On a problem of Bauschke and Borwein

Dániel Virosztek

Department of Mathematical Analysis,
Budapest University of Technology and Economics,
Egry József u. 1., Budapest, 1111 Hungary

Abstract

Consider a differentiable convex function \( f : \mathbb{R}^n \supset \text{dom} f \to \mathbb{R} \). The induced spectral function \( F \) is given by \( F = f \circ \lambda \), where \( \lambda : \mathbb{M}_n^{sa} \to \mathbb{R}^n \) is the eigenvalue map. Let us denote by \( D_f \) and \( D_F \) the Bregman distances associated with \( f \) and \( F \), respectively. In the paper Joint and separate convexity of the Bregman distance written by H. Bauschke and J. Borwein [BB01] the following open problem has been suggested. Is \( D_f \) jointly convex if and only if \( D_F \) is? In this short note we provide a negative answer to this question.

Keywords: Bregman divergence, joint convexity

Mathematics subject classification (2010): 15B57, 46N10

Introduction. The Bregman distance (or Bregman divergence) was introduced by Lev Bregman [BR67] for differentiable convex functions \( f : \mathbb{R}^n \supset \text{dom} f \to \mathbb{R} \) with nonempty open convex domain as follows:

\[
D_f(x, y) = f(x) - f(y) - d_f[y](x - y),
\]

where \( x, y \in \text{dom} f \) and \( d\phi[a] \) denotes the Fréchet derivative of the function \( \phi \) at the point \( a \). We say that the Bregman distance \( D_f \) is jointly convex if \( (x, y) \mapsto D_f(x, y) \) is convex on \( \text{dom} f \times \text{dom} f \).

Throughout this note \( \mathbb{R}^+ (\mathbb{R}^{++}) \) denotes the set of all nonnegative (positive) numbers and \( \mathbb{M}_n (\mathbb{M}_n^{sa}, \mathbb{M}_n^+, \mathbb{M}_n^{++}) \) denotes the set of \( n \times n \) complex (self-adjoint, positive semidefinite, positive definite) matrices.

Let \( \lambda : \mathbb{M}_n^{sa} \to \mathbb{R}^n \) be the eigenvalue map which collects the eigenvalues of a self-adjoint matrix ordered decreasingly. The spectral function induced by \( f \) is defined by

\[
F = f \circ \lambda
\]

and the domain of \( F \) is the preimage of \( \text{dom} f \), i.e., \( \text{dom} F = \lambda^{-1}(\text{dom} f) \subset \mathbb{M}_n^{sa} \). (Remark that \( \mathbb{M}_n^{sa} \) can be canonically identified with \( \mathbb{R}^{n^2} \).)

\[\text{1E-mail: virosz@math.bme.hu}\]
The question of Bauschke and Borwein was: “Is $D_f$ jointly convex if and only if $D_F$ is?” [BB01]. ($D_F$ denotes the Bregman divergence induced by the function $\mathbb{R}^n \ni M \mapsto Mx \in \text{dom} F \to \mathbb{R}$.) We show that the joint convexity of $D_f$ does not imply the joint convexity of $D_F$, but the converse is true under some assumptions.

**Proposition 1.** The function $h : \mathbb{R}^+ \to \mathbb{R}$, $x \mapsto h(x) := \frac{1}{2-e^{-x}}$ is not operator convex.

**Proof:** $h$ is operator convex if and only if $g(x) = 1 - h(x) = 1 - \frac{1}{2-e^{-x}}$ is operator concave. $g$ is an $\mathbb{R}^+ \to \mathbb{R}^+$ map, hence the operator concavity is equivalent to the operator monotonicity [BH96, Thm. V.2.5]. For $0 < x < y$ the divided difference matrix is the following:

$$D = \begin{pmatrix} \frac{g'(x)}{g(x) - g(y)} & \frac{g(x) - g(y)}{x-y} \\ \frac{g'(y)}{g(x) - g(y)} & \frac{g(x) - g(y)}{x-y} \end{pmatrix} = \begin{pmatrix} \frac{e^{-x}}{(2-e^{-x})^2} & \frac{e^{-y} - e^{-x}}{(2-e^{-x})(2-e^{-y})(x-y)} \\ \frac{e^{-y}}{(2-e^{-x})(2-e^{-y})(x-y)} & \frac{e^{-y}}{(2-e^{-y})^2} \end{pmatrix}.$$  

The determinant is

$$\text{Det}(D) = \frac{1}{(2-e^{-x})^2(2-e^{-y})^2} \left( e^{-x}e^{-y} - \left( \frac{e^{-x} - e^{-y}}{y-x} \right)^2 \right).$$  

(2)

The logarithmic mean of two different positive numbers $a$ and $b$ is $L(a, b) = \frac{a-b}{\log a - \log b}$ and this is larger than the geometric mean $G(a, b) = \sqrt{ab}$ [NE95]. Therefore, the expression $e^{-x}e^{-y} - \left( \frac{e^{-x} - e^{-y}}{y-x} \right)^2$ is negative by the inequality of the geometric and the logarithmic mean: $G(e^{-x}, e^{-y}) < L(e^{-x}, e^{-y})$. It follows that the determinant of the divided difference matrix (2) is negative, hence by [HP14, Thm 4.5], $g$ is not operator monotone, thus the proof is complete. 

\[\square\]

**Remark.** A standard continuity argument shows that if $h$ is not operator convex on $\mathbb{R}^+$ — that is, $h(\alpha A + (1 - \alpha)B) \not\leq \alpha h(A) + (1 - \alpha)h(B)$ holds for some $A, B \in \mathbb{M}^+_n$ and $\alpha \in (0, 1)$ — then it is not operator convex on the smaller set $\mathbb{R}^{++}$ either (which means that we have $h(\alpha A + (1 - \alpha)B) \not\leq \alpha h(A) + (1 - \alpha)h(B)$ for some invertible matrices $A, B \in \mathbb{M}^+_n$ and $\alpha \in (0, 1)$).

**The counterexample.** Consider the function $h : \mathbb{R}^{++} \to \mathbb{R}$, $h(x) = \frac{1}{2-e^{-x}}$. Let $\tilde{h} : \mathbb{R}^{++} \to \mathbb{R}$ be a function such that $\tilde{h}'' = h$. (For example, $m(x) := \int_0^x h(t)dt$ for $x > 0$ and $\tilde{f}(x) := \int_0^x m(t)dt$ for $x > 0$.) Now we can define the function

$$f : \mathbb{R}^n \ni \text{dom} f \to \mathbb{R}, \ x = (x_1, x_2, \ldots, x_n) \mapsto f(x) := \sum_{j=1}^n \tilde{f}(x_j),$$

where $\text{dom} f = \{x \in \mathbb{R}^n | x_j > 0 \forall j\}$ is a nonempty open convex set in $\mathbb{R}^n$. $f$ is a separable symmetric function, hence the inverse of the second derivative matrix (Hessian) of $f$ is clearly

$$\text{Diag}(2 - e^{-x_1}, \ldots, 2 - e^{-x_n}).$$
This matrix valued function is concave with respect to the Löwner ordering\(^2\) on \(\text{dom} \, f\) by the concavity of the scalar function \(x \mapsto 2 - e^{-x}\). By [BB01, Corollary 6.2], it follows that \(D_f\) is jointly convex.

Observe that the trace function associated with \(\tilde{f}\) coincides with the spectral function induced by \(f\), that is, \(\text{Tr} \, \tilde{f}(\cdot) = f \circ \lambda =: F\) and the domain of \(F\) is

\[
\lambda^{-1}(\text{dom} \, f = \{x \in \mathbb{R}^n | x_j > 0 \forall j\}) = M^{++}_n.
\]

For positive definite matrices \(X\) and \(Y\) the Bregman divergence associated with the function \(F\) is the following:

\[
D_F(X, Y) = F(X) - F(Y) - dF[Y](X - Y) = \text{Tr} \, \tilde{f}(X) - \text{Tr} \, \tilde{f}(Y) - d(\text{Tr} \, \tilde{f})[Y](X - Y). \quad (3)
\]

By the linearity of the trace, \(d(\text{Tr} \, \tilde{f})[Y] = \text{Tr} \, \left( d\tilde{f}[Y] \right)\) for any \(Y \in M^{++}_n\), where \(d\tilde{f}[Y]\) denotes the Fréchet derivative of the standard matrix function\(^3\) \(\tilde{f} : M^{++}_n \rightarrow M^{sa}_n\) at \(Y\). Therefore, (3) can be written as

\[
D_F(X, Y) = \text{Tr} \left( \tilde{f}(X) - \tilde{f}(Y) - d\tilde{f}[Y](X - Y) \right),
\]

so it is equal to the Bregman \(\tilde{f}\)-divergence \(H_{\tilde{f}}(X, Y)\) defined in [PV14]. The solution of the suggested problem is based substantially on our recent work with József Pitrik [PV14], where the main theorem is the following.

**Theorem** ([PV14]). Let \(k \in C^2((0, \infty))\) be a convex function. The following conditions are equivalent.

(A) \(k''\) is operator convex and numerically non-increasing.

(B) The Bregman \(k\)-divergence

\[
H_k : M^{++}_n \times M^{++}_n \rightarrow \mathbb{R}^+; \quad (X, Y) \mapsto H_k(X, Y) = \text{Tr} \left( k(X) - k(Y) - d_k[Y](X - Y) \right)
\]

is jointly convex.

By this theorem, the fact that \(\tilde{f}''\) is not operator convex on \(\mathbb{R}^{++}\) (Proposition 1) means that the Bregman divergence \(D_f\) (which was shown to be equal to \(H_{\tilde{f}}\)) is not jointly convex on \(M^{++}_n \times M^{++}_n\).

So the joint convexity of \(D_f\) does not imply the joint convexity of \(D_F\), hence we can give a negative answer to the Open Problem 7.6 of [BB01].

---

\(^2\) \(A \leq B\) if and only if \(B - A\) is positive semidefinite for \(A, B \in M^{sa}_n\)

\(^3\) If \(l\) is an \(\mathbb{R} \supset I \rightarrow \mathbb{R}\) function then the corresponding **standard matrix function** is the following map:

\[
l : \{A \in M^{sa}_n : \sigma(A) \subset I\} \rightarrow M^{sa}_n, \quad A = \sum \lambda_j P_j \mapsto l(A) := \sum l(\lambda_j) P_j,
\]

where \(\sigma(A)\) is the spectrum and \(\sum \lambda_j P_j\) is the spectral decomposition of \(A\).
The converse statement. On the other hand, the joint convexity of $D_F$ implies the joint convexity of $D_f$ (on a restricted domain). Let $\{|\varphi_j\rangle\}_{j=1}^n$ be an orthonormal basis of $\mathbb{C}^n$ (with respect to the Euclidean inner product) and let us denote by $P_j$’s the corresponding orthoprojections, that is, $P_j := |\varphi_j\rangle \langle \varphi_j|$. Then the map

$$i : \mathbb{R}^n \to M_n^{sa} : x = (x_1, x_2, \ldots, x_n) \mapsto i(x) := \sum_{j=1}^n x_j P_j$$

is an isometric linear embedding — with respect to the metric defined by the Hilbert-Schmidt inner product $\langle X, Y \rangle = \text{Tr} XY$ on $M_n^{sa}$ — and $\lambda \circ i$ is the identity map of $\text{ran} \lambda = \{x \in \mathbb{R}^n | x_1 \geq x_2 \geq \cdots \geq x_n\}$. Therefore, it is easy to check that for any $x, y \in \text{int} (\text{dom} f \cap \text{ran} \lambda)$ we have

$$D_f(x, y) = D_F(i(x), i(y)).$$

Indeed,

$$D_f(x, y) = f(x) - f(y) - df[y](x - y) = f \circ \lambda \circ i(x) - f \circ \lambda \circ i(y) - d(f \circ \lambda \circ i)[y](x - y)$$

$$= F(i(x)) - F(i(y)) - dF[i(y)] \circ di[y](x - y) = F(i(x)) - F(i(y)) - dF[i(y)](i(x) - i(y))$$

$$= D_F(i(x), i(y)),$$

where we used that $f \circ \lambda = F$, the chain rule for $F \circ i$ and the fact that $i$ is linear, hence it coincides with its derivative. By (4), if the joint convexity of $D_f$ fails on $\text{int} (\text{dom} f \cap \text{ran} \lambda)$, then so does the joint convexity of $D_F$. In other words, the joint convexity of $D_F$ implies the joint convexity of $D_f$ on $\text{int} (\text{dom} f \cap \text{ran} \lambda)$.

Acknowledgments. This work was partially supported by the Hungarian Research Grant OTKA K104206. The author would like to thank József Pitrik for comments and remarks, Dénes Petz for his support and Heinz Bauschke for communication.

References

[BB01] H. Bauschke and J. Borwein, Joint and separate convexity of the Bregman distance, Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (Haifa 2000), D. Butnariu, Y. Censor, S. Reich (editors), Elsevier, pp. 23-36, 2001.

[BH96] R. Bhatia, Matrix analysis, Springer, 1996.

[BR67] L. M. Bregman, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, USSR Computational Mathematics and Mathematical Physics 7(3)(1967), 200-217.

[HP14] F. Hiai, D. Petz, Introduction to Matrix Analysis and Applications, Hindustan Book Agency and Springer Verlag, 2014.
[NE95] R. B. Nelson, Proof without Words: The Arithmetic-Logarithmic-Geometric Mean Inequality. Math. Mag. 68, 305, 1995.

[PV14] J. Pitrik, D. Virosztek, On the joint convexity of the Bregman divergence of matrices, arXiv: 1405.7885