Acta Math. Hungar. **116**(2007), 127-131.

Bregman divergence as relative operator entropy

Dedicated to Professor Slava Belavkin on the occasion of his 60th birthday

Dénes Petz [1]

Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
H-1053 Budapest, Reáltanoda u. 13-15, Hungary

In this paper the Bregman operator divergence is introduced for density matrices by differentiation of the matrix-valued function $x \mapsto x \log x$. This quantity is compared with the relative operator entropy of Fujii and Kamei. It turns out that the trace is the usual Umegaki's relative entropy which is the only intersection of the classes of quasi-entropies and Bregman divergences.

MSC2000: 94A17 (47A63, 47B18)

Keywords: Bregman divergence, f-divergence, quantum relative entropy.

Let C be a convex set in a Banach space and $\mathcal{H}$ be a Hilbert space. For an operator-valued smooth functional $\Psi : C \to B(\mathcal{H})$, we define the **Bregman divergence** of $x, y \in C$ as

$$D_\Psi(x,y) := \Psi(x) - \Psi(y) - \lim_{t \to +0} t^{-1}\Big(\Psi(y + t(x-y)) - \Psi(y)\Big).$$

Recall that originally Lev Bregman introduced this concept for real-valued convex functions Ψ [2]. On the self-adjoint operators, there is a standard partial ordering, $A \le B$, if $\langle \xi, A\xi \rangle \le \langle \xi, B\xi \rangle$ for any vector ξ in the Hilbert space. Hence the convexity of $\Psi : C \to B(\mathcal{H})$ makes sense and $D_\Psi(x,y) \ge 0$ remains true.

We are basically interested in the case when C is the set of positive semidefinite matrices of trace 1 and $\Psi(x) = \eta(x)$, where $\eta(t) = t \log t$ is a common function defined on $\mathbb{R}^+$ and $\eta(x)$ is defined by the functional calculus. However, we also consider $\Psi(x) = f(x)$ for a general smooth function f.

Remember that positive semidefinite matrices of trace 1 are known as density matrices and play central role in quantum information theory [6, 7]. Concerning Umegaki's quantum relative entropy Chap. 1 of [7] is our main reference.

---

[1] E-mail: petz@math.bme.hu.

Lemma 1 *Let* $f : \mathbb{R}^+ \to \mathbb{R}$ *be a continuously differentiable function and let* y *be a positive definite matrix.*

If y *and* $z = z^*$ *commute, then*

$$\lim_{t \to +0} t^{-1}\Big(f(y+tz) - f(y)\Big) = f'(y)z\,.$$

If w *is skew-adjoint (that is,* $w^* = -w$*), then*

$$\lim_{t \to +0} t^{-1}\Big(f(y+t[y,w]) - f(y)\Big) = [f(y), w]\,.$$

Both differentiation rules are verified by simple computation if $f(t) = t^k$. For a general f one can benefit from a polynomial approximation.

Recall that $\langle t, u\rangle = \operatorname{Tr} t^* u$ is an inner product on matrices.

Lemma 2 *Let* t *and* u *be self-adjoint matrices. Then* $t = z + z^\perp$*, where* z *commutes with* u *and* $z^\perp = [u, w]$ *for some skew-adjoint* w*.*

Proof: For the linear mapping $A : r \mapsto \mathrm{i}(ur - ru) \equiv \mathrm{i}[u, r]$, we have $\operatorname{Rng}(A)^\perp = \operatorname{Ker}(A)$. This means that an self-adjoint matrix t is a sum $z + z^\perp$, where z commutes with u and $z^\perp = \mathrm{i}[u, r]$ is a commutator.

If $u = \sum_i \lambda_i p_i$ is the spectral decomposition, then the pinching $E(r) = \sum_i p_i r p_i$ is the conditional expectation onto the subalgebra generated by u. Since E is the orthogonal projection onto the subalgebra, we have $z = E(t)$. $\square$

In order to compute $D_\Psi(x, y)$, we decompose

$$x - y = z + z^\perp,$$

where z commutes with y and $\langle z, z^\perp\rangle := \operatorname{Tr} z z^\perp = 0$ is the orthogonality relation. Note that this orthogonal decomposition is unique and $z^\perp = [y, w]$ for some skew-adjoint w. If E denotes the conditional expectation onto the subalgebra generated by y, then $z = E(x - y) = Ex - y$.

The differentiation rules in Lemma 1 give the following.

Theorem 1 *Let* $\Psi(x) = f(x)$ *for a continuously differentiable function* $f : [0, 1] \to \mathbb{R}$*. If* x *and* y *are invertible density matrices, then*

$$D_\Psi(x, y) = f(x) - f(y) - zf'(y) - [f(y), w]\,.$$

f is called operator convex if $f(\lambda A + (1-\lambda)B) \le \lambda f(A) + (1-\lambda)f(B)$ for any real number $0 \le \lambda \le 1$ and for self-adjoint operators A and B with spectrum in $[0, 1]$.

When f is operator convex, then Ψ is convex and $D_\Psi(x,y) \geq 0$. The function η is operator convex and in the case $f(t) = \eta(t)$, we have

$$D_\Psi(x,y) = (y - Ex) + x\log x - (\log y)Ex - [\eta(y), w]. \tag{1}$$

This operator-valued divergence is positive and can be called **Bregman operator divergence**. Note that for commuting x and y, we have

$$D_\Psi(x,y) = y - x + x(\log x - \log y).$$

Example 1 Let

$$y = \begin{bmatrix} \lambda & 0 \\ 0 & 1-\lambda \end{bmatrix} \qquad \text{and} \qquad x = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Then

$$w = \frac{1}{2(2\lambda - 1)}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

and we have

$$D_\Psi(x,y) = \begin{bmatrix} \lambda - \frac{1}{2} - \frac{\log\lambda}{2} & \frac{1}{2(2\lambda-1)}\Big(\eta(\lambda) - \eta(1-\lambda)\Big) \\ \frac{1}{2(2\lambda-1)}\Big(\eta(\lambda) - \eta(1-\lambda)\Big) & (1-\lambda) - \frac{1}{2} - \frac{\log(1-\lambda)}{2} \end{bmatrix}$$

for $\Psi(x) = \eta(x)$. □

Next we consider a number valued functional $\psi(x) := \mathrm{Tr}\,\eta(x)$ on the convex set of density matrices.

Theorem 2 *Let $\psi(x) = Tr\,\eta(x)$ for a density matrix x. If x and y are invertible density matrices, then*

$$D_\psi(x,y) = Tr\,x(\log x - \log y)$$

is Umegaki's relative entropy.

Proof: We continue to use the above notation and need to compute $\mathrm{Tr}\,D_\Psi(x,y)$. Since the trace wanishes on a commutator and $\mathrm{Tr}\,(\log y)Ex = \mathrm{Tr}\,x\log y$, our statement follows from Theorem 1.

Note that the statement can be proved directly, without the use of the operator-valued version. □

Let $\Psi(x) = \eta(x)$. It follows that in the operator inequality

$$D_\Psi(x,y) \geq 0$$

the equality holds if and only if $x = y$. Indeed, from $D_\Psi(x,y) = 0$, we have $\mathrm{Tr}\,D_\Psi(x,y) = \mathrm{Tr}\,x(\log x - \log y) = 0$. Then the inequality

$$\mathrm{Tr}\,(x-y)^2 \leq 2\mathrm{Tr}\,x(\log x - \log y)$$

gives $x = y$. (Concerning this inequality, see [7], Prop. 1.1.)

Another relative operator entropy was introduced by Fujii and Kamei in 1989 [4]. In order to have a positive operator we write it in a slightly modified form as

$$S_{FK}(x,y) := y - x + x^{1/2}\big(\log(x^{1/2}y^{-1}x^{1/2}\big)x^{1/2}.$$

It was proven that

$$\operatorname{Tr} S_{FK}(x,y) \geq \operatorname{Tr} D_\Psi(x,y)$$

by Hiai and Petz. The left-hand-side is the relative entropy introduced by Belavkin and Staszewski [1], see Prop. 7.11 in [7].

Example 2 Let

$$y = \begin{bmatrix} \lambda & 0 \\ 0 & 1-\lambda \end{bmatrix} \qquad \text{and} \qquad x = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Then we have

$$S_{FK}(x,y) = \begin{bmatrix} \lambda & 0 \\ 0 & 1-\lambda \end{bmatrix} + \frac{1}{2}\log\left(\frac{\frac{1}{\lambda}+\frac{1}{1-\lambda}}{4}\right)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

From this computation and from Example 1, one can conclude that $S_{FK}(x,y) \leq D_\Psi(x,y)$ does not hold. (There is no inequality between the (1,1)-elements of the two matrices.) □

It was proved in [3] that the Kullback-Leibler divergence is the only Bregman divergence which is an f-divergence at the same time. Csiszár's f-divergence was generalized to the non-commutative algebraic setting in [8, 9]. One can characterize Umegaki's relative entropy as the only Bregman divergence in the sense of this paper which is also a quasi-entropy in the sense of [8].

Acknowledgement. The author is grateful to Prof. Imre Csiszár for proposing the problem treated in the paper and for useful discussions. The work was partially supported by the grant OTKA T032662.

References

[1] V.P. Belavkin and P. Staszewski, C*-algebraic generalization of relative entropy and entropy, Ann. Inst. Henri Poincaré, Sec. A **37**(1982), 51–58.

[2] L.M. Bregman, The relaxation method to find the common points of convex sets and its application to the solution of problems in convex programming, USSR Compt. Math. and Math. Phys. **7**(1967), 200–217.

[3] I. Csiszár, Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems, Ann. Statist. **19**(1991), 2032–2066.

[4] J.I. Fujii and E. Kamei, Relative operator entropy in noncommutative information theory, Math. Japon. **34**(1989), 341–348.

[5] T. Furuta,Two reverse inequalities associated with Tsallis relative operator entropy via generalized Kantorovich constant and their applications, Linear Algebra Appl., **412**(2006), 526–537.

[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.

[7] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Heidelberg, 1993, 2nd edition 2004.

[8] D. Petz, Quasi-entropies for states of a von Neumann algebra, Publ. RIMS. Kyoto Univ. **21**(1985), 781–800.

[9] D. Petz, Quasi-entropies for finite quantum systems, Rep. Math. Phys., **23**(1986), 57-65.