

Convergence of generalized entropy minimizers in sequences of convex problems

Imre Csiszár

A. Rényi Institute of Mathematics
Hungarian Academy of Sciences
H-1364 Budapest, P.O.Box 127, Hungary
Email: csiszar@renyi.hu

František Matúš

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
18208 Prague, P.O.Box 18, Czech Republic
Email: matus@utia.cas.cz

Abstract—Integral functionals based on convex normal integrands are minimized over convex constraint sets. Generalized minimizers exist under a boundedness condition. Sequences of the minimization problems are studied when the constraint sets are nested. The corresponding sequences of generalized minimizers are related to the minimization over limit convex sets. Martingale theorems and moment problems are discussed.

I. INTRODUCTION

Let (Z, \mathcal{Z}, μ) be a σ -finite measure space. For real \mathcal{Z} -measurable functions g on Z , let

$$H_\beta: g \mapsto \int_Z \beta(z, g(z)) \mu(dz).$$

The functional H_β is based on an integrand β on $Z \times \mathbb{R}$ with values in $(-\infty, +\infty]$. Assumptions on β , summarized at the beginning of Section II, include strict convexity in the second coordinate, finiteness of $\beta(z, t)$ for $t > 0$ and $\beta(z, t) = +\infty$ for $t < 0$. Thus, H_β is finite only when g is nonnegative, μ -a.e.

This work generalizes and strengthens existing results on minimization of the integral functional H_β over convex sets \mathcal{C} of functions g for which $\inf_{\mathcal{C}} H_\beta \triangleq \inf_{g \in \mathcal{C}} H_\beta(g)$ is finite. If a minimizer g in \mathcal{C} exists, $H_\beta(g) = \inf_{\mathcal{C}} H_\beta$, then it is unique, μ -a.e., by strict convexity. Otherwise, the minimizing sequences g_n in \mathcal{C} , $H_\beta(g_n) \rightarrow \inf_{\mathcal{C}} H_\beta$, are of interest. The minimization of H_β over \mathcal{C} has a *generalized minimizer* if all minimizing sequences have a common limit, denoted here by $\hat{g}_{\mathcal{C}}$. Theorem 1 implies existence of $\hat{g}_{\mathcal{C}}$ when a single minimizing sequence is bounded. Corollary 1 presents equivalent assumptions.

For a sequence \mathcal{C}_n of convex sets of functions g let

$$(P_n) \quad J_\beta^{[n]} = \inf_{\mathcal{C}_n} H_\beta, \quad n \geq 1.$$

Assuming monotonicity, either $\mathcal{C}_n \supseteq \mathcal{C}_{n+1}$ or $\mathcal{C}_n \subseteq \mathcal{C}_{n+1}$, let \mathcal{C}_∞ denote the intersection or union of the sets \mathcal{C}_n , respectively. The goal is to relate the problems (P_n) to

$$(P_\infty) \quad J_\beta^{[\infty]} = \inf_{\mathcal{C}_\infty} H_\beta.$$

Theorems 2 and 3, formulated in Section III, deal with the limit behavior of the sequence $\hat{g}_{\mathcal{C}_n}$ of generalized minimizers in the problems (P_n) . Under some conditions the convergence to the generalized minimizer $\hat{g}_{\mathcal{C}_\infty}$ in (P_∞) is established in Bregman distance based on β .

In Section IV, the results are applied to extend martingale theorems. Discussion of moment problems is presented in Section V. Proofs are postponed to Appendix.

II. PRELIMINARIES AND ASSUMPTIONS

It is assumed throughout that β is a normal integrand [7, Chapter 14] such that $\beta(z, \cdot)$, $z \in Z$, belongs to the class Γ of functions γ on \mathbb{R} that are finite and strictly convex for $t > 0$ and equal to $+\infty$ for $t < 0$. The integrand is *autonomous* when $\gamma \in \Gamma$ exists such that $\beta(z, \cdot) = \gamma$ for all $z \in Z$.

A function γ is asymptotically nonlinear if either the limit $\gamma'(+\infty)$ of $\gamma'(t)$ when $t \uparrow +\infty$ is infinite or the increasing function $t \mapsto t\gamma'(+\infty) - \gamma(t)$ is unbounded. The integrand β is asymptotically nonlinear if $\beta(z, \cdot)$ has the property for μ -a.a. $z \in Z$.

From now on, all functions f, g, h on Z are assumed to be nonnegative and \mathcal{Z} -measurable. Equalities between them are considered μ -a.e. If neither the positive nor the negative part of the function $z \mapsto \beta(z, g(z))$ is μ -integrable, the integral of this function over Z is $+\infty$ by convention.

The *Bregman distance* of g and h , based on β , is given by

$$B_\beta(g, h) \triangleq \int_Z \Delta_\beta(z, g(z), h(z)) \mu(dz)$$

where Δ_β is a nonnegative integrand on $Z \times \mathbb{R}^2$ such that $\Delta_\beta(z, s, t)$ for $z \in Z$ and $s, t \geq 0$ is equal to

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

and $s \cdot (+\infty)$ otherwise. Here, γ abbreviates $\beta(z, \cdot)$, $\text{sgn}(r)$ denotes $+$ if $r \geq 0$ and $-$ if $r < 0$, and γ'_\pm are the one-sided derivatives. The Bregman distance is nonnegative and vanishes if and only if $g = h$. For more technical details see [5].

A sequence of functions g_n on Z converges *locally in the measure* μ to a function h , in symbols $g_n \rightsquigarrow h$, if it converges in measure on each set C of finite μ -measure, thus

$$\mu(C \cap \{|g_n - h| > \varepsilon\}) \rightarrow 0, \quad \varepsilon > 0.$$

Either of the convergences $B_\beta(g_n, h) \rightarrow 0$ or $B_\beta(h, g_n) \rightarrow 0$ implies $g_n \rightsquigarrow h$ [5, Corollary 2.14].

A set of functions \mathcal{H} on Z is *bounded locally in the measure* μ if for each set $C \in \mathcal{Z}$ with finite $\mu(C)$ to every $\varepsilon > 0$ there exists K such that

$$\sup_{h \in \mathcal{H}} \mu(C \cap \{h > K\}) \leq \varepsilon.$$

A set of functions \mathcal{G} or \mathcal{H} is β -*bounded* or *reversely β -bounded* if there exist functions g, h such that, respectively,

$$\sup_{\mathcal{G}} B_\beta(\cdot, h) < +\infty \quad \text{or} \quad \sup_{\mathcal{H}} B_\beta(g, \cdot) < +\infty.$$

III. MAIN RESULTS

The following assertion extends [3, Theorem 1(c)] that is confined to autonomous integrands. There, an assumption of boundedness was missing, though implicitly used in a proof, see also [5, Example 10.5].

Theorem 1. *Let \mathcal{C} be a convex set of functions on Z with finite $\inf_{\mathcal{C}} H_{\beta}$. If a minimizing sequence $g_n \in \mathcal{C}$ is bounded locally in measure then there exists a unique function $\hat{g}_{\mathcal{C}}$ such that*

$$(1) \quad H_{\beta}(g) \geq \inf_{\mathcal{C}} H_{\beta} + B_{\beta}(g, \hat{g}_{\mathcal{C}}), \quad g \in \mathcal{C}.$$

Remark 1. If the finite infimum in (1) is attained, the function $\hat{g}_{\mathcal{C}}$ from Theorem 1 equals the minimizer and

$$H_{\beta}(g) \geq H_{\beta}(\hat{g}_{\mathcal{C}}) + B_{\beta}(g, \hat{g}_{\mathcal{C}}), \quad g \in \mathcal{C}.$$

Otherwise, $\hat{g}_{\mathcal{C}}$ is the generalized minimizer in the sense that $B_{\beta}(g_n, \hat{g}_{\mathcal{C}}) \rightarrow 0$ for every minimizing sequence g_n in \mathcal{C} .

The assumption of boundedness has equivalent reformulations.

Corollary 1. *Assuming $\inf_{\mathcal{C}} H_{\beta}$ is finite, the following assertions are equivalent.*

- (i) *A minimizing sequence is bounded locally in measure.*
- (ii) *Every minimizing sequence is eventually β -bounded.*
- (iii) *The generalized minimizer exists.*
- (iv) *The level sets of H_{β} intersected with \mathcal{C} are β -bounded.*

Let \mathcal{C}_n be a sequence of convex sets of nonnegative functions and (P_n) the corresponding minimization problems with values $J_{\beta}^{[n]}$. When nonincreasing/nondecreasing, in symbols $\mathcal{C}_n \searrow / \mathcal{C}_n \nearrow$, the intersection/union \mathcal{C}_{∞} gives rise to the problem (P_{∞}) . Correspondingly, the sequence of values $J_{\beta}^{[n]}$ is nondecreasing/nondecreasing and upper/lower bounded by $J_{\beta}^{[\infty]}$.

Theorem 2. *Let $\mathcal{C}_n \searrow$ be a sequence of convex sets such that $J_{\beta}^{[n]}$ is finite. Let H_{β} have a minimizing sequence in \mathcal{C}_1 that is bounded locally in measure. If the sequence $J_{\beta}^{[n]}$ has a finite limit then there exists a unique function h_{∞} such that $B_{\beta}(h_{\infty}, \hat{g}_{\mathcal{C}_n}) \rightarrow 0$ and*

$$(2) \quad H_{\beta}(g) \geq \lim_{n \rightarrow \infty} J_{\beta}^{[n]} + B_{\beta}(g, h_{\infty}), \quad g \in \mathcal{C}_{\infty}.$$

If the finite limit of $J_{\beta}^{[n]}$ equals $J_{\beta}^{[\infty]}$ then $h_{\infty} = \hat{g}_{\mathcal{C}_{\infty}}$.

Remark 2. In Theorem 2, let the infima in the problems (P_n) be attained. By Remark 1, $\hat{g}_{\mathcal{C}_n}$ are the unique minimizers. If the sequence $J_{\beta}^{[n]}$ is bounded then a unique function h_{∞} exists such that $\hat{g}_{\mathcal{C}_n}$ converges in the sense $B_{\beta}(h_{\infty}, \hat{g}_{\mathcal{C}_n}) \rightarrow 0$ and

$$(3) \quad H_{\beta}(g) \geq \lim_{n \rightarrow \infty} H_{\beta}(\hat{g}_{\mathcal{C}_n}) + B_{\beta}(g, h_{\infty}), \quad g \in \mathcal{C}_{\infty}.$$

By Lemma 4 in Appendix, $\hat{g}_{\mathcal{C}_n} \rightsquigarrow h_{\infty}$. Hence, along a subsequence $\hat{g}_{\mathcal{C}_n} \rightarrow h_{\infty}$, μ -a.e. Often H_{β} is lower semicontinuous, $\liminf_{n \rightarrow \infty} H_{\beta}(g_n) \geq H_{\beta}(g)$ whenever $g_n \rightarrow g$, μ -a.e. Then,

$$J_{\beta}^{[\infty]} \geq \lim_{n \rightarrow \infty} H_{\beta}(\hat{g}_{\mathcal{C}_n}) \geq H_{\beta}(h_{\infty}).$$

If $h_{\infty} \in \mathcal{C}_{\infty}$ then the inequalities are tight by (3), $h_{\infty} = \hat{g}_{\mathcal{C}_{\infty}}$ is the minimizer in the problem (P_{∞}) , $B_{\beta}(\hat{g}_{\mathcal{C}_{\infty}}, \hat{g}_{\mathcal{C}_n}) \rightarrow 0$ and $H_{\beta}(\hat{g}_{\mathcal{C}_n}) \rightarrow H_{\beta}(\hat{g}_{\mathcal{C}_{\infty}})$.

Example 1. Let μ be the measure on \mathbb{R}^2 expressible as sum of the probability measure (pm) sitting in the point $(0, 1)$, the pm sitting in $(0, 2)$ and the pm on $(0, +\infty)^2$ with the density $(x_1, x_2) \mapsto e^{-x_1 - x_2}$, see [4, Example 1]. The functional H_{γ} is based on the autonomous integrand $\gamma: t \mapsto t \ln t$.

Given $s \geq 1$, the functional H_{γ} is minimized over

$$\mathcal{C}_n = \bigcup \{ \mathcal{G}(\varepsilon, s) : 0 \leq \varepsilon \leq \frac{1}{n} \}$$

where $\mathcal{G}(a_1, a_2)$ for $(a_1, a_2) \in \mathbb{R}^2$ denotes the set of nonnegative functions g that satisfy the three moment constraints

$$\int_{\mathbb{R}^2} (1, z_1, z_2) g(z_1, z_2) \mu(dz_1, dz_2) = (1, a_1, a_2).$$

The minimum of H_{γ} over $\mathcal{G}(\varepsilon, s)$, $\varepsilon > 0$, is attained uniquely at the function $f_{\varepsilon, s}: (z_1, z_2) \mapsto e^{\vartheta_1^* z_1 + \vartheta_2^* z_2 - \Lambda(\vartheta_1^*, \vartheta_2^*)}$ where

$$\Lambda: (\vartheta_1, \vartheta_2) \mapsto \ln \left[e^{\vartheta_2} + e^{2\vartheta_2} + \frac{1}{1-\vartheta_1} \frac{1}{1-\vartheta_2} \right], \quad \vartheta_1, \vartheta_2 < 1,$$

is the log-Laplace transform of μ and $(\vartheta_1^*, \vartheta_2^*)$ is the unique solution of $\nabla \Lambda(\vartheta_1^*, \vartheta_2^*) = (\varepsilon, s)$. Thus,

$$\begin{aligned} \frac{1}{(1-\vartheta_1^*)^2} \frac{1}{1-\vartheta_2^*} &= \varepsilon \left[e^{\vartheta_1^*} + e^{2\vartheta_1^*} + \frac{1}{1-\vartheta_1^*} \frac{1}{1-\vartheta_2^*} \right] \\ e^{\vartheta_2^*} + 2e^{2\vartheta_2^*} + \frac{1}{1-\vartheta_1^*} \frac{1}{(1-\vartheta_2^*)^2} &= s \left[e^{\vartheta_1^*} + e^{2\vartheta_1^*} + \frac{1}{1-\vartheta_1^*} \frac{1}{1-\vartheta_2^*} \right]. \end{aligned}$$

For fixed $s > s^* \triangleq \frac{1+2e}{1+e}$, the second equation and $\vartheta_2^* \nearrow 1$ imply

$$\frac{1}{1-\vartheta_1^*} \frac{1}{(1-\vartheta_2^*)^2} \rightarrow (e + e^2)(s - s^*) \quad \text{and} \quad \vartheta_1^* \rightarrow -\infty.$$

By the first equation, $\vartheta_1^* \varepsilon \rightarrow 0$ and $\varepsilon \rightarrow 0$. Hence, when $\varepsilon \searrow 0$ the functions $f_{\varepsilon, s}$ converge to zero on $(0, +\infty)^2$, to $\frac{1}{1+e}$ at $(0, 1)$ and to $\frac{e}{1+e}$ at $(0, 2)$. Further,

$$\min_{\mathcal{G}(\varepsilon, s)} H_{\gamma} = \vartheta_1^* \varepsilon + \vartheta_2^* s - \Lambda(\vartheta_1^*, \vartheta_2^*) \rightarrow s - 1 - \ln(1 + e).$$

This minimum is actually $\Lambda^*(\varepsilon, s)$ where Λ^* is the conjugate of Λ . By calculus, for $s > s^*$

$$\begin{aligned} \Lambda^*(\varepsilon, s) &\leq \sup_{\vartheta_1, \vartheta_2 < 1} \left[\vartheta_1 \varepsilon + \vartheta_2 s - \ln \left[e^{\vartheta_2} + e^{2\vartheta_2} + \frac{1}{1-\vartheta_1} e^{\vartheta_2} \right] \right] \\ &= s - 1 + \sup_{\vartheta_1 < 1} \left[\vartheta_1 \varepsilon - \ln \left[1 + e + \frac{1}{1-\vartheta_1} \right] \right] \\ &= s - 1 - \ln(1+e) - 2\sqrt{\varepsilon/(1+e)} + O(\varepsilon). \end{aligned}$$

The sets \mathcal{C}_n intersect to $\mathcal{C}_{\infty} = \mathcal{G}(0, s)$. For $s \leq 2$ this family contains the single function $f_{0, s}$ that equals $2-s$ at $(0, 1)$, $s-1$ at $(0, 2)$, and 0 otherwise. Hence,

$$J_{\beta}^{[\infty]} = (s-1) \ln(s-1) + (2-s) \ln(2-s), \quad 1 \leq s \leq 2.$$

If $s > 2$ then $\mathcal{G}(0, s)$ is empty, thus $J_{\beta}^{[\infty]} = +\infty$.

For $s > s^*$ it follows that the infimum in $J_{\beta}^{[n]}$ is attained and

$$\lim_{n \rightarrow \infty} J_{\beta}^{[n]} = s - 1 - \ln(1 + e) < J_{\beta}^{[\infty]}.$$

Thus, the limit is finite but smaller than $J_{\beta}^{[\infty]}$, which is even infinite for $s > 2$. When n is sufficiently large the minimizer $\hat{g}_{\mathcal{C}_n}$ in the problem (P_n) is equal to the minimizer $f_{1/n, s}$ of H_{γ} over $\mathcal{G}(1/n, s)$, using the upper bound on Λ^* . Theorem 2 applies. The function $h_{\infty} = f_{0, s^*}$ is different from $\hat{g}_{\mathcal{C}_{\infty}} = f_{0, s}$ for $s^* < s \leq 2$, while $\hat{g}_{\mathcal{C}_{\infty}}$ is not defined for $s > 2$.

Theorem 3. Let β be asymptotically nonlinear and $\mathcal{C}_n \nearrow$ be a sequence of convex sets such that $J_\beta^{[n]}$ is finite. Let H_β have a minimizing sequence in \mathcal{C}_n that is bounded locally in measure, $n \geq 1$. If the sequence $J_\beta^{[n]}$ has a finite limit then the limit equals $J_\beta^{[\infty]}$, the minimizing sequences in \mathcal{C}_∞ are β -bounded and $B_\beta(\hat{g}_{\mathcal{C}_n}, \hat{g}_{\mathcal{C}_\infty}) \rightarrow 0$.

Example 2. Let μ be the Lebesgue measure on $[0, 1]$. The functional H_γ is based on $\gamma: t \mapsto t \ln t$ and is minimized over

$$\mathcal{C}_n = \{g \geq 0: \int_0^1 g d\mu = 1, g(t) = 2t \text{ for } t \leq \frac{1}{n}\}.$$

The minimizer exists, $\hat{g}_{\mathcal{C}_n}(t) = 1 + \frac{1}{n}$, $t \geq \frac{1}{n}$, and

$$J_\beta^{[n]} = \int_0^{1/n} 2t \ln(2t) dt + (1 - \frac{1}{n^2}) \ln(1 + \frac{1}{n}) \searrow 0.$$

By Theorem 3, $J_\beta^{[\infty]} = 0$ and $\hat{g}_{\mathcal{C}_\infty} \equiv 1$ is the generalized minimizer in the problem (P_∞) , which has however no minimizer because $\hat{g}_{\mathcal{C}_\infty}$ is not in $\mathcal{C}_\infty = \bigcup_{n \geq 1} \mathcal{C}_n$. Instead, $\hat{g}_{\mathcal{C}_\infty}$ is a minimizer when minimizing over all functions integrating to 1.

IV. MARTINGALE THEOREMS

In this section Theorems 2 and 3 are related to convergence of conditional expectations in Bregman distances.*

A function f has the covering property w.r.t. a sub- σ -algebra \mathcal{Y} of \mathcal{Z} (and μ) if Z is the union of at most countably many sets $Y \in \mathcal{Y}$ with $\int_Y f d\mu$ finite. By Radon-Nikodym theorem, there exists a \mathcal{Y} -measurable function $f|_{\mathcal{Y}} = f|_{\mathcal{Y}, \mu}$ that satisfies $\int_Y f|_{\mathcal{Y}} d\mu = \int_Y f d\mu$ for $Y \in \mathcal{Y}$. It is nonnegative and unique. If $f > 0$ then $f|_{\mathcal{Y}} > 0$.

Let sub- σ -algebras \mathcal{Y}_n of \mathcal{Z} be nested as $\mathcal{Y}_n \subseteq \mathcal{Y}_{n+1}$, in symbols $\mathcal{Y}_n \nearrow$, and generate \mathcal{Y}_∞ . A martingale theorem asserts that if f has the covering property w.r.t. \mathcal{Y}_1 then $f|_{\mathcal{Y}_n} \rightarrow f|_{\mathcal{Y}_\infty}$, μ -a.e. In a backward martingale theorem, sub- σ -algebras are nested as $\mathcal{Y}_n \supseteq \mathcal{Y}_{n+1}$ and intersect to \mathcal{Y}_∞ .

Let $\mathcal{D}_{\mathcal{Y}, f}$ denote the set of nonnegative functions g that satisfy $\int_Y g d\mu = \int_Y f d\mu$ for $Y \in \mathcal{Y}$. The minimization of the functionals H_β over the convex set $\mathcal{C} = \mathcal{D}_{\mathcal{Y}, f}$ is not considered here in full generality but only for integrands

$$\beta: (z, t) \mapsto h(z) \gamma\left(\frac{t}{h(z)}\right), \quad z \in Z, t \in \mathbb{R},$$

where $\gamma \in \Gamma$ and h is a positive function on Z . If $h \equiv 1$ then the integrand is autonomous, which can be arranged replacing μ by ν with $d\nu = h d\mu$ and g by gh , see [5, Appendix C].

It is further assumed that γ is differentiable, nonnegative and $\gamma(1) = 0$. Then, $H_\beta(g)$ becomes the γ -divergence $D_\gamma(g, h)$ of a function g from h

$$H_\beta(g) = \int_Z h \gamma\left(\frac{g}{h}\right) d\mu = D_\gamma(g, h).$$

Lemma 1. If $f \geq 0$ and $h > 0$ have the covering property w.r.t. a sub- σ -algebra \mathcal{Y} of \mathcal{Z} and μ then

$$\min_{g \in \mathcal{D}_{\mathcal{Y}, f}} D_\gamma(g, h) = D_\gamma(f|_{\mathcal{Y}}, h|_{\mathcal{Y}}).$$

If finite, the minimum is attained at $g = f|_{\mathcal{Y}} h/h|_{\mathcal{Y}}$.

*Proofs of the assertions presented here are not included. They are available for reviewer's purposes at <http://staff.utia.cas.cz/matus/>

When the minimum in Lemma 1 is finite, Theorem 1 for $H_\beta = D_\gamma(\cdot, h)$ and $\mathcal{C} = \mathcal{D}_{\mathcal{Y}, f}$ features the minimizer $\hat{g}_{\mathcal{D}_{\mathcal{Y}, f}}$ equal to $f|_{\mathcal{Y}} h/h|_{\mathcal{Y}}$. Ineq. (1) is tight

$$D_\gamma(g, h) = D_\gamma(f|_{\mathcal{Y}}, h|_{\mathcal{Y}}) + B_\beta(g, f|_{\mathcal{Y}} h/h|_{\mathcal{Y}}), \quad g \in \mathcal{D}_{\mathcal{Y}, f}.$$

In fact, if the right hand side is finite the Bregman distance is

$$\int_Z \left[h \gamma\left(\frac{g}{h}\right) - h \gamma\left(\frac{f|_{\mathcal{Y}}}{h|_{\mathcal{Y}}}\right) - \gamma'\left(\frac{f|_{\mathcal{Y}}}{h|_{\mathcal{Y}}}\right) \left[g - \frac{f|_{\mathcal{Y}} h}{h|_{\mathcal{Y}}} \right] \right] d\mu.$$

The third term can be omitted since $g|_{\mathcal{Y}} = f|_{\mathcal{Y}}$. For the autonomous integrand, $h \equiv 1$, $\hat{g}_{\mathcal{D}_{\mathcal{Y}, f}} = f|_{\mathcal{Y}}$ and (1) takes the form

$$H_\gamma(g) = H_\gamma(f|_{\mathcal{Y}}) + B_\gamma(g, f|_{\mathcal{Y}}), \quad g \in \mathcal{D}_{\mathcal{Y}, f}.$$

Lemma 2. If sub- σ -algebras $\mathcal{Y}_n \nearrow$, generate \mathcal{Y}_∞ , and $f \geq 0$ and $h > 0$ have the covering property w.r.t. \mathcal{Y}_1 then

$$D_\gamma(f|_{\mathcal{Y}_n}, h|_{\mathcal{Y}_n}) \nearrow D_\gamma(f|_{\mathcal{Y}_\infty}, h|_{\mathcal{Y}_\infty}).$$

Theorem 2 restricted to the autonomous integrands applies to sequences of sets of the type $\mathcal{D}_{\mathcal{Y}, f}$. Lemmas 1 and 2 are invoked in the proof.

Corollary 2. Under the assumptions of Lemma 2 with $h \equiv 1$, if $H_\gamma(f|_{\mathcal{Y}_\infty})$ is finite then $B_\gamma(f|_{\mathcal{Y}_\infty}, f|_{\mathcal{Y}_n}) \rightarrow 0$.

The choice $\gamma: t \mapsto t \ln t - t + 1$ gives the convergence of relative entropies in Lemma 2, see [6, Theorem 2], while the assertion of Corollary 2 seems to be new.

A consequence of Theorem 3 is analogous to Corollary 2.

Corollary 3. Let γ be asymptotically nonlinear. If $\mathcal{Y}_n \searrow$, intersect to \mathcal{Y}_∞ , $f \geq 0$ and $h > 0$ have the covering property w.r.t. \mathcal{Y}_∞ , and $D_\gamma(f|_{\mathcal{Y}_1}, h|_{\mathcal{Y}_1})$ is finite then

$$D_\gamma(f|_{\mathcal{Y}_n}, h|_{\mathcal{Y}_n}) \searrow D_\gamma(f|_{\mathcal{Y}_\infty}, h|_{\mathcal{Y}_\infty}),$$

and if $H_\gamma(f|_{\mathcal{Y}_1})$ is finite then $B_\gamma(f|_{\mathcal{Y}_n}, f|_{\mathcal{Y}_\infty}) \rightarrow 0$.

Choosing $\gamma: t \mapsto t \ln t - t + 1$ the convergence of relative entropies follows, see [6, Theorem 3].

V. DISCUSSION

The minimization of integral functionals studied in this contribution includes a number of special situations that appeared before. The relations to literature are discussed at length in [5, Section 11]. The focus here is on sequences of generalized minimizers for entropy-like functionals, which is of novelty even for the autonomous integrands. The generalization matters, see the simple situation in Example 2.

To make Theorem 2 operational, conditions for the convergence $J_\beta^{[n]} \rightarrow J_\beta^{[\infty]} < +\infty$ are desirable. By Remark 2, it is sufficient to have $\beta \geq 0$, making H_β lower semicontinuous, and $h_\infty \in \mathcal{C}_\infty$. An alternative is to assume that μ is finite, and β bounded below. Likewise, for Theorem 3 boundedness of $J_\beta^{[n]}$ should be under control. For more general sequences of convex sets, see [2], [8].

In moment problems, \mathcal{C} is the set of functions g satisfying

$$\int_Z \varphi_j g d\mu = a_j, \quad j = 1, \dots, d,$$

where $\varphi_j: Z \rightarrow \mathbb{R}$ are given moment functions and $a_j \in \mathbb{R}$ are prescribed moments. Feasibility is resolved by the concept of conic core [5]. If $\inf_C H_\beta$ is finite then the boundedness condition from Theorem 1 is equivalent to the modified dual constraint qualification. The generalized minimizer can be explicitly described avoiding primary constraint qualification [5].

A well-understood special limiting situation is when μ is finite, H_γ is based on $\gamma \in \Gamma$ with $\gamma(t)/t \rightarrow +\infty$,

$$\mathcal{C}_n = \{g \in L_1(\mu): \int_Z \varphi_j g d\mu = a_j, j = 1, \dots, n\}$$

where $\varphi_j \in L_\infty(\mu)$, and $\mathcal{C}_\infty = \bigcap_{n \geq 1} \mathcal{C}_n$ contains a function g with $H_\gamma(g)$ finite. Then, H_γ is bounded below, \mathcal{C}_n are weakly closed, H_γ has weakly compact level sets and is strongly lower semicontinuous in $L_1(\mu)$. Therefore, $\hat{g}_{\mathcal{C}_n}$ and $\hat{g}_{\mathcal{C}_\infty}$ are minimizers [8, Theorem 1]. The convergence $J_\beta^{[n]} \rightarrow J_\beta^{[\infty]} < +\infty$ takes place. By Theorem 2, $B_\gamma(\hat{g}_{\mathcal{C}_\infty}, \hat{g}_{\mathcal{C}_n}) \rightarrow 0$. If $\gamma(t) = t \ln t$ this improves the previously known convergence in $L_1(\mu)$ [1, Corollary 3.3]. See also [3] for a sufficient condition on γ to conclude this convergence.

APPENDIX

Any β -bounded set is bounded locally in measure and so is any reversely β -bounded set provided β is asymptotically nonlinear, see Corollary 5.

The following appeared previously as [5, Lemma 2.13].

Lemma 3. *To any set $C \in \mathcal{Z}$ of finite μ -measure and K, ξ, ε positive there exists $\delta > 0$ such that for functions g and h either of $B_\beta(g, h) \leq \delta$ or $B_\beta(h, g) \leq \delta$ implies*

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

Lemma 4. *If a sequence g_n is bounded locally in measure, $h_{n,m}$ is an array of functions and $B_\beta(g_n, h_{n,m}) \rightarrow 0$ with $n, m \rightarrow \infty$ ($n \geq m$ or $m \geq n$) then $g_n - h_{n,m} \rightsquigarrow 0$ with $n, m \rightarrow \infty$ ($n \geq m$ or $m \geq n$).*

Proof: Since the sequence is bounded, if $C \in \mathcal{Z}$ has finite μ -measure and $\xi > 0$ then there exists $K > 0$ such that $\mu(C \cap \{g_n > K\}) < \xi$ for all $n \geq 1$. By Lemma 3, for $\varepsilon > 0$ there exists $\delta > 0$ such that if $B_\beta(g_n, h_{n,m}) \leq \delta$ then

$$\mu(C \cap \{|h_{n,m} - g_n| > \varepsilon\}) < \xi + \mu(C \cap \{g_n > K\}) < 2\xi.$$

The assumption on convergence implies that this holds eventually in n, m . Hence, the local convergence follows. ■

Corollary 4. *If a sequence g_n is bounded locally in measure and $B_\beta(g_n, g_m) \rightarrow 0$ with $n \geq m \rightarrow \infty$ or $m \geq n \rightarrow \infty$ then g_n converges locally in measure.*

Proof: Lemma 4 is applied with $h_{n,m} = g_m$ to conclude that $g_n - g_m \rightsquigarrow 0$ with n, m as above. In either case, g_n is Cauchy locally in measure, and the convergence follows. ■

Proof of Theorem 1: The main ingredient is the identity involving integral functionals and Bregman distances

$$\begin{aligned} & tH_\beta(g_1) + (1-t)H_\beta(g_2) - H_\beta(tg_1 + (1-t)g_2) \\ &= tB_\beta(g_1, h) + (1-t)B_\beta(g_2, h) - B_\beta(tg_1 + (1-t)g_2, h) \end{aligned}$$

where g_1, g_2 and h are nonnegative functions, $0 < t < 1$ and the H_β -terms and $B_\beta(tg_1 + (1-t)g_2, h)$ are finite.

By assumption, there exists a minimizing sequence g_n in \mathcal{C} that is bounded locally in measure and finite $H_\beta(g_n)$ converge to finite $\inf_C H_\beta$. Then, the identity implies

$$\begin{aligned} & \frac{1}{2}H_\beta(g_n) + \frac{1}{2}H_\beta(g_m) - H_\beta(h_{n,m}) \\ &= \frac{1}{2}B_\beta(g_n, h_{n,m}) + \frac{1}{2}B_\beta(g_m, h_{n,m}) \end{aligned}$$

where $h_{n,m} = \frac{1}{2}(g_n + g_m)$. The left-hand side tends to 0 with $n, m \rightarrow \infty$ as $H_\beta(h_{n,m})$ converges to $\inf_C H_\beta$ due to convexity of H_β . Then, $B_\beta(g_n, h_{n,m}) \rightarrow 0$ and $B_\beta(g_m, h_{n,m}) \rightarrow 0$. By Lemma 4, $g_n - h_{n,m} \rightsquigarrow 0$ and $g_m - h_{n,m} \rightsquigarrow 0$. Hence, $g_n - g_m \rightsquigarrow 0$ which expresses that the sequence g_n is Cauchy locally in measure. Going to a subsequence if necessary, $g_n \rightarrow \hat{g}_C$, μ -a.e., for some function \hat{g}_C .

Taking $g \in \mathcal{C}$ with finite $H_\beta(g)$ and $h_n = t_n g + (1-t_n)g_n$ with $0 < t_n < 1$, the identity implies

$$\begin{aligned} & t_n H_\beta(g) + (1-t_n)H_\beta(g_n) - H_\beta(h_n) \\ &= t_n B_\beta(g, h_n) + (1-t_n)B_\beta(g_n, h_n). \end{aligned}$$

If $t_n \downarrow 0$ then $H_\beta(h_n) \rightarrow \inf_C H_\beta$. Hence, t_n can decrease slowly enough to make $\frac{1}{t_n}[H_\beta(g_n) - H_\beta(h_n)]$ converging to zero. Dividing by t_n ,

$$H_\beta(g) - \inf_C H_\beta \geq \limsup_{n \rightarrow \infty} \left[\frac{H_\beta(h_n) - H_\beta(g_n)}{t_n} + B_\beta(g, h_n) \right].$$

Since $h_n \rightarrow \hat{g}_C$, μ -a.e., ineq. (1) follows by lower semicontinuity of the Bregman distance [5, Lemma 2.12]. ■

Proof of Theorem 2: By assumptions, Theorem 1 applied to \mathcal{C}_1 and Corollary 1, the level sets of H_β intersected with \mathcal{C}_1 are β -bounded. Then, Theorem 1 and Corollary 1 apply to all \mathcal{C}_n and the generalized minimizers $\hat{g}_{\mathcal{C}_n}$ exist. For $n \geq m$, ineq. (1) and $\mathcal{C}_n \subseteq \mathcal{C}_m$ imply

$$(4) \quad H_\beta(g) \geq J_\beta^{[m]} + B_\beta(g, \hat{g}_{\mathcal{C}_m}), \quad g \in \mathcal{C}_n.$$

If $g_k \in \mathcal{C}_n$ is a minimizing sequence in the problem (P_n) then (4) with g_k instead of g is limited in $k \rightarrow \infty$ and gives

$$(5) \quad J_\beta^{[n]} \geq J_\beta^{[m]} + \limsup_{k \rightarrow \infty} B_\beta(g_k, \hat{g}_{\mathcal{C}_m}) \geq J_\beta^{[m]} + B_\beta(\hat{g}_{\mathcal{C}_n}, \hat{g}_{\mathcal{C}_m})$$

due to $g_k \rightsquigarrow \hat{g}_{\mathcal{C}_n}$ and lower semicontinuity of the Bregman distance [5, Lemma 2.12]. Hence, $B_\beta(\hat{g}_{\mathcal{C}_n}, \hat{g}_{\mathcal{C}_m}) \rightarrow 0$ when $n, m \rightarrow \infty$ and $n \geq m$. In addition, $\hat{g}_{\mathcal{C}_n}$ is β -bounded as

$$\lim_{n \rightarrow \infty} J_\beta^{[n]} - J_\beta^{[1]} \geq \sup_n B_\beta(\hat{g}_{\mathcal{C}_n}, \hat{g}_{\mathcal{C}_1}).$$

By Corollary 4, $\hat{g}_{\mathcal{C}_n} \rightsquigarrow h_\infty$ for a function h_∞ . Limiting in (5) with $n \rightarrow \infty$ and lower semicontinuity imply

$$\lim_{m \rightarrow \infty} J_\beta^{[m]} \geq J_\beta^{[m]} + B_\beta(h_\infty, \hat{g}_{\mathcal{C}_m}), \quad m \geq 1.$$

Hence, $B_\beta(h_\infty, \hat{g}_{\mathcal{C}_m}) \rightarrow 0$ with $m \rightarrow \infty$. It follows from (4) restricted to $\mathcal{C}_\infty \subseteq \mathcal{C}_n$ that

$$H_\beta(g) \geq \lim_{m \rightarrow \infty} J_\beta^{[m]} + \limsup_{m \rightarrow \infty} B_\beta(g, \hat{g}_{\mathcal{C}_m}), \quad g \in \mathcal{C}_\infty.$$

This, $\hat{g}_{\mathcal{C}_m} \rightsquigarrow h_\infty$ and lower semicontinuity imply ineq. (2).

When the value $J_\beta^{[\infty]}$ in (P_∞) is finite then ineq. (2) implies that the level sets of H_β intersected with \mathcal{C}_∞ are β -bounded. Theorem 1 together with Corollary 1 imply existence of a function $\hat{g}_{\mathcal{C}_\infty}$ such that

$$(6) \quad H_\beta(g) \geq J_\beta^{[\infty]} + B_\beta(g, \hat{g}_{\mathcal{C}_\infty}), \quad g \in \mathcal{C}_\infty.$$

When $\lim_{n \rightarrow \infty} J_\beta^{[n]} = J_\beta^{[\infty]}$, ineqs. (6) and (2) differ only in h_∞ and $\hat{g}_{\mathcal{C}_\infty}$. Then, the two functions coincide due to the uniqueness stated in Theorem 1. ■

Remark 3. In Theorem 2, if $g_n \in \mathcal{C}_n$ and $H_\beta(g_n) - J_\beta^{[n]} \rightarrow 0$ then $g_n \rightsquigarrow h_\infty$. In fact, ineq. (4) implies

$$\lim_{n \rightarrow \infty} J_\beta^{[n]} \geq J_\beta^{[m]} + \limsup_{n \rightarrow \infty} B_\beta(g_n, \hat{g}_{\mathcal{C}_m}).$$

Then, g_n is eventually β -bounded. Since $B_\beta(g_n, \hat{g}_{\mathcal{C}_n}) \rightarrow 0$, due to ineq. (4) with $n = m$, it follows that $g_n - \hat{g}_{\mathcal{C}_n} \rightsquigarrow 0$ by Lemma 4. This and $g_n \rightsquigarrow \hat{g}_{\mathcal{C}_\infty}$ imply $g_n \rightsquigarrow h_\infty$.

The following is an extended version of [5, Lemma 9.9]†.

Lemma 5. *Let $C \in \mathcal{Z}$ have finite μ -measure and $L, \xi, \delta > 0$. There exists $K > L$ such that if $B_\beta(g, h) \leq \delta$ for functions g, h then*

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

The assertion holds also assuming instead $B_\beta(h, g) \leq \delta$ and the asymptotic nonlinearity of β .

Proof: Let $M = 2\delta/\xi$. By monotonicity, for $K > L$

$$\begin{aligned} B_\beta(g, h) &\geq \int_{\{g > K, h \leq L\}} \Delta_\beta(z, g(z), h(z)) \mu(dz) \\ &\geq \int_{\{g > K, h \leq L\}} \Delta_\beta(z, K, L) \mu(dz) \\ &\geq M \cdot \mu(C \cap \{\Delta_\beta(\cdot, K, L) \geq M, g > K, h \leq L\}). \end{aligned}$$

Then, $\mu(C \cap \{g > K\})$ is at most

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

If $B_\beta(g, h) \leq \delta$ then $\frac{1}{M} B_\beta(h, g) \leq \frac{1}{2}\xi$. By strict convexity, $\Delta_\beta(z, K, L)$ increases to $+\infty$ when $K \uparrow +\infty$. Therefore, for $K > L$ sufficiently large $\mu(C \cap \{\Delta_\beta(\cdot, K, L) < M\}) < \frac{1}{2}\xi$, and the assertion follows.

If $B_\beta(h, g) \leq \delta$ then for $K > L$

$$\begin{aligned} \delta &\geq \int_{\{g > K, h \leq L\}} \Delta_\beta(z, h(z), g(z)) \mu(dz) \\ &\geq \int_{\{g > K, h \leq L\}} \Delta_\beta(z, L, K) \mu(dz) \\ &\geq M \cdot \mu(C \cap \{\Delta_\beta(\cdot, L, K) \geq M, g > K, h \leq L\}). \end{aligned}$$

Therefore, $\mu(C \cap \{g > K\})$ is at most

$$\frac{1}{2}\xi + \mu(C \cap \{\Delta_\beta(\cdot, L, K) < M\}) + \mu(C \cap \{h > L\}).$$

To conclude $\Delta_\beta(z, L, K) \uparrow +\infty$ with $K \uparrow +\infty$, the asymptotic nonlinearity of $\beta(z, \cdot)$ suffices. The modified assertion follows as above. ■

Corollary 5. *If a set \mathcal{G} of functions is β -bounded then is bounded locally in measure. The conclusion holds also for*

†In Lemma 9.9 the arguments in $B_\beta(h, g)$ are erroneously interchanged.

sets \mathcal{H} that are reversely β -bounded assuming the asymptotic nonlinearity of β .

Proof of Theorem 3: By Theorem 1, the generalized minimizers $\hat{g}_{\mathcal{C}_n}$ exist. For $n \geq m$ it follows from ineq. (1) and $\mathcal{C}_m \subseteq \mathcal{C}_n$ that

$$(7) \quad H_\beta(g) \geq J_\beta^{[n]} + B_\beta(g, \hat{g}_{\mathcal{C}_n}), \quad g \in \mathcal{C}_m.$$

Since the left-hand side of (7) can be finite the sequence $\hat{g}_{\mathcal{C}_n}$ is reversely β -bounded. Then, it is bounded locally in measure by Corollary 5, using that β is asymptotically nonlinear.

For a minimizing sequence g_k in \mathcal{C}_m , Theorem 1 implies $g_k \rightsquigarrow \hat{g}_{\mathcal{C}_m}$, and then a subsequence converges μ -a.e. Limiting in (7) along the subsequence

$$(8) \quad J_\beta^{[m]} \geq J_\beta^{[n]} + B_\beta(\hat{g}_{\mathcal{C}_m}, \hat{g}_{\mathcal{C}_n}), \quad n \geq m,$$

on account of the lower semicontinuity of Bregman distance [5, Lemma 2.12]. Since the sequence $J_\beta^{[n]}$ is bounded and monotone $B_\beta(\hat{g}_{\mathcal{C}_n}, \hat{g}_{\mathcal{C}_m}) \rightarrow 0$, $n \geq m \rightarrow \infty$. By Lemma 4, there exists a function h^∞ such that $\hat{g}_{\mathcal{C}_n} \rightsquigarrow h^\infty$.

Limiting in (7) with $n \rightarrow \infty$, along a subsequence,

$$H_\beta(g) \geq \lim_{n \rightarrow \infty} J_\beta^{[n]} + B_\beta(g, h^\infty), \quad g \in \mathcal{C}_m.$$

analogously as above. Since $m \geq 1$ is arbitrary this inequality holds even for $g \in \mathcal{C}_\infty$. Then, $\inf_{\mathcal{C}_\infty} H_\beta$ is at least the finite limit, and $J_\beta^{[\infty]} = \lim_{n \rightarrow \infty} J_\beta^{[n]}$ by monotonicity. Hence,

$$H_\beta(g) \geq J_\beta^{[\infty]} + B_\beta(g, h^\infty), \quad g \in \mathcal{C}_\infty,$$

and the minimizing sequences in \mathcal{C}_∞ are β -bounded. Theorem 1 applies to \mathcal{C}_∞ and, by uniqueness, $h^\infty = \hat{g}_{\mathcal{C}_\infty}$.

Limiting in (8) with $n \rightarrow \infty$, along a subsequence that gives $\hat{g}_{\mathcal{C}_n} \rightarrow \hat{g}_{\mathcal{C}_\infty}$, μ -a.e.,

$$J_\beta^{[m]} \geq \lim_{n \rightarrow \infty} J_\beta^{[n]} + B_\beta(\hat{g}_{\mathcal{C}_m}, \hat{g}_{\mathcal{C}_\infty}), \quad m \geq 1.$$

Then, $B_\beta(\hat{g}_{\mathcal{C}_m}, \hat{g}_{\mathcal{C}_\infty}) \rightarrow 0$ when $m \rightarrow \infty$. ■

ACKNOWLEDGEMENT

The work on this contribution was supported by the Hungarian National Foundation for Scientific Research under Grant K 105840 and by Grant Agency of the Czech Republic under Grant 16-12010S.

REFERENCES

- [1] Borwein, J.M. and Lewis, A.S., Convergence of best entropy estimates. *SIAM J. Optimization* **1** (1991) 191–205.
- [2] Borwein, J.M. and Lewis, A.S., Strong rotundity and optimization. *SIAM J. Optimization* **4** (1994) 146–158.
- [3] Csiszár, I., Generalized projections for non-negative functions. *Acta Math. Hungar.* **68** (1995), 1–2, 161–185.
- [4] Csiszár, I. and Matúš, F., Information projections revisited. *IEEE Trans. Inform. Theory* **49** (2003) 1474–1490.
- [5] Csiszár, I. and Matúš, F., Generalized minimizers of convex integral functionals, Bregman distance, Pythagorean identities. *Kybernetika* **47** (2012) 637–689.
- [6] Harremoës, P. and Holst, K.K., Convergence of Markov chains in information divergence. *J. Theoretical Probability* **22** (2009) 186–202. Press, Princeton 1970.
- [7] Rockafellar, R.T. and Wets, R.J-B., *Variational Analysis*. Springer Verlag, Berlin, Heidelberg, New York 2004.
- [8] Teboulle, M. and Vajda, I., Convergence of best φ -entropy estimates. *IEEE Transactions on Information Theory* **39** (1993) 297–301.

PROOFS OF THE RESULTS FROM SECTION IV

This is an appendix which is not a part of the publication. It provides supplementary material to Section IV.

Proof of Lemma 1: Let ν denote the measure with the μ -density h . Assuming the γ -divergence is finite,

$$\begin{aligned} D_\gamma(g, h) &= \int_Z h \gamma\left(\frac{g}{h}\right) d\mu = \int_Z \gamma\left(\frac{g}{h}\right) d\nu = \int_Z \gamma\left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} d\nu \\ &= \int_Z \gamma\left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} h d\mu = \int_Z \gamma\left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} h_{|\mathcal{Y}, \mu} d\mu. \end{aligned}$$

Conditional Jensen's inequality with $\gamma \geq 0$ implies

$$\gamma\left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} \geq \gamma\left(\left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu}\right).$$

By assumptions, $f_{|\mathcal{Y}}$ and $h_{|\mathcal{Y}}$ are well-defined and so is $g_{|\mathcal{Y}}$ for $g \in \mathcal{D}_{\mathcal{Y}, f}$. For such g and $Y \in \mathcal{Y}$

$$\begin{aligned} \int_Y g_{|\mathcal{Y}, \mu} d\mu &= \int_Y g d\mu = \int_Y \frac{g}{h} d\nu \\ &= \int_Y \left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} d\nu = \int_Y \left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} h d\mu \\ &= \int_Y \left(\left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} h\right)_{|\mathcal{Y}, \mu} d\mu = \int_Y \left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} h_{|\mathcal{Y}, \mu} d\mu \end{aligned}$$

Using that $h_{|\mathcal{Y}, \mu} > 0$,

$$\left(\frac{g}{h}\right)_{|\mathcal{Y}, \nu} = \frac{g_{|\mathcal{Y}, \mu}}{h_{|\mathcal{Y}, \mu}}.$$

It follows by $g_{|\mathcal{Y}, \mu} = f_{|\mathcal{Y}, \mu}$ that

$$D_\gamma(g, h) \geq \int_Z \gamma\left(\frac{g_{|\mathcal{Y}, \mu}}{h_{|\mathcal{Y}, \mu}}\right) h_{|\mathcal{Y}, \mu} d\mu = D_\gamma(f_{|\mathcal{Y}, \mu}, h_{|\mathcal{Y}, \mu}).$$

Thus, the minimum is at least $D_\gamma(f_{|\mathcal{Y}}, h_{|\mathcal{Y}})$.

Denoting $f_{|\mathcal{Y}} h/h_{|\mathcal{Y}}$ by f^h , for $Y \in \mathcal{Y}$

$$\int_Y f d\mu = \int_Y f_{|\mathcal{Y}} d\mu = \int_Y (f^h)_{|\mathcal{Y}} d\mu = \int_Y f^h d\mu$$

whenever the integral on the left is finite. Therefore, f^h belongs to $\mathcal{D}_{\mathcal{Y}, f}$. As $D_\gamma(f^h, h) = D_\gamma(f_{|\mathcal{Y}}, h_{|\mathcal{Y}})$ the minimum equals $D_\gamma(f_{|\mathcal{Y}}, h_{|\mathcal{Y}})$ and f^h is a minimizer. ■

Proof of Lemma 2: By Lemma 1, $D_\gamma(f_{|\mathcal{Y}_n}, h_{|\mathcal{Y}_n})^\uparrow$ and the limit of the sequence is at most $D_\gamma(f_{|\mathcal{Y}_\infty}, h_{|\mathcal{Y}_\infty})$. A martingale theorem implies μ -a.e. convergences $f_{|\mathcal{Y}_n} \rightarrow f_{|\mathcal{Y}_\infty}$ and $h_{|\mathcal{Y}_n} \rightarrow h_{|\mathcal{Y}_\infty}$. Hence,

$$\liminf_{n \rightarrow \infty} D_\gamma(f_{|\mathcal{Y}_n}, h_{|\mathcal{Y}_n}) \geq D_\gamma(f_{|\mathcal{Y}_\infty}, h_{|\mathcal{Y}_\infty}).$$

by Fatou lemma and lower semicontinuity of $\gamma \geq 0$. ■

Proof of Corollary 2: By monotonicity of \mathcal{Y}_n , $\mathcal{D}_{\mathcal{Y}_n, f} \searrow$. A function g belongs to the intersection of $\mathcal{D}_{\mathcal{Y}_n, f}$ if and only if $\int_Y g d\mu = \int_Y f d\mu$ for $Y \in \bigcup_{n \geq 1} \mathcal{Y}_n$. The equality holds also for $Y \in \mathcal{Y}_\infty$. Hence, $\mathcal{C}_n = \mathcal{D}_{\mathcal{Y}_n, f}$ intersect to $\mathcal{C}_\infty = \mathcal{D}_{\mathcal{Y}_\infty, f}$. If $H_\gamma(f_{|\mathcal{Y}_\infty}) = D_\gamma(f_{|\mathcal{Y}_\infty}, 1)$ is finite then, by Lemma 1 with $h \equiv 1$, $\hat{g}_{\mathcal{C}_n} = f_{|\mathcal{Y}_n}$ and $\hat{g}_{\mathcal{C}_\infty} = f_{|\mathcal{Y}_\infty}$. Lemma 2 and Theorem 2 with $h_\infty = \hat{g}_{\mathcal{C}_\infty}$ imply the convergence. ■

Proof of Corollary 3: The integrand β is asymptotically nonlinear and $\mathcal{C}_n = \mathcal{D}_{\mathcal{Y}_n, f}^\uparrow$ and have the union \mathcal{C}_∞ contained in $\mathcal{D}_{\mathcal{Y}_\infty, f}^\ddagger$. Lemma 1 implies that $J_\beta^{[n]} = D_\gamma(f_{|\mathcal{Y}_n}, h_{|\mathcal{Y}_n}) \searrow$. Then, the sequence is finite and has a nonnegative limit because $\gamma \geq 0$. It follows by Theorem 3 that $B_\beta(\hat{g}_{\mathcal{C}_n}, \hat{g}_{\mathcal{C}_\infty}) \rightarrow 0$ where $\hat{g}_{\mathcal{C}_n} = f_{|\mathcal{Y}_n} h/h_{|\mathcal{Y}_n}$. Then, by Fatou lemma

$$\begin{aligned} 0 &\geq \int_Z \liminf_{n \rightarrow \infty} \left[h \gamma\left(\frac{f_{|\mathcal{Y}_n}}{h_{|\mathcal{Y}_n}}\right) - h \gamma\left(\frac{\hat{g}_{\mathcal{C}_\infty}}{h}\right) \right. \\ &\quad \left. - \gamma'\left(\frac{\hat{g}_{\mathcal{C}_\infty}}{h}\right) \left[\frac{f_{|\mathcal{Y}_n} h}{h_{|\mathcal{Y}_n}} - \hat{g}_{\mathcal{C}_\infty}\right] \right] d\mu \geq 0. \end{aligned}$$

In turn, $B_\beta(f_{|\mathcal{Y}_\infty} h/h_{|\mathcal{Y}_\infty}, \hat{g}_{\mathcal{C}_\infty}) = 0$, using that $f_{|\mathcal{Y}_n} \rightarrow f_{|\mathcal{Y}_\infty}$ and $h_{|\mathcal{Y}_n} \rightarrow h_{|\mathcal{Y}_\infty}$, μ -a.e. Hence, $\hat{g}_{\mathcal{C}_\infty} = f_{|\mathcal{Y}_\infty} h/h_{|\mathcal{Y}_\infty}$ and

$$B_\beta(\hat{g}_{\mathcal{C}_n}, \hat{g}_{\mathcal{C}_\infty}) = D_\gamma(f_{|\mathcal{Y}_n}, h_{|\mathcal{Y}_n}) - D_\gamma(f_{|\mathcal{Y}_\infty}, h_{|\mathcal{Y}_\infty}) \rightarrow 0.$$

This rewrites to $B_\gamma(f_{|\mathcal{Y}_n}, f_{|\mathcal{Y}_\infty}) = H_\gamma(f_{|\mathcal{Y}_n}) - H_\gamma(f_{|\mathcal{Y}_\infty}) \rightarrow 0$ when $h \equiv 1$. ■

‡In Example 2, the inclusion was strict.