Universität des Saarlandes

Fachbereich 9 – Mathematik

Mathematischer Preprint

On generalized Csiszár–Kullback inequalities

Anton Arnold, Peter Markowich,
Giuseppe Toscani and Andreas Unterreiter

Preprint No. 10

Saarbrücken 2000

Universität des Saarlandes

Fachbereich 9 – Mathematik

On generalized Csiszár–Kullback inequalities

Anton Arnold

Fachbereich Mathematik
Universität des Saarlandes
D-66123 Saarbrücken
Germany
arnold@num.uni-sb.de

Peter Markowich

Institut für Mathematik
Universität Wien
A-1090 Wien
Austria
peter.markowich@univie.ac.at

Giuseppe Toscani

Dipartimento di Matematica
Università di Pavia
I-27100 Pavia
Italy
toscani@dimat.unipv.it

Andreas Unterreiter

Fachbereich Mathematik
Universität Kaiserslautern
D-67653 Kaiserslautern
Germany
unterreiter@mathematik.uni-kl.de

submitted: February 15, 2000

Preprint No. 10

Saarbrücken 2000

Edited by
FB 9 – Mathematik
Im Stadtwald
D–66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Abstract

The classical Csiszár–Kullback inequality bounds the L^1–distance of two probability densities in terms of their relative (convex) entropies. Here we generalize such inequalities to not necessarily normalized and possibly non-positive L^1 functions. Also, our generalized Csiszár–Kullback inequalities are in many important cases sharper than the classical ones (in terms of the functional dependence of the L^1 bound on the relative entropy). Moreover our construction of these bounds is rather elementary.

Acknowledgements

This research was partially supported by the grants ERBFMRXCT970157 (TMR-Network) from the EU, the bilateral DAAD-Vigoni Program, the DFG under Grant-No. MA 1662/1-3, and the NSF (DMS-9500852).

AMS 1991 Subject Classification: 28A33, 28D20, 52A10, 52A40, 94A17, 46E30

1 Introduction

In this paper we shall be concerned with proving Csiszár–Kullback inequalities, which are estimates of the L^1-distance of two functions in terms of their relative entropy. These inequalities have at least 30 years of history in probability and information theory [Csi63, Csi66, Csi67, Kul59, KuLei51, Per57, Per65, BaNi64]. Even if the main fields of application of these inequalities are traditionally probability and information theory, there has been new interest in the last years in the context of PDE's. This recent interest in Csiszár–Kullback inequalities stems from their importance in the study of thermo-dynamical evolution equations (see [AMTU99] for an extensive discussion of the pertinent literature). Here, the dissipative nature of the system is often reflected in the time decay to zero of the relative entropy, which is a convex functional relating the state of the system to its thermo-dynamical equilibrium state.

It is well known that the relative entropy of functions with equal mass does not define a topology on L^1 ([Csi67]). Nevertheless it provides an upper bound for the L^1-distance of those functions via the classical Csiszár–Kullback inequality ([Csi63]). For determining the optimal L^1–time decay rate towards equilibrium (using the entropy method, see [AMTU99]) one needs optimal Csiszár–Kullback inequalities.

In [Csi67] (Theorem 3.1, page 309, and Section 4, page 314) the following result (*classical Csiszár–Kullback inequality*) was obtained:

Theorem 1.1. *Let $\psi : (0, \infty) \to I\!R$ be bounded below, convex on $(0, \infty)$, strictly convex at 1 with $\psi(1) = 0$. Then there exists a function $\mathsf{W}_\psi : I\!R \to [0, \infty)$ such that*

1. W_ψ *is increasing.*
2. $\mathsf{W}_\psi(0) = 0$.
3. W_ψ *is continuous at* 0.
4. *For all non-negative $u \in L^1(\Omega, \mathcal{S}, \mu)$ (where $(\Omega, \mathcal{S}, \mu)$ is a probability space) with $\int_\Omega u \, d\mu = 1$ the* Csiszár-Kullback inequality *holds:*
$$\|u - 1\|_{L^1(d\mu)} \leq \mathsf{W}_\psi(e_\psi(u)), \tag{1.1}$$
where $e_\psi(u) := \int_\Omega \psi(u) \, d\mu$ is the entropy of u (relative to the state 1) generated by the function ψ.

For the generating function $\psi(\sigma) = \sigma \ln \sigma - \sigma + 1$, e.g., it is known [Csi67] that
$$\mathsf{W}_\psi(c) \leq \sqrt{2c}. \tag{1.2}$$

By a rescaling argument the inequality (1.1) can be extended to functions $u \in L^1(d\mu)$, $u \geq 0$, which are not necessarily probability densities: Excluding the case $u \equiv 0$, we replace u by $u/(\int_\Omega u \, d\mu)$, apply (1.1) and multiply by $\int_\Omega u \, d\mu$ to obtain
$$\left\| u - \int_\Omega u \, d\mu \right\|_{L^1(d\mu)} \leq \left(\int_\Omega u \, d\mu \right) \mathsf{W}_\phi \left(e_\psi(u) \right),$$
where $\phi(\sigma) = \psi(\int_\Omega u d\mu \, \sigma)$.

The analysis of this paper aims at generalizing the above estimate. In particular we shall establish a generalized Csiszár-Kullback inequality of the form
$$\left\| u - \int_\Omega u \, d\mu \right\|_{L^1(d\mu)} \leq \mathsf{U} \left(e_\psi(u), \int_\Omega u \, d\mu \right), \tag{1.3}$$

where the function U only depends on ψ.

This main result deepens the insight into Csiszár-Kullback inequalities in four ways. Firstly, inequality (1.3) is valid for *all* functions u with $e_\psi(u) < \infty$, i.e. there is no normalizing condition $\int u d\mu = 1$ assumed and u is possibly non-positive. Secondly, we shall prove an optimality result for the function U and give an explicit construction of U for certain convex functions ψ. Thirdly, we shall prove that U is *strictly* increasing. Fourthly, the construction of U is rather straightforward and requires no measure-theoretic background. Since it is somewhat lengthy we defer it to the Appendix.

2 The generalized Inequality

Throughout this Section we shall make use of the following assumptions and notations:

B.1 $(\Omega, \mathcal{S}, \mu)$ is a measure space and μ is a probability measure on $\mathcal{S}$, i.e. $\mu(\Omega) = 1, \mu \geq 0$.

B.2 $\psi : J \to I\!\!R$ is strictly convex, continuous and non-negative. The (possibly unbounded) interval J has a non-empty interior J°. If $\alpha := \inf J \notin J$, then $\lim_{\sigma\to\alpha} \psi(\sigma) = \infty$. If $\beta := \sup J \notin J$, then $\lim_{\sigma\to\beta} \psi(\sigma) = \infty$.

For functions $u \in L^1(d\mu) := L^1(\Omega; d\mu)$ we define the *entropy* of u relative to $[u] := \int_\Omega u \, d\mu$ as

$$e_\psi(u) := \begin{cases} \displaystyle\int_\Omega \psi(u) \, d\mu - \psi([u]) & \text{if } u(x) \in J \ \mu\text{-a.e.,} \\ \infty & \text{else.} \end{cases} \tag{2.1}$$

Remark 2.1. *Suppose $u \in L^1(d\mu)$ with $u(x) \in J \ \mu - a.e.$ Since ψ is continuous on J the mapping $x \mapsto (\psi \circ u)(x)$ is $\mathcal{S}$-measurable whenever u is $\mathcal{S}$-measurable. Furthermore we have due to Jensen's inequality*

$$\int_\Omega \psi(u) \, d\mu \geq \psi\left(\int_\Omega u \, d\mu\right).$$

Thus the right-hand side of (2.1) has a well-defined value in $[0, +\infty]$.

Remark 2.2. *a) The required continuity of* ψ *as well as the assumed behavior of* $\psi(\sigma)$ *as* σ *tends to the boundary of* J *deserve a few more comments. Let* $\hat{\psi} : J \to \mathbb{R}$ *be convex on the interval* J*, which has a non-empty interior. Then* $\hat{\psi}$ *is continuous on* J° *and possible points of discontinuity are contained in* $\partial J \cap J$*. If* $\alpha \in \mathbb{R}$ *and if* $\psi(\sigma)$ *remains bounded as* σ *approaches* α*, then we can (re-)define* $\hat{\psi}(\alpha) := \lim_{\sigma \to \alpha+} \hat{\psi}(\sigma)$*. We obtain a convex function, less or equal to the original one (i.e. the entropy decreases), which is continuous at* α*. The same procedure applies in cases where* β *is finite and* $\psi(\sigma)$ *remains bounded as* σ *approaches* β*. The cases* $\alpha = -\infty$ *or* $\beta = \infty$ *are discussed in b).*
b) Not every strictly convex, continuous function $\hat{\psi} : J \to \mathbb{R}$ *on the (possibly unbounded) interval* J *with non-empty interior satisfies* **B**.*2. However, we can normalize the generating function* $\hat{\psi}$ *as follows. Due to the strict convexity of* $\hat{\psi}$ *there exist for each* $\sigma_0 \in J^\circ$ *a constant* $b \in \mathbb{R}$ *such that the function*

$$\psi(\sigma) := \hat{\psi}(\sigma) - \hat{\psi}(\sigma_0) - b(\sigma - \sigma_0)$$

is non-negative on J *and* $\psi(\sigma_0) = 0$*. Furthermore we have for all* $u \in L^1(d\mu)$ *the equality* $e_\psi(u) = e_{\hat{\psi}}(u)$ *which shows that* ψ *and* $\hat{\psi}$ *generate the same entropy.*

Remark 2.3. *Let us assume that* $\Omega \subseteq \mathbb{R}^n$*,* $n \in \mathbb{N}$ *is a Borel set,* $\mathcal{S}$ *is the Borel sigma algebra on* Ω*, and* μ *is a probability measure on* $\mathcal{S}$*, which is absolutely continuous with respect to the Lebesgue measure. In this case the Radon-Nikodym derivative of* μ *with respect to the Lebesgue measure exists:*

$$g(x) := \frac{d\mu(x)}{dx}.$$

To keep things simple let us assume that $g(x) > 0$ *for almost all* $x \in \Omega$*. Now consider a function* $f \in L^1(\Omega, dx)$ *which satisfies*

$$u(x) := \frac{f(x)}{g(x)} \in J \text{ a.e. on } \Omega.$$

Then $e_\psi(u)$ *is the relative entropy of* f *w.r.t.* g *(see [AMTU99]):*

$$e_\psi(u) = e_\psi\left(\frac{f}{g}\right) = \int_\Omega \psi(f/g) g \, dx - \psi\left(\int_\Omega f \, dx\right).$$

$\int_\Omega f \, dx$ *needs not to be normalized here. In the case of* $\psi(\sigma) = \sigma \ln \sigma - \sigma + 1$*, and* $f(x) \geq 0$ *with* $\int_\Omega f \, dx = 1$ *the inequalities (1.1) and (1.2) combine to*

$$\|f - g\|_{L^1(\Omega)} \leq \sqrt{2 \int_\Omega f \ln \frac{f}{g} \, dx}.$$

In the sequel we shall need the following ψ-dependent constants. For any $\sigma \in J^\circ$ we define:

$$Q_+\psi(\sigma) := \lim_{z\to\infty} \frac{\psi(\sigma+z)-\psi(\sigma)}{z} \in (0,\infty], \quad \text{if } \beta=\infty,$$

$$Q_-\psi(\sigma) := \lim_{z\to\infty} \frac{\psi(\sigma-z)-\psi(\sigma)}{z} \in (0,\infty], \quad \text{if } \alpha=-\infty,$$

$$c_0(\sigma) := \begin{cases} \psi(\alpha)-\psi(\sigma)+(\sigma-\alpha)Q_+\psi(\sigma) & \text{if } \alpha, \psi(\alpha) \in I\!R, \beta=\infty, \\ \psi(\beta)-\psi(\sigma)+(\beta-\sigma)Q_-\psi(\sigma) & \text{if } \beta, \psi(\beta) \in I\!R, \alpha=-\infty, \\ \dfrac{(\sigma-\alpha)\psi(\beta)+(\beta-\sigma)\psi(\alpha)}{\beta-\alpha} - \psi(\sigma) & \text{if } \alpha, \beta, \psi(\alpha), \psi(\beta) \in I\!R, \\ \infty & \text{else,} \end{cases}$$

(with the convention $\delta\,\infty = \infty$ for all $\delta > 0$), and

$$\mathsf{U}_{sup}(\sigma) := \begin{cases} 2(\sigma-\alpha) & \text{if } \alpha \in I\!R, \beta=\infty \\ 2(\beta-\sigma) & \text{if } \alpha=-\infty, \beta \in I\!R \\ \dfrac{2(\beta-\sigma)(\sigma-\alpha)}{\beta-\alpha} & \text{if } \alpha, \beta \in I\!R \\ \infty & \text{if } \alpha=-\infty, \beta=\infty \end{cases}$$

We now collect the essential properties of U in the following theorem; its elementary proof is given in the Appendix.

Theorem 2.4. *Under the assumption* **B**.*2 there exists a function* $\mathsf{U} : [0,\infty)\times J \to I\!R$ *such that*

P1. U *is continuous.*

P2. For all $\sigma \in J$*:* $\mathsf{U}(0,\sigma) = 0$.

P3. For all $\sigma \in J$*: The mapping* $c \mapsto \mathsf{U}(c,\sigma)$, $c \in [0,\infty)$, *is increasing.*

P4. For all $\sigma \in \partial J \cap J$ *and all* $c \in [0,\infty)$ *the identity* $\mathsf{U}(c,\sigma) = 0$ *holds.*

P5. For all $\sigma \in J^\circ$*: The mapping* $c \mapsto \mathsf{U}(c,\sigma)$, $c \in [0,\infty)$, *is strictly increasing on* $[0, c_0(\sigma))$.

P6. For all $\sigma \in J^\circ$*:* $\lim_{c\to\infty} \mathsf{U}(c,\sigma) = \mathsf{U}_{sup}(\sigma)$.

P7. For all $\sigma \in J^\circ$ and all $c \in [c_0(\sigma), \infty)$: $\mathsf{U}(c,\sigma) = \mathsf{U}_{sup}(\sigma)$.

We note that $\partial J \cap J$ (see P4.) may be empty. In P7. we use the convention $[\infty, \infty) = \emptyset$. The main result of this Section is:

Theorem 2.5. *Assume* **B**.*1* - **B**.*2. Then we have for all $u \in L^1(d\mu)$ with $e_\psi(u) < \infty$:*

$$\left\| u - \int_\Omega u \, d\mu \right\|_{L^1(d\mu)} \leq \mathsf{U}\left(e_\psi(u), \int_\Omega u \, d\mu \right). \tag{2.2}$$

PROOF: We introduce some abbreviations. Recall that $[u] := \int_\Omega u \, d\mu$. Since $e_\psi(u) < \infty$ we have $u(x) \in J$ μ-a.e. Hence $[u] \in J$. We set

$$\Omega_0^+ := \{x \in \Omega : u(x) \geq [u]\}, \quad \Omega^- := \{x \in \Omega : u(x) < [u]\},$$

and define

$$w := u - [u] \quad , \quad a := \int_{\Omega_0^+} w \, d\mu \quad , \quad \vartheta := \mu(\Omega_0^+).$$

We note $\vartheta \in (0,1]$ and $\mu(\Omega^-) = 1 - \vartheta$. Since $\int_\Omega w \, d\mu = 0$, we have $\int_{\Omega^-} w \, d\mu = -a$ and

$$\left\| u - \int_\Omega u \, d\mu \right\|_{L^1(d\mu)} = \int_\Omega |w| \, d\mu = \int_{\Omega_0^+} w \, d\mu - \int_{\Omega^-} w \, d\mu = 2a.$$

(This factor 2 also appears in the definition of U.) We proceed by a case-distinction. The first case is probably the most interesting one.

<u>Case 1:</u> $\vartheta \neq 1$. If $[u] \in \partial J \cap J$, then $u = [u]$ and therefore $\vartheta = 1$. This is a contradiction since we have assumed $\vartheta < 1$. Hence $[u] \in J^\circ$. Furthermore we obtain from the definition of $a, [u]$ and ϑ:

$$a \leq \min\{\vartheta(\beta - [u]), (1-\vartheta)([u] - \alpha)\},$$

and therefore

$$a \leq \frac{\mathsf{U}_{sup}([u])}{2} = \frac{(\beta - [u])([u] - \alpha)}{2} \quad , \quad \vartheta \in \left[\frac{a}{\beta - [u]}, 1 - \frac{a}{[u] - \alpha} \right] \cap (0,1).$$

Now we apply Jensen's inequality:

$$
\begin{aligned}
\infty &> e_\psi(u) \\
&= \int_\Omega [\psi(u) - \psi([u])] \; d\mu \\
&= \int_\Omega [\psi(w + [u]) - \psi([u])] \; d\mu \\
&= \int_{\Omega_0^+} [\psi(w + [u]) - \psi([u])] \; d\mu + \int_{\Omega^-} [\psi(w + [u]) - \psi([u])] \; d\mu \\
&\geq \vartheta \left(\psi\left([u] + \frac{a}{\vartheta}\right) - \psi([u]) \right) + (1-\vartheta) \left(\psi\left([u] - \frac{a}{1-\vartheta}\right) - \psi([u]) \right) \\
&= a \left(\frac{\psi\left([u] + \frac{a}{\vartheta}\right) - \psi([u])}{\frac{a}{\vartheta}} + \frac{\psi\left([u] - \frac{a}{1-\vartheta}\right) - \psi([u])}{\frac{a}{1-\vartheta}} \right) \\
&=: \hat{\phi}([u], a, \vartheta) \\
&\geq \inf\{\hat{\phi}([u], a, \vartheta) : \vartheta \in J([u], a) \cap (0,1)\},
\end{aligned}
$$

where

$$
J([u], a) := \begin{cases}
\left[\frac{a}{\beta - [u]}, 1 - \frac{a}{[u] - \alpha}\right] & \text{if } \alpha, \beta \in J \\
\left[\frac{a}{\beta - [u]}, 1 - \frac{a}{[u] - \alpha}\right) & \text{if } \alpha \notin J, \beta \in J \\
\left(\frac{a}{\beta - [u]}, 1 - \frac{a}{[u] - \alpha}\right] & \text{if } \alpha \in J, \beta \notin J \\
\left(\frac{a}{\beta - [u]}, 1 - \frac{a}{[u] - \alpha}\right) & \text{if } \alpha, \beta \notin J
\end{cases}
$$

We distinguish three cases.

<u>Case 1a:</u> $0 < a < \mathsf{U}_{sup}([u])/2$. In this case the interval $J([u], a) \cap (0,1)$ has a non-empty interior. We have

$$
\begin{aligned}
e_\psi(u) &\geq \inf\left\{\hat{\phi}([u], a, \vartheta) : \vartheta \in J([u], a) \cap (0,1)\right\} \\
&= \inf\left\{\hat{\phi}([u], a, \vartheta) : \vartheta \in \left(\frac{a}{\beta - [u]}, 1 - \frac{a}{[u] - \alpha}\right)\right\},
\end{aligned}
$$

and therefore with the notations of the proof of Theorem 2.4

$$I(e_\psi(u),[u]) \geq a = \frac{\|u-[u]\|_{L^1(d\mu)}}{2},$$

which gives due to the definition of U

$$\mathsf{U}(e_\psi(u),[u]) \geq \|u-[u]\|_{L^1(d\mu)}.$$

Case 1b: $a=0$. This case need not to be considered, because $a=0$ implies $u=[u]$ and therefore $\vartheta=1$.

Case 1c: $a=\mathsf{U}_{sup}([u])/2$. Since $a \in I\!R$ we have $\alpha \in I\!R$ or $\beta \in I\!R$. If $\alpha \in I\!R$ and $\beta=\infty$ we will obtain $\vartheta \in [0,0)=\emptyset$, and if $\alpha=-\infty$ and $\beta \in I\!R$ we will obtain $\vartheta \in [1,1)=\emptyset$. Hence, the only remaining case is $\alpha,\beta \in I\!R$. We obtain $\vartheta=([u]-\alpha)/(\beta-\alpha)$ which gives

$$\begin{aligned} e_\psi(u) &\geq \frac{([u]-\alpha)\psi(\beta)+(\beta-[u])\psi(\alpha)}{\beta-\alpha}-\psi([u]) \\ &\geq \frac{([u]-\alpha)\psi(\beta-)+(\beta-[u])\psi(\alpha+)}{\beta-\alpha}-\psi([u]) \\ &= c_0([u]), \end{aligned}$$

and therefore by the definition of U and $\mathsf{U}_{sup}([u])$

$$\mathsf{U}(e_\psi([u]),[u])=\mathsf{U}(c_0([u]),[u])=\mathsf{U}_{sup}([u])=2a=\|u-[u]\|_{L^1(d\mu)}.$$

Case 2: $\vartheta=1$. In this case $u=[u]$. Hence, $e_\psi(u)=0=\|u-[u]\|_{L^1(d\mu)}$ and therefore

$$\mathsf{U}(e_\psi(u),[u])=\mathsf{U}(0,[u])=0=\|u-[u]\|_{L^1(d\mu)}.$$

□

We shall discuss inequality (1.3) now.

2.1 Optimality

A natural question in connection with (1.3) is the following. Is it possible to improve (1.3), i.e. is there a function $\mathsf{U}^*:\{(c,\sigma):\sigma\in J^\circ, c\in[0,c_0(\sigma))\}$ (for reasons explained in Remark 2.7 below we exclude the cases $c\in[c_0(\sigma),\infty)$ and $\sigma\in\partial J\cap J$ here) such that

1. for all $u \in L^1(d\mu)$ with $e_\psi(u) < \infty$: $\|u - [u]\|_{L^1(d\mu)} \le \mathsf{U}^*(e_\psi(u), [u])$,

2. there exists a $u^* \in L^1(d\mu)$ with $e_\psi(u^*) < \infty$ and $\mathsf{U}^*(e_\psi(u^*), [u^*]) < \mathsf{U}(e_\psi(u^*), [u^*])$?

Under the assumption **B**.3 (see below) the answer to this question is *no.*

Theorem 2.6. *Assume* **B**.*1,* **B**.*2, and furthermore*

B.*3 For all $\vartheta \in (0,1)$ there is $S_\vartheta \in \mathcal{S}$ with $\mu(S_\vartheta) = \vartheta$.*

Then for all $\sigma \in J^\circ$ and $c \in [0, c_0(\sigma))$ there exists a sequence $(u_m)_{m\in I\!N}$ in $L^1(d\mu)$ such that

1. *For all $m \in I\!N$: $[u_m] = \sigma$.*
2. *For all $m \in I\!N$: $e_\psi(u_m) = c$.*
3. $\lim\limits_{m\to\infty} \|u_m - [u_m]\|_{L^1(d\mu)} = \mathsf{U}(c, \sigma)$.

Proof: Let us consider the case $0 < c < c_0(\sigma)$ first: There exists a (uniquely determined) $a \in (0, \mathsf{U}_{sup}/2)$ such that

$$c = \inf_{\vartheta \in (a/(\beta-\sigma), 1-(a/(\sigma-\alpha)))} \hat{\phi}(\sigma, a, \vartheta),$$

where $\hat{\phi}$ is defined as in the proof of Theorem 2.5. Clearly, $2a = \mathsf{U}(c, \sigma)$. We note that the mapping $\vartheta \mapsto \hat{\phi}(\sigma, a, \vartheta)$ is not constant and the mapping $a \mapsto \inf_{\vartheta \in (a/(\beta-\sigma), 1-(a/(\sigma-\alpha)))} \hat{\phi}(\sigma, a, \vartheta)$ is increasing. Furthermore, $\hat{\phi}$ is continuous. We can therefore choose a sequence $(\vartheta_m)_{m\in I\!N}$ in $(0,1)$ and a sequence $(a_m)_{m\in I\!N}$ in $(0, a)$ such that $\hat{\psi}(\sigma, a_m, \vartheta_m) = c$ and $\lim_{m\to\infty} a_m = a$. Now we choose a sequence $(S_m)_{m\in I\!N}$ in $\mathcal{S}$ such that $\mu(S_m) = \vartheta_m$ for all $m \in I\!N$. We set for all $m \in I\!N$

$$\begin{array}{rcl} u_m & : & \Omega \rightarrow I\!R \\ & & x \mapsto \begin{cases} (a_m/\vartheta_m) + \sigma & \text{if } x \in S_m \\ -(a_m/(1-\vartheta_m)) + \sigma & \text{if } x \in \Omega \setminus \sigma \end{cases} \end{array}$$

We easily verify that $[u_m] = \sigma$ and $e_\psi(u_m) = c$. Furthermore,

$$\lim_{m\to\infty} \|u_m - [u_m]\|_{L^1(d\mu)} = 2a = \mathsf{U}(c, \sigma).$$

This finishes the proof for the case $0 < c < c_0(\sigma)$. If $c = 0$ we choose $u_m = \sigma$, which gives $[u_m] = \sigma$, $e_\psi(u_m) = 0 = c$ and

$$\|u_m - [u_m]\|_{L^1(d\mu)} = 0 = \mathsf{U}(0, \sigma) = \mathsf{U}(c, \sigma).$$

□

Remark 2.7. *In Theorem 2.6 the cases* $\sigma \in \partial J \cap J$ *and* $\sigma \in J^\circ$ *with* $c \in [c_0(\sigma), \infty)$, $c_0(\sigma) < \infty$ *are not included. Indeed, if* $\sigma \in \partial J \cap J$, *then* $u = [u]$. *Hence,* $e_\psi(u) = 0$ *and in this case the estimate (1.3) is optimal, too:*

$$\|u - [u]\|_{L^1(d\mu)} = 0 = \mathsf{U}(0, \sigma) = \mathsf{U}(e_\psi(u), [u]).$$

If $\sigma \in J^\circ$ *and* $e_\psi(u) \geq c_0(\sigma)$ *we will trivially obtain for all* $u \in L^1(d\mu)$ *with* $u(x) \in J$ μ*-a.e. and* $[u] = \sigma$

$$\|u - [u]\|_{L^1(d\mu)} \leq \mathsf{U}_{sup}([u]) = \mathsf{U}(c_0([u]), [u]) = \mathsf{U}(c, [u]).$$

We note that this inequality is strict in cases where $\alpha = -\infty$ *or* $\beta = \infty$.

2.2 An extension of the inequality

A closer screening of the proof of Theorems 2.4 and 2.5 shows that the core of all estimates is the fact that for all $\sigma \in J^\circ$ and all $a \in (0, \mathsf{U}_{sup}/2)$ the *strict* estimate

$$\inf_{\vartheta \in (a/(\beta-\sigma), 1-(a/(\sigma-\alpha)))} \hat{\phi}(\sigma, a, \vartheta) > 0$$

holds. For fixed $\sigma_\circ \in J^\circ$ this is the case iff ψ is *strictly convex at* $\sigma_\circ$, i.e. iff there exists a constant $b \in \mathbb{R}$ such that the *strict* estimate

$$\psi(\sigma) > \psi(\sigma_\circ) + b(\sigma - \sigma_\circ)$$

holds for all $\sigma \in (\alpha, \beta)$, $\sigma \neq \sigma_\circ$. We can therefore extend Theorems 2.4 and 2.5 in the following way:

Theorem 2.8. *Assume* **B**.*1 and*

B.*2'* $\tilde{\psi} : J \to \mathbb{R}$ *is convex, continuous and non-negative. The (possibly unbounded) interval* J *has a non-empty interior* J°. *If* $\alpha := \inf J \notin J$, *then* $\lim_{\sigma \to \alpha} \tilde{\psi}(\sigma) = \infty$. *If* $\beta := \sup J \notin J$, *then* $\lim_{\sigma \to \beta} \tilde{\psi}(\sigma) = \infty$. *Furthermore assume that* $\tilde{\psi}$ *is strictly convex at* $\sigma_\circ \in J^\circ$.

Then:

a. *There exists a function* $\mathsf{U}_{\sigma_\circ} : [0, \infty) \to \mathbb{R}$ *such that*

   P1. $\mathsf{U}_{\sigma_\circ}$ *is continuous.*

P2. $\mathsf{U}_{\sigma_\circ}(0) = 0$.

P3. $\mathsf{U}_{\sigma_\circ}$ *is increasing.*

P4. $\mathsf{U}_{\sigma_\circ}$ *is strictly increasing on* $[0, c_0(\sigma_\circ))$.

P5. $\lim_{c\to\infty} \mathsf{U}_{\sigma_\circ}(c) = \mathsf{U}_{sup}(\sigma_\circ)$.

P6. For all $c \in [c_0(\sigma_\circ), \infty)$*:* $\mathsf{U}_{\sigma_\circ}(c) = \mathsf{U}_{sup}(\sigma_\circ)$.

b. Let $u \in L^1(d\mu)$*. If* $e_\psi(u) < \infty$ *and* $[u] = \sigma_\circ$*, then*

$$\left\| u - \int_\Omega u \, d\mu \right\|_{L^1(d\mu)} \leq \mathsf{U}_{\sigma_\circ}(e_{\tilde{\psi}}(u)).$$

We note that an optimality result as Theorem 2.6 also holds for $\mathsf{U}_{\sigma_\circ}$.

As an example consider the function $\tilde{\psi}(\sigma) = |\sigma|$ which is strictly convex at $\sigma_\circ = 0$. In this case we obtain $e_{|.|}(u) = \int |u| d\mu$ and $\mathsf{U}_0(c) = c$ which trivially gives for all $u \in L^1(d\mu)$ with $[u] = 0$

$$\|u\|_{L^1(d\mu)} \leq \mathsf{U}_0(e_{|.|}(u)) = \mathsf{U}_0(\int_\Omega |u| d\mu) = \int_\Omega |u| d\mu.$$

2.3 An example

The calculation of U involves a minimizing procedure. If ψ is differentiable this will lead to the problem of finding roots of transcendental equations. Hence there is no hope for an explicit presentation of U in general.

In some cases, however, the function U can be found explicitly, at least for special values of $[u] = \int_\Omega u \, d\mu$. If $\psi(\sigma) = |\sigma - 1|^{2p}$, $p \in [1, \infty)$, on $J = \mathbb{R}$, we have for all admissible u with $[u] = 1$:

$$\|u - 1\|_{L^1(d\mu)} \leq \left(\int_\Omega \left(|u|^{2p} - 1 \right) \, d\mu \right)^{1/2p}.$$

In the situation described in Remark 2.3 this becomes with the additional requirement $\int_\Omega f \, dx = 1$:

$$\|f - g\|_{L^1(dx)} \leq \left(\int_\Omega |f - g|^{2p} \, g^{1-2p} \, dx \right)^{1/2p}.$$

For the special case of $p = 1$ and $[u] = 1$ the above two inequalities read:

$$\|u - 1\|_{L^1(d\mu)} \leq \sqrt{\int_\Omega (u^2 - 1) \, d\mu},$$

and respectively,

$$\|f-g\|_{L^1(dx)} \le \sqrt{\int_\Omega (f-g)^2\, g^{-1}\, dx}.$$

2.4 Asymptotic behavior of U at 0

As mentioned before, U can in general *not* be expressed in terms of elementary functions. For 'small' values of $e_\psi(u)$, however, one can obtain an asymptotic expansion of U. Let us recall the following result of [Csi67]: If $\psi \in C^2(J)$, $1 \in J^\circ$ and $\psi''(1) > 0$, then the Cszisár-Kullback inequality holds:

$$\begin{cases} \exists K_{1,2} > 0 : \forall u \in L^1(d\mu), u \ge 0, [u] = 1 : \text{ If } e_\psi(u) \le K_2 \text{ then} \\ \qquad \|u-1\|_{L^1(d\mu)} \le K_1 \sqrt{e_\psi(u)}. \end{cases} \tag{2.3}$$

In [Csi67] the investigations are specialized to the case $\psi(\sigma) = \sigma\ \ln\sigma$. In this case the constants can be specified:

For all $u \in L^1(d\mu)$ with $u \ge 0$, $\int_\Omega u\ d\mu = 1$:

$$\|u-1\|_{L^1(d\mu)} \le \sqrt{2\int_\Omega u\ \ln u\ d\mu}.$$

We note that $\psi''(1) > 0$ was essential to obtain (2.3). Otherwise one can, in general, *not* obtain an estimate of the form

$$\|u-1\|_{L^1(d\mu)} \le K\,(e_\psi(u))^\eta\,, \quad K, \eta \in (0,\infty),$$

even for small values of $e_\psi(u)$ (to keep things simple we consider only the case $\int_\Omega u\ d\mu = 1$).

Example 2.9. *Let $\Omega = [0,1]$, $d\mu = dx$. Let*

$$\psi : I\!R \to I\!R \quad , \quad \sigma \mapsto \int_1^\sigma \phi(s)\ ds$$

with

$$\begin{aligned} \phi \ :\ & I\!R \to I\!R \\ \sigma \ \mapsto\ & \begin{cases} -(\sigma-1) & , \ \sigma \in (-\infty, 0], \\ -e^{-\sigma/(1-\sigma)} & , \ \sigma \in (0,1), \\ 0 & , \ \sigma = 1, \\ e^{-\sigma/(\sigma-1)} & , \ \sigma \in (1,2), \\ e^{-2}(\sigma-1) & , \ \sigma \in [2,+\infty). \end{cases} \end{aligned}$$

Certainly, ψ is bounded below, strictly convex at 1 and belongs to $C^2(\mathbb{R})$ with $\psi(1) = \psi'(1) = \psi''(1) = 0$. For $k \in \mathbb{N}$ let

$$\begin{array}{rcll} u_k & : & [0,1] & \to \mathbb{R} \\ & & \sigma & \mapsto \left\{ \begin{array}{ll} 1 + (1/k) & , \ \sigma \in [0, 1/2] \\ 1 - (1/k) & , \ \sigma \in (1/2, 1] \end{array} \right. . \end{array}$$

For all $k \in \mathbb{N}$ we have $u_k \in L^1(dx)$, $u_k \geq 0$, $\int_\Omega u_k \, dx = 1$ and $\|u_k - 1\|_{L^1(d\mu)} = 1/k$. Furthermore, for all $k \in \mathbb{N}$:

$$e_\psi(u_k) = \frac{1}{2}\left(\psi\left(\frac{k+1}{k}\right) + \psi\left(\frac{k-1}{k}\right)\right).$$

Hence,

$$\lim_{k\to\infty} \|u_k - 1\|_{L^1(d\mu)} = \lim_{k\to\infty} e_\psi(u_k) = 0.$$

Now let $\eta \in (0, \infty)$. Then a simple application of the de l'Hospital rule gives

$$\lim_{k\to\infty} \frac{\|u_k - 1\|_{L^1(d\mu)}^{1/\eta}}{e_\psi(u_k)} = \frac{2}{\eta(1+e)} \lim_{\sigma\to\infty} a^{1-1/\eta} e^{\sigma+1} = \infty.$$

The core of this example is the fact that *all* derivatives of ψ vanish at 1.

However, if ψ is m-times, $m \geq 2$, continuously differentiable in a neighborhood of $[u] \in J^\circ$ with

$$\psi'([u]) = \ldots = \psi^{(m-1)}([u]) = 0, \psi^{(m)}([u]) > 0$$

(due to the convexity of ψ, $\psi^{(m)}([u])$ has to be non-negative), we can use the observation that for all $[u] \in J^\circ$, all sufficiently small $a \in (0, \infty)$ and $\vartheta \in (a/(\beta - [u]), 1 - (a/([u] - \alpha)))$ (see the proofs of Theorems 2.4 and 2.5):

$$\begin{aligned} &\frac{\psi\left([u] + \frac{a}{\vartheta}\right) - \psi([u])}{\frac{a}{\vartheta}} + \frac{\psi\left([u] - \frac{a}{1-\vartheta}\right) - \psi([u])}{\frac{a}{1-\vartheta}} \\ &\qquad\qquad \geq \frac{\psi([u] + a) - \psi([u])}{a} + \frac{\psi([u] - a) - \psi([u])}{a}. \end{aligned} \tag{2.4}$$

Now we can use a Taylor-expansion argument to obtain

$$\begin{array}{lrcl} \text{if } m \text{ is even:} & \left\|u - \int_\Omega u \, d\mu\right\|_{L^1(d\mu)} (1 + o(1)) & \leq & 2\left(\dfrac{m! \, e_\psi(u)}{4\psi^{(m)}\left(\int_\Omega u \, d\mu\right)}\right)^{1/m}, \\ \text{if } m \text{ is odd:} & \left\|u - \int_\Omega u \, d\mu\right\|_{L^1(d\mu)} & = & o((e_\psi(u)^{1/(m+1)}), \end{array} \tag{2.5}$$

as $e_\psi(u) \to 0$.

The upper estimate can be refined in the following cases.

Case 1: Let $\psi([u]) = \psi'([u]) = 0$, $\psi''([u]) > 0$ and let ψ have a logarithmic sub-entropy χ, i.e.

$$\psi(\sigma) \geq \chi(\sigma) := a(\sigma + b) \ln \frac{\sigma + b}{[u] + b} - a(\sigma - [u])$$

for some $a > 0$ and $b \geq 0$ (see §2.2 of [AMTU99] for details). Then the term $o(1)$ in (2.5) can be ignored and the estimate

$$\|u - [u]\|_{L^1(d\mu)} \leq \sqrt{\frac{2\, e_\psi(u)}{\psi''([u])}} \tag{2.6}$$

holds.

Case 2: $\psi([u]) = \psi'([u]) = 0$, $\psi''([u]) > 0$, and $\psi''' \geq 0$ in a neighborhood of $[u] = \int_\Omega u\, d\mu$. Then the term $o(1)$ is nonnegative and we obtain (2.6) for all admissible u with $[u] \in J^\circ$ as $e_\psi(u) \to 0$. We note that this estimate holds for *all* $e_\psi(u) \in [0, \infty]$ whenever $\psi \in C^3(J)$ with $\psi''([u]) > 0$ and $\psi''' \geq 0$ on J. As an example consider $\psi(\sigma) = e^\sigma - e\,\sigma$. We have

$$\|u - 1\|_{L^1(d\mu)} \leq \sqrt{2 \int_\Omega (e^{u-1} - 1)\ d\mu},$$

for all admissible u with $[u] = 1$.

Let us keep the assumption that ψ is three times differentiable in a neighborhood of $[u]$ with $\psi''([u]) > 0$. Then the estimate (2.5) is in general not optimal. Consider e.g. $\psi(\sigma) = \sigma\ (\ln \sigma - \ln[u] - 1)$, where $[u] \in (0, \infty)$. We obtain from (2.5) for all admissible u:

$$\|u - [u]\|_{L^1(d\mu)}\, (1 + o(1)) \leq \sqrt{2[u] \left(\int_\Omega u \ln u\, d\mu \right) - 2[u]^2 \ln[u]},$$

where due to $\psi''' \leq 0$ the term $o(1)$ is negative, while on the other hand (as mentioned above) we have for all admissible u with $[u] = 1$:

$$\|u - 1\|_{L^1(d\mu)} \leq \sqrt{2 \int_\Omega u \ln u\, d\mu}.$$

The reason for this is the fact that the Taylor series argument used to obtain (2.5) provides in general only a (not necessarily sharp) upper bound for U.

If the function ψ is not differentiable at $[u] = \int_\Omega u \, d\mu \in J^\circ$, then $\partial^+\psi([u]) - \partial^-\psi([u]) > 0$ and we obtain from (2.4)

$$\left\| u - \int_\Omega u \, d\mu \right\|_{L^1(d\mu)} \leq \frac{2 \, e_\psi(u)}{\partial^+\psi([u]) - \partial^-\psi([u])}$$

as $e_\psi(u) \to 0$.

We can draw the following conclusion from this discussion: The more derivatives of $\psi([u])$ vanish (where $[u] \in J^\circ$), the slower u converges to $[u]$ in $L^1(d\mu)$ as $e_\psi(u) \to 0$.

2.5 Passing from weak to strong convergence in $L^1(d\mu)$

Inequality (1.3) allows to pass from weak $L^1(d\mu)$-convergence to strong $L^1(d\mu)$-convergence in the following sense. Given a sequence $(u_k)_{k\in\mathbb{N}}$ in $L^1(d\mu)$ and a $K \in \mathbb{R}$ with

i) $\int_\Omega u_k \, d\mu \to K$ as $k \to \infty$ (this is, e.g. the case if $u_k \rightharpoonup u$ as $k \to \infty$ weakly in $L^1(d\mu)$, where $u \in L^1(d\mu)$).

ii) There exists a strictly convex and bounded below function $\psi : J \to \mathbb{R}$ (J is not necessarily open) such that:

   ii.a) For all sufficiently large $k \in \mathbb{N}$: $u_k(x) \in J$ μ-a.e.,

   ii.b) $\lim\limits_{k\to\infty} \int_\Omega \psi(u_k) \, d\mu = \psi(K)$.

Then

$$u_k \to K \quad \text{strongly in } L^1(d\mu) \text{ as } k \to \infty.$$

The non-trivial fact is that no growth conditions on ψ are imposed. Consider e.g. the case where $\psi(\sigma) = e^{-\sigma}$ and let $(h_k)_{k\in\mathbb{N}}$ be a sequence in $L^1(dx)$ and let $h \in L^1(dx)$ such that $h(x) > 0$ for almost all $x \in \Omega$. Then by setting $d\mu \equiv h(x)dx$ we see that

$$\lim_{k\to\infty} \int_\Omega h_k \, dx = \int_\Omega h \, dx$$

together with

$$\lim_{k\to\infty}\int_\Omega e^{-h_k[h]/h}\ h\ dx = e^{-[h]}\ [h]$$

implies $h_k \to h$ strongly in $L^1(dx)$ as $k \to \infty$.

2.6 The optimal inequality for the logarithmic entropy

Now we consider the case $\psi(\sigma) = \sigma\ln\sigma - \sigma + 1$ with $J = [0,\infty) = [\alpha,\beta)$, $\psi(0) = 1$ and $[u] = 1$. We recall that in this case we have for all $u \in L^1(d\mu)$ with $e_\psi(u) < \infty$ the estimates

$$\|u-1\|_{L^1(d\mu)} \leq \mathsf{U}\left(\int_\Omega u\ \ln u\ d\mu, 1\right), \tag{2.7}$$

(from (1.3)) and

$$\|u-1\|_{L^1(d\mu)} \leq \sqrt{2\int_\Omega u\ \ln u\ d\mu}, \tag{2.8}$$

(Csiszár-Kullback inequality). In the following plot (Figure 1) we compare the functions $\mathsf{U}(c,1)$ (solid line) and $\sqrt{2\ c}$ (dashed line).

We observe that for all $u \in L^1(d\mu)$ with $[u] = 1$ and $u(x) \geq 0$ μ- a.e. the estimate

$$\|u-1\|_{L^1(d\mu)} < 2 = \mathsf{U}_{sup}(1) = 2(1-0)$$

holds. Hence the estimate given by (2.8) provides no information when $e_\psi(u) \geq 2$. From (2.7), however, one obtains a non-trivial bound of $\|u - 1\|_{L^1(d\mu)}$ for *all* finite values of $e_\psi(u)$.

2.7 Entropy-type estimates on $\|f-g\|_{L^1(d\mu)}$

In this subsection we investigate possibilities to estimate $\|f-g\|_{L^1(d\mu)}$ in terms of

$$e_\psi(f|g) = \int_\Omega \psi(f/g)g\ d\mu,$$

where we make the assumptions

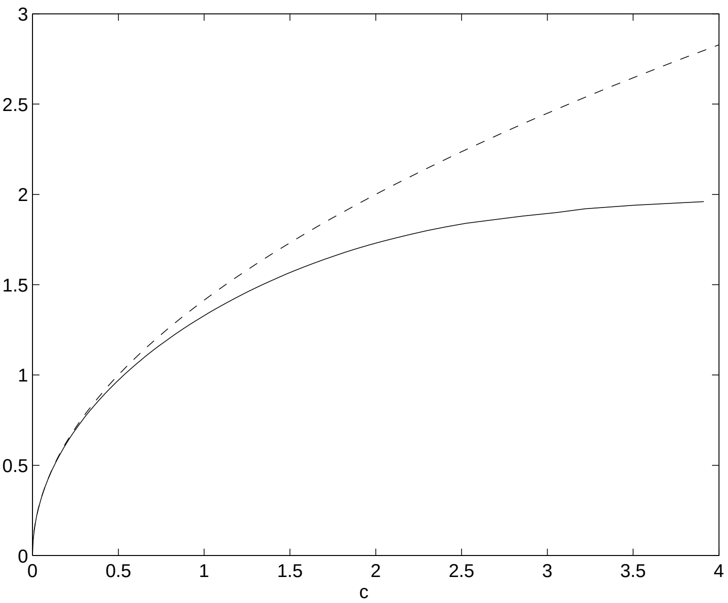

Figure 1: $\mathsf{U}(c,1)$ (solid) and $\sqrt{2\,c}$ (dashed) for $\psi(\sigma) = \sigma\ln\sigma - \sigma + 1$

B.1 $(\Omega, \mathcal{S}, \mu)$ is a measure space and μ is a probability measure on $\mathcal{S}$, i.e. $\mu(\Omega) = 1$, $\mu \geq 0$.

B.2" $\psi : J \to I\!R$ is convex, continuous and non-negative. The (possibly unbounded) interval J has a non-empty interior J°. If $\alpha := \inf J \notin J$, then $\lim_{\sigma\to\alpha} \psi(\sigma) = \infty$. If $\beta := \sup J \notin J$, then $\lim_{\sigma\to\beta} \psi(\sigma) = \infty$. $1 \in J^\circ$, ψ is strictly convex at 1, and $\psi(1) = 0$.

B.3 $g \in L^1(d\mu)$ is positive for μ-almost all $x \in \Omega$ and $\int_\Omega g\; d\mu = 1$.

B.4 $f \in L^1(d\mu)$ satisfies $f(x)/g(x) \in J$ for μ-almost all $x \in \Omega$.

If we additionally assume that ψ is *strictly* convex, then we obtain from Theorem 2.4

$$\|f - g\|_{L^1(d\mu)} \leq \mathsf{U}\left(e_\psi(f|g) - \psi([f]), [f]\right) + |[f] - 1|\,, \tag{2.9}$$

with the notation $[f] = \int_\Omega f\; d\mu$.
(2.9) can be applied whenever $e_\psi(f|g)$ *and* $[f]$ are known. The question arises whether $\|f-g\|_{L^1(d\mu)}$ can be estimated *just in terms of* $e_\psi(f|g)$. We will give the surprisingly simple, affirmative answer to this question below. To prepare the discussion we shall introduce some abbreviations first. We define

$$J^+ := \{\sigma \in [0,\infty) : 1 + \sigma \in J\} \quad , \quad J^- := \{\sigma \in [0,\infty) : 1 - \sigma \in J\},$$

and set

$$\begin{array}{rcl} \psi^+ & : & J^+ \to [0,\infty) \\ & & \sigma \mapsto \psi^+(\sigma) := \psi(1+\sigma) \end{array}, \qquad \begin{array}{rcl} \psi^- & : & J^- \to [0,\infty) \\ & & \sigma \mapsto \psi^-(\sigma) := \psi(1-\sigma) \end{array}.$$

Clearly $\psi^\pm$ are continuous, non-negative, convex functions which are strictly convex at 0. Their respective epigraphs

$$\mathsf{epi}(\psi^\pm) = \{(\sigma, z) : z \geq \psi^\pm(\sigma)\}$$

are due to **B**.2″ closed and convex subsets of $I\!R^2$. Following [EkTe74] it is easy to see that the closure of the convex hull of $\mathsf{epi}(\psi^+) \cup \mathsf{epi}(\psi^-)$ is given by

$$\overline{co\,(\mathsf{epi}(\psi^+) \cup \mathsf{epi}(\psi^-))} = \mathsf{epi}(co(\psi^+, \psi^-)),$$

where the function $co(\psi^+, \psi^-) : J^+ \cup J^- \to [0,\infty)$ is the envelope of all affine functions lower or equal to $\psi^\pm$, respectively. Some properties of $co(\psi^+, \psi^-)$ are collected in the following lemma.

Lemma 2.10. *Assume* **B**.*2". Then:*

a) $co(\psi^+, \psi^-)$ *is convex.*

b) $co(\psi^+, \psi^-)$ *is continuous.*

c) $co(\psi^+, \psi^-)$ *is strictly increasing.*

d) $co(\psi^+, \psi^-)(0) = 0$ *and* $co(\psi^+, \psi^-)(\sigma) > 0$ *for all* $\sigma \in J^+ \cup J^-$*,* $\sigma > 0$.

The proof of Lemma 2.10 is deferred to the appendix.

Remark 2.11. *Due to a) and d) of Lemma 2.10 the function* $co(\psi^+, \psi^-)$ *is strictly convex at* 0*. One may ask whether* $co(\psi^+, \psi^-)$ *is strictly convex on* $J^+ \cup J^-$ *whenever* ψ^+ *and* ψ^- *are strictly convex. The following example gives a negative answer to this question: Let* $\psi^+ : [0,\infty) \to I\!R, \sigma \mapsto \sigma^2$ *and let* $\psi^- : [0,\infty) \to I\!R, \sigma \mapsto \sigma^4$*. Then it is easy to see that*

$$\begin{array}{rcl} co(\psi^+, \psi^-) & : & [0,\infty) \to I\!R \\ & & \sigma \mapsto \left\{ \begin{array}{ll} \sigma^4 & , \ \sigma \in \left[0, \frac{1}{2}\sqrt{3}\right) \\ \frac{3\sqrt{3}}{4}\sigma - \frac{27}{16} & , \ \sigma \in \left[\frac{1}{2}\sqrt{3}, \frac{3}{4}\sqrt{3}\right) \\ \sigma^2 & , \ \sigma \in \left[\frac{3}{4}\sqrt{3}, \infty\right) \end{array} \right. \end{array},$$

i.e. $co(\psi^+, \psi^-)$ *is* not *strictly convex.*

We set

$$e_\psi^* := \sup_{\sigma \in J^+ \cup J^-} co(\psi^+, \psi^-)(\sigma).$$

The main result of this subsection is

Theorem 2.12. *Assume* **B**.*1,* **B**.*2" and* **B**.*3. Then:*

a) For all $f \in L^1(d\mu)$ satisfying **B**.*4 the following estimate holds:*

$$\|f-g\|_{L^1(d\mu)} \le \begin{cases} [co(\psi^+,\psi^-)]^{-1}(e_\psi(f|g)) & , \quad e_\psi(f|g) < e_\psi^* \\ \max\{1-\alpha, \beta-1\} & , \quad e_\psi(f|g) \ge e_\psi^* \end{cases}$$

b) If there exists for all $\vartheta \in (0,1)$ a set $S_\vartheta \in \mathcal{S}$ such that $\int_{S_\vartheta} g\, d\mu = \vartheta$, then there exists for all $\sigma \in J^+ \cup J^-$ a sequence $(f_n)_{n\in I\!N}$ in $L^1(d\mu)$ such that **B**.*4 is satisfied for all $n \in I\!N$ and*

$$\|f_n - g\|_{L^1(d\mu)} = \sigma \quad \text{for all} \quad n \in I\!N \quad ,$$

$$\sigma = \lim_{n\to\infty} \left[co(\psi^+,\psi^-)\right]^{-1}(e_\psi(f_n|g)).$$

The proof of Theorem 2.12 is deferred to the appendix. Part b) of Theorem 2.12 is an optimality result (the knowledge of $e_\psi(f|g)$ does not allow to conclude more about $\|f-g\|_{L^1(d\mu)}$ than stated in Theorem 2.12) which applies, e.g. for the n-dimensional Lebesgue-measure $d\mu = dx$. The calculation of $co(\psi^+,\psi^-)$ is very simple in cases where $\psi^- \ge \psi^+$ with $J^- \subseteq J^+$ (or $\psi^+ \ge \psi^-$ with $J^+ \subseteq J^-$) holds. In such situations we have

$$co(\psi^+,\psi^-) = \psi^+ \quad (\text{or} \quad co(\psi^+,\psi^-) = \psi^-).$$

As an example consider $\psi(\sigma) = \sigma\,(\ln\sigma - 1) + 1$. Then $J^- = [0,1]$, $J^+ = [0,\infty)$ with

$$\psi^+(\sigma) = (1+\sigma)\,(\ln(1+\sigma)-1)+1 \le (1-\sigma)\,(\ln(1-\sigma)-1)+1 = \psi^-(\sigma),$$

for all $\sigma \in [0,1]$. Thus $co(\psi^+,\psi^-)(\sigma) = (1+\sigma)\,(\ln(1+\sigma)-1)+1$ and we obtain

Proposition 2.13. *Assume* **B**.*1,* **B**.*2" and* **B**.*3. Then for all $f \in L^1(d\mu)$ with $f \ge 0$ the estimate*

$$(1+\|f-g\|_{L^1(d\mu)})(\ln(1+\|f-g\|_{L^1(d\mu)})-1)+1 \le \int_\Omega f \ln(f/g)\, d\mu + \int_\Omega (f-g)\, d\mu$$

holds.

3 Appendix: Proofs of Section 2

Proof of Theorem 2.4:

Step 1: We shall first construct an auxiliary function ϕ. We introduce

$M := \{(\sigma, a, \vartheta) \in J^\circ \times (0,\infty) \times (0,1) : a < \min\{\vartheta(\beta - \sigma), (1-\vartheta)(\sigma - \alpha)\}\}.$

We note that $M \neq \emptyset$ consists exactly of those $(\sigma, a, \vartheta) \in J^\circ \times (0,\infty) \times (0,1)$ for which $\sigma + (a/\vartheta)$ and $\sigma - (a/(1-\vartheta))$ belongs to J°. We define

$$\phi : M \to I\!R,$$

$$\phi(\sigma, a, \vartheta) = \vartheta \left(\psi\left(\sigma + \frac{a}{\vartheta}\right) - \psi(\sigma)\right) + (1-\vartheta)\left(\psi\left(\sigma - \frac{a}{1-\vartheta}\right) - \psi(\sigma)\right).$$

For fixed $\sigma \in J^\circ$ let $\mathbf{Pr}_a(M;\sigma)$ be the projection of M onto the a-axes and for fixed $\sigma \in J^\circ$ and fixed $a \in \mathbf{Pr}_a(M;\sigma)$ let $\mathbf{Pr}_\vartheta(M;\sigma,a)$ be the projection of M onto the ϑ-axis. Then we can easily verify that

$$\mathbf{Pr}_a(M;\sigma) = \left(0, \frac{\mathsf{U}_{sup}(\sigma)}{2}\right) \quad , \quad \mathbf{Pr}_\vartheta(M;\sigma,a) = \left(\frac{a}{\beta-\sigma}, 1 - \frac{a}{\sigma-\alpha}\right).$$

Due to the assumed strict convexity of ψ we observe that for all $(\sigma, a) \in J^\circ \times \left(0, \frac{\mathsf{U}_{sup}(\sigma)}{2}\right)$:

$$\begin{aligned} & \inf_{\vartheta \in \mathbf{Pr}_\vartheta(M;\sigma,a)} \phi(\sigma,a,\vartheta) \\ = \; & \inf_{\vartheta \in \mathbf{Pr}_\vartheta(M;\sigma,a)} a \left(\frac{\psi\left(\sigma + \frac{a}{\vartheta}\right) - \psi(\sigma)}{\frac{a}{\vartheta}} + \frac{\psi\left(\sigma - \frac{a}{1-\vartheta}\right) - \psi(\sigma)}{\frac{a}{1-\vartheta}}\right) \\ > \; & 0 \end{aligned}$$

Step 2: We define

$$\hat{H} : J^\circ \times \left(0, \frac{\mathsf{U}_{sup}(\sigma)}{2}\right) \to (0,\infty), \quad \hat{H}(\sigma,a) = \inf_{\vartheta \in \mathbf{Pr}_\vartheta(M;\sigma,a)} \phi(\sigma,a,\vartheta).$$

We note that $\hat{H}$ is defined as the infimum of a bounded-below, continuous function (recall that each convex function ψ is continuous on the interior of its domain) over an open set where two entries are fixed. Hence $\hat{H}$ is continuous. Furthermore, by the *strict* convexity of ψ, the function

$$\hat{H}(\sigma,.) : \left(0, \frac{\mathsf{U}_{sup}(\sigma)}{2}\right) \to (0,\infty), \quad a \mapsto \hat{H}(\sigma,a)$$

is for all fixed $\sigma \in J^\circ$ *strictly* increasing. It is not very difficult to check that for all $\sigma \in J^\circ$:

$$\lim_{a\to 0} \hat{H}(\sigma,a) = 0 \quad , \quad \lim_{a\to \frac{\mathsf{U}_{sup}(\sigma)}{2}} \hat{H}(\sigma,a) = c_0(\sigma).$$

For fixed $\sigma \in J^\circ$ we can therefore define the generalized inverse mapping (see [BeLo76])

$$I(.,\sigma) : (0,\infty) \to \left(0, \frac{\mathsf{U}_{sup}(\sigma)}{2}\right), I(c,\sigma) = \begin{cases} \left(\hat{H}(\sigma,.)\right)^{-1}(c) & \text{if } c \in (0, c_0(\sigma)) \\ \dfrac{\mathsf{U}_{sup}(\sigma)}{2} & \text{if } c \in [c_0(\sigma),\infty) \end{cases} .$$

We note that $I(.,\sigma)$ is increasing, strictly increasing on $(0,c_0(\sigma))$ and that the mapping $(\sigma,c) \mapsto I(c,\sigma)$, $\sigma \in J^\circ$, $c \in (0,\infty)$ is continuous. Furthermore it is easy to check that $\lim_{c\to 0} I(c,\sigma) = 0$. Let us note that

$$\lim_{c\to c_0(\sigma)} I(c,\sigma) = \frac{\mathsf{U}_{sup}(\sigma)}{2}.$$

<u>Step 3:</u> Now we are in the position to define U.

$$\mathsf{U} : [0,\infty) \times J \to [0,\infty), \quad \mathsf{U}(c,\sigma) = \begin{cases} 2I(c,\sigma) & \text{if } c \in (0,c_0(\sigma)), \sigma \in J^\circ \\ \mathsf{U}_{sup}(\sigma) & \text{if } c \in [c_0(\sigma),\infty), \sigma \in J^\circ \\ 0 & \text{if } c = 0 \text{ or if } \sigma \in \partial J \cap J \end{cases} .$$

The verification of P1. -P7. follows from the previous remarks and can therefore be left to the reader. □

Proof of Lemma 2.10:

a), b) $co(\psi^+,\psi^-)$ is the envelope of a family of affine functions. Hence $co(\psi^+,\psi^-)$ is convex and continuous, see [EkTe74].

d) $\psi^\pm \geq 0$ and $\psi^\pm(0) = 0$ implies $co(\psi^+,\psi^-) = 0$. We easily deduce from **B**.2" that $(\partial^+\psi^\pm)(\sigma) \geq \psi^\pm(\sigma)/\sigma$ for all $\sigma \in J^\pm$, $\sigma > 0$. For each $\sigma \in J^+ \cup J^-$, $\sigma > 0$, there exists a $\vartheta = \vartheta(\sigma) \in (0,1)$ such that $\vartheta\sigma \in J^+ \cap J^-$. We set $K_\sigma := \min\{(\partial^+\psi^+)(\vartheta\sigma), (\partial^+\psi^-)(\vartheta\sigma)\}$. Then it is easy to verify that for all $s_1 \in J^+$ and all $s_2 \in J^-$:

$$\psi^+(s_1) \geq K_\sigma\ (s_1 - \vartheta\sigma) \quad , \quad \psi^-(s_2) \geq K_\sigma\ (s_2 - \vartheta\sigma).$$

Hence, by definition, $co(\psi^+,\psi^-)(\sigma) \geq K_\sigma\ (\sigma - \vartheta\sigma) = (1-\vartheta)\ K_\sigma\ \sigma > 0$.

c) Let $\sigma_1, \sigma_2 \in J^+ \cup J^-$ with $\sigma_1 < \sigma_2$. We have to prove $co(\psi^+,\psi^-)(\sigma_1) < co(\psi^+,\psi^-)(\sigma_2)$. If $\sigma_1 = 0$, this estimate will follow from d). If $\sigma_1 > 0$, we set $\vartheta := 1 - (\sigma_1/\sigma_2)$. Certainly $\vartheta \in (0,1)$. We calculate $co(\psi^+,\psi^-)(\sigma_1) = co(\psi^+,\psi^-)(\vartheta\ 0 + (1-\vartheta)\ \sigma_2) \leq \vartheta co(\psi^+,\psi^-)(0) + (1-\vartheta) co(\psi^+,\psi^-)(\sigma_2) < co(\psi^+,\psi^-)(\sigma_2)$, since $co(\psi^+,\psi^-)(0) = 0$, $1-\vartheta \in (0,1)$ and $co(\psi^+,\psi^-)(\sigma_2) > 0$. $\square$

Proof of Theorem 2.12:

a) After the previous remarks it suffices to prove: For all $f \in L^1(d\mu)$ satisfying **B**.4 and $e_\psi(f|g) < \infty$, the pair $(\|.\|, e_\psi(.)) := (\|f-g\|_{L^1(d\mu)}, e_\psi(f|g)) \in co(\mathsf{epi}(\psi^+) \cup \mathsf{epi}(\psi^-))$. Since both $\mathsf{epi}(\psi^+)$ and $\mathsf{epi}(\psi^-)$ are convex (subsets of $\mathbb{R}^2$) we have

$$
\begin{aligned}
&co(\mathsf{epi}(\psi^+) \cup \mathsf{epi}(\psi^-)) \\
&\qquad = \big\{(\vartheta\sigma^+ + (1-\vartheta)\sigma^-, z) : \sigma^+ \in J^+, \sigma^- \in J^-, \vartheta \in [0,1], \\
&\qquad\qquad\qquad z \geq \vartheta\psi^+(\sigma^+) + (1-\vartheta)\psi^-(\sigma^-)\big\}\,.
\end{aligned}
$$

Let $\Omega_0^+ := \{f \geq g\}$ and $\Omega^- := \{f < g\}$. We define $\vartheta := \int_{\Omega_0^+} g\ d\mu$. Then $\vartheta \in [0,1]$ and $1-\vartheta = \int_{\Omega^-} g\ d\mu$. We introduce $w := (f/g) - 1$ and set $a := \int_{\Omega_0^+} w\, g\ d\mu$, as well as $b := -\int_{\Omega^-} w\, g\ d\mu$. Then $a, b \in [0,\infty)$ and

$$\|.\| = a + b.$$

We assume for the time being $\vartheta \neq 0, 1$. Then

$$
\begin{aligned}
e_\psi(.) &= \vartheta \int_{\Omega_0^+} \psi(1+w)\ \frac{g}{\vartheta}\ d\mu + (1-\vartheta)\int_{\Omega^-} \psi(1+w)\ \frac{g}{1-\vartheta}\ d\mu \\
&\geq \vartheta\psi\left(\frac{1}{\vartheta}\int_{\Omega_0^+} (g+wg)\ d\mu\right) + (1-\vartheta)\psi\left(\frac{1}{1-\vartheta}\int_{\Omega^-} (g+wg)\ d\mu\right) \\
&\qquad\qquad = \vartheta\psi\left(1+\frac{a}{\vartheta}\right) + (1-\vartheta)\psi\left(1-\frac{b}{1-\vartheta}\right).
\end{aligned}
$$

Setting

$$\sigma^+ := \frac{a}{\vartheta} \quad , \quad \sigma^- := \frac{b}{1-\vartheta},$$

we obtain

$$\|.\| = \vartheta\sigma^+ + (1-\vartheta)\sigma^- \quad , \quad e_\psi(.) \geq \vartheta\psi^+(\sigma^+) + (1-\vartheta)\psi^-(\sigma^-), \tag{3.1}$$

where $\sigma^+ \in J^+$ and $\sigma^- \in J^-$. The reader may wish to verify (3.1) in case of $\vartheta = 0$ or $\vartheta = 1$ for some $\sigma^+ \in J^+$ and $\sigma^- \in J^-$. Hence $(\|.\|, e_\psi(.)) \in co(\mathsf{epi}(\psi^+) \cup \mathsf{epi}(\psi^-))$.

b) For all $\sigma \in J^+ \cup J^-$ we have

$$\begin{aligned} &co(\psi^+, \psi^-)(\sigma) \\ &\quad = \inf\{\vartheta\psi^+(\sigma^+) + (1-\vartheta)\psi^-(\sigma^-) : \sigma^+ \in J^+, \sigma^- \in J^-, \vartheta \in [0,1], \\ &\qquad\qquad \vartheta\sigma^+ + (1-\vartheta)\sigma^- = \sigma\}. \end{aligned}$$

Hence it is possible to choose a sequence $(\vartheta_n, \sigma_n^+, \sigma_n^-)_{n \in I\!N}$ in $[0,1] \times J^+ \times J^-$ with

$$\vartheta_n \sigma_n^+ + (1-\vartheta_n)\sigma_n^- - = \sigma \quad \text{for all} \quad n \in I\!N \quad ,$$

$$\lim_{n\to\infty} \vartheta_n \psi^+(\sigma_n^+) + (1-\vartheta_n)\psi^-(\sigma_n^-) = co(\psi^+, \psi^-)(\sigma).$$

By assumption we can choose for all $n \in I\!N$ a set $S_n \in \mathcal{S}$ with $\int_{S_n} g \, d\mu = \vartheta_n$. Setting for $n \in I\!N$

$$\begin{array}{rcccl} f_n & : & \Omega & \to & I\!R \\ & & x & \mapsto & \begin{cases} (1+\sigma_n^+)\, g(x) & , \quad x \in S_n \\ (1-\sigma_n^-)\, g(x) & , \quad x \in \Omega \setminus S_n \end{cases} \end{array}$$

finishes the proof. □

References

[AMTU99] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations, submitted to *Comm. PDE*, 1999.

[BaNi64] O. Barndorff-Nielsen, Subfields and loss of information, *Z. Wahrsch. Verw. Geb.*, 2:369–379, 1964.

[BeLo76] J. Bergh, J. Lofstrom, Interpolation Spaces, *Springer*, 1976.

[Bla65] N.M. Blachman, The convolution inequality for entropy powers, *IEEE Trans. Info. Thy.* 2:267–271, 1965.

[Che81] H. Chernoff, A note on an inequality involving the normal distribution, *Ann Probab.* 9:533–535, 1981.

[Csi63] I. Csiszár, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis von Markoffschen Ketten, *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:85–108, 1963.

[Csi66] I. Csiszár, A note on Jensen's inequality, *Stud. Sc. Math. Hung.*, 1:227–230, 1966.

[Csi67] I. Csiszár, Information-type measures of difference of probability distributions, *Stud. Sc. Math. Hung.*, 2:299–318, 1967.

[EkTe74] I. Ekeland, R. Temam, Analyse convexe et problèmes variationnels *Dunod Gauthier-Villars*, 1974.

[Kul59] S. Kullback, Information Theory and Statistics, *John Wiley*, 1959.

[KuLei51] S. Kullback and R. Leibler, On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[Per57] A. Pérez, Notions généralisées d'intertitude, d'entropie et d'inform-
ation du point de vue de la théorie de martingales, *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague*, 183–208, 1957.

[Per65] A. Pérez, Information, ϵ-sufficiency and data reduction problems, *Kybernetika*, 1:297–322, 1965.

[Rud87] W. Rudin, Real and Complex Analysis, *Mc-Graw Hill*, 1987.

[Sta59] A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Info. Contr.*, 2:101–112, 1959.