}$ and gives rise to the same function $f_{F,\vartheta}$, arguing as in Remark 4.9. This function is referred to as the *effective F-dual solution* $g^*_{F,a}$ for a.

For $a \in \mathit{cn}_\varphi(\mu)$ let $F(a)$ denote the unique face of $\mathit{cn}_\varphi(\mu)$ whose relative interior contains a.

Theorem 7.6. *For $a \in \mathbb{R}^d$ such that $J_\beta(a)$ is finite*

(i) the $F(a)$-dual value $K^*_{F(a),\beta}(a)$ is attained and $J_\beta(a) = \omega_{F(a),\beta} + K^*_{F(a),\beta}(a)$,

(ii) the primal solution g_a exists if and only if $\Theta_{F(a),\beta} \neq \emptyset$ and the moment vector of the effective $F(a)$-dual solution $g^*_{F(a),a}$ exists and equals a, in which case $g_a = g^*_{F(a),a}$,

(iii) the generalized primal solution exists if and only if $\Theta_{F(a),\beta} \neq \emptyset$, in which case $\hat{g}_a = g^*_{F(a),a}$.

Proof. By finiteness of $J_\beta(a)$ and Lemma 7.4, $\mathit{ri}(\mathit{dom}(J_{F(a),\beta})) = \mathit{ri}(F(a))$. Then, eq. (21) implies that $J_\beta(a)$ equals $\omega_{F(a),\beta} + J_{F(a),\beta}(a)$ where both quantities are finite. Since $J_{F(a),\beta}(a)$ is finite, the PCQ in the $F(a)$-primal problem for a holds. By Lemma 4.2, $J_{F(a),\beta}(a) = K^*_{F(a),\beta}(a)$ and an $F(a)$-dual solution for a exists. These observations imply *(i)*.

By Lemma 7.3, g_a exists if and only if the $F(a)$-primal solution $g_{F(a),a}$ does, in which case they coincide. Knowing that the PCQ holds in the $F(a)$-primal problem for a, the latter existence is equivalent by Lemma 4.10 to $\Theta_{F(a),\beta} \neq \emptyset$ and $g^*_{F(a),a} \in \mathcal{G}_{F(a),a}$, in which case $g_{F(a),a} = g^*_{F(a),a}$ μ-a.e. on $\{\varphi \in \mathit{cl}(F(a))\}$. The incidence means $g^*_{F(a),a} \in \mathcal{G}_a$. These observations imply *(ii)*.

By Remark 7.5 and Theorem 4.17, $\Theta_{F(a),\beta} \neq \emptyset$ is equivalent to existence of the generalized $F(a)$-primal solution $\hat{g}_{F(a),a}$. In this case, $\hat{g}_{F(a),a} = g^*_{F(a),a}$. Lemma 7.3 implies that $\hat{g}_{F(a),a}$ exists if and only if $\hat{g}_a$ does, in which case they coincide. Hence *(iii)* follows. □

Corollary 7.7. *Existence of the primal solution g_a implies that the generalized primal solution $\hat{g}_a$ exists and equals g_a.*

Theorem 7.6 makes sense also when the value function J_β equals $-\infty$ at some point, thus the PCQ holds for no a, as in Example 10.7. There, J_β^* is identically $+\infty$ and the dual values equal $-\infty$, so that the dual problems (3) bear no information on the primal ones. However, J_β can be yet finite at some point a and, due to Theorem 7.6, the $F(a)$-dual problem of Definition 7.1 provides complete understanding of the primal problem for this a.

Definition 7.8. The *extension* $\mathit{exn}(\mathcal{F}_\beta)$ of the family $\mathcal{F}_\beta$ is defined as union of the families $\mathcal{F}_{F,\beta} = \{f_{F,\vartheta} \colon \vartheta \in \Theta_{F,\beta}\}$ over the faces F of $\mathit{cn}_\varphi(\mu)$.

The necessary and sufficient condition for existence of a primal solution can be reformulated by means of the extension, without ever mentioning convex duality.

Corollary 7.9. *Let $a \in \mathbb{R}^d$ and $J_\beta(a)$ be finite. The families $\mathcal{G}_a$ and $\mathit{exn}(\mathcal{F}_\beta)$ intersect if and only if the primal solution for a exists, in which case the intersection equals $\{g_a\}$.*

Proof. By Theorem 7.6*(ii)*, if the primal solution g_a for a exists then $\Theta_{F(a),\beta} \neq \emptyset$, the effective $F(a)$-dual solution $g^*_{F(a),a}$ is defined, and $g_a = g^*_{F(a),a}$ belongs to $\mathcal{G}_a \cap \mathcal{F}_{F(a),\beta}$, contained in $\mathcal{G}_a \cap \mathit{exn}(\mathcal{F}_\beta)$.

In the opposite direction, if $\mathcal{G}_a \cap \mathit{exn}(\mathcal{F}_\beta)$ contains a function $f_{G,\vartheta}$, where G is a face of $\mathit{cn}_\varphi(\mu)$ and $\vartheta \in \Theta_{G,\beta}$, then $f_{G,\vartheta} \in \mathcal{G}_a$ implies $a \in G$, by Lemma 6.5. Therefore, $F(a) \subseteq G$. It follows from $\Theta_{G,\beta} \subseteq \Theta_{F(a),\beta}$ that $\vartheta \in \Theta_{F(a),\beta}$. By Lemma 6.5, $f_{G,\vartheta}$ equals $f_{F(a),\vartheta}$. Hence, $\mathcal{G}_a$ intersects $\mathcal{F}_{F(a),\beta}$. By Lemma 4.10, the $F(a)$-primal solution $g_{F(a),a}$ exists and equals $f_{F(a),\vartheta}$. Therefore, g_a exists by Lemma 7.3, and $f_{G,\vartheta} = g_a$. Thus, $\mathcal{G}_a \cap \mathit{exn}(\mathcal{F}_\beta)$ equals $\{g_a\}$. □

Corollary 7.9 practically amounts to solving the equation $\int_Z \varphi f_{F,\vartheta}\, \mathrm{d}\mu = a$ over the faces F of $\mathit{cn}_\varphi(\mu)$ and $\vartheta \in \Theta_{F,\beta}$, which is within the framework of the last inference principle of Subsection 1.B.

The correction functional C_β has been temporarily defined in eq. (18) under certain conditions which are now relaxed, adapting the former definition to the F-problems. Analogously to (16), for $\vartheta \in \Theta_{F,\beta}$ and $g \geqslant 0$ $\mathcal{Z}$-measurable let

$$D^\vartheta_{F,\beta}(g) \triangleq \int_{\{\varphi \in \mathit{cl}(F)\}} \left[\beta'_{\mathit{sgn}(s - f_{F,\vartheta}(z))}(z, f_{F,\vartheta}(z)) - \langle \vartheta, \varphi(z) \rangle\right] \left[g(z) - f_{F,\vartheta}(z)\right] \mu(\mathrm{d}z)\,.$$

In turn, for any function $g \in \mathcal{G}^+$ with the moment vector $\int_Z \varphi g\, \mathrm{d}\mu$ denoted by a, let

$$C_\beta(g) \triangleq D^\vartheta_{F(a),\beta}(g) \quad \text{where } \vartheta \in \Theta_{F(a),\beta} \text{ is any } F(a)\text{-dual solution for } a\,, \tag{22}$$

provided that $J_\beta(a)$ is finite and $\Theta_{F(a),\beta} \neq \emptyset$. Recalling that in the $F(a)$-problem the PCQ holds for a by the finiteness, the correction functional is thereby well defined as it has been in eq. (18). By (17), if β is differentiable then for g and ϑ as above

$$C_\beta(g) = \int_Z |\beta'_+(\cdot, 0) - \langle \vartheta, \varphi \rangle|_+\, g\, \mathrm{d}\mu \tag{23}$$

where the integral is indeed over Z because $g \in \mathcal{G}^+_a$ vanishes on $\{\varphi \notin \mathit{cl}(F(a))\}$.

Theorem 7.10. *For every $a \in \mathbb{R}^d$ with $J_\beta(a)$ finite and $\Theta_{F(a),\beta} \neq \emptyset$, there exists a ($\mu$-a.e.) unique $\mathcal{Z}$-measurable function $\tilde{g}_a$ such that*

$$H_\beta(g) = J_\beta(a) + B_\beta(g, \tilde{g}_a) + C_\beta(g)\,, \qquad g \in \mathcal{G}_a^+\,. \tag{24}$$

This function $\tilde{g}_a$ equals the effective solution $g^*_{F(a),a}$ of the $F(a)$-dual problem.

Proof. If $\tilde{g}_a$ satisfying (24) exists then its uniqueness follows by considering minimizing sequences that necessarily converge to $\tilde{g}_a$ locally in measure, similarly to arguments at the end of the proof of Theorem 4.17.

It suffices to prove (24) for $g^*_{F(a),a}$ in the role of $\tilde{g}_a$. As in the proof of Theorem 7.6, the first two hypotheses imply that the PCQ holds in the $F(a)$-primal problem for a. Since $\Theta_{F(a),\beta} \neq \emptyset$, the DCQ holds in the $F(a)$-dual problems by Remark 7.5. Hence, Lemma 4.16 implies

$$H_{F(a),\beta}(g) = J_{F(a),\beta}(a) + B_{F(a),\beta}(g, g^*_{F(a),a}) + C_{F(a),\beta}(g)\,, \qquad g \in \mathcal{G}^+_{F(a),a}\,.$$

Since the functions $g \in \mathcal{G}_a^+$ and $g^*_{F(a),a}$ vanish on $\{\varphi \notin cl(F(a))\}$, see Lemma 6.5 and Remark 7.5, the above Bregman distance equals $B_\beta(g, g^*_{F(a),a})$. By the definition (22), $C_{F(a),\beta}(g) = C_\beta(g)$ for $g \in \mathcal{G}_a^+$. The assertion follows by Lemma 7.3 knowing that $\omega_{F(a),\beta}$ is finite. □

Comparing Theorems 7.6*(iii)* and 7.10, the generalized primal solution $\hat{g}_a$ exists if and only if eq. (24) is available, in which case $\tilde{g}_a = \hat{g}_a$. If a primal solution g_a exists then the hypothesis of Theorem 7.10 holds by Theorem 7.6*(ii)*, eq. (24) takes the form

$$H_\beta(g) = H_\beta(g_a) + B_\beta(g, g_a) + C_\beta(g)\,, \qquad g \in \mathcal{G}_a^+\,,$$

and $g_a = \tilde{g}_a$. This implies again Corollary 7.7.

8. BREGMAN PROJECTIONS

For any integrand $\beta \in B$ and $\mathcal{Z}$-measurable function h, the mapping

$$(z,t) \mapsto \Delta_\beta(t, h(z))\,, \qquad z \in Z,\, t \in \mathbb{R}\,,$$

is denoted by $[\beta h]$. It is a normal integrand, see Lemma 2.10 and [50, Proposition 14.45(c)]. It is always *assumed* that $h \geqslant 0$, and $h(z) > 0$ whenever $\beta'_+(z,0) = -\infty$, $z \in Z$. Then for $t \geqslant 0$

$$[\beta h](z,t) = \beta(z,t) - \beta(z,h(z)) - \beta'_{sgn(t-h(z))}(z,h(z))[t-h(z)]\,,$$

and $[\beta h] \in B$, by Lemma 2.6. Since $B_\beta(g,h) = H_{[\beta h]}(g)$ for $\mathcal{Z}$-measurable functions g on Z, the Bregman distance $B_\beta(g,h)$ as a function of g is an integral functional of the form (1).

In this section, the results on the problem (2) are specialized to the minimization in

$$J_{[\beta h]}(a) = \inf\nolimits_{g \in \mathcal{G}_a^+} B_\beta(g,h)\,, \qquad a \in \mathbb{R}^d\,. \tag{25}$$

A (generalized) primal solution of this problem is renamed to a *(generalized) Bregman projection* of h to $\mathcal{G}_a^+$ or to $\mathcal{G}_a$.

The dual problem to (25) features the function $K_{[\beta h]}$ that is equal at $\vartheta \in \mathbb{R}^d$ to the μ-integral of

$$[\beta h]^*(\cdot, \langle\vartheta, \varphi\rangle) = \beta^*(\cdot, \langle\vartheta, \varphi\rangle + \beta'_{\mathit{sgn}(\langle\vartheta,\varphi\rangle)}(\cdot, h)) - \beta^*(\cdot, \beta'_{\mathit{sgn}(\langle\vartheta,\varphi\rangle)}(\cdot, h)) \tag{26}$$

using Lemma 2.6. In particular, $K_{[\beta h]}(\vartheta) = 0$ at $\vartheta = \mathbf{0}$, by the assumption on h. In (26), the missing arguments $z \in Z$ of functions are the same, for example the left hand side denotes the function $z \mapsto [\beta h]^*(z, \langle\vartheta, \varphi(z)\rangle)$. This convention is applied below without any further comments. By Lemma 2.6, $[\beta h]'(\cdot, +\infty)$ equals $\beta'(\cdot, +\infty) - \beta'_+(\cdot, h)$. Referring to (9), the crucial set $\Theta_{[\beta h]}$ consists of those $\vartheta \in \mathit{dom}(K_{[\beta h]})$ that satisfy

$$\langle\vartheta, \varphi(z)\rangle < \beta'(z, +\infty) - \beta'_+(z, h(z)) \qquad \text{for } \mu\text{-a.a. } z \in Z. \tag{27}$$

Since $\beta \in B$, the difference is positive whence $\vartheta = \mathbf{0}$ always belongs to $\Theta_{[\beta h]}$. For $\vartheta \in \Theta_{[\beta h]}$ the functions given by

$$f_{[\beta h],\vartheta} = \begin{cases} (\beta^*)'\big(\cdot\,, \langle\vartheta, \varphi\rangle + \beta'_{\mathit{sgn}(\langle\vartheta,\varphi\rangle)}(\cdot\,, h)\big)\,, & \text{when the ineq. in (27) holds,} \\ 0\,, & \text{otherwise,} \end{cases} \tag{28}$$

form the family $\mathcal{F}_{[\beta h]}$, see (4) and Lemma 2.6. The family contains the function h, parameterized by $\vartheta = \mathbf{0}$, see Lemma 2.1.

Remark 8.1. Equations (26) and (28) admit simplifications on the set $\{h = 0\}$. Namely, if $h = 0$, it is possible to write $\beta'_+(\cdot, h)$ instead of $\beta'_{\mathit{sgn}(\langle\vartheta,\varphi\rangle)}(\cdot, h)$ even if $\langle\vartheta, \varphi\rangle < 0$, due to the fact that $\beta^*(\cdot, r) = -\beta(\cdot, 0)$ and $(\beta^*)'(\cdot, r) = 0$ for all $r \leqslant \beta'_+(\cdot, 0)$. In particular, if β is differentiable and conventionally $\beta'(\cdot, 0) = \beta'_+(\cdot, 0)$, the indices $\mathit{sgn}(\langle\vartheta, \varphi\rangle)$ can be omitted in (26) and (28).

Since the integrand $[\beta h]$ is nonnegative, the PCQ of the problem (25) for $a \in \mathbb{R}^d$ reduces to $a \in \mathit{ri}(\mathit{dom}(J_{[\beta h]}))$. Assuming $J_{[\beta h]} \not\equiv +\infty$, thus existence of $g \in \mathcal{G}^+$ with $B_\beta(g, h)$ finite, the relative interiors of $\mathit{dom}(J_{[\beta h]})$ and $\mathit{cn}_\varphi(\mu)$ coincide, by Lemma 6.6. Then, the PCQ is equivalent to $a \in \mathit{ri}(\mathit{cn}_\varphi(\mu))$, not depending on h.

Theorems 7.6 and 7.10 can be reformulated as follows. In these reformulations, in addition to restricting μ, the integrand β is replaced by $[\beta h]$, as indicated in indices. Accordingly, $(F, [\beta h])$-problems, $(F, [\beta h])$-solutions, etc., come into play.

Recall the running assumption on $h \geqslant 0$, thus finiteness of $\beta'_+(z, h(z))$, $z \in Z$.

Theorem 8.2. *For every* $a \in \mathit{dom}(J_{[\beta h]})$

(i) the $(F(a), [\beta h])$*-dual value is attained and* $J_{[\beta h]}(a) = \omega_{F(a),[\beta h]} + K^*_{F(a),[\beta h]}(a)$*,*

(ii) the Bregman projection $g_{[\beta h],a}$ *of* h *to* $\mathcal{G}_a$ *exists if and only if the moment vector of the effective* $(F(a), [\beta h])$*-dual solution* $g^*_{F(a),[\beta h],a}$ *exists and equals* a*, in which case* $g_{[\beta h],a} = g^*_{F(a),[\beta h],a}$*,*

(iii) the generalized Bregman projection $\hat{g}_{[\beta h],a}$ *of* h *to* $\mathcal{G}_a$ *exists and equals* $g^*_{F(a),[\beta h],a}$*.*

Theorem 8.3. *For every $a \in \mathsf{dom}(J_{[\beta h]})$ there exists a unique $\mathcal{Z}$-measurable function $\tilde{g}_{[\beta h],a}$ such that*

$$B_\beta(g,h) = J_{[\beta h]}(a) + B_{[\beta h]}(g,\tilde{g}_{[\beta h],a}) + C_{[\beta h]}(g)\,, \qquad g \in \mathcal{G}_a^+\,. \tag{29}$$

This function $\tilde{g}_{[\beta h],a}$ equals the effective dual solution $g^*_{F(a),[\beta h],a}$ of the $(F(a),[\beta h])$-dual problem.

As a consequence, the generalized Bregman projection $\hat{g}_{[\beta h],a}$ of h to $\mathcal{G}_a$ equals $\tilde{g}_{[\beta h],a}$. The genuine Bregman projection $g_{[\beta h],a}$ exists if and only if $\hat{g}_{[\beta h],a} \in \mathcal{G}_a$, in which case they coincide and (29) reduces to

$$B_\beta(g,h) = B_{[\beta h]}(g,g_{[\beta h],a}) + B_\beta(g_{[\beta h],a},h) + C_{[\beta h]}(g)\,, \qquad g \in \mathcal{G}_a^+\,. \tag{30}$$

A new feature of eqs. (29) and (30) is the presence of two kinds of Bregman distances, the original one based on β, and another one based on $[\beta h]$. The following lemma presents a comparison.

Lemma 8.4. *For any nonnegative $\mathcal{Z}$-measurable functions $g, \tilde{g}$,*

$$B_\beta(g,\tilde{g}) = B_{[\beta h]}(g,\tilde{g}) + \int_{\{\tilde{g}\neq h\}} \left[\beta'_{\mathsf{sgn}(g-h)}(z,h(z)) - \beta'_{\mathsf{sgn}(\tilde{g}-h)}(z,h(z))\right][g(z)-h(z)]\,\mu(\mathrm{d}z).$$

The integral is nonnegative and vanishes if $\beta(z,\cdot)$ is differentiable at $t = h(z)$ for μ-a.a. $z \in Z$ with $h(z) > 0$.

P r o o f. Applying Lemma 2.7 to $\gamma = \beta(z,\cdot)$, $s = g(z)$, $t = h(z)$ and $r = \tilde{g}(z)$, the above identity follows by integration and implies the remaining assertions. $\square$

On account of Lemma 8.4, if $B_{[\beta h]}(g,g_{[\beta h],a})$ were replaced by $B_\beta(g,g_{[\beta h],a})$ in eq. (30) then an (explicitly specified) nonnegative term had to be subtracted on the right-hand side. This term is not necessarily canceled by the correction term $C_{[\beta h]}(g)$ and it may happen, see Example 10.10, that although the Bregman projection exists, the inequality $B_\beta(g,h) \geqslant B_\beta(g,g_{[\beta h],a}) + B_\beta(g_{[\beta h],a},h)$ does not hold for some $g \in \mathcal{G}_a^+$.

On the other hand, if $\beta(z,\cdot)$, $z \in Z$, is differentiable at each positive number then the two kinds of Bregman distances coincide and in eqs. (29) and (30) the nuisance of Bregman distance based on $[\beta h]$ disappears. By Theorem 8.3, (23) and Lemma 2.6, for $a \in \mathsf{dom}(J_{[\beta h]})$ and $g \in \mathcal{G}_a^+$ (see Remark 8.1 for $\beta'(z,0)$)

$$B_\beta(g,h) = J_{[\beta h]}(a) + B_\beta(g,\tilde{g}_{[\beta h],a}) + \int_Z |\beta'(z,0) - \beta'(z,h(z)) - \langle\vartheta,\varphi(z)\rangle|_+ \cdot g(z)\,\mu(\mathrm{d}z), \tag{31}$$

where $\vartheta \in \Theta_{F(a),[\beta h]}$ is any solution of the $(F(a),[\beta h])$-dual problem. The integral accounts for the lack of essential smoothness of the functions $\beta(z,\cdot)$ at 0.

The results below deal with the special situation when the function h projected to $\mathcal{G}_a$ belongs to the family $\mathcal{F}_\beta$. In this situation, regularity assumptions enable to relate directly the Bregman distance minimization (25), the original primal problem (2), and even its dual (3).

Lemma 8.5. *Let β be essentially smooth and $a \in \mathbb{R}^d$.*

(i) If $\theta \in \Theta_\beta$ and $K_\beta(\theta)$ is finite then the primal problem (2) for a is equivalent to minimization of $B_\beta(g, f_\theta)$ subject to $g \in \mathcal{G}_a^+$.

(ii) If $g \in \mathcal{G}_a^+$, $H_\beta(g)$ is finite, and the DCQ *holds then the dual problem (3) for a is equivalent to minimization of $B_\beta(g, f_\theta)$ subject to $\theta \in \Theta_\beta$.*

Proof. By essential smoothness, Lemma 4.15 implies that for $g \in \mathcal{G}_a^+$ and $\theta \in \Theta_\beta$

$$H_\beta(g) = \langle \theta, a \rangle - K_\beta(\theta) + B_\beta(g, f_\theta) \tag{32}$$

because the term $D_\beta^\theta(g)$ vanishes. Then, *(i)* follows and likewise *(ii)*, by Lemmas 3.4 and 4.7. □

Lemma 8.5 is well known, and so are also the following results when β is essentially smooth. Below, however, only the differentiability of this integrand is required, thus the correction term need not vanish. The role of h is played by f_θ with $\theta \in \Theta_\beta^+$. Recall that Θ_β^+ consists of those $\theta \in \Theta_\beta$ for which $\langle \theta, \varphi \rangle \geqslant \beta'(\cdot, 0)$ $[\mu]$. When β is not essentially smooth, this is, in general, a proper subset of Θ_β, and the results below need not hold for all $\theta \in \Theta_\beta$, see Example 10.11.

Theorem 8.6. *Suppose β is differentiable and $\Theta_\beta^+ \neq \emptyset$.*

(i) If $J_\beta \not\equiv +\infty$ then J_β and K_β are proper. If, in addition, $\theta \in \Theta_\beta^+$ then $dom(J_{[\beta f_\theta]})$ equals $dom(J_\beta)$ and for a in this domain

$$H_\beta(g) - J_\beta(a) = B_\beta(g, f_\theta) - J_{[\beta f_\theta]}(a) = B_\beta(g, \hat{g}_a) + C_\beta(g), \qquad g \in \mathcal{G}_a^+, \ \theta \in \Theta_\beta^+. \tag{33}$$

Further, for such a the generalized Bregman projection $\hat{g}_{[\beta f_\theta],a}$ exists and equals $\hat{g}_a$, and the condition $\hat{g}_a \in \mathcal{G}_a^+$ is necessary and sufficient both for the existence of the primal solution g_a and of the Bregman projection $g_{[\beta f_\theta],a}$, in which case both are equal to $\hat{g}_a$.

(ii) If $J_\beta \equiv +\infty$ then $dom(J_{[\beta f_\theta]}) = dom(J_\beta) = \emptyset$ for $\theta \in \Theta_\beta^+$ with $K_\beta(\theta)$ finite.

For the last assertion, the finiteness hypothesis is essential, see Example 10.2.

Proof. *(i)* By the hypotheses and Theorem 1.1, J_β and $K_\beta = J_\beta^*$ have nonempty effective domains, hence both are proper. Lemma 4.15 implies eq. (32) also under the current hypotheses, because if β is differentiable and $\theta \in \Theta_\beta^+$ then $D_\beta^\theta(g) = 0$ due to eq. (17). It follows minimizing in eq. (32) over $g \in \mathcal{G}_a^+$, or trivially if $\mathcal{G}_a^+ = \emptyset$, that

$$J_\beta(a) = \langle \theta, a \rangle - K_\beta(\theta) + J_{[\beta f_\theta]}(a), \qquad a \in \mathbb{R}^d, \ \theta \in \Theta_\beta^+,$$

where $K_\beta(\theta)$ is finite since K_β is proper and $\theta \in dom(K_\beta)$. This identity implies the claimed equality of domains. In case $a \in dom(J_\beta)$ it also implies, by subtraction from eq. (32), the first equality in (33). The second equality in (33) follows from (24), as $\tilde{g}_a$ in (24) equals the generalized primal solution $\hat{g}_a$, see also Theorem 7.6(iii). The assertions about (generalized) Bregman projections immediately follow from the first equality in eq. (33).

(ii) The above proof of $dom(J_\beta) = dom(J_{[\beta f_\theta]})$ goes through also when $J_\beta \equiv +\infty$ but $K_\beta(\theta)$ is finite, since the hypothesis $J_\beta \not\equiv +\infty$ has been used only to guarantee that finiteness. □

The following lemma relates the dual functions K_β and $K_{[\beta h]}$, as well as the families $\mathcal{F}_\beta$ and $\mathcal{F}_{[\beta h]}$, corresponding to the problems (2) and (25).

Lemma 8.7. *If β is differentiable then for $\theta \in \Theta_\beta^+$ with $K_\beta(\theta)$ finite*

$$K_{[\beta f_\theta]}(\vartheta) = K_\beta(\vartheta + \theta) - K_\beta(\theta)\,, \qquad \vartheta \in \mathbb{R}^d\,, \tag{34}$$

and $\vartheta \in \Theta_{[\beta f_\theta]}$ is equivalent to $\vartheta + \theta \in \Theta_\beta$, in which case $f_{[\beta f_\theta],\vartheta} = f_{\vartheta+\theta}$. The families $\mathcal{F}_\beta$ and $\mathcal{F}_{[\beta f_\theta]}$ coincide, and so do their extensions $\mathsf{exn}(\mathcal{F}_\beta)$ and $\mathsf{exn}(\mathcal{F}_{[\beta f_\theta]})$.

P r o o f. The value $K_{[\beta f_\theta]}(\vartheta)$ is obtained by integrating the function $[\beta f_\theta]^*(\cdot, \langle \vartheta, \varphi\rangle)$ which is equal to $\beta^*(\cdot, \langle \vartheta, \varphi\rangle + \beta'(\cdot, f_\theta)) - \beta^*(\cdot, \beta'(\cdot, f_\theta))$ by (26) and Remark 8.1. Here, $\beta'(\cdot, f_\theta) = \langle \theta, \varphi\rangle$ by the definition (4) of f_θ, differentiability, Lemma 2.2*(i)* and the assumption $\theta \in \Theta_\beta^+$. Therefore, $K_{[\beta f_\theta]}(\vartheta)$ is obtained by integrating the difference $\beta^*(\cdot, \langle \vartheta, \varphi\rangle + \langle \theta, \varphi\rangle) - \beta^*(\cdot, \langle \theta, \varphi\rangle)$. This proves (34), using that $K_\beta(\theta)$ is finite.

By (9), $\vartheta \in \Theta_{[\beta f_\theta]}$ is equivalent to $\vartheta \in \mathsf{dom}(K_{[\beta f_\theta]})$ and $\langle \vartheta, \varphi\rangle < [\beta f_\theta]'(\cdot, +\infty)\,[\mu]$. This takes place if and only if $\vartheta + \theta \in \mathsf{dom}(K_\beta)$ and $\langle \vartheta, \varphi\rangle < \beta'(\cdot, +\infty) - \beta'(\cdot, f_\theta)\,[\mu]$, by (34) and Lemma 2.6. Hence, the second assertion is proved. By (28) and Remark 8.1, $f_{[\beta f_\theta],\vartheta}$ is equal to $(\beta^*)'(\cdot, \langle \vartheta, \varphi\rangle + \beta'(\cdot, f_\theta))$ where $\beta'(\cdot, f_\theta) = \langle \theta, \varphi\rangle$ as above. This implies that $f_{[\beta f_\theta],\vartheta} = f_{\vartheta+\theta}$ and $\mathcal{F}_\beta = \mathcal{F}_{[\beta f_\theta]}$. The same proof works for the families built upon faces F of $\mathsf{cn}_\varphi(\mu)$ whose unions define the extensions of $\mathcal{F}_\beta$ and $\mathcal{F}_{[\beta f_\theta]}$, using that the hypothesis $\theta \in \Theta_\beta^+$ implies $\theta \in \Theta_{F,\beta}^+$ for each face F of $\mathsf{cn}_\varphi(\mu)$. Thus, also these extensions coincide. $\square$

Corollary 8.8. *If β is differentiable then for $\theta \in \Theta_\beta^+$ with $K_\beta(\theta)$ finite*

$$K^*_{[\beta f_\theta]}(a) = K^*_\beta(a) + K_\beta(\theta) - \langle \theta, a\rangle\,, \qquad a \in \mathbb{R}^d\,.$$

In above results, the (generalized) Bregman projections of f_θ, to $\mathcal{G}_a$, $\theta \in \Theta_\beta^+$, are related to the original primal and dual problems, (2) and (3), not depending on θ. This section is concluded by analogous results for the Bregman projection problem (25) with the function h arbitrary, subject to the running assumption. They feature the set $\Theta^+_{[\beta h]}$ consisting of those $\theta \in \Theta_{[\beta h]}$ for which

$$\langle \theta, \varphi\rangle \geqslant [\beta h]'(\cdot, 0) = \beta'(\cdot, 0) - \beta'(\cdot, h)\ \ [\mu]\,,$$

using Lemma 2.6. To simplify the notation in Theorem 8.9, the function $f_{[\beta h],\theta}$, see eq. (28) and Remark 8.1, will be denoted by h_θ. Note that $\Theta^+_{[\beta h]}$ contains the origin $\mathbf{0}$ and $h = h_{\mathbf{0}}$.

Theorem 8.9. *Suppose β is differentiable.*

For $\theta \in \Theta^+_{[\beta h]}$ with $K_{[\beta h]}(\theta)$ finite, $\Theta_{[\beta h_\theta]}$ coincides with $\Theta_{[\beta h]} - \theta$. For ϑ in that set $f_{[\beta h_\theta],\vartheta}$ equals $f_{[\beta h],\vartheta+\theta} = h_{\vartheta+\theta}$. Further, $\mathcal{F}_{[\beta h]}$ coincides with $\mathcal{F}_{[\beta h_\theta]}$ and so do their extensions.

If $J_{[\beta h]} \not\equiv +\infty$ then $\mathsf{dom}(J_{[\beta h]}) = \mathsf{dom}(J_{[\beta h_\theta]})$ for $\theta \in \Theta^+_{[\beta h]}$, and for $a \in \mathsf{dom}(J_{[\beta h]})$

$$B_\beta(g, h) - J_{[\beta h]}(a) = B_\beta(g, h_\theta) - J_{[\beta h_\theta]}(a) = B_\beta(g, \hat{g}_{[\beta h],a}) + C_{[\beta h]}(g)\,, \quad g \in \mathcal{G}_a^+\,,\ \theta \in \Theta^+_{[\beta h]}\,,$$

where $C_{[\beta h]}(g)$ *equals the integral in eq.* (31)*. Each* h_θ *above, including* $h_{\mathbf{0}} = h$*, has the same generalized Bregman projection to* $\mathcal{G}_a$*. This generalized projection belongs to* $\mathcal{G}_a$ *if and only if all the projections exist, in which case they coincide.*

P r o o f. By the assumption, the integrand $[\beta h]$ is differentiable. By Lemma 2.7 and the differentiability of β, $[[\beta h]h_\theta] = [\beta h_\theta]$. It suffices to apply Lemma 8.7 and Theorem 8.6 to the integrand $[\beta h]$ in the role of β, which gives rise to the same Bregman distance as β, by Lemma 8.4. $\square$

9. GENERALIZED SOLUTIONS OF THE DUAL PROBLEM

The results of Section 7 addressed the primal problem in absence of the PCQ. This section approaches the dual problem (3) in this very respect.

When the value function J_β is proper, the biconjugate J^{**}_β is equal to the lsc envelope of J_β, and $J^{**}_\beta = K^*_\beta$ by Theorem 1.1. In this case, $\mathit{dom}(K^*_\beta)$ contains $\mathit{dom}(J_\beta)$ and is contained in its closure which is equal to the closure of $\mathit{cn}_\varphi(\mu)$, by Lemma 6.6. In general, both inclusions can be strict. Proposition 9.4 gives a sufficient condition for equality in the first one. First, dual attainment is briefly addressed.

Lemma 9.1. *Let* $H = \{x\colon \langle\theta, x\rangle = 0\}$ *be a hyperplane such that* $H_< = \{x\colon \langle\theta, x\rangle < 0\}$ *contains* $\mathit{ri}(\mathit{cn}_\varphi(\mu))$*. Then, for* $a \in H$ *and* $\vartheta \in \mathbb{R}^d$ *with* $K_\beta(\vartheta)$ *finite*

$$K^*_\beta(a) \geqslant \langle\vartheta, a\rangle - K_\beta(\vartheta) + \int_{\{\varphi\in H_<\}} \left[\beta^*\big(z, \langle\vartheta, \varphi(z)\rangle\big) - \beta^*(z, -\infty)\right] \mu(\mathrm{d}z)\,.$$

P r o o f. By Corollary 6.2, $\varphi \in \mathit{cl}(\mathit{cn}_\varphi(\mu))$ μ-a.e. The closure is contained in $H \cup H_<$ by assumption. Hence, using that $\langle\theta, \varphi\rangle$ vanishes on the set $\{\varphi \in H\}$ and $K_\beta(\vartheta)$ is finite,

$$K_\beta(\vartheta + t\theta) = \int_{\{\varphi\in H\}} \beta^*\big(z, \langle\vartheta, \varphi(z)\rangle\big)\, \mu(\mathrm{d}z) \; + \int_{\{\varphi\in H_<\}} \beta^*\big(z, \langle\vartheta + t\theta, \varphi(z)\rangle\big)\, \mu(\mathrm{d}z)\,.$$

Here, the first integral is finite and equals $K_\beta(\vartheta) - \int_{\{\varphi\in H_<\}} \beta^*\big(z, \langle\vartheta, \varphi(z)\rangle\big)\, \mu(\mathrm{d}z)$. When $t \to +\infty$, the second one converges to the integral over $\{\varphi \in H_<\}$ of $\beta^*(z, -\infty)$, by monotone convergence. Since $K^*_\beta(a)$ is lower bounded by $\lim_{t\to+\infty} [\langle\vartheta + t\theta, a\rangle - K_\beta(\vartheta + t\theta)]$ and $\langle\theta, a\rangle = 0$, the assertion follows. $\square$

Proposition 9.2. *If* $J_\beta \not\equiv +\infty$ *and a dual solution* ϑ *for some* $a \in \mathbb{R}^d$ *exists, then either the* PCQ *holds for* a *or else* H *and* $H_<$ *exist as in Lemma 9.1 such that* $a \in H$ *and* $\langle\vartheta, \varphi\rangle \leqslant \beta'_+(\cdot, 0)$ μ*-a.e. on* $\{\varphi \in H_<\}$*.*

P r o o f. The hypotheses imply that J_β is proper and $a \in \mathit{cl}(\mathit{cn}_\varphi(\mu))$. If the PCQ for a fails then $a \notin \mathit{ri}(\mathit{cn}_\varphi(\mu))$, see Lemma 6.6. Therefore, there exists a hyperplane $H = \{x\colon \langle\theta, x\rangle = 0\}$ containing a and the origin such that $\mathit{ri}(\mathit{cn}_\varphi(\mu))$ is contained in $H_<$ as in Lemma 9.1 [49, Theorem 11.2]. Since ϑ is a dual solution for a, the nonnegative difference in the integral of Lemma 9.1 equals zero μ-a.e. on the set $\{\varphi \in H_<\}$. As that difference vanishes if and only if $\langle\vartheta, \varphi(z)\rangle \leqslant \beta'_+(z, 0)$, this completes the proof. $\square$

Corollary 9.3. *If* $J_\beta \not\equiv +\infty$ *and* $\beta'_+(\cdot, 0) = -\infty\,[\mu]$*, in particular if* β *is essentially smooth, then the* PCQ *for* a *is necessary and sufficient for existence of a dual solution for* a*.*

P r o o f. By Lemma 4.2, sufficiency holds for any $\beta \in B$. Necessity under the additional hypothesis follows from Proposition 9.2, as the hypothesis on β rules out the second contingency there. □

Proposition 9.4. *If J_β is proper and its effective domain is equal to $\mathsf{ri}(\mathsf{cn}_\varphi(\mu))$ then $\mathsf{dom}(\mathsf{K}^*_\beta)$ coincides with $\mathsf{dom}(\mathsf{J}_\beta)$.*

P r o o f. By the facts sent forward at the beginning of this section, it suffices to show that if a belongs to the closure but not to the relative interior of $\mathsf{cn}_\varphi(\mu)$ then $\mathsf{K}^*_\beta(a)$ equals $+\infty$. Let H, $H_<$ and $a \in H$ be as in Lemma 9.1. Since J_β is proper there exists ϑ with $\mathsf{K}_\beta(\vartheta)$ finite. The hypothesis $\mathsf{dom}(\mathsf{J}_\beta) = \mathsf{ri}(\mathsf{cn}_\varphi(\mu))$ implies by Corollary 6.12 that the integral of $\beta(\cdot, 0)$ over $H_<$ equals $+\infty$. Since $\beta^*(z, -\infty) = -\beta(z, 0)$, Lemma 9.1 implies $\mathsf{K}^*_\beta(a) = +\infty$. □

Analogously to the generalized primal solutions, a generalized dual solution is introduced for each $a \in \mathbb{R}^d$ with finite dual value $\mathsf{K}^*_\beta(a)$, attained or not. More precisely, this concept generalizes that of the effective dual solution rather than that of a dual solution proper. It requires the DCQ which is assumed throughout the remaining part of this section. By Lemma 3.4, there exist sequences $\vartheta_n \in \Theta_\beta$ with $\langle \vartheta_n, a\rangle - \mathsf{K}_\beta(\vartheta_n)$ tending to $\mathsf{K}^*_\beta(a)$. A $\mathcal{Z}$-measurable function h_a is a *generalized dual solution* for a if for each sequence ϑ_n as above the functions f_{ϑ_n} converge to h_a locally in measure. Existence of generalized dual solutions follows from the main result of this section.

Theorem 9.5. *Assuming the DCQ, for every $a \in \mathbb{R}^d$ with $\mathsf{K}^*_\beta(a)$ finite there exists a unique $\mathcal{Z}$-measurable function h_a such that*

$$\mathsf{K}^*_\beta(a) - \big[\langle \vartheta, a\rangle - \mathsf{K}_\beta(\vartheta)\big] \geqslant \mathsf{B}_\beta(h_a, f_\vartheta)\,, \qquad \vartheta \in \Theta_\beta\,. \tag{35}$$

If the effective dual solution $g^*_a = f_\vartheta$ exists, where $\vartheta \in \Theta_\beta$ is a dual solution for a, then $g^*_a = h_a$ by ineq. (35). If $\vartheta_n \in \Theta_\beta$ is a maximizing sequence for $\langle \vartheta, a\rangle - \mathsf{K}_\beta(\vartheta)$, ineq. (35) implies that the Bregman distances $\mathsf{B}_\beta(h_a, f_{\vartheta_n})$ tend to zero, and then $f_{\vartheta_n} \rightsquigarrow h_a$ by Corollary 2.14. Thus, h_a is the generalized dual solution for a. This establishes also the uniqueness in Theorem 9.5.

First, a special case of Theorem 9.5 is established, for its simplicity and independent interest. In this case, the generalized primal and dual solutions for a coincide.

Proposition 9.6. *Assuming the DCQ, for $a \in \mathbb{R}^d$ with $\mathsf{K}^*_\beta(a)$ finite and zero duality gap, ineq.* (35) *holds with $h_a = \tilde{g}_a$, see Theorem 7.10.*

P r o o f. By assumptions, $\mathsf{J}_\beta(a)$ is finite. Let g_n be a sequence in $\mathcal{G}^+_a$ with $\mathsf{H}_\beta(g_n)$ converging to $\mathsf{J}_\beta(a)$. Limiting along g_n in Theorem 7.10, $\mathsf{B}_\beta(g_n, \tilde{g}_a) \to 0$ whence $g_n \rightsquigarrow \tilde{g}_a$, by Corollary 2.14. Lemma 4.15 implies

$$\mathsf{H}_\beta(g) \geqslant \langle \vartheta, a\rangle - \mathsf{K}_\beta(\vartheta) + \mathsf{B}_\beta(g, f_\vartheta)\,, \quad g \in \mathcal{G}^+_a\,,\ \vartheta \in \Theta_\beta\,.$$

Limiting here along g_n, $\mathsf{J}_\beta(a) \geqslant \langle \vartheta, a\rangle - \mathsf{K}_\beta(\vartheta) + \mathsf{B}_\beta(\tilde{g}_a, f_\vartheta)$ for $\vartheta \in \Theta_\beta$, by Lemma 2.12. This and the hypothesis $\mathsf{J}_\beta(a) = \mathsf{K}^*_\beta(a)$ imply that $\mathsf{B}_\beta(\tilde{g}_a, f_{\vartheta_n}) \to 0$ for each sequence

ϑ_n in Θ_β with $\langle \vartheta_n, a\rangle - K_\beta(\vartheta_n)$ converging to $K^*_\beta(a)$. The assertion $f_{\vartheta_n} \rightsquigarrow \tilde g_a$ follows by Corollary 2.14. $\square$

Example 10.8 illustrates a situation when the PCQ fails, the primal and dual values are finite but different and the primal solution g_a is different from h_a. Additionally, μ is finite and φ bounded.

To prove Theorem 9.5 in general, the following lemmas and corollary are needed. The inequality below compares the Jensen difference of K_β with Bregman distances.

Lemma 9.7. *If K_β is proper then for θ_1, θ_2 in Θ_β and $0 < t < 1$*

$$\begin{aligned} tK_\beta(\theta_1)+&(1-t)K_\beta(\theta_2) - K_\beta(t\theta_1 + (1-t)\theta_2) \\ &\geqslant tB_\beta(f_{t\theta_1+(1-t)\theta_2}, f_{\theta_1}) + (1-t)B_\beta(f_{t\theta_1+(1-t)\theta_2}, f_{\theta_2})\,. \end{aligned}$$

Proof. The left-hand side is equal to

$$t[K_\beta(\theta_1) - K_\beta(t\theta_1 + (1-t)\theta_2)] + (1-t)[K_\beta(\theta_2) - K_\beta(t\theta_1 + (1-t)\theta_2)]$$

where all values are finite. The left bracket takes the form

$$\begin{aligned} &\int_Z \Big[\beta^*\big(z, \langle\theta_1, \varphi(z)\rangle\big)\,\mu(\mathrm{d}z) - \beta^*\big(z, \langle t\theta_1 + (1-t)\theta_2, \varphi(z)\rangle\big)\Big]\,\mu(\mathrm{d}z) \\ &= \int_Z \Big[\Delta_{\beta^*(z,\cdot)}(\langle\theta_1, \varphi(z)\rangle, \langle t\theta_1 + (1-t)\theta_2, \varphi(z)\rangle) \\ &\qquad + (1-t)\langle\theta_1 - \theta_2, \varphi(z)\rangle f_{t\theta_1+(1-t)\theta_2}(z)\Big]\,\mu(\mathrm{d}z) \end{aligned}$$

and the right bracket

$$\int_Z \Big[\Delta_{\beta^*(z,\cdot)}(\langle\theta_2, \varphi(z)\rangle, \langle t\theta_1 + (1-t)\theta_2, \varphi(z)\rangle) + t\langle\theta_2 - \theta_1, \varphi(z)\rangle f_{t\theta_1+(1-t)\theta_2}(z)\Big]\,\mu(\mathrm{d}z)\,.$$

Then, the left-hand side of the inequality rewrites to

$$\begin{aligned} &\int_Z \Big[t\Delta_{\beta^*(z,\cdot)}(\langle\theta_1, \varphi(z)\rangle, \langle t\theta_1 + (1-t)\theta_2, \varphi(z)\rangle) \\ &\qquad + (1-t)\Delta_{\beta^*(z,\cdot)}(\langle\theta_2, \varphi(z)\rangle, \langle t\theta_1 + (1-t)\theta_2, \varphi(z)\rangle)\Big]\,\mu(\mathrm{d}z)\,. \end{aligned}$$

The assertion follows by the consequence $\Delta_{\gamma^*}(r_2, r_1) \geqslant \Delta_\gamma(u(r_1), u(r_2))$ of Lemma 2.9, where $u(r) = {\gamma^*}'(r)$. $\square$

Corollary 9.8. *If $K^*_\beta(a)$ is finite, $\theta_1, \theta_2 \in \Theta_\beta$ and $0 < t < 1$ then*

$$\begin{aligned} t\big[K^*_\beta(a) - [\langle\theta_1, a\rangle - K_\beta(\theta_1)]\big]+&(1-t)\big[K^*_\beta(a) - [\langle\theta_2, a\rangle - K_\beta(\theta_2)]\big] \\ &\geqslant tB_\beta(f_{t\theta_1+(1-t)\theta_2}, f_{\theta_1}) + (1-t)B_\beta(f_{t\theta_1+(1-t)\theta_2}, f_{\theta_2})\,. \end{aligned}$$

Proof. The Jensen difference in Lemma 9.7 is equal to

$$\begin{aligned} \big[\langle t\theta_1 + (1-t)\theta_2, a\rangle-&K_\beta(t\theta_1 + (1-t)\theta_2)\big] \\ &- t\big[\langle\theta_1, a\rangle - K_\beta(\theta_1)\big] - (1-t)\big[\langle\theta_2, a\rangle - K_\beta(\theta_2)\big] \end{aligned}$$

where the first bracket is dominated by $K^*(a)$. $\square$

Lemma 9.9. *Let $C \in \mathcal{Z}$ have finite μ-measure, let L, ξ, and δ be positive numbers. Then there exists $K > L$ such that if $B_\beta(h,g) \leqslant \delta$ for some nonnegative $\mathcal{Z}$-measurable functions g, h then*

$$\mu(C \cap \{g > K\}) < \xi + \mu(C \cap \{h > L\}).$$

P r o o f. Let $M = 2\delta/\xi$. By monotonicity, for $K > L$

$$\begin{aligned} B_\beta(g,h) \geqslant \int_{\{g>K, h\leqslant L\}} \Delta_\beta(z, g(z), h(z))\, \mu(\mathrm{d}z) \geqslant \int_{\{g>K, h\leqslant L\}} \Delta_\beta(z, K, L)\, \mu(\mathrm{d}z) \\ \geqslant M \cdot \mu\big(C \cap \{\Delta_\beta(\cdot, K, L) \geqslant M\} \cap \{g > K, h \leqslant L\}\big) \end{aligned}$$

whence

$$\mu(C \cap \{g > K\}) \leqslant \tfrac{1}{M} B_\beta(g,h) + \mu(C \cap \{\Delta_\beta(\cdot, K, L) < M\}) + \mu(C \cap \{h > L\}).$$

Since $\Delta_\beta(z, K, L) \uparrow +\infty$ if $K \uparrow +\infty$ due to strict convexity, there exists $K > L$ such that $\mu\big(C \cap \{\Delta_\beta(\cdot, K, L) < M\}\big) < \frac{1}{2}\xi$. With this K the assertion follows by the choice of M, implying $\frac{1}{M} B_\beta(h,g) \leqslant \frac{1}{2}\xi$ whenever $B_\beta(h,g) \leqslant \delta$. □

P r o o f o f T h e o r e m 9.5. By assumptions and Lemma 3.4, K_β is proper and there exists a sequence τ_n in Θ_β such that $\langle \tau_n, a\rangle - K_\beta(\tau_n)$ converges to $K^*(a)$.

Let $\vartheta \in \Theta_\beta$. Applying Corollary 9.8 to $\theta_1 = \vartheta$, $\theta_2 = \tau_n$ and $0 < t_n < 1$ yields

$$\begin{aligned} \big[K^*_\beta(a) - [\langle \vartheta, a\rangle - K_\beta(\vartheta)]\big] + \tfrac{1-t_n}{t_n}\big[K^*_\beta(a) - [\langle \tau_n, a\rangle - K_\beta(\tau_n)]\big] \\ \geqslant B_\beta(f_{t_n\vartheta+(1-t_n)\tau_n}, f_\vartheta) + \tfrac{1-t_n}{t_n} B_\beta(f_{t_n\vartheta+(1-t_n)\tau_n}, f_{\tau_n}). \end{aligned} \tag{36}$$

Let $t_n \to 0$ sufficiently slowly to make the second term on the left hand side go to zero. Then, (36) implies that the sequence $B_\beta(f_{t_n\vartheta+(1-t_n)\tau_n}, f_\vartheta)$ is bounded, the sequence $B_\beta(f_{t_n\vartheta+(1-t_n)\tau_n}, f_{\tau_n})$ tends to zero and

$$K^*_\beta(a) - [\langle \vartheta, a\rangle - K_\beta(\vartheta)] \geqslant \liminf_{n\to\infty} B_\beta(f_{t_n\vartheta+(1-t_n)\tau_n}, f_\vartheta), \quad \vartheta \in \Theta_\beta. \tag{37}$$

By Lemma 2.12, it suffices to prove that the sequence $f_{t_n\vartheta+(1-t_n)\tau_n}$ converges locally in measure.

Let $C \in \mathcal{Z}$ have finite μ-measure and $\xi > 0$. Then, $\mu(C \cap \{f_\vartheta > L\}) < \xi$ for some $L > 0$. Since $B_\beta(f_{t_n\vartheta+(1-t_n)\tau_n}, f_\vartheta)$ is bounded, by Lemma 9.9 there exists $K > L$ such that

$$\mu(C \cap \{f_{t_n\vartheta+(1-t_n)\tau_n} > K\}) < \xi + \mu(C \cap \{f_\vartheta > L\}) < 2\xi \tag{38}$$

for all n. Since $B_\beta(f_{t_n\vartheta+(1-t_n)\tau_n}, f_{\tau_n}) \to 0$, Lemma 2.13 and (38) imply that for any $\varepsilon > 0$

$$\mu(C \cap \{|f_{t_n\vartheta+(1-t_n)\tau_n} - f_{\tau_n}| > \varepsilon\}) < 3\xi, \qquad \text{eventually in } n. \tag{39}$$

Combining (38) and (39),

$$\mu(C \cap \{f_{\tau_n} > K + \varepsilon\}) < 5\xi, \qquad \text{eventually in } n. \tag{40}$$

By Corollary 9.8 applied to $\theta_1 = \tau_m$, $\theta_2 = \tau_n$ and $t = \frac{1}{2}$,

$$B_\beta(f_{(\tau_n+\tau_m)/2}, f_{\tau_m}) + B_\beta(f_{(\tau_n+\tau_m)/2}, f_{\tau_n}) \to 0 \qquad \text{as } n, m \to \infty.$$

This convergence, Lemma 2.13 and (40) imply that for any $\varepsilon > 0$ and $\xi > 0$

$$\mu(C \cap \{|f_{\tau_n} - f_{\tau_m}| > \varepsilon\}) < \xi + \mu(C \cap \{f_{\tau_n} > K + \varepsilon\}) < 6\xi \tag{41}$$

provided n, m are sufficiently large. Thus, the sequence f_{τ_n} is Cauchy, locally in μ-measure. Hence, $f_{\tau_n} \rightsquigarrow h$ for some $\mathcal{Z}$-measurable nonnegative function h. This and (39) imply that also $f_{t_n\vartheta+(1-t_n)\tau_n} \rightsquigarrow h$, needed to complete the proof. $\square$

Remark 9.10. In Theorem 9.5, it can happen that $J_\beta(a)$ is not finite, even $J_\beta \equiv +\infty$ is allowed. Assuming the DCQ, if $J_\beta(a)$ is finite then Lemma 4.15 implies for $g_n \in \mathcal{G}_a^+$ with $H_\beta(g_n) \to J_\beta(a)$ and $\vartheta_n \in \Theta_\beta$ with $\langle \vartheta_n, a\rangle - K_\beta(\vartheta_n) \to K_\beta^*(a)$ that

$$J_\beta(a) - K_\beta^*(a) \geqslant \limsup_{n\to\infty}\ B_\beta(g_n, f_{\vartheta_n})\,.$$

Here, if β is essentially smooth then the equality takes place and the limit exists. Since $g_n \rightsquigarrow \tilde{g}_a = \hat{g}_a$ by Theorem 7.10, and $f_{\vartheta_n} \rightsquigarrow h_a$ by Theorem 9.5, it follows by Lemma 2.12 that the duality gap $J_\beta(a) - K_\beta^*(a)$ majorizes the Bregman distance $B_\beta(\hat{g}_a, h_a)$ of the generalized solutions. Conditions making this bound tight remain elusive.

10. EXAMPLES

This section demonstrates that ‘irregular’ behavior may occur in the primal and dual problems, even in the autonomous case $\beta(z,t) = \gamma(t)$, $z \in Z$, $t \in \mathbb{R}$, with γ differentiable. In this situation, $\gamma \in \Gamma$ replaces β in notations like $H_\beta(g)$, J_β, etc. Fig. 1 summarizes some properties of the first eight examples. It contains a region marked by $\emptyset$, see Lemma 4.10. Example 10.9 describes the situation when the moment mapping vanishes identically, showing how the problem of unconstrained minimization of convex integral functionals fits into our framework. The remaining three examples illustrate Bregman projections and closure.

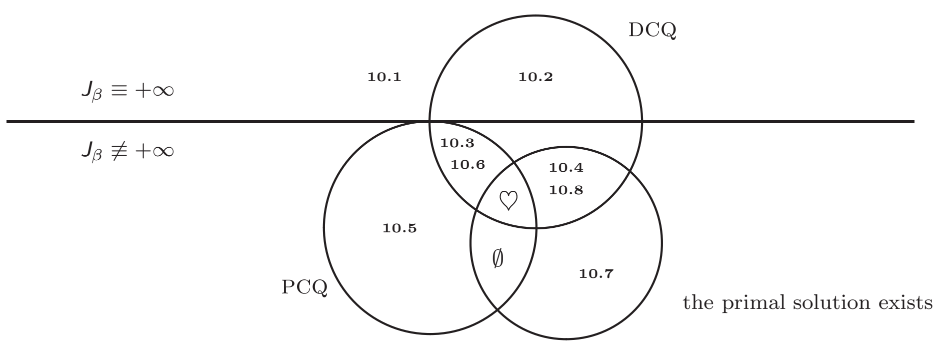

Fig. 1.

Example 10.1. Let μ be the counting measure and φ the identity mapping on the set Z of integers. The functional H_γ is considered with $\gamma \in \Gamma$ given by $\gamma(t) = (2t)^{-1}$ for $t > 0$. Then, $H_\gamma(g) < +\infty$ implies that g is positive and $\sum_{z\in Z} 1/g(z)$ converges. In this case, $g(z) \geqslant 1$ if $|z|$ is sufficiently large whence both the positive and negative parts of $\int_Z \varphi g \, \mathrm{d}\mu$ are infinite. Therefore, $J_\gamma \equiv +\infty$ and $J_\gamma^* \equiv -\infty$. The conjugate of γ is given by $\gamma^*(r) = -\sqrt{-2r}$ for $r \leqslant 0$ and $\gamma^*(r) = +\infty$ otherwise. Then, $K_\gamma(\vartheta) = \int_Z \gamma^*(\vartheta\varphi)\, \mathrm{d}\mu$ equals 0 for $\vartheta = 0$ and $+\infty$, otherwise, both the positive and negative parts being infinite. In particular, K_γ is proper, but its effective domain has empty interior. The DCQ fails, $\Theta_\gamma = \emptyset$. For $a = 0$ the dual value is 0, different from the primal one $+\infty$.

Example 10.2. Let μ be the counting measure and φ the identity mapping on the set $Z = \{1, 2, \ldots\}$. Let $\gamma(t) = t \ln t - t + 1$, $t \geqslant 0$, so that $\gamma^*(r) = e^r - 1$, $r \in \mathbb{R}$. Then, $K_\gamma(\vartheta) = \sum_{z\in Z} \gamma^*(\vartheta z) = \vartheta \cdot (+\infty)$, $\vartheta \in \mathbb{R}$, and $dom(K_\gamma)$ coincides with $\Theta_\gamma = (-\infty, 0]$. Although $dom(K_\gamma)$ has nonempty interior and the DCQ holds, $J_\gamma \equiv +\infty$ for otherwise K_γ could not take the value $-\infty$. For $\vartheta < 0$ the function f_ϑ belongs to $\mathcal{G}$, i.e., the moment $\int_Z \varphi f_\vartheta \, \mathrm{d}\mu = \sum_{z=1}^{\infty} z e^{\vartheta z}$ exists, while $K_\gamma(\vartheta) = -\infty$. For a equal to that moment, $\inf_{g\in\mathcal{G}_a^+} B_\gamma(g, f_\vartheta) = 0$, hence $dom(J_{[\gamma f_\vartheta]})$ contains a, see (25). This shows that in the last assertion of Theorem 8.6 the finiteness assumption is essential.

Example 10.3. Let μ be the Borel measure on $Z = \mathbb{R}$ given by $\mathrm{d}\mu = \frac{\mathrm{d}z}{1+z^2}$, φ the identity mapping on Z and $\gamma(t) = t \ln t$, $t \geqslant 0$. Since $\mu(Z) = \pi$ and $\gamma \geqslant \gamma(\frac{1}{e}) = -\frac{1}{e}$, the value function J_γ is lower bounded by $-\frac{\pi}{e}$. The functional H_γ is finite for functions $g \geqslant 0$ on Z that are nonzero and bounded on bounded sets. Then the PCQ holds for every $a \in \mathbb{R}$. In dual problems, $\gamma^*(r) = e^{r-1}$, $r \in \mathbb{R}$, and $K_\gamma(\vartheta) = \int_{\mathbb{R}} e^{\vartheta z - 1}\, \mu(\mathrm{d}z)$ is equal to $\frac{\pi}{e}$ for $\vartheta = 0$ and $+\infty$, otherwise. Therefore, the DCQ holds, $\Theta_\gamma = \{0\}$ and $J_\gamma = K_\gamma^* \equiv -\frac{\pi}{e}$. For each $a \in \mathbb{R}$ the effective dual solution g_a^* is identically equal to $\frac{1}{e}$. The moment of g_a^* does not exist whence the primal problem has no solution. Nevertheless, the generalized primal solution $\hat{g}_a$ exists and equals g_a^*. A modification of this example with information theoretical interpretation appeared in [31, Example 1].

Example 10.4. On $Z = \{0, 1\}$, let μ be the counting measure and φ the identity mapping. Let $\gamma(t)$ equal $\frac{1}{2}t^2$ for $t \geqslant 0$. Then, $J_\gamma = \gamma$ has the effective domain $[0, +\infty)$. Since $\gamma^*(r)$ equals $\frac{1}{2}r^2$ for $r \geqslant 0$ and 0 otherwise, the DCQ holds and $\Theta_\gamma = \mathbb{R}$. For $a = 0$ not enjoying the PCQ, each $\vartheta \leqslant 0$ is a dual solution and the effective dual solution g_a^* is identically equal to 0. It belongs to $\mathcal{G}_a$, and thus coincides with the primal solution g_a.

Example 10.5. On $Z = \{-1, 1\}$, let μ be the counting measure and φ the identity mapping. Let γ be the same as in Example 10.1. In the primal problem

$$J_\gamma(a) = \inf\left\{ \tfrac{1}{2g(1)} + \tfrac{1}{2g(-1)} : g(1), g(-1) > 0\,,\ g(1) - g(-1) = a \right\} = 0\,, \quad a \in \mathbb{R}\,,$$

the infimum is not attained. Though the PCQ holds for each $a \in \mathbb{R}$, no primal solution exists. No generalized primal solution exists either, since any minimizing sequence g_n in the primal problem for a satisfies $g_n(1) \to +\infty$ and $g_n(1) - g_n(-1) = a$, and thus g_n cannot converge in any standard sense. Since $dom(J_\gamma) = \mathbb{R}$, the effective domain of $K_\gamma = J_\gamma^*$ equals the singleton $\{0\}$. For $\vartheta = 0$ the function γ^* is not finite around $\langle \vartheta, \varphi(z) \rangle = 0$ whence the DCQ does not hold.

Example 10.6. (a modification of the example in [10, p. 263]) Let μ be the Borel measure on $Z = [0,1]$ with $\mathrm{d}\mu = 2z\,\mathrm{d}z$, $\varphi(z) = (1, z)$, $z \in Z$, and $\gamma(t) = -\ln t$, $t > 0$. By Theorem 6.8, $\mathsf{dom}(J_\gamma)$ consists of the pairs $a = (a_1, a_2)$ such that $0 < a_2 < a_1$. Since $\gamma^*(r)$ equals $-1 - \ln(-r)$ for $r < 0$ and $+\infty$ otherwise, for $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2$

$$K_\gamma(\vartheta) = \begin{cases} -1 - \int_0^1 2z \ln(-\vartheta_1 - \vartheta_2 z)\,\mathrm{d}z\,, & \vartheta_1 \leqslant 0\,,\ \vartheta_1 + \vartheta_2 < 0\,, \\ +\infty\,, & \text{otherwise,} \end{cases}$$

where the integral is finite. The set Θ_γ is equal to $\mathsf{dom}(K_\gamma)$, given by the two above inequalities. The dual problem takes the form

$$K_\gamma^*(a) = \sup_{(\vartheta_1,\vartheta_2)\in\Theta_\gamma} \Big[\vartheta_1 a_1 + \vartheta_2 a_2 + 1 + \textstyle\int_0^1 2z \ln(-\vartheta_1 - \vartheta_2 z)\,\mathrm{d}z\Big]\,, \qquad a \in \mathbb{R}^2\,,$$

and $\mathsf{dom}(K_\gamma^*)$ equals $\mathsf{dom}(J_\gamma)$ determined above, by Proposition 9.4. In the interior of Θ_γ, the derivative $\vartheta_1 \frac{\partial}{\partial \vartheta_1} + \vartheta_2 \frac{\partial}{\partial \vartheta_2}$ of the above bracket is equal to $\vartheta_1 a_1 + \vartheta_2 a_2 + 1$. Assuming $a \in \mathsf{dom}(K_\gamma^*)$, the derivative vanishes if and only if ϑ belongs to the relatively open segment between $\tau = (0, -\frac{1}{a_2})$ and $(-\frac{1}{a_1 - a_2}, \frac{1}{a_1 - a_2})$. The supremum can be restricted to this segment, parallel to the direction $(-a_2, a_1)$. The directional derivative of the bracket at τ in this direction is equal to $a_2(2a_2 - a_1)$, by a direct computation. Therefore, if $2a_2 \leqslant a_1$ then τ is the unique dual solution for a, not depending on a_1. In this case, the dual value $K_\gamma^*(a)$ is equal to $-\frac{1}{2} - \ln a_2$ and the effective dual solution is $g_a^* \colon z \mapsto \frac{a_2}{z}$. When even $2a_2 < a_1$, the first coordinate $\int_0^1 1 \cdot \frac{a_2}{z} \cdot 2z\,\mathrm{d}z$ of the moment vector of g_a^* is less than a_1. Thus, the primal solution for a does not exist. Nevertheless, g_a^* provides the generalized primal solution $\hat{g}_a$.

Example 10.7. Let $Z = \mathbb{R}^2$, μ be the sum of the Lebesgue measure on the horizontal axis and the unit point mass at $(0,1)$, $\varphi(z) = (1, z_1, z_2)$ for $z = (z_1, z_2) \in Z$, and $\gamma(t) = t \ln t$, $t \geqslant 0$. The φ-cone of μ is the convex hull of two of its proper faces $F = \{(t, 0, t) \colon t \geqslant 0\}$ and $\{(0,0,0)\} \cup \{(t, r, 0) \colon t > 0\}$. Then, $J_\gamma(a) = \gamma(t)$ for $a = (t, 0, t) \in F$ and $J_\gamma \equiv -\infty$ on $\mathsf{cn}_\varphi(\mu) \setminus F$, using the fact that the Shannon functional can be explicitly evaluated at the Gaussian densities. Hence, $K_\gamma \equiv +\infty$ and the DCQ fails. Nevertheless, by Definition 7.1, for $a \in \mathsf{ri}(F)$

$$K_{F,\gamma}(\vartheta) = \int_{\{z_1 = 0\,,\, z_2 = 1\}} e^{\vartheta_0 + \vartheta_1 z_1 + \vartheta_2 z_2 - 1}\, \mu(\mathrm{d}z_1, \mathrm{d}z_2) = e^{\vartheta_0 + \vartheta_2 - 1}\,, \qquad \vartheta = (\vartheta_0, \vartheta_1, \vartheta_2)\,,$$

whence $\Theta_{F,\gamma} = \mathbb{R}^3$. The F-dual problem for $a = (t, 0, t)$, $t > 0$, has many solutions, e.g. $(1 + \ln t, 0, 0)$, and $K_{F,\gamma}^*(a) = \gamma(t)$. By Theorem 7.6, the primal solution g_a exists and equals $g_{F,a}^*$, a function equal to t at $(0,1) \in Z$ and zero otherwise.

Example 10.8. On $Z = [0,1]^2$, let μ be the sum of the Lebesgue measure and the unit masses at $(0, \frac{1}{3})$ and $(0, \frac{2}{3})$. Let $\varphi(z) = (1, z_1, z_2)$, $z = (z_1, z_2) \in Z$, and $\gamma(t) = -2\sqrt{t}$ for $t \geqslant 0$ whence $\gamma^*(r) = -r^{-1}$ for $r < 0$. If $a = (3, 0, 1)$ then $J_\gamma(a) = H_\gamma(g_a) = -2\sqrt{3}$ where $g_a \in \mathcal{G}_a$ is equal to 3 at $(0, \frac{1}{3})$ and to 0 otherwise. If $\vartheta = (\vartheta_0, \vartheta_1, \vartheta_2) \in \mathbb{R}^3$ has ϑ_0, $\vartheta_0 + \vartheta_1$, $\vartheta_0 + \vartheta_2$ and $\vartheta_0 + \vartheta_1 + \vartheta_2$ negative then

$$K_\gamma(\vartheta) = -\frac{3}{3\vartheta_0 + \vartheta_2} - \frac{3}{3\vartheta_0 + 2\vartheta_2} - \int_{[0,1]^2} \frac{\mathrm{d}z_1\,\mathrm{d}z_2}{\vartheta_0 + \vartheta_1 z_1 + \vartheta_2 z_2}$$

where the integral is finite. Otherwise, $K_\gamma(\vartheta) = +\infty$. Hence, the DCQ holds. The maximization in the dual problem for $a = (3,0,1)$ includes the limiting $\vartheta_1 \downarrow -\infty$, thus

$$K^*_\gamma(a) = \sup_{\vartheta_0<0,\ \vartheta_0+\vartheta_2<0} \left[3\vartheta_0 + \vartheta_2 + \tfrac{3}{3\vartheta_0+\vartheta_2} + \tfrac{3}{3\vartheta_0+2\vartheta_2} \right].$$

The bracket is increasing when $(\vartheta_0,\vartheta_2)$ moves in the direction $(1,-3)$ which implies that $K^*_\gamma(a)$ is equal to $\max_{\vartheta_2<0}[\vartheta_2 + \frac{9}{2\vartheta_2}] = -3\sqrt{2}$. Hence, the primal value is strictly greater than the dual one. The sequence $-(\frac{1}{n}, n, \frac{3}{\sqrt{2}})$ is maximizing in the dual problem. By Theorem 9.5, the generalized dual solution h_a is the limit in measure of the sequence of functions $(\frac{1}{n} + nz_1 + \frac{3}{\sqrt{2}}z_2)^{-1}$. Since h_a is equal to $\sqrt{2}$ at $(0,\frac{1}{3})$ it differs from the primal solution g_a.

Example 10.9. Let $\varphi \equiv \mathbf{0}$. Then $\mathcal{G}_a$ consists of all $\mathcal{Z}$-measurable functions if $a = \mathbf{0}$, and is empty otherwise. Thus, $\mathit{dom}(J_\beta) \subseteq \{\mathbf{0}\}$ for each $\beta \in B$. For the equality, i.e., for the existence of a measurable function g with $H_\beta(g) < +\infty$, the obvious necessary condition $\int_Z \inf_t \beta(\cdot,t)\,\mathrm{d}\mu < +\infty$ is sufficient, as well, by Lemma A.5, also implying that $J_\beta(\mathbf{0})$ equals that integral. Further, $K_\beta(\vartheta) = \int_Z \beta^*(\cdot,0)\,\mathrm{d}\mu$ for each $\vartheta \in \mathbb{R}^d$, whence $K^*_\beta(\mathbf{0}) = -\int_Z \beta^*(\cdot,0)\,\mathrm{d}\mu$. If the integral is finite then each $\vartheta \in \mathbb{R}^d$ is a dual solution for $a = \mathbf{0}$. Since $\beta^*(\cdot,0)$ is equal to $-\inf_t \beta(\cdot,t)$, in case $J_\beta(\mathbf{0}) < +\infty$ the primal and dual values are equal, $J_\beta(\mathbf{0}) = K^*_\beta(\mathbf{0})$. The latter may fail if $J_\beta(\mathbf{0}) = +\infty$, for the adopted convention admits both $\pm\inf_t \beta(\cdot,t)$ to have integral $+\infty$. The finiteness of $J_\beta(\mathbf{0})$ is equivalent to the PCQ for $a = \mathbf{0}$, in which case the primal solution for $a = \mathbf{0}$ exists if and only if $\inf_t \beta(\cdot,t)$ is attained μ-a.e. [50, Theorem 14.60]. This is equivalent to $\beta'(\cdot,+\infty) > 0\,[\mu]$, thus $\mathbf{0} \in \mathit{\Theta}_\beta$ by (9), hence the mentioned result is contained in Corollary 4.12. By Theorem 4.17, if the DCQ fails then no generalized primal solution exists, either.

Example 10.10. Let $\gamma \in \mathit{\Gamma}$ be differentiable except at $t = 1$, μ be a pm on $(Z,\mathcal{Z})$, $d = 1$ and $\varphi \equiv 1$. Then $\mathcal{G}_a$ consists of the $\mathcal{Z}$-measurable functions whose μ-integral equals $a \in \mathbb{R}$. The Bregman projection of $h \equiv 1$ to $\mathcal{G}_a$ features the integrand

$$[\gamma h](s) = \mathit{\Delta}_\gamma(s,1) = \gamma(s) - \gamma(1) - \gamma'_{\mathit{sgn}(s-1)}(1)[s-1]\,, \qquad s \geqslant 0\,.$$

For any $a > 0$ the minimum subject to $g \in \mathcal{G}_a$ of $B_\gamma(g,h) = \int_Z \mathit{\Delta}_\gamma(g,1)\,\mathrm{d}\mu$ is attained when $g \equiv a$, by Jensen inequality. In other words, the Bregman projection of h to $\mathcal{G}_a$ exists and $g_{[\gamma h],a} \equiv a$. For $0 < a < 1$ and $g \in \mathcal{G}_a$ with $\int_Z \gamma(g)\,\mathrm{d}\mu$ finite,

$$\begin{aligned} &B_\gamma(g, g_{[\gamma h],a}) + B_\gamma(g_{[\gamma h],a}, h) \\ &= \int_Z \big[\gamma(g) - \gamma(a) - \gamma'(a)[g-a]\big]\,\mathrm{d}\mu + \gamma(a) - \gamma(1) - \gamma'_-(1)[a-1] \\ &= \int_Z \gamma(g)\,\mathrm{d}\mu - \gamma(1) - \gamma'_-(1)[a-1] \end{aligned}$$

while

$$B_\gamma(g,h) = \int_Z \gamma(g)\,\mathrm{d}\mu - \gamma(1) - \gamma'_-(1)[a-1] + \int_{\{g>1\}} [\gamma'_-(1) - \gamma'_+(1)][g-1]\mathrm{d}\mu\,.$$

This shows that $B_\gamma(g,h) < B_\gamma(g,g_{[\gamma h],a}) + B_\gamma(g_{[\gamma h],a},h)$ when $g \in \mathcal{G}_a$ and the set $\{g > 1\}$ is not μ-negligible.

Example 10.11. Let μ be the counting measure on $Z = \{1, 2, 3\}$, and φ have the values $\varphi(1) = (1, 1)$, $\varphi(2) = (1, -1)$ and $\varphi(3) = (1, 0)$. Functions g on Z are identified with points in $\mathbb{R}^3$. Thus,

$$\mathcal{G}_a = \left\{ \left(\tfrac{t+a_2}{2}, \tfrac{t-a_2}{2}, a_1 - t\right) : t \in \mathbb{R} \right\}, \qquad a = (a_1, a_2) \in \mathbb{R}^2,$$

and the φ-cone $cn_\varphi(\mu) \subseteq \mathbb{R}^2$ is given by $|a_2| \leqslant a_1$. If $\gamma(t) = \frac{t^2}{2}$, $t \geqslant 0$, then $\Theta_\gamma = \mathbb{R}^2$ and, using that $(\gamma^*)' = |\cdot|_+$, the family

$$\mathcal{F}_\gamma = \left\{ f_\vartheta = (|\vartheta_1 + \vartheta_2|_+ \, , |\vartheta_1 - \vartheta_2|_+ \, , |\vartheta_1|_+) : \ \vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2 \right\}$$

is the union of three two-dimensional cones in $\mathbb{R}^3$. The set Θ_γ^+ of $\vartheta \in \Theta_\gamma$ with $\langle \vartheta, \varphi \rangle \geqslant \gamma'(0) = 0$ is determined by $|\vartheta_2| \leqslant \vartheta_1$ and coincides with one of the three cones. Theorem 8.6 fails if Θ_γ^+ is replaced by the whole Θ_γ, noting that the Bregman distance B_γ of two points $g, h \in \mathbb{R}^3$ equals their squared Euclidean distance divided by two. Lemma 8.7 fails as well, e. g., if $\theta = (0, 1) \notin \Theta_\gamma^+$ then $f_\theta = (1, 1, 0)$,

$$\begin{aligned} f_{\vartheta+\theta} &= (|\vartheta_1 + \vartheta_2 + 1|_+ \, , |\vartheta_1 - \vartheta_2 - 1|_+ \, , |\vartheta_1|_+) \, , \\ f_{[\beta f_\theta], \vartheta} &= (|\vartheta_1 + \vartheta_2 + 1|_+ \, , |\vartheta_1 - \vartheta_2 + 1|_+ \, , |\vartheta_1|_+) \, , \end{aligned}$$

by Lemma 2.6.

Example 10.12. On $Z = (1, +\infty)$ let μ be the pm with density $2z^{-3}dz$, $\varphi(z) = (1, z)$, $z \in Z$, and $\gamma(t) = \frac{1}{2}t^2$, $t \geqslant 0$. Then, $cn_\varphi(\mu) = \{(a_1, a_2) : a_2 > a_1 > 0\} \cup \{(0, 0)\}$ coincides with $dom(J_\gamma)$. For $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2$,

$$K_\gamma(\vartheta) = \begin{cases} \int_1^{+\infty} |\vartheta_1 + \vartheta_2 z|_+^2 \, \frac{2}{z^3} \, \mathrm{d}z \, , & \vartheta_2 \leqslant 0 \, , \\ +\infty \, , & \text{otherwise.} \end{cases}$$

Hence, Θ_β is also given by the above inequality. The function $f_\vartheta(z) = |\vartheta_1 + \vartheta_2 z|_+$, $\vartheta \in \Theta_\beta$, identically vanishes on Z if $\vartheta_1 \leqslant -\vartheta_2$. Otherwise, $\vartheta = t(1, -r)$ with $t > 0$ and $0 \leqslant r < 1$, and f_ϑ has the moment vector $t\big((1-r)^2, \ 2(1 - r + r \ln r)\big)$. By a straightforward calculation, the moment vectors of f_ϑ, $\vartheta \in \Theta_\beta$, exhaust the subcone of $cn_\varphi(\mu)$ given by $2a_1 \geqslant a_2$. It follows that for a in this cone the primal solution g_a exists.

In the case $a_2 > 2a_1 > 0$ Proposition 4.18 is employed. Since $B_\gamma(g, h) = \frac{1}{2}\|g - h\|^2_{L_2(\mu)}$ for g, h nonnegative measurable, Bregman closure equals the $L_2(\mu)$-closure. The $L_2(\mu)$-closure of $\mathcal{G}_a^+$ with $0 < a_1 < a_2$ contains each function f_ϑ, $\vartheta \in \Theta_\gamma$, whose moment vector (b_1, b_2) satisfies $b_1 = a_1$, $b_2 < a_2$. Indeed, for $n \geqslant 1$ there exists $x_n > 1$ such that

$$(a_2 - b_2)n = \int_1^{x_n} \frac{\mathrm{d}z}{z \ln z} - \frac{b_2}{a_1} \int_1^{x_n} \frac{\mathrm{d}z}{z^2 \ln z} \, ,$$

by continuity. Then, $x_n \uparrow +\infty$, the function

$$z \mapsto \left[1 - \frac{1}{a_1 n} \int_1^{x_n} \frac{\mathrm{d}r}{r^2 \ln r} \right] f_\vartheta(z) + \frac{z}{2n \ln z} \mathbf{1}_{(1, x_n)}(z)$$

belongs to $\mathcal{G}_a$, and $L_2(\mu)$-converges to f_ϑ as $n \to +\infty$ because $\frac{z}{\ln z} \in L_2(\mu)$.

The set Θ_β^+ consists of those $\vartheta \in \Theta_\gamma$ for which $\vartheta_1 + \vartheta_2 z \geqslant 0,\ z \in Z$. Thus, it is given by $\vartheta_1 \geqslant 0$ and $\vartheta_2 = 0$. For such $(\vartheta_1, \vartheta_2)$, the function f_ϑ equals identically ϑ_1 and has the moment $(\vartheta_1, 2\vartheta_1)$. By the above result, if $a_2 > 2a_1 > 0$ then the $L_2(\mu)$-closure of $\mathcal{G}_a$ contains f_ϑ with $\vartheta_1 = a_1$. Hence, Proposition 4.18 implies that the effective dual solution g_a^* equals the constant a_1. Note that the last assertion of Proposition 4.18 fails if the restriction $\vartheta \in \Theta_\beta^+$ is dropped.

11. RELATION OF THIS WORK TO PREVIOUS ONES

The subject addressed in this paper has one of its origins in the principle of maximum entropy (MaxEnt) which comes from statistical physics and has been promoted as a general principle of inference primarily by Jaynes [37] and Kullback [39]. While MaxEnt calls for maximizing Shannon entropy or for minimizing I-divergence (Kullback–Leibler distance [40]), maximization of Burg entropy [14, 15] and other 'entropy functionals' is also widely used in sciences. These applications motivated the formulation of the general minimization problem in eq. (2) with autonomous integrands. It is for convenience that minimization of convex integral functionals is addressed, maximization of concave ones as Shannon or Burg entropy is covered by taking their negatives.

The literature of the subject is extensive, some pointers are given here to works that have influenced ours. The integral functional (1) with an autonomous integrand $\beta(z,t) = \Phi(t)$ is called Φ-entropy in [9, 51], where for a growing number of constraints, convergence of the solutions of the problem (2) is established under suitable conditions. Divergences of form (44) have been introduced by Csiszár [18, 19], called f-divergences, and by Ali and Silvey [1]. They, as well as I-divergence, were originally defined for probability densities g and h only. Substantial developments in their theory and applications are due to I. Vajda, see e. g. [45]. More recent references include [24, 13, 3]. Bregman distances were introduced in [12] as non-metric distances between vectors in $\mathbb{R}^d$, associated with a convex function on $\mathbb{R}^d$, for numerous applications in convex programming problems see the book [16]. The subclass of separable Bregman distances is the one whose infinite dimensional extension is used in this paper, see Remark 2.11. Their statistical applications, initiated by Jones and Byrne [38], are currently wide ranging, see e. g. Murata et al [46]. For more general Bregman distances see e. g. [6, 35]. The axiomatic study [22] highlights (in the finite dimensional case) the distinguished role of γ-divergences and separable Bregman distances, and primarily that of I-divergence. The problem of minimizing convex integral functionals arises also in large deviations theory, for interplay with this field see e. g. [32, 36, 24, 44]. Other fields could also be mentioned, such as in control theory where, typically, also derivatives of the unknown function are involved. Regarding a possible interplay see [11].

A large body of the literature on the minimization problem of eq. (2) is application oriented and mathematically non-rigorous. For example, while the form of the solution like (4) is derived via Lagrange multipliers, little attention is payed to conditions under which a solution exists, and is indeed of this form when it does. Early rigorous results about I-divergence minimization and associated Pythagorean (in)equalities were obtained by Chentsov [17] and Csiszár [20]. Recent works typically employ convex duality, following the lead of Borwein and Lewis, see [8, 9, 10]. Previously, convex duality had been applied to the I-divergence minimization problem in [7]. Advanced tools from functional analysis appear indispensable to efficiently deal with the case, not treated here, when the range of the moment mapping is infinite dimensional, see Léonard [41]–[44]. Léonard's results are strong and general also when restricted to finite dimensional mappings. Still, they appear to require assumptions on the integrand and the moment mapping not needed here, e. g., that $\beta(z,t)$ is nonnegative and equals 0 for some $t = t_z$. Another tools are provided by differential geometry [2, 3], first applied in the MaxEnt context by Chentsov [17]. They require strong regularity conditions but lead to impressive 'geometric' results for example about Pythagorean identities, going beyond those obtainable otherwise.

This paper generalizes the results obtained for the Shannon case in [26]. Convex duality is used, as there, for the value function only, a convex conjugate of the integral functional is not needed. Accordingly, the functional is not restricted to a 'good' space that has a manageable dual space, as frequently done in the literature. A key tool in [26] has been the convex core of a measure on $\mathbb{R}^d$, introduced in [25]. In the present generality, its role is played by the concept of conic core, introduced here. The framework is in several respects more general than usual: (i) non-differentiable integrands are allowed (ii) the integrands need not be autonomous (iii) there are no restrictions beyond measurability, neither on the functions g over which the functional is minimized nor on the moment mapping φ, other than that φ has finite dimensional range. While feature (i) is not unique for this paper, in the literature often stronger assumptions are adopted on β than differentiability (in addition to strict convexity which is assumed also here). Typical ones are essential smoothness plus cofiniteness, or the equivalent assumption that β^* is strictly convex on $\mathbb{R}$, as in [46]. Non-autonomous integrands (also admitted in [43, 44]) do not cause conceptual difficulties but do cause technical ones concerning measurability. These are handled here via the theory of normal integrands initiated in [47, 48] and summarized in detail in the recent book [50]. The latter is relied upon in the paper also elsewhere.

The limitations of our framework are, in addition to restricting attention to moment mappings of finite dimensional range, that only equality constraints are considered, and the integrand value $\beta(z,t)$ has to be finite for $t > 0$ and $+\infty$ for $t < 0$. To consider only equality constraints does not seem a serious restriction. It should not be difficult to extend the results to constraints of the form that the moment vector belongs to a convex subset of $\mathbb{R}^d$, see e. g. [24]. Our restriction on the integrand is not needed for the mere extension of familiar results to the generality of (i)–(iii). It is, however, essential for the main results, viz. the geometric characterization of the effective domain of the value function and its implications that extend several results previously proved only under the PCQ beyond that assumption. These main results are relevant in those cases when the effective domain includes a nontrivial boundary, which is typical when the underlying measure μ has discrete components. In 'classical' moment problems involving Lebesgue measure on $\mathbb{R}^k$ and moment mappings formed by polynomials or trigonometric polynomials, the need for going beyond the PCQ does not arise (while the problem of nonexistence of primal solution does).

As genuine primal and dual solutions do not always exist, also generalized ones are studied which are universal limits of minimizing (maximizing) sequences. Another concept of generalized solution, not used in this paper, involves relaxation of the problem (2) to minimization over a larger space. The latter is typically obtained regarding the functional to be defined over a specific linear space with a manageable dual space, which paves a road to extend the functional to the second dual. In [41, 42, 43] this extension is to the topological dual of an Orlicz space. In [44] it is also shown in considerable generality that the 'absolutely continuous component' of this kind of generalized solution coincides with the generalized solution in our sense; a similar but more special result appeared previously in [23]. Note that nonexistence of a minimizer in eq. (2) had emerged as a practical problem in the context of three dimensional density reconstruction via Burg entropy maximization, see references in [10] where the mathematical background of this phenomenon has been clarified.

By key results of this paper, generalized primal and dual solutions in our sense exist, subject to the DCQ, in all nontrivial cases, and their Bregman distance is a lower bound to the duality gap. The adopted concept of generalized solution dates back to Topsoe [52] who, for Shannon entropy maximization over a convex set of probability distributions, established a Pythagorean inequality involving a 'center of attraction' perhaps not in that set. Actually, the existence of generalized I-projections is implicit already in [20]. They were studied in detail in [21, 26]; in [21] also their relevance for large deviations theory is demonstrated. Generalized minimizers of integral functionals with (differentiable) autonomous integrands $\gamma \in \mathit{\Gamma}$, for arbitrary convex sets of functions, were introduced and corresponding Pythagorean inequalities established in [23]. Generalized primal solutions for the minimization problem in eq. (2), assuming the PCQ and implicitly the DCQ, are treated in [24]. Note that the existence result [23, Theorem 1(c)] does not hold in full generality, its proof contains a gap of implicitly assuming that functions in a minimizing sequence always have bounded Bregman distance from some fixed function. For the minimization problem in eq. (2) the latter is true if the DCQ holds, due to Lemma 4.15, hence the existence of a generalized primal solution subject to the DCQ does follow from [23, Theorem 1(c)]. In this respect, the new feature of the result here is that it explicitly specifies the generalized primal solution.

Generalized dual solutions for the I-divergence case, viz. generalized maximum likelihood estimates, have been treated in [27]; stronger results including their explicit description appear in [28]. In the present paper the existence of generalized dual solution is proved, for any integral functional with integrand $\beta \in B$ subject to the DCQ, via a nontrivial updating of the technique used in [27], similar to that in [23], going back to [52]. This proof works also when the moment mapping has infinite dimensional range, but gives no indication how to construct the generalized dual solution. The latter remains an open problem. Note that when the duality gap is zero, a direct proof shows that, subject to the DCQ, the generalized dual solution exists and coincides with the generalized primal solution, which has been explicitly described.

Finally, let us comment on the family of functions $\mathcal{F}_\beta = \{f_\vartheta \colon \vartheta \in \mathit{\Theta}_\beta\}$. It has been observed many times that $\mathcal{F}_\beta$ plays the same role as a classical exponential family does in the Shannon case, see Appendix B for formal details. For statistical applications see e. g. [46] where our $\mathcal{F}_\beta$ is called U-model (with $U = \beta^*$). For the classical theory of exponential families see [17, 4]. For a very general concept see [33]. Apparently, it has not been pointed out before that some key properties of exponential families extend to $\mathcal{F}_\beta$ only when β is essentially smooth. If β is merely differentiable, those properties extend only partially, to functions in $\mathcal{F}_\beta$ parameterized by ϑ in a specific set $\mathit{\Theta}^+_\beta \subseteq \mathit{\Theta}_\beta$, see Proposition 4.18 and Theorems 8.6, 8.9.

APPENDICES

A. INTEGRAL REPRESENTATION

The proof of Theorem 1.1 depends on an interchange of minimization and integration, see below. We are not aware of a reference that would guarantee admissibility of this interchange in the required generality, though in the special case of finite μ and μ-integrable φ, [50, Theorem 14.60] suffices. An extension of the latter, Theorem A.4, will be proved below and applied to cover the general case.

A linear space $\mathcal{H}$ of real $\mathcal{Z}$-measurable functions is *decomposable* w.r.t. μ [50, Definition 14.59] if $g\boldsymbol{1}_{Z\setminus A}+h\boldsymbol{1}_A$ belongs to $\mathcal{H}$ whenever $g \in \mathcal{H}$, $A \in \mathcal{Z}$ has finite μ-measure and h is bounded $\mathcal{Z}$-measurable. For μ finite, $\mathcal{H}$ is decomposable if and only if it contains all bounded $\mathcal{Z}$-measurable functions. A weaker notion is introduced as follows.

Definition A.1. A space $\mathcal{H}$ is *σ-decomposable* w.r.t. μ if Z can be covered by a countable family of sets $Z_n \in \mathcal{Z}$ with $\mu(Z_n)$ finite such that $g\boldsymbol{1}_{Z\setminus A} + h\boldsymbol{1}_A \in \mathcal{H}$ whenever $g \in \mathcal{H}$, $A \in \mathcal{Z}$ is contained in some Z_n and h is bounded $\mathcal{Z}$-measurable.

Remark A.2. There is no loss of generality in assuming that $Z_n \subseteq Z_{n+1}$ in Definition A.1.

Remark A.3. If $\varphi\colon Z \to \mathbb{R}^d$ is any moment mapping then the space $\mathcal{G}$, consisting of those functions g for which the moment vector $\int_Z \varphi g \,\mathrm{d}\mu$ exists, is σ-decomposable w.r.t. μ. Namely, by σ-finiteness, Z can be covered by sets $Y_n \in \mathcal{Z}$ of finite measure, and the countable family $Z_{n,m} = Y_n \cap \{\|\varphi\| \leqslant m\}$ indexed by n, m is suitable. In fact, if $A \in \mathcal{Z}$ is contained in some $Z_{n,m}$ and h is bounded $\mathcal{Z}$-measurable then $g \in \mathcal{G}$ implies existence of $\int_{Z\setminus A} \varphi g \,\mathrm{d}\mu + \int_A \varphi h \,\mathrm{d}\mu$. The space $\mathcal{G}$ need not be decomposable w.r.t. μ even if μ is finite. For example, if $Z = \mathbb{R}$, $\mathrm{d}\mu = \frac{\mathrm{d}z}{1+z^2}$ and $\varphi(z) = z$ then $\mathcal{G}$ does not contain the constant functions.

The following assertion on interchange of minimization and integration is an extension of [50, Theorem 14.60] to the σ-decomposable spaces.

Theorem A.4. *Let $\mathcal{H}$ be a σ-decomposable linear space of $\mathcal{Z}$-measurable functions on a σ-finite measure space $(Z, \mathcal{Z}, \mu)$, and let $\alpha\colon Z\times\mathbb{R} \to [-\infty, +\infty]$ be a normal integrand such that the integral functional $\mathsf{H}_\alpha(g) \triangleq \int_Z \alpha(z, g(z))\, \mu(\mathrm{d}z)$ does not identically equal $+\infty$ for $g \in \mathcal{H}$. Then*

$$\inf_{g\in\mathcal{H}} \int_Z \alpha(z, g(z))\, \mu(\mathrm{d}z) = \int_Z \inf_{t\in\mathbb{R}} \alpha(z, t)\, \mu(\mathrm{d}z)\,. \tag{42}$$

The function on Z that is integrated on the right is denoted by $\alpha_{\inf}$. It is $\mathcal{Z}$-measurable by [50, Theorem 14.37]. The following lemma does not involve $\mathcal{H}$.

Lemma A.5. *If $\int_Z \alpha_{\inf} \,\mathrm{d}\mu < t$ then $\mathsf{H}_\alpha(h) < t$ for some real $\mathcal{Z}$-measurable function h.*

P r o o f. This is proved neatly on the lines 7–17 of the proof of [50, Theorem 14.60]. □

Lemma A.6. *If $\mathcal{H}$ is σ-decomposable, $\mathsf{H}_\alpha(f)$ finite for some $f \in \mathcal{H}$, and $\int_Z \alpha_{\inf} \,\mathrm{d}\mu < t$ then $\mathsf{H}_\alpha(g) < t$ for some $g \in \mathcal{H}$.*

Proof. By Lemma A.5, $H_\alpha(h) < t$ for some $\mathcal{Z}$-measurable h. Let $0 < 2\varepsilon < t - H_\alpha(h)$. By σ-decomposability and Remark A.2, Z is covered by a countable increasing sequence Z_n with $\mu(Z_n)$ finite that has the property from Definition A.1. Since $H_\alpha(f)$ is finite, for n sufficiently large

$$\varepsilon \geqslant \int_{Z \setminus Z_n} \alpha(z, f(z))\, \mu(\mathrm{d}z) \quad \text{and} \quad t - 2\varepsilon > \int_{Z_n} \alpha(z, h(z))\, \mu(\mathrm{d}z)\,.$$

For m sufficiently large

$$\varepsilon \geqslant \int_{Z_n \setminus \{h \leqslant m\}} \alpha(z, f(z))\, \mu(\mathrm{d}z) \quad \text{and} \quad t - 2\varepsilon > \int_{Z_n \cap \{h \leqslant m\}} \alpha(z, h(z))\, \mu(\mathrm{d}z)\,.$$

Since h is bounded on $A = Z_n \cap \{h \leqslant m\} \subseteq Z_n$, the function $g = f\boldsymbol{1}_{Z\setminus A} + h\boldsymbol{1}_A$ belongs to $\mathcal{H}$ by σ-decomposability. Combining the above inequalities, $H_\alpha(g) < t$. $\square$

Proof of Theorem A.4. The inequality $\geqslant$ in (42) follows from $\alpha(z, g(z)) \geqslant \alpha_{\inf}(z)$, $z \in Z$, by integration. By assumption, $H_\alpha(f) < +\infty$ for some f in $\mathcal{H}$. If $H_\alpha(f) = -\infty$ then (42) has $-\infty$ on both sides. Otherwise, the inequality $\leqslant$ in (42) follows from Lemma A.6. $\square$

Proof of Theorem 1.1. By definition,

$$J^*_\beta(\vartheta) = -\inf_{a \in \mathbb{R}^d} \big[-\langle \vartheta, a\rangle + \inf_{g \in \mathcal{G}_a} H_\beta(g) \big]\,, \qquad \vartheta \in \mathbb{R}^d\,.$$

The expression on the right rewrites to

$$-\inf_{a \in \mathbb{R}^d} \inf_{g \in \mathcal{G}_a} \int_Z \big[-\langle \vartheta, \varphi(z) g(z)\rangle + \beta(z, g(z)) \big] \mu(\mathrm{d}z)$$

where two infima reduce to one, over $g \in \mathcal{G}$. By Remark A.3, the linear space $\mathcal{G}$ defined by any moment mapping is σ-decomposable. Hence, by Theorem A.4, if $J_\beta \not\equiv +\infty$ then

$$J^*_\beta(\vartheta) = -\int_Z \inf_{t \in \mathbb{R}} \big[-\langle \vartheta, \varphi(z)\rangle t + \beta(z, t) \big] \mu(\mathrm{d}z)\,, \qquad \vartheta \in \mathbb{R}^d\,,$$

and thus $J^*_\beta = K_\beta$. $\square$

The assumption in Theorem 1.1 that J_β is not identically $+\infty$ does matter, see Example 10.1.

B. RESTRICTED VALUE FUNCTION

In this appendix, some details are discussed for the Shannon's integral functional defined by the autonomous integrand $\gamma(t) = t \ln t$, $t > 0$. If g is nonnegative and $\int_Z g\, \mathrm{d}\mu = t > 0$ then $H_\gamma(g) = \gamma(t) + tH_\gamma(g/t)$. Hence, the integral functional H_γ is determined by its values on the probability densities g w.r.t. μ. For such a density, $H_\gamma(g)$ is the negative entropy of the corresponding probability measure w.r.t. μ, or its I-divergence from μ when $\mu(Z) = 1$.

Further, it is assumed that the moment mapping φ has first coordinate identically equal to 1, which is nonrestrictive in many applications, see Section 1.B. Let $\varphi = (1, \psi)$ where $\psi\colon Z \to \mathbb{R}^{d-1}$. For any vector $a \in \mathbb{R}^d$ with a positive first component, writing it as (t,b) with $t>0$ and $b\in\mathbb{R}^{d-1}$, $J_\gamma(a)=\gamma(t)+t\,J_\gamma(1,b/t)$. Hence, the value function J_γ is uniquely determined by its restriction. Let $I_\gamma(b)\triangleq J_\gamma(1,b)$, $b\in\mathbb{R}^{d-1}$. The conjugate of the value function at any point $\vartheta=(r,\tau)$, where $r\in\mathbb{R}$ and $\tau\in\mathbb{R}^{d-1}$, is

$$\begin{aligned} J^*_\gamma(\vartheta) &= \sup\nolimits_{t>0,\, b\in\mathbb{R}^{d-1}} \big[rt+\langle\tau,b\rangle-\gamma(t)-t\,I_\gamma(b/t)\big] \\ &= \sup\nolimits_{t>0}\big[rt-\gamma(t)+tI^*_\gamma(\tau)\big]=\gamma^*(r+I^*_\gamma(\tau))=\exp[r+I^*_\gamma(\tau)-1] \end{aligned} \tag{43}$$

using that the convex conjugate of γ is $\gamma^*(r)=e^{r-1}$, $r\in\mathbb{R}$. By Theorem 1.1, knowing that $H_\gamma(g)=0$ for $g\equiv 0$, J^* admits the integral representation, and hence

$$I^*_\gamma(\tau)=1-r+\ln J^*_\gamma(r,\tau)=\ln\int_Z e^{\langle\tau,\psi\rangle}\,\mathrm{d}\mu\,,\quad \tau\in\mathbb{R}^{d-1}\,.$$

This formula is well-known and has been a key tool when minimizing the negative Shannon entropy $H_\gamma(g)$ of a probability density g subject to moment constraints, e. g. in [26].

The set Θ_γ equals $\mathit{dom}(J^*_\gamma)$, consisting of all $\vartheta\in\mathbb{R}^d$ with $e^{\langle\vartheta,\varphi\rangle}$ μ-integrable. The functions (4) of the family $\mathcal{F}_\gamma$ are given by $f_\vartheta=e^{r+\langle\tau,\psi\rangle-1}$ where $\vartheta=(r,\tau)$. The family of f_ϑ that integrate to 1 is known as the exponential family based on μ with canonical statistic ψ, see [4, 17].

The original and restricted dual problems are also simply related, for $a=(t,b)$ with $t>0$

$$\begin{aligned} J^{**}_\gamma(a) &= \sup\nolimits_{\tau\in\mathbb{R}^{d-1}}\Big[\langle\tau,b\rangle+\sup\nolimits_{r\in\mathbb{R}}\big[rt-\gamma^*(r+I^*_\gamma(\tau))\big]\Big] \\ &= \sup\nolimits_{\tau\in\mathbb{R}^{d-1}}\big[\langle\tau,b\rangle-tI^*_\gamma(\tau)+\gamma(t)\big]=\gamma(t)+t\,I^{**}_\gamma(b/t)\,, \end{aligned}$$

using (43). For example, the duality gap of the original problem at a is t times the duality gap $I^{**}_\gamma(b/t)-I_\gamma(b/t)$ of the restricted problem at b/t.

Note that other integral functionals do not admit such simple formulas that would relate unrestricted and restricted value functions, and their conjugates and biconjugates.

C. γ-DIVERGENCES

Let $\gamma\in\Gamma$ be nonnegative with $\gamma(1)=0$. The γ-divergence of a function $g:Z\to[0,+\infty)$ from $h:Z\to(0,+\infty)$, both $\mathcal{Z}$-measurable, is defined by

$$D_\gamma(g,h)\triangleq\int_Z h\,\gamma(g/h)\,\mathrm{d}\mu\,. \tag{44}$$

This divergence is nonnegative, and equals 0 only if $g=h$ $[\mu]$. If $\gamma(t)=t\ln t-t+1$ then $D_\gamma(g,h)$ is equal to the I-divergence of g from h. As mentioned in subsection 1.B., the minimization of a γ-divergence $D_\gamma(g,h)$ subject to $g\in\mathcal{G}_a$, for fixed h, is a frequently studied instance of the minimization problem addressed in this paper. The integrand $\beta\colon(z,t)\mapsto h(z)\,\gamma(t/h(z))$, $z\in Z$, $t\in\mathbb{R}$, in (44) is in general non-autonomous, but belongs to B and $D_\gamma(g,h)=H_\beta(g)$. Lemma C.1 below shows that a reformulation of the minimization with an autonomous integrand is possible.

A general idea is to modify simultaneously a measure μ and integrand $\beta\in B$ to $\tilde\mu$ and $\tilde\beta$ given by

$$\mathrm{d}\tilde\mu=h\,\mathrm{d}\mu\qquad\text{and}\qquad\tilde\beta(z,t)=\beta(z,t\,h(z))/h(z)\,,\quad z\in Z\,,t\in\mathbb{R}\,,$$

where h is a given positive $\mathcal{Z}$-measurable function. By [50, Proposition 14.45], $\tilde{\beta} \in B$. As earlier, the dependence on μ in H, J, $\mathcal{G}_a$ is added to indices while the moment mapping φ is not changed.

Lemma C.1. *Given a positive $\mathcal{Z}$-measurable function h, let $\tilde{g} = g/h$ for any $\mathcal{Z}$-measurable function g. Then $H_{\mu,\beta}(g) = H_{\tilde{\mu},\tilde{\beta}}(\tilde{g})$ and $\int_Z g\varphi \,\mathrm{d}\mu = \int_Z \tilde{g}\varphi \,\mathrm{d}\tilde{\mu}$ if one of the integrals exists. Further, $g \in \mathcal{G}_{\mu,a}$ if and only if $\tilde{g} \in \mathcal{G}_{\tilde{\mu},a}$, and $J_{\mu,\beta} = J_{\tilde{\mu},\tilde{\beta}}$.*

A simple proof based on substitutions in integrals is omitted.

When considering the minimization of $D_\gamma(g, h)$ subject to a moment constraint on g, Lemma C.1 applies with the very function h from the divergence. The corresponding integrand $\tilde{\beta}$ is autonomous and coincides with γ. Hence, the minimization of $D_\gamma(\cdot, h)$ over $\mathcal{G}_a$ is equivalent to that of $H_{\tilde{\mu},\gamma}$ over $\mathcal{G}_{\tilde{\mu},a}$, whence an autonomous integrand suffices.

Remark C.2. The function $h > 0$ in Lemma C.1 can always be chosen to make the measure $\tilde{\mu}$ finite. Therefore, the finiteness of the underlying measure could have been assumed throughout this paper, without any loss of generality.

ACKNOWLEDGEMENT

This work was supported by the Hungarian National Foundation for Scientific Research under Grant K76088, and by Grant Agency of the Czech Republic under Grants P202/10/0618 and 201/08/0539. Preliminary results on this research were published in *Proceedings ISIT* 2008 [29] and *Proceedings ITW* 2009 [30]. An abridged version of the presented results appears in *Proceedings ISIT* 2012 [31].

(Received December 3, 2011)

REFERENCES

[1] S. M. Ali and S. D. Silvey: A general class of coefficients of divergence of one distribution from another. J. Roy. Statist. Soc. Ser. B *28* (1966) 131–142.

[2] S. Amari and H. Nagaoka: Methods of Information Geometry. Transl. Math. Monographs *191*, Oxford Univ. Press, 2000.

[3] S. Amari and A. Cichocki: Information geometry of divergence functions. Bull. Polish Acad. Sci. *58* (2010) 183–194.

[4] O. Barndorff-Nielsen: Information and Exponential Families in Statistical Theory. Wiley, 1978.

[5] H. H. Bauschke and J. M. Borwein: Legendre functions and the method of random Bregman projections. J. Convex Anal. *4* (1997), 27–67.

[6] H. H. Bauschke, J. M. Borwein, and P. L. Combettes: Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. Comm. Contemp. Math. *3* (2001), 615–647.

[7] A. Ben-Tal and A. Charnes: A dual optimization framework for some problems of information theory and statistics. Problems Control Inform. Theory *8* (1979), 387–401.

[8] J. M. Borwein and A. S. Lewis: Duality relationships for entropy-like minimization problems. SIAM J. Control Optim. *29* (1991), 325–338.

[9] J. M. Borwein and A. S. Lewis: Convergence of best entropy estimates. SIAM J. Optim. *1* (1991), 191–205.

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

[11] J. M. Borwein, A. S. Lewis, and D. Noll: Maximum entropy spectral analysis using derivative information. Part I: Fisher information and convex duality. Math. Oper. Res. *21* (1996), 442–468.

[12] L. M. Bregman: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. and Math. Phys. *7* (1967), 200–217.

[13] M. Broniatowski and A. Keziou: Minimization of ϕ-divergences on sets of signed measures. Studia Sci. Math. Hungar. *43* (2006), 403–442.

[14] J. P. Burg: Maximum entropy spectral analysis. Paper presented at 37th Meeting of Soc. Explor. Geophysicists, Oklahoma City 1967.

[15] J. P. Burg: Maximum entropy spectral analysis. Ph.D. Thesis, Dept. Geophysics, Stanford Univ., Stanford 1975.

[16] Y. Censor and S. A. Zenios: Parallel Optimization. Oxford University Press, New York 1997.

[17] N. N. Chentsov: Statistical Decision Rules and Optimal Inference. Transl. Math. Monographs *53*, American Math. Soc., Providence 1982. Russian original: Nauka, Moscow 1972.

[18] I. Csiszár: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci. *8* (1963), 85–108.

[19] I. Csiszár: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. *2* (1967), 299–318.

[20] I. Csiszár: I-divergence geometry of probability distributions and minimization problems. Ann. Probab. *3* (1975), 146–158.

[21] I. Csiszár: Sanov property, generalized I-projection and a conditional limit theorem. Ann. Probab. *12* (1984), 768–793.

[22] I. Csiszár: Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. Ann. Statist. *19* (1991), 2031–2066.

[23] I. Csiszár: Generalized projections for non-negative functions. Acta Math. Hungar. *68* (1995), 1–2, 161–185.

[24] I. Csiszár, F. Gamboa, and E. Gassiat: MEM pixel correlated solutions for generalized moment and interpolation problems. IEEE Trans. Inform. Theory *45* (1999), 2253–2270.

[25] I. Csiszár and F. Matúš: Convex cores of measures on $\mathbb{R}^d$. Studia Sci. Math. Hungar. *38* (2001), 177–190.

[26] I. Csiszár and F. Matúš: Information projections revisited. IEEE Trans. Inform. Theory *49* (2003), 1474–1490.

[27] I. Csiszár and F. Matúš: Generalized maximum likelihood estimates for infinite dimensional exponential families. In: Proc. Prague Stochastics'06, Prague 2006, pp. 288–297.

[28] I. Csiszár and F. Matúš: Generalized maximum likelihood estimates for exponential families. Probab. Theory Related Fields *141* (2008), 213–246.

[29] I. Csiszár and F. Matúš: On minimization of entropy functionals under moment constraints. In: Proc. ISIT 2008, Toronto, pp. 2101–2105.

[30] I. Csiszár and F. Matúš: On minimization of multivariate entropy functionals. In: Proc. ITW 2009, Volos, Greece, pp. 96–100.

[31] I. Csiszár and F. Matúš: Minimization of entropy functionals revisited. In: Proc. ISIT 2012, Cambridge, MA, pp. 150–154.

[32] D. Dacunha-Castelle and F. Gamboa: Maximum d'entropie et problème des moments. Ann. Inst. H. Poincaré Probab. Statist. *26* (1990), 567–596.

[33] A. P. Dawid and P. D. Grünwald: Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. Ann. Statist. *32* (2004), 1367–1433.

[34] S. Eguchi: Information geometry and statistical pattern recognition. Sugaku Expositions, Amer. Math. Soc. *19* (2006), 197–216.

[35] B. A. Frigyik, S. Srivastava, and M. R. Gupta: Functional Bregman divergence and Bayesian estimation of distributions. IEEE Trans. Inform. Theory *54* (2008), 5130–5139.

[36] F. Gamboa and E. Gassiat: Bayesian methods and maximum entropy for ill-posed inverse problems. Ann. Statist. *25* (1997), 1, 328–350.

[37] E. T. Jaynes: Information theory and statistical mechanics. Physical Review Ser. II *106* (1957), 620–630.

[38] L. Jones and C. Byrne: General entropy criteria for inverse problems with application to data compression, pattern classification and cluster analysis. IEEE Trans. Inform. Theory *36* (1990), 23–30.

[39] S. Kullback: Information Theory and Statistics. John Wiley and Sons, New York 1959.

[40] S. Kullback and R. A. Leibler: On information and sufficiency. Ann. Math. Statist. *22* (1951), 79–86.

[41] C. Léonard: Minimizers of energy functionals. Acta Math. Hungar. *93* (2001), 281–325.

[42] C. Léonard: Minimizers of energy functionals under not very integrable constraints. J. Convex Anal. *10* (2003), 63–68.

[43] C. Léonard: Minimization of entropy functionals. J. Math. Anal. Appl. *346* (2008), 183–204.

[44] C. Léonard: Entropic projections and dominating points. ESAIM: Probability and Statistics *14* (2010), 343–381.

[45] F. Liese and I. Vajda: Convex Statistical Distances. Teubner Texte zur Mathematik *95*, Teubner Verlag, Leipzig 1986.

[46] N. Murata, T. Takenouchi, T. Kanamori, and S. Eguchi: Information geometry of U-Boost and Bregman divergence. Neural Computation *16* (2004), 1437–1481.

[47] R. T. Rockafellar: Integrals which are convex functionals. Pacific J. Math. *24* (1968), 525–539.

[48] R. T. Rockafellar: Convex integral functionals and duality. In: Contributions to Nonlinear Functional Analysis (E. H. Zarantonello, ed.), Academic Press, New York 1971, pp. 215–236.

[49] R. T. Rockafellar: Convex Analysis. Princeton University Press, Princeton 1970.

[50] R. T. Rockafellar and R. J.-B. Wets: Variational Analysis. Springer Verlag, Berlin – Heidelberg – New York 2004.

[51] M. Teboulle and I. Vajda: Convergence of best ϕ-entropy estimates. IEEE Trans. Inform. Theory *39* (1993), 297–301.

[52] F. Topsoe: Information-theoretical optimization techniques. Kybernetika *15* (1979), 8–27.

[53] I. Vajda: Theory of Statistical Inference and Information. Kluwer Academic Puplishers, Dordrecht 1989.

INDEX

Imre Csiszár, A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, H-1364 Budapest, P.O.Box 127. Hungary.
e-mail: csiszar.imre@renyi.mta.hu

František Matúš, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: matus@utia.cas.cz