Refinements of Pinsker's Inequality

Alexei A. Fedotov, Peter Harremoës and Flemming Topsøe

Abstract—**Let V and D denote, respectively, total variation and divergence. We study lower bounds of D with V fixed. The theoretically best (i.e. largest) lower bound determines a function $L = L(V)$, *Vajda's tight lower bound*, cf. Vajda, [?]. The main result is an exact parametrization of L. This leads to Taylor polynomials which are lower bounds for L, and thereby extensions of the classical *Pinsker inequality* which has numerous applications, cf.Pinsker, [?] and followers.**

Keywords— **Divergence, total variation, Pinsker's inequality, Vajda's tight lower bound.**

I. Introduction and survey of results

Let $M_+^1(n)$ be the set of probability measures on an "alphabet" A with n elements. Denote by $D = D(P \| Q)$ the divergence

$$D(P \| Q) = \sum_{i \in A} p_i \log \frac{p_i}{q_i} ,$$

and by $V = V(P,Q)$ the total variation

$$V(P,Q) = \sum_{i \in A} |p_i - q_i| .$$

We are interested in lower bounds of D in terms of V. The start of research in this direction is *Pinsker's inequality*

$$D \geq \frac{1}{2} V^2,$$

cf. Pinsker [?], and a later improvement by Csiszár [?], where the best constant ($\frac{1}{2}$ as stated) was determined. The best two-term inequality of this type is

$$D \geq \frac{1}{2} V^2 + \frac{1}{36} V^4 ,$$

as proved by Krafft [?].

A further term $c_6 V^6$ was added by Krafft and Schmitz [?], Toussaint [?] and by Topsøe [?], where the best constant $c_6 = \frac{1}{270}$ was determined.

By the *best constants* $c_\nu^{\max}$, $\nu = 0, 1, 2, \ldots$, we shall understand the constants defined recursively by taking $c_\nu^{\max}$ to be the largest constant c for which the inequality

$$D \geq \sum_{i < \nu} c_i^{\max} V^i + c V^\nu$$

holds generally. Clearly, $c_\nu^{\max}$, $\nu = 0, 1, 2, \ldots$, are well defined non-negative real constants.

Another kind of bound was found by Vajda in [?] who proved that

$$D \geq \log\left(\frac{2+V}{2-V}\right) - \frac{2V}{2+V} .$$

This bound, *Vajda's lower bound*, is almost as good as Pinsker's inequality for values of V near 0, and has the added advantage that it gives the "right" bound (∞) when V approaches 2. Vajda suggested a closer study of the function L defined by

$$L(V_0) = \inf_{V(P,Q) = |V_0|} D(P \| Q) \quad \text{for } V_0 \in]-2; 2[.$$

This function we shall refer to as *Vajda's tight lower bound.* Note that by definition, L is an even function of V. In the original definition given by Vajda, only non-negative values of V_0 were considered. All the above inequalities may be seen as lower bounding approximations to L.

Fig. 1 Range of the map $(P,Q) \curvearrowright (D,V)$ and the function L.

In the main section, Section 2, we state the key result, Theorem 1, which offers a parametrization of L expressed in terms of elementary functions, with a pronounced occurrence of hyperbolic functions. Figure 1 shows the graph of L, situated in the (V,D)-plane. The two corollaries which follow immediately after Theorem 1 show how the parametrization of L may be used to derive Pinsker's inequality as well as Vajda's lower bound.

We continue with an investigation of the nature of the function L. Corollary 5 offers an integral representation which makes the calculation of approximating Taylor polynomials easy, at least in principle. In Theorem 6 we point out that L is analytic but, surprisingly enough, with a radius of convergence strictly less than 2 (for the power series

A. A. Fedotov is with the Department of Information Technologies, Institute of Computational Technologies, Novosibirsk-90, Russia (e-mail: lesha@adm.ict.nsc.ru). A. A. Fedotov have been supported by a grant from NATO Scientific Affairs Division

P. Harremoës is with Dept. of Mathematics, University of Copenhagen, Denmark (e-mail: moes@math.ku.dk). P. Harremoës by a grant from the Cowi Foundation and a post-doc fellowship from the Willum Kann Rasmussen Foundation.

F. Topsøe is with Department of Mathematics, University of Copenhagen, DK-2100 Copenhagen, Denmark (e-mail: topsoe@math.ku.dk).

F. Topsøe has been supported by a grant from the Danish Natural Science Research Council.

expansion centred at $V=0$). Parts of the proof of this result amounts to technical computations which are shown in detail in the appendix.

Conjectures regarding refinements of Pinsker's original inequality come up naturally but are difficult to decide. As a specific result in this direction, the general validity of the inequality

$$D \geq \frac{1}{2}V^2 + \frac{1}{36}V^4 + \frac{1}{270}V^6 + \frac{221}{340200}V^8$$

is asserted in Theorem 7 where it is also pointed out that the constants are best possible in the sense introduced above. However, the proof we present of this result depends on new ideas which are explained in Section 3. The final technical details of the proof of Theorem 7 are given in the appendix. The last section contains a discussion of our results.

II. Parametrization of Vajda's tight lower bound

By a well known data reduction inequality, cf. Kullback and Leibler [?], it follows that the determination of lower bounds of the type considered only depends on the interrelationship between D and V for distributions P,Q in $M_+^1(2)$. We therefore limit the further discussion to distributions of this type.

Two distributions, P and Q, occur in our investigations and we shall write $P=(p_1,p_2)$ and $Q=(q_1,q_2)$. By the *signed total variation* we understand the quantity $V=V(p_1,q_1)=2(q_1-p_1)$. We shall study the variation of $D=D(P\parallel Q)$ in terms of V. We may conceive D, just as V, as a function of (p_1,q_1). Both functions are defined on the $(p_1,q_1)-$unit square and are finite valued except for D which is infinite along two of the edges.

Apart form the obvious symmetry of D with respect to the point $(p_1,q_1)=(\frac{1}{2},\frac{1}{2})$ we note that the function $(P,Q) \curvearrowright D(P\parallel Q)$ is strictly convex, jointly in P and Q, except on the diagonal. In more detail, what we mean by this, is, firstly, that the inequality

$$D\left(\overline{P}\parallel\overline{Q}\right) \leq \alpha\, D\left(P_1\parallel Q_1\right) + \beta\, D\left(P_2\parallel Q_2\right)$$

holds for all distributions P_1,P_2,Q_1,Q_2 in $M_+^1(2)$ and all $\alpha>0\,,\beta>0$ with $\alpha+\beta=1$ where we have put $\overline{P}=\alpha P_1+\beta P_2$ and $\overline{Q}=\alpha Q_1+\beta Q_2$, and that, secondly, strict inequality holds *unless* either $(P_1,Q_1)=(P_2,Q_2)$ or else $P_1=Q_1$ and $P_2=Q_2$. [1]

By the convexity result just quoted, for each $V\in\,]-2;2[\setminus\{0\}$, there exists a unique pair (P_V,Q_V) of probability distributions such that $D(P\parallel Q)$ is minimal among all distributions with signed total variation equal to V. Define $(P_0,Q_0)=\left(\left(\frac{1}{2},\frac{1}{2}\right),\left(\frac{1}{2},\frac{1}{2}\right)\right)$. Then Vajda's tight lower bound is the function $V\curvearrowright L(V)$ given by $L(V)=D(P_V\parallel Q_V)$, $V\in\,]-2;2[$. By γ we denote the curve $V\curvearrowright(V,L(V)), V\in\,]-2;2[$. This curve we conceive as a curve in the (V,D)-plane. It has already been shown in Figure 1. Figure 2 gives an impression of the variation of (P_V,Q_V) by showing the curve determined by the first coordinates p_1 and q_1 of these distributions. This curve varies in the unit square. Each line parallel to the main diagonal determines a specific value of V as indicated on the figure. Thus, using the figure, you can determine approximately the distributions P_V and Q_V, given any value of $V\in]-2,2[$.

Fig. 2 The curve $V\curvearrowright(p_1,q_1)$ and contours of V.

The parameter V cannot be used to give an explicit parametrization of γ. As we shall see below both D and L are convex functions, and the convex conjugate (also called the Fenchel transform, see [?]) of both these functions can be calculated explicitly. The idea is now to use the parameter $t=\frac{dL}{dV}$ from the convex conjugate of L to parametrize L.

We shall express all functions which enter the analysis as functions of t. Apart from V this concerns L, i.e. the function $t\curvearrowright L(V(t))$ and then the coordinate functions $t\curvearrowright p_1(t)$ and $t\curvearrowright q_1(t)$ determined by the equations $P_{V(t)}=(p_1(t),1-p_1(t))$ and $Q_{V(t)}=(q_1(t),1-q_1(t))$.

We can now state our main result:

Theorem 1: The curve γ is a differentiable curve in the (V,D)-plane, symmetric around the D-axes. With $t=\frac{dL}{dV}$, the relationship $t\leftrightarrow V$ is a diffeomorphism between $\mathbb{R}$ and $]-2;2[$.

Using $t\in\mathbb{R}$ as parameter, γ is parametrized by

$$V\left(t\right)=2\coth t-\frac{t}{\sinh^2\left(t\right)}-t^{-1} \qquad (1)$$

$$L(V(t))=\log\left(\frac{t}{\sinh\left(t\right)}\right)+t\coth\left(t\right)-\frac{t^2}{\sinh^2\left(t\right)}\,.$$

[1] This result does not seem to be standard, e.g. in [?, theorem 2.7.2], only the inequality is deduced. The "strictness" – which is important for our purposes – can be deduced in the general case (i.e. with an arbitrary alphabet) from the log-sum inequality but, more expediently in our case of a two-letter alphabet, by observing that the determinant of the Hessian of the map $(p_1,q_1)\curvearrowright D(P\parallel Q)$ is $(p_1-q_1)^2/(p_1p_2q_1^2q_2^2)$.

Furthermore, the curve $V \curvearrowright (p_1(V), q_1(V))$ in the unit square (which characterizes the curve $V \curvearrowright (P_V, Q_V)$ in $M_+^1(2) \times M_+^1(2)$) has the parametrization

$$\begin{aligned} p_1(t) &= \frac{1}{2} + \frac{\frac{t}{\sinh^2(t)} - \coth(t)}{2} \\ q_1(t) &= \frac{1}{2} + \frac{\coth(t) - t^{-1}}{2} \end{aligned}$$

with $t \in \mathbb{R}$ as above and expressions defined by continuity for $t = 0$.

Proof: Knowing that D is convex we are able to determine the convex conjugate $D^*$of D. The convex conjugate of D is defined by

$$D^*(x,y) = \sup_{p_1,q_1} \left(\begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} - D(p_1,q_1) \right) .$$

We have

$$\begin{aligned} \frac{\partial}{\partial p_1} \left(\begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} - D(p_1,q_1) \right) &= x - \log \frac{p_1}{q_1} + \log \frac{p_2}{q_2} \\ \frac{\partial}{\partial q_1} \left(\begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} - D(p_1,q_1) \right) &= y + \frac{p_1}{q_1} - \frac{p_2}{q_2} \end{aligned}$$

To find the point where these partial derivatives are 0 we have to solve the simultaneous equations

$$\begin{aligned} \log \frac{p_1}{q_1} - \log \frac{p_2}{q_2} &= x \\ -\frac{p_1}{q_1} + \frac{p_2}{q_2} &= y \end{aligned}$$

which has the solution

$$\begin{aligned} p_1 &= e^x \frac{y + e^x - 1}{(e^x - 1)^2} \\ q_1 &= \frac{1}{1 - e^x} - \frac{1}{y} \end{aligned} \tag{2}$$

for x and y different from 0. For

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2t \\ 2t \end{pmatrix} \tag{3}$$

we get

$$\begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} = tV .$$

Therefore

$$\begin{aligned} D^*(-2t, 2t) &= \sup_{x,y} (t \cdot V - D(p,q)) \\ &= \sup_{V} (t \cdot V - L(V)) \end{aligned}$$

is the convex conjugate of L, and t must be the derivative of L. We see that (2) and (3) exactly solves our minimization problem. Convex conjugation transforms differentiable functions into differentiable functions. The parametrizations of p_1, q_1, V and $L \circ V$ are obtained by direct evaluation of the quantities involved. ∎

Remark 2: For the proof of Theorem 1 it appears natural to consider the convex conjugate of the function concerned, however in our case this is not necessary. Indeed, one may simply check directly that the suggested solution has the properties required.

Corollary 3 (Pinsker's inequality) For all $V \in [0; 2[$,

$$D \geq \frac{1}{2} V^2 . \tag{4}$$

Proof: Consider the difference $E = L(V) - \frac{1}{2}V^2$. Clearly, $E(0) = 0$. Accordingly, if we can show that $\frac{dE}{dV} \geq 0$ we are done. Now note that $\frac{dE}{dV} = t(V) - V = t - V$. The non-negativity of this quantity follows immediately upon noting that we may rewrite the parametrization for V in Theorem 1 in the following form:

$$V = t \left(1 - \left(\coth(t) - \frac{1}{t} \right)^2 \right) . \tag{5}$$

∎

Corollary 4 (Vajda's lower bound) For all $V \in [0; 2[$,

$$D \geq \log \frac{2+V}{2-V} - \frac{2V}{2+V} . \tag{6}$$

Proof: We use the same approach as in the previous proof and consider the function

$$E = L(V) - \left(\log \frac{2+V}{2-V} - \frac{2V}{2+V} \right) .$$

Then $E(0) = 0$ and

$$\frac{dE}{dV} = t - \frac{8V}{(2-V)(2+V)^2} .$$

If $t < 1$ then $V < 1$, and

$$\frac{8}{(2-V)(2+V)^2} \leq 1,$$

and $\frac{dE}{dV} \geq t - V \geq 0$.

If $t \geq 1$ then, using also the general inequality $\frac{8V}{(2+V)^2} \leq 1$, we find that

$$\begin{aligned} \frac{dE}{dV} &\geq t - \frac{1}{2-V} = t - \frac{1}{2 - t + t(\coth(t) - \frac{1}{t})^2} \\ &\geq t - \frac{1}{2 - t + t(1 - \frac{1}{t})^2} = 0, \end{aligned}$$

hence $\frac{dE}{dV} \geq 0$ also holds in this case.

All things considered, we conclude, as desired, that $E \geq 0$. ∎

We then turn to a closer study of Vajda's tight lower bound, L. Clearly, L is infinitely often differentiable. We shall show that L is in fact analytic. We start by a trivial but useful integral representation which allows easy exact calculation of the Taylor coefficients of L, at least in principle. Here the function $t \curvearrowright V(t)$ and its inverse $V \curvearrowright t(V)$ play the key role.

Fig. 3 Graph of the function $t \curvearrowright V(t)$.

Corollary 5 (Integral representation) For all $V_0 \in \left]-2;2\right[$ Vajda's tight lower bound L can be written as

$$L(V_0) = \int_0^{V_0} t(V)\, dV. \tag{7}$$

Proof: This follows as $L(0) = 0$ and as $\frac{dL}{dV} = t(V)$. ■

Now, using (7) in conjunction with either (1) or, simpler perhaps, (5), it is straightforward to calculate approximating Taylor polynomials of any degree, and we get

$$\begin{aligned} L(V) = {} & \frac{1}{2}V^2 + \frac{1}{36}V^4 + \frac{1}{270}V^6 + \frac{221}{340\,200}V^8 \\ & + \frac{299}{2296\,350}V^{10} + \frac{5983}{212\,182\,740}V^{12} \\ & + \frac{9953\,639}{1551\,586\,286\,250}V^{14} + \frac{24\,080\,603}{15\,959\,173\,230\,000}V^{16} \\ & + \frac{258\,692\,351}{712\,178\,105\,388\,750}V^{18} \\ & + \frac{125\,041\,974\,165\,263}{1406\,587\,367\,048\,050\,687\,500}V^{20} \\ & + \frac{195\,059\,968\,637\,159}{8861\,500\,412\,402\,719\,331\,250}V^{22} \\ & + \frac{79\,414\,742\,287\,586\,653}{14\,452\,301\,581\,682\,253\,163\,875\,000}V^{24} \\ & + \frac{12\,332\,430\,212\,594\,640\,377}{8942\,361\,603\,665\,894\,145\,147\,656\,250}V^{26} \\ & + \frac{38\,690\,559\,172\,885\,033\,903}{111\,435\,583\,061\,067\,296\,270\,301\,562\,500}V^{28} \\ & + \frac{1102\,997\,556\,766\,204\,706\,333}{12\,603\,364\,444\,206\,711\,208\,171\,106\,718\,750}V^{30} \\ & + O\left(V^{32}\right). \end{aligned} \tag{8}$$

We see that the first 3 coefficients are the same as the ones found in the lower bounding polynomials known from the literature [**?**], [**?**].

Theorem 6: Vajda's tight lower bound $L = L(V)$ is analytic and the radius of convergence r, for the power series expansion around $V = 0$ is $r \approx 1.8285$.

Fig. 4 Radius of convergence.

Proof: By Corollary 5 we have to show that $V \curvearrowright t(V)$ is analytic and that the radius of convergence for the power expansion centred at $V = 0$ is approximately 1.8285.

The function $t \curvearrowright V(t)$ has a unique holomorfic continuation as a function from $\mathbb{C} \backslash \{ni\pi, n \in \mathbb{Z}\backslash 0\}$ into $\mathbb{C}$. The derivative is

$$\frac{dV(z)}{dz} = \frac{2z^3\cosh(z) - 3z^2\sinh(z) + \sinh^3(z)}{z^2 \cdot \sinh^3 z}.$$

The derivative has no zeroes in a neighbourhood of the real axis. Let z_0 be a solution of $\frac{dV(z)}{dz} = 0$ such that $|\mathrm{Im}\, z_0|$ is minimal. We see that $\overline{z_0}$, $-z_0$ and $-\overline{z_0}$ are also solutions to this equation. Therefore we may assume that $\mathrm{Re}\, z_0 \geq 0$ and $\mathrm{Im}\, z_0 \geq 0$.

By Lemma 12 which is proved in the Appendix, $z_0 \approx 3.0682 + 2.8568i$ and $|V(z_0)| \approx 1.8285$.

We will show that $V \curvearrowright t(V)$ has a holomorfic continuation to $D = \{z \mid |z| < |V(z_0)|\}$. Let U be the set $\{z \mid -\mathrm{Im}\, z_0 < \mathrm{Im}\, z < \mathrm{Im}\, z_0\}$.

By a careful inspection, cf. the proof of Lemma 13 in the appendix, we see that the image of the boundary of U under the mapping V has no points in D. This implies that $D \subseteq V(U)$ because $V(t) \to \pm 2$ for $\mathrm{Re}\, t \to \pm\infty$.

By our choice of z_0, $\frac{dV(z)}{dz} \neq 0$ on U and therefore t is locally conformal as a complex function on D. Hence t can be continued from a neighbourhood of a zero to D by a holomorfic and continuous extension. This completes the proof. ■

We see that all the coefficients in (8) are positive, but the coefficient of V^{62} is $-3.263 \times 10^{-21} < 0$. Actually this is the first negative coefficient but there are infinitely many - otherwise the radius of convergence would be 2. The power series expansion of $L(V)$ can be used to suggest more terms in the lower bounding polynomials for $L(V)$.

Theorem 7: $L(V) \geq \frac{1}{2}V^2 + \frac{1}{36}V^4 + \frac{1}{270}V^6 + \frac{221}{340\,200}V^8$ and the constants are best possible.

The power expansion (8) implies that if the inequality is satisfied then the constants are best possible. The proof of Theorem 7 is based on a special expansion of D and will be outlined in the next section.

III. The Kambo–Kotz expansion

We shall now work with the parametrization (ρ, V) where

$$\rho = \frac{\frac{1}{2} - p_1}{\frac{1}{2} - q_2} \tag{9}$$

in order to characterize P and Q. Denote by Ω the subset of the (ρ, V)-plane defined by

$$\Omega = \{(-1, 0)\} \cup \Omega_1 \cup \Omega_2 \cup \Omega_3$$

with

$$\begin{aligned}
\Omega_1 &= \{(\rho, V) \mid \rho < -1, 0 < V \leq 1 + 1/\rho\},\\
\Omega_2 &= \{(\rho, V) \mid -1 < \rho \leq 1, 0 < V \leq 1 + \rho\},\\
\Omega_3 &= \{(\rho, V) \mid 1 < \rho, 0 < V \leq 1 + 1/\rho\}.
\end{aligned}$$

¿From Kambo and Kotz [?], we have (adapting notation etc. to our setting):

Theorem 8: Consider P and Q where $q_1 > \frac{1}{2}$, and define ρ by (9). Then $(\rho, V) \in \Omega$ and

$$D(P\|Q) = \sum_{\nu=1}^{\infty} \frac{f_\nu(\rho)}{2\nu(2\nu-1)} V^{2\nu},$$

where f_ν, $\nu \geq 1$, are rational functions defined by

$$f_\nu(\rho) = \frac{\rho^{2\nu} + 2\nu\rho + 2\nu - 1}{(\rho+1)^{2\nu}}, \quad \rho \neq -1.$$

We shall refer to the functions f_ν as the *Kambo–Kotz functions*. Let us state some basic properties of these functions, taken from [?]:

Lemma 9: The Kambo-Kotz functions f_ν, $\nu \geq 1$, are everywhere positive, f_1 is the constant function 1 and all other functions f_ν assume their minimal value at a uniquely determined point ρ_ν which is the only stationary point of f_ν. We have $\rho_2 = 2$, $1 < \rho_\nu < 2$ for $\nu \geq 3$ and $\rho_\nu \to 1$ as $\nu \to \infty$. For $\nu \geq 2$, f_ν is strictly increasing in the two intervals $]-\infty, -1[$ and $[2, \infty[$ and f_ν is strictly decreasing in $]-1, 1]$. Furthermore, f_ν is strictly convex in $[1, 2]$ and, finally, $f_\nu(\rho) \to 1$ for $\rho \to \pm\infty$.

In the sequel, we shall write $D(\rho, V)$ in place of $D(P\|Q)$. Motivated by the lemma, we define the *critical domain* as the set

$$\begin{aligned}
\Omega^* &= \{(\rho, V) \in \Omega \mid 1 \leq \rho \leq 2\}\\
&= \{(\rho, V) \in \Omega \mid 1 \leq \rho \leq 2, 0 < V < 1 + 1/\rho\}.
\end{aligned}$$

We then realize that in the search for lower bounds of D in terms of V we may restrict the attention to the critical domain. In particular:

Corollary 10: For each $\nu_0 \geq 1$

$$c_{\nu_0}^{\max} = \inf\left\{V^{-\nu_0}\left(D(\rho, V) - \sum_{\nu<\nu_0} c_\nu^{\max} V^\nu\right) \mid (\rho, V) \in \Omega^*\right\}.$$

We may use this result as a basis for a proof of Theorem 7. First, we note that

$$D - \frac{1}{2}V^2 - \frac{1}{36}V^4 - \frac{1}{270}V^6 =$$
$$\frac{1}{18}\left(\frac{2-\rho}{1+\rho}\right)^2 V^4 - \frac{(2-\rho)q(\rho)}{270(1+\rho)^4}V^6 + \sum_{\nu=4}^{\infty} \frac{f_\nu(\rho)}{2\nu(2\nu-1)}V^{2\nu},$$

where $q(\rho) = 8\rho^3 - 6\rho^2 + 9\rho - 22$. We may lower bound this expression by only including the two first terms in the infinite sum. Doing this and appealing to Corollary 10, we find that

$$c_8^{\max} \geq \inf\left\{\begin{array}{l} \frac{1}{18}\left(\frac{2-\rho}{1+\rho}\right)^2 V^{-4} - \frac{(2-\rho)q(\rho)}{270(1+\rho)^4}V^{-2} \\ +\frac{f_4(\rho)}{56} + \frac{f_5(\rho)}{90}V^2 \end{array}\right\}$$

where we may restrict attention to values of ρ and V with $1 \leq \rho \leq 2$. Theorem 7 now follows from the following lemma:

Lemma 11: With $x = \frac{2-\rho}{V^2}$ the following inequality holds for all $(\rho, V) \in \Omega^*$:

$$\frac{1}{18(1+\rho)^2}\left(x^2 - \frac{q(\rho)}{15(1+\rho)^2}x\right) + \frac{f_4(\rho)}{56} + \frac{f_5(\rho)}{90}V^2 \geq \frac{221}{340200}.$$

The proof of this lemma is elementary but somewhat technical. The details are given in the appendix.

IV. Discussion

Problem solved?

In a sense, Theorem 1 completely settles the problem of lower bounding D in terms of V, research initiated by Pinsker [?] and clearly formulated by Vajda [?]. On the other hand the solution provided is rather complex and raises a number of new problems, not solved in the present paper. We shall comment on the wider perspectives below.

The power expansion of L and lower bounding polynomials

In the power expansion of L the first many coefficients are positive. Therefore one should expect that more of the higher degree Taylor polynomials, denote them by T_ν, $\nu \geq$

0, actually lower bound L. On the other hand we know that $L^{(62)}(0) < 0$ so T_{60} definitely does not lower bound L.

Theorem 7 states that $L \geq T_8$. The proof of this fact was complicated enough and one should expect that the eventual proofs that higher degree polynomials lower bound L will be more and more complicated. The fact that the power series of L has 1.8285 as radius of convergence indicates that there might not exist simple proofs for such results. Possibly, our main result, Theorem 1, may be more helpful than shown here for such proofs. It is a bit discouraging that we had to recourse to a second type of expansion (the simple Kambo-Kotz expansion discussed in Section 3) in order to prove Theorem 8. Of course, the positivity of the coefficient functions f_ν occurring there is, qualitatively speaking, what we need to compensate for the "catastrophe" of negative coefficients in the power expansion of L.

To be specific, let us formulate some, partly interrelated, questions concerning the problem of finding lower bounding polynomials to D, i.e. to Vajda's tight lower bound, L:

Q1: Does $L \geq T_{58}$ hold or, equivalently, is $\sum_{i \leq 59} c_i^{\max} V^i = T_{58}$?

Q2: What is the largest ν for which $L \geq T_\nu$ holds?

Q3: Are the best constants $c_\nu^{\max}$ eventually zero?

Q4: Do all best constants $c_\nu^{\max}$ with ν uneven vanish?

Other lower bounds

The classical Pinsker inequality is very useful in a number of situations (the literature on such applications will surely contain some hundred references). And in most cases refinements in terms of polynomials, though fascinating, are not really needed.

It was pointed out previously that the polynomial bounds are not precise when the total variation is large. As demonstrated by Vajda [?], there exist other interesting lower bounding functions which also give a good lower bound for $V \approx 2$. Using the exact parametrization given in this article it should be possible to find functions which give better uniform bounds than the ones found until now. We shall not formulate any precise question in this direction only point to a natural further request of such bounds, namely that they expand naturally to even functions of the argument V. Vajda's lower bound, cf. Corollary 4, does not fulfil this request.

Information diagrams

In Harremoës and Topsøe [?],[?] a notion of *information diagrams* was introduced. In the simplest case such a diagram gives the range of two quantities of interest. Actually, what we have determined is the *V/D-diagram*, i.e. the range of the map $(P,Q) \curvearrowright (V(P,Q), D(P \parallel Q))$. Indeed, the upper bound is easy to determine (∞) and therefore the V/D-diagram consists of the point $(0,0)$ and all points (x,y) with $0 < x \leq 2$ and $L(x) \leq y \leq \infty$ (with $L(2) = \infty$). The present method of proof is very different from the topological method in Harremoës and Topsøe [?],[?]. This is of course related to the very special feature that the relevant data reduction procedure reduces divergence but keeps total variation fixed, and reduces the problem to one involving distributions over a two-element set.

V. Appendix

Now we present some computations for Theorem 6 which estimates the radius of convergence of $t = t(V)$. We use interval arithmetic as implemented in Maple to provide a strict basis for our calculations. Let us introduce the following notation: By $[a+b\,i; c+d\,i]$ we denote the rectangle with corners $a+b\,i$ and $c+d\,i$.

The following lemma locates an exceptional point of V.

Lemma 12: Let z_0 be a solution to the equation $\frac{dV}{dt} = 0$ such that $\operatorname{Re} z_0 \geq 0$, $\operatorname{Im} z_0 \geq 0$, and $\operatorname{Im} z_0$ is minimal. Then $z_0 \in [3.068161 + 2.856781\,i; 3.068162 + 2.856782\,i]$ which implies $V(z_0) \in [1.823738 + .1319510\,i; 1.823745 + .1319586\,i]$ and $|V(z_0)| \in [1.828505; 1.828513]$.

Proof: If we equate the derivative $\frac{dV}{dt}$ to a zero and convert all hyperbolic functions to exponents, we obtain the following:

$$1 + \frac{-12\,{z_0}^2 + 8\,{z_0}^3}{e^{2\,z_0} - 1} + \frac{-12\,{z_0}^2 + 24\,{z_0}^3}{\left(e^{2\,z_0} - 1\right)^2} + \frac{16\,{z_0}^3}{\left(e^{2\,z_0} - 1\right)^3} = 0.$$

Now we have the following estimate for a module and a real part of z_0:

$$1 \geq \frac{12\,|z_0|^2 + 8\,|z_0|^3}{e^{2\,\operatorname{Re} z_0} - 1} + \frac{12\,|z_0|^2 + 24\,|z_0|^3}{\left(e^{2\,\operatorname{Re} z_0} - 1\right)^2} + \frac{16\,|z_0|^3}{\left(e^{2\,\operatorname{Re} z_0} - 1\right)^3}.$$

It is easy to see that for $\operatorname{Re} z_0 \geq 3.5$ and $\operatorname{Im} z_0 \leq 2.9$ the previous inequality does not hold. So to find z_0 we can split $[0; 3.5 + 2.9\,i]$ into a finite number of rectangles and find a zero of the derivative using interval computations. ■

Now we show that D is inscribed into the image of U.

Lemma 13: Consider the sets

$$\partial U = \{z \mid \operatorname{Im} z = \pm \operatorname{Im} z_0\},$$
$$D = \{z \mid |z| < |V(z_0)|\},$$

where z_0 is given by Lemma 12. Then $|V(z)| \geq |V(z_0)|$ for $z \in \partial U$.

Proof: We only have to show that $|V(t + i\ \operatorname{Im} z_0)| \geq |V(z_0)|$ (that $|V(t - i\ \operatorname{Im} z_0)| \geq |V(z_0)|$ follows by conjugation). By symmetry we may assume that $t \geq 0$.

The inequality $|V(t + i\ \operatorname{Im} z_0)| \geq |V(z_0)|$ is obviously satisfied for $t \geq 6$ due to the following decomposition:

$$|2 - V| = \left| \frac{1}{t} + \frac{4\,t}{(e^{2t} - 1)^2} + \frac{-4 + 4\,t}{e^{2t} - 1} \right|,$$

so we just have to prove that $|V(t + i\ \operatorname{Im} z_0)|$ achieve a local minimum in a neighbourhood ω of z_0. Let $\omega = (3.00, 3.13)$. Taking an interval representation of z_0 from Lemma 12 we obtain the following:

$$\operatorname{Re} \frac{d^2V(z)/dz^2(z_0)}{V(z_0)} = [.018655; .018658]$$
$$\operatorname{Re} \frac{d^3V(z)/dz^3(t + i\ \operatorname{Im}(z_0))}{V(z_0)} = [-.803832; .645122]$$

for $t \in \omega$. Using Taylor's formula around z_0 we get

$$
\begin{aligned}
&|V(t+i\,\mathrm{Im}\,z_0)| \\
&= \left| \begin{array}{c} V(z_0) + (t-t_0)\frac{dV(z)}{dz}(z_0) \\ +\frac{(t-t_0)^2}{2}\frac{d^2V(z)}{dz^2}(z_0) + \frac{(t-t_0)^3}{6}\frac{d^3V(z)}{dz^3}\left(\tilde{t}+i\,\mathrm{Im}\,z_0\right) \end{array} \right| \\
&\geq |V(z_0)| \times \left| 1 + \frac{(t-t_0)^2}{2} \left(\begin{array}{c} \frac{\frac{d^2V(z)}{dz^2}(z_0)}{V(z_0)} - \\ \frac{(t-t_0)}{3}\frac{\frac{d^3V(z)}{dz^3}(\tilde{t}+i\,\mathrm{Im}(z_0))}{V(z_0)} \end{array} \right) \right|
\end{aligned}
$$

which is greater than $|V(z_0)|$ in ω, because a real part of the expression in brackets in $[.002086; .036922]$.

The proof could be completed by splitting a rest of a segment $[0;6]\backslash\omega$ to a bunch of smaller segments and estimating $|V(z)|$ there by means of interval arithmetic. ∎

We end this technical appendix by giving the details in the proof of Lemma 11.

Proof: Recall that we have put $q(\rho) = 8\rho^3 - 6\rho^2 + 9\rho - 22$ and $x = \frac{2-\rho}{V^2}$. Now note that

$$
\begin{aligned}
&x^2 - \frac{q(\rho)}{15(1+\rho)^2}x \\
&= \left(x - \frac{2}{15}\right)^2 - \frac{4}{225} + \frac{(2-\rho)(8\rho^2+6\rho+13)}{15(1+\rho)^2}x \\
&\geq -\frac{4}{225} + \frac{(2-\rho)(8\rho^2+6\rho+13)}{15(1+\rho)^2}x,
\end{aligned}
$$

hence it suffices to show that

$$
\begin{aligned}
&\left(\frac{f_4(\rho)}{56} - \frac{2}{2025(1+\rho)^2}\right) \\
&\quad + \left(\frac{f_5(\rho)}{90}V^2 + \frac{(2-\rho)^2(8\rho^2+6\rho+13)}{270(1+\rho)^4}\frac{1}{V^2}\right) \\
&\qquad\qquad \geq \frac{221}{340200}
\end{aligned}
$$

or that

$$
\begin{aligned}
&\left(\frac{f_4(\rho)}{56} - \frac{2}{2025(1+\rho)^2}\right) \\
&\quad + \frac{1}{45\cdot 3^{\frac{1}{2}}}\left(\frac{f_5(\rho)(2-\rho)^2(8\rho^2+6\rho+13)}{(1+\rho)^4}\right)^{\frac{1}{2}} \\
&\qquad\qquad \geq \frac{221}{340200}
\end{aligned}
$$

which may be further transformed into

$$
\begin{aligned}
&\frac{1}{56(1+\rho)^6}\sum_{\nu=0}^{6}(-1)^\nu(7-\nu)\rho^\nu - \frac{2}{2025(1+\rho)^2} \\
&+\frac{2-\rho}{45\cdot 3^{\frac{1}{2}}(1+\rho)^6}\left(\sum_{\nu=0}^{8}(-1)^\nu(9-\nu)\rho^\nu\cdot\left(8\rho^2+6\rho+13\right)\right)^{\frac{1}{2}} \\
&\qquad\qquad \geq \frac{221}{340200}.
\end{aligned}
$$

Numerical evidence shows that the function on the left hand side is decreasing in the interval $[1,2]$ and as the function value at 2 equals the fraction on the right hand side, the proof could be terminated here. However, we continue with a more formal analytical proof by rearranging terms and factoring out $2-\rho$, thus arriving at the inequality

$$
\begin{aligned}
&5854\rho^5 - 1768\rho^4 + 11038\rho^3 - 7988\rho^2 + 9068\rho - 20984 \\
&\qquad \leq \frac{7560}{3^{\frac{1}{2}}}(\cdots)^{\frac{1}{2}},
\end{aligned}
$$

where the last square root is the one occurring in the previous expression. In fact, this inequality holds for all $\rho \in$ and even in the stronger form that the square of the left-hand side is dominated by the square of the right-hand side. Working out the details, this amounts to the inequality

$$\sum_{\nu=0}^{10}\alpha_\nu\rho^\nu \geq 0$$

with the alphas defined in Table 1.

ν	α_ν	ν	α_ν
0	1341496608	6	593673633
1	−428995632	7	−472120596
2	1330064244	8	257940954
3	−972755136	9	−127359192
4	960869448	10	88605213
5	−659621844		

Table 1

Defining $\beta_{10}, \beta_8, \beta_6, \beta_4, \beta_2$ and β_0 recursively by $\beta_{10} = \alpha_{10}$ and

$$\beta_{2\nu} = \alpha_{2\nu} - \frac{\alpha_{2\nu+1}^2}{4\beta_{2\nu+2}},$$

one finds that all β's are positive and then that

$$
\begin{aligned}
\sum_{\nu=0}^{10}\alpha_\nu\rho^\nu &\geq \beta_8\rho^8 + \alpha_7\rho^7 + \cdots + \alpha_0 \\
&\geq \beta_6\rho^6 + \alpha_5\rho^5 + \cdots + \alpha_0 \\
&\geq \cdots \geq \beta_0.
\end{aligned}
$$

One finds that $\beta_8 \approx 2.1\cdot 10^8, \cdots, \beta_0 \approx 1.3\cdot 10^9$. Backtracing our steps, we see that $c_8^{\max} \geq 221/340200$. The opposite inequality was established before. Thus, we have finally determined $c_8^{\max}$ and thereby finished the proof of Theorem 7. ∎