JOINT AND SEPARATE CONVEXITY OF THE BREGMAN DISTANCE

Heinz H. Bauschke[a] and Jonathan M. Borwein[b]

[a]Department of Mathematics and Statistics, Okanagan University College,
Kelowna, British Columbia V1V 1V7, Canada.
Email: bauschke@cecm.sfu.ca. Research supported by NSERC.

[b]Centre for Experimental and Constructive Mathematics, Simon Fraser University,
Burnaby, British Columbia V5A 1S6, Canada.
Email: jborwein@cecm.sfu.ca. Research supported by NSERC.

Algorithms involving Bregman projections for solving optimization problems have been receiving much attention lately. Several of these methods rely crucially on the joint convexity of the Bregman distance. In this note, we study joint and separate convexity of Bregman distances.

To bring out the main ideas more clearly, we consider first functions defined on an open interval. Our main result states that the Bregman distance of a given function is jointly convex if and only if the reciprocal of its second derivative is concave. We observe that Bregman distances induced by the two most popular choices — the *energy* and the *Boltzmann-Shannon entropy* — are limiting cases in a profound sense. This result is generalized by weakening assumptions on differentiability and strict convexity.

We then consider general, not necessarily separable, convex functions. The characterization of joint convexity has a natural and beautiful analog. Finally, we discuss spectral functions, where the situation is less clear. Throughout, we provide numerous examples to illustrate our results.
Keywords: Bregman distance, convex function, joint convexity, separate convexity.

1. INTRODUCTION

Unless stated otherwise, we assume throughout that

$$\boxed{I \text{ is a nonempty open interval in } \mathbb{R}, \text{ and that } f \in C^3(I) \text{ with } f'' > 0 \text{ on } I.}$$

Clearly, f is strictly convex. Associated with f is the *Bregman "distance"* D_f:

$$\boxed{D_f : I \times I \to [0, +\infty) : (x, y) \mapsto f(x) - f(y) - f'(y)(x - y).}$$

For further information on Bregman distances and their applications, see [1], [6], [13], [21–23], and [33, Section 2]. (See also [14] for pointers to related software.)

Since f is convex, it is clear that the Bregman distance is convex in the first variable:

- $x \mapsto D_f(x, y)$ is convex, for every $y \in I$.

In this note, we are interested in the following two stronger properties:

- D_f is *jointly convex*: $(x, y) \mapsto D_f(x, y)$ is convex on $I \times I$;
- D_f is *separately convex*: $y \mapsto D_f(x, y)$ is convex, for every $x \in I$.

Clearly, if D_f is jointly convex, then it is separately convex. Joint convexity lies at the heart of the analysis in many recent papers. It was used explicitly by Butnariu, Censor, and Reich [9, Section 1.5], by Butnariu, Iusem, and Burachik [12, Section 6], by Butnariu and Iusem [11, Section 2.3], by Byrne and Censor [7,8], by Csiszár and Tusnády [16], by Eggermont and LaRiccia [17], by Iusem [19]. Separate convexity is a sufficient condition for results on the convergence of certain algorithms; see Butnariu and Iusem's [10, Theorem 1], and Bauschke et al.'s [2].

Despite the usefulness of joint and separate convexity of D_f, we are not aware of any work that studies these concepts in their own right — except for an unpublished manuscript [4] on which the present note is based.

The objective of this note is to systematically study separate and joint convexity of the Bregman distance.

The material is organized as follows.

In Section 2, we collect some preliminary results. Joint and separate convexity of D_f for a one-dimensional convex function f are given in Section 3. Our main result states that D_f is jointly convex if and only if $1/f''$ is concave. The well-known examples of functions inducing jointly convex Bregman distances — the energy and the Boltzmann-Shannon entropy — are revealed as limiting cases in a profound sense. Section 4 discusses asymptotic behavior whereas in Section 5 we relax some of our initial assumptions. We turn to the general discussion of (not necessarily separable) convex function in the final Section 6. Our main result has a beautiful analog: D_f is jointly convex if and only if the inverse of the Hessian of f is (Loewner) concave. Finally, we discuss spectral functions where the situation appears to be less clear.

2. PRELIMINARIES

The results in this section are part of the folklore and listed only for the reader's convenience.

Fact 2.1. [30, Theorem 13.C] Suppose ψ is a convex function, and ϕ is an increasing and convex function. Then the composition $\phi \circ \psi$ is convex.

Corollary 2.2. Suppose g is a positive function. Consider the following three properties:

(i) $1/g$ is concave.

(ii) g is *log-convex*: $\ln \circ g$ is convex.

(iii) g is convex.

Then: (i)$\Rightarrow$(ii)$\Rightarrow$(iii).

Proof. "(i)$\Rightarrow$(ii)": Let $\psi = -1/g$ and $\phi : (-\infty, 0) \to \mathbb{R} : x \mapsto -\ln(-x)$. Then ψ is convex, and ϕ is convex and increasing. By Fact 2.1, $\ln \circ g = \phi \circ \psi$ is convex.
"(ii)$\Rightarrow$(iii)": Let $\psi = \ln \circ g$ and $\phi = \exp$. Then $g = \phi \circ \psi$ is convex, again by Fact 2.1. $\square$

Remark 2.3. It is well-known that the implications in Corollary 2.2 are not reversible: • exp is log-convex on $\mathbb{R}$, but $1/\exp$ is not concave; • $x \mapsto x$ is convex on $(0 + \infty)$, but not log-convex.

Fact 2.4. Suppose g is a differentiable function on I. Then:

(i) g is convex $\Leftrightarrow D_g(x, y) \geq 0$, for all $x, y \in I$.

(ii) g is affine $\Leftrightarrow D_g(x, y) = 0$, for all $x, y \in I$.

Proof. (i): [30, Theorem 42.A]. (ii): use (i) with g and $-g$, and $D_{-g} = -D_g$. $\square$

Proposition 2.5. If g is convex and proper, then $\lim_{z \to +\infty} g(z)/z$ exists in $(-\infty, +\infty]$.

Proof. Fix x_0 in $\operatorname{dom} g$ and check that $\Psi : x \mapsto \big(g(x_0 + x) - g(x_0)\big)/x$ is increasing. Hence $\lim_{x \to +\infty} \Psi(x)$ exists in $(-\infty, +\infty]$. The result follows from

$$\frac{g(x_0 + x)}{x_0 + x} = \frac{g(x_0 + x) - g(x_0)}{x} \frac{x}{x_0 + x} + \frac{g(x_0)}{x_0 + x}$$

and the change of variables $z = x_0 + x$. $\square$

The next two results characterize convexity of a function.

Theorem 2.6. Suppose U is a convex nonempty open set in $\mathbb{R}^N$ and $g : U \to \mathbb{R}$ is continuous. Let $A := \{x \in U : \nabla g(x) \text{ exists}\}$. Then the following are equivalent.

(i) g is convex.

(ii) $U \setminus A$ is a set of measure zero, and $\nabla g(x)(y-x) \leq g(y) - g(x)$, for all $x \in A$ and $y \in U$.

(iii) A is dense in U, and $\nabla g(x)(y-x) \leq g(y) - g(x)$, for all $x \in A$ and $y \in U$.

Proof. "(i)⇒(ii)": g is almost everywhere differentiable, either by [30, Theorem 44.D] or by local Lipschitzness and Rademacher's Theorem ([15, Theorem 3.4.19]). On the other hand, the subgradient inequality holds on U. Altogether, (ii) follows.

"(ii)⇒(iii)": trivial.

"(iii)⇒(i)": Fix u, v in U, $u \neq v$, and $t \in (0,1)$. It is not hard to see that there exist sequences (u_n) in U, (v_n) in U, (t_n) in $(0,1)$ with $u_n \to u$, $v_n \to v$, $t_n \to t$, and $x_n := t_n u_n + (1-t_n)v_n \in A$, for every n. By assumption, $\nabla g(x_n)(u_n - x_n) \leq g(u_n) - g(x_n)$ and $\nabla g(x_n)(v_n - x_n) \leq g(v_n) - g(x_n)$. Equivalently, $(1-t_n)\nabla g(x_n)(v_n - u_n) \leq g(u_n) - g(x_n)$ and $t_n \nabla g(x_n)(u_n - v_n) \leq g(v_n) - g(x_n)$. Multiply the former inequality by t_n, the latter by $1 - t_n$, and add. We obtain that $g(x_n) \leq t_n g(u_n) + (1-t_n)g(v_n)$. Let n tend to $+\infty$ to deduce that $g\big(tu + (1-t)v\big) \leq tg(u) + (1-t)g(v)$. The convexity of g follows and the proof is complete. ☐

Theorem 2.7. Suppose U is a convex nonempty open set in $\mathbb{R}^N$ and $g : U \to \mathbb{R}$ is continuously differentiable on U. Let $A := \{x \in U : g''(x) \text{ exists}\}$. Then g is convex ⇔ $U \setminus A$ is a set of measure zero, and g'' is positive semidefinite on A.

Proof. "⇒": Aleksandrov's Theorem (see [32, Theorem 13.51]) states that the set $U \setminus A$ is of measure zero. Fix $x \in U$ and $y \in \mathbb{R}^N$ arbitrarily. The function $t \mapsto g(x+ty)$ is convex in a neighborhood of 0; consequently, its second derivative $\big\langle y, \big(g''(x+ty)\big)(y)\big\rangle$ is nonnegative whenever it exists. Since $g''(x)$ does exist, it must be that $g''(x)$ is positive semidefinite, for every $x \in A$.

"⇐": Fix x and y in U. By assumption and a Fubini argument, we obtain two sequences (x_n) and (y_n) in U with $x_n \to x$, $y_n \to y$, and $t \mapsto g''(x_n + t(y_n - x_n))$ exists almost everywhere on $[0,1]$. Integrating $t \mapsto \big\langle y_n - x_n, \big(g''(x_n + t(y_n - x_n))\big)(y_n - x_n)\big\rangle \geq 0$ from 0 to 1, we deduce that $\big(g'(y_n) - g'(x_n)\big)(y_n - x_n) \geq 0$. Taking limits and recalling that g' is continuous, we see that g' is monotone and so g is convex [30, Theorem 42.B]. ☐

3. JOINT AND SEPARATE CONVEXITY ON $\mathbb{R}$

BASIC RESULTS

Definition 3.1. We say that D_f is:

(i) *separately convex*, if $y \mapsto D_f(x,y)$ is convex, for every $x \in I$.

(ii) *jointly convex*, if $(x,y) \mapsto D_f(x,y)$ is convex on $I \times I$.

The following result will turn out to be useful later.

Lemma 3.2. Suppose $h : I \to (0,+\infty)$ is a differentiable function. Then:

(i) $1/h$ is concave $\Leftrightarrow h(y) + h'(y)(y-x) \geq \big(h(y)\big)^2/h(x)$, for all x, y in I.

(ii) $1/h$ is affine $\Leftrightarrow h(y) + h'(y)(y-x) = \big(h(y)\big)^2/h(x)$, for all x, y in I.

(iii) $1/h$ is concave $\Rightarrow h$ is log-convex $\Rightarrow h$ is convex.

(iv) If h is twice differentiable, then: $1/h$ is concave $\Leftrightarrow hh'' \geq 2(h')^2$.

Proof. Let $g := -1/h$ so that $g' = h'/h^2$.
"(i)": $1/h$ is concave $\Leftrightarrow g$ is convex $\Leftrightarrow D_g$ is nonnegative (Lemma 2.4) $\Leftrightarrow 0 \leq -1/h(x) + 1/h(y) - \big(h'(y)/h^2(y)\big)(x-y)$, $\forall x, y \in I \Leftrightarrow 0 \leq h(x)h(y) - h^2(y) - h(x)h'(y)(x-y)$, $\forall x, y \in I$. "(ii)": is similar to (i). "(iii)": restates Corollary 2.2. "(iv)": $1/h$ is concave $\Leftrightarrow g$ is convex $\Leftrightarrow g'' = \big(h^2h'' - 2h(h')^2\big)/h^4 \geq 0 \Leftrightarrow hh'' \geq 2(h')^2$. ☐

Theorem 3.3. Let $h := f''$. Then:

(i) D_f is jointly convex $\Leftrightarrow 1/h$ is concave $\Leftrightarrow$

$$h(y) + h'(y)(y-x) \geq \big(h(y)\big)^2/h(x), \quad \text{for all } x, y \text{ in } I. \tag{j}$$

In particular, if f'''' exists, then: D_f is jointly convex $\Leftrightarrow hh'' \geq 2(h')^2$.

(ii) D_f is separately convex $\Leftrightarrow$

$$h(y) + h'(y)(y-x) \geq 0, \quad \text{for all } x, y \text{ in } I. \tag{s}$$

Proof. "(i)": $\nabla^2 D_f(x,y)$, the Hessian of D_f at $(x,y) \in I \times I$, is equal to

$$\begin{pmatrix} f''(x) & -f''(y) \\ -f''(y) & f''(y) + f'''(y)(y-x) \end{pmatrix} = \begin{pmatrix} h(x) & -h(y) \\ -h(y) & h(y) + h'(y)(y-x) \end{pmatrix}.$$

Using [30, Theorem 42.C], we have the following equivalences: D_f is jointly convex $\Leftrightarrow$ $\nabla^2 D_f(x,y)$ is positive semidefinite, $\forall x, y \in I \Leftrightarrow h(x) \geq 0$, $h(y) + h'(y)(y-x) \geq 0$, and $\det \nabla^2 D_f(x,y) \geq 0$, $\forall x, y \in I \Leftrightarrow 1/h$ is concave (using $h > 0$ and Lemma 3.2.(i)). The "In particular" part is now clear from Lemma 3.2.(iv).

"(ii)": for fixed x, the second derivative of $y \mapsto D_f(x,y)$ equals $h(y) + h'(y)(y-x)$. Hence the result follows. ☐

Separate convexity is genuinely less restrictive than joint convexity:

Example 3.4. Let $f(x) = \exp(-x)$. Then $f''(x) = \exp(-x)$, and $1/f''(x) = \exp(x)$ is nowhere concave. Set $I = (0,1)$. By Theorem 3.3.(i), D_f *is not jointly convex on* I. On the other hand, fix $x, y \in I$ arbitrarily. Then $y - x \leq |y - x| \leq 1$ and hence

$$f''(y) + f'''(y)(y-x) = \exp(-y)\big(1 - (y-x)\big) > 0 > -\exp(-y) = f'''(y).$$

The first inequality and Theorem 3.3.(ii) imply that D_f *is separately convex.*

Example 3.5. (Fermi-Dirac entropy) Let $f(x) = x\ln(x) + (1-x)\ln(1-x)$ on $I = (0,1)$. Then $f''(x) = 1/\big(x(1-x)\big)$. Thus $1/f''(x) = x(1-x)$ is concave (but not affine). By Theorem 3.3.(i), D_f is jointly convex.

LIMITING CASES

Remark 3.6. A limiting case occurs when f as in Theorem 3.3 "barely" produces a jointly convex Bregman distance, by which we mean that the inequality (j) is always an equality. By the proof of Theorem 3.3 and Lemma 3.2.(ii), this occurs precisely when $1/h = 1/f''$ is affine, i.e., there exist real α, β such that

$$f''(x) = \frac{1}{\alpha x + \beta} > 0, \quad \text{for every } x \in I.$$

This condition determines f *up to additive affine perturbations.* We discover that either

$$f(x) = \frac{x^2}{2\beta}, \quad \text{if } \alpha = 0 \text{ and } \beta > 0;$$

or

$$f(x) = \frac{(\alpha x + \beta)\big(-1 + \ln(\alpha x + \beta)\big)}{\alpha^2}, \quad \text{if } \alpha \neq 0.$$

Hence f must be either a "general energy" ($\alpha = 0$ and $\beta = 1$ yields the *energy* $f(x) = \frac{1}{2}x^2$ on the real line) or a "general entropy" ($\alpha = 1$ and $\beta = 0$ results in the *Boltzmann-Shannon entropy* $f(x) = x \ln x - x$ on $(0, +\infty)$).

We saw in Remark 3.6 how limiting the requirement of joint convexity of D_f on the entire real line is — only quadratic functions have this property. Is this different for the weaker notion of separate convexity? The answer is negative:

Corollary 3.7. Suppose $I = \mathbb{R}$. Then D_f is separately convex if and only if f is essentially the *energy*: there exist $a, b, c \in \mathbb{R}$, $a > 0$, such that $f(x) = a\frac{1}{2}x^2 + bx + c$.

Proof. Consider inequality (s) of Theorem 3.3.(ii) and let x tend to $\pm\infty$. We conclude that $h'(y) = f'''(y) = 0$, for all $y \in \mathbb{R}$, and the result follows. $\square$

Remark 3.8. If $I = \mathbb{R}$, then the above yields following characterization:

$$D_f \text{ is jointly convex} \Leftrightarrow D_f \text{ is separately convex} \Leftrightarrow f \text{ is a quadratic.}$$

Hence an example such as Example 3.4 requires I to be a proper subset of the real line.

THE CASE WHEN $I = (0, +\infty)$

We now provide a more usable characterization of separate convexity for the important case when $I = (0, +\infty)$.

Corollary 3.9. If $I = (0, +\infty)$, then: D_f is separately convex $\Leftrightarrow$ $f''(x) + xf'''(x) \geq 0 \geq f'''(x)$, for every $x > 0$.

Proof. By Theorem 3.3.(ii), D_f is separately convex if and only if

$$f''(y) + f'''(y)(y - x) \geq 0, \quad \text{for all } x, y > 0. \tag{s}$$

"$\Rightarrow$": Consider (s). Let x tend to $+\infty$ to learn that $f'''(y) \leq 0$. Then let x tend to 0 from the right to deduce that $f''(y) + yf'''(y) \geq 0$. Altogether, $f''(y) + yf'''(y) \geq 0 \geq f'''(y)$, for every $y > 0$. "$\Leftarrow$": straight-forward. $\square$

Discussing limiting cases leads to no new class of functions:

Remark 3.10. Suppose D_f is separately convex and $I = (0, +\infty)$; equivalently, by Corollary 3.9,

$$f''(x) + xf'''(x) \geq 0 \geq f'''(x), \quad \text{for every } x > 0. \tag{$*$}$$

When considering limiting solutions of $(*)$, we have two choices: Either we require the right inequality in $(*)$ to be an equality throughout — this results in $f''' = 0$, and we obtain (as in Corollary 3.7) essentially the *energy*. Or we impose equality in the left inequality of $(*)$: $f''(x) + xf'''(x) = 0$, for all $x > 0$. But this differential equation readily leads to (essentially) the *Boltzmann-Shannon entropy*:

$$f(x) = a\big(x \ln(x) - x\big) + bx + c,$$

where $a, b, c \in \mathbb{R}$ and $a > 0$.

Remark 3.8 and Remark 3.10 may make the reader wonder whether separate convexity differs from joint convexity when $I = (0, +\infty)$. Indeed, they do differ, and a counter-example will be constructed (see Example 3.13) with the help of the following result.

Theorem 3.11. Suppose $I = (0, +\infty)$ and f'''' exists. Let $\psi := \ln \circ (f'')$. Then:

(i) D_f is jointly convex $\Leftrightarrow \psi''(x) \geq \big(\psi'(x)\big)^2$, $\forall x > 0$.

(ii) D_f is separately convex $\Leftrightarrow 0 \geq \psi'(x) \geq -1/x$, $\forall x > 0$.

(iii) f'' is log-convex $\Leftrightarrow \psi''(x) \geq 0$, $\forall x > 0$.

(iv) f'' is convex $\Leftrightarrow \psi''(x) \geq -\big(\psi'(x)\big)^2$, $\forall x > 0$.

Proof. "(i)": clear from Theorem 3.3.(i). "(ii)": use Corollary 3.9. "(iii)": f'' is log-convex $\Leftrightarrow \psi$ is convex $\Leftrightarrow \psi'' \geq 0$. "(iv)": use f'' is convex $\Leftrightarrow h'' \geq 0$. $\square$

Remark 3.12. *Joint (or separate) convexity of D_f is* not preserved *under Fenchel conjugation*: indeed, if $f(x) = x \ln(x) - x$ is the Boltzmann-Shannon entropy, then $f^* = \exp$ on $\mathbb{R}$. Now D_f is jointly convex, whereas D_{f^*} is not separately convex on $(0, +\infty)$ (by Theorem 3.11.(ii)).

Example 3.13. On $I = (0, +\infty)$, let

$$\psi(x) := \frac{-\ln(x) - \mathrm{Si}(x)}{2}, \quad \text{where} \quad \mathrm{Si}(x) := \int_0^x \frac{\sin(t)}{t} dt$$

denotes the sine integral function. Let f be a second anti-derivative of $\exp\circ\psi$. Then $\psi'(x) = -\big(1+\sin(x)\big)/\big(2x\big)$; therefore, by Theorem 3.11.(ii), *D_f is separately convex.* However, since condition (iv) of Theorem 3.11 fails at $x = 2\pi$, f'' cannot be convex and *D_f is not jointly convex.* (It appears there is no elementary closed form for f.)

Example 3.14. (Burg entropy) Suppose $f(x) = -\ln(x)$ on $(0, +\infty)$. Then $f'(x) = -1/x$ and $f''(x) = 1/x^2$. Hence f'' is convex. Let $\psi(x) := \ln\big(f''(x)\big) = -2\ln(x)$. Since $\psi'(x) = -2/x < -1/x$, Theorem 3.11.(ii) implies that D_f is not separately convex. (In fact, there is no $x > 0$ such that $y \mapsto D_f(x,y)$ is convex on $(0,+\infty)$.) However, $\psi''(x) = 2/x^2 > 0$, so f'' is log-convex by Theorem 3.11.(iii).

Remark 3.15. Example 3.13 and Example 3.14 show that *separate convexity of D_f and (log-) convexity of f'' are independent properties.*

Although not directly related to joint or separate convexity of D_f, we would nonetheless like to mention a result due to Iusem.

Lemma 3.16. (Iusem; [20]) If D_f is symmetric, then f is a quadratic.

Proof. Differentiate both sides of $D_f(x,y) = D_f(y,x)$ with respect to x to learn that the gradient map $y \mapsto f'(y)$ is affine. It follows that f is a quadratic. $\square$

4. ASYMPTOTIC RESULTS

This subsection once again shows the importance of the energy and the Boltzmann-Shannon entropy — they appear naturally when studying asymptotic behavior.

Lemma 4.1. Suppose $\sup I = +\infty$ and D_f is separately convex. Then f *does not grow faster than the energy*: $0 \leq \lim_{x\to+\infty} f(x)/x^2 < +\infty$.

Proof. Let $L := \lim_{x\to+\infty} f(x)/x^2$. We must show that L exists as a nonnegative finite real number. Since D_f is separately convex, Theorem 3.3.(ii) yields $f''(y) + f'''(y)(y-x) \geq 0$. Letting tend x to $+\infty$ shows that $f'''(y) \leq 0$. Hence f' is concave and f'' is decreasing. Also, $f'' > 0$. It follows that

$$f''(+\infty) := \lim_{x\to+\infty} f''(x) \in [0, +\infty).$$

We will employ similar notation when there is no cause for confusion. Since f is convex, $f(+\infty)$ exists in $[-\infty, +\infty]$. If $f(+\infty)$ is finite, then $L = 0$ and we are done. Thus assume $f(+\infty) = \pm\infty$. Consider the quotient

$$q(x) := \frac{f'(x)}{2x}$$

for large x. Since f' is concave, $q(+\infty)$ exists by Proposition 2.5. L'Hospital's Rule (see [34, Theorem 4.22]) yields $q(+\infty) = L$. Since f' is increasing, $f'(+\infty)$ exists in $(-\infty, +\infty]$. Thus if $f'(+\infty)$ is finite, then $L = 0$. So we assume $f'(+\infty) = +\infty$. Now $f''(+\infty)/2 \in [0, +\infty)$; hence, again by L'Hospital Rule, so is $q(+\infty) = L$. $\square$

Example 4.2. (p-norms) For $1 < p$, let $f(x) = \frac{1}{p}x^p$ on $I = (0, +\infty)$. Then:

$$D_f \text{ is jointly convex} \Leftrightarrow D_f \text{ is separately convex} \Leftrightarrow 1 < p \leq 2.$$

Proof. Case 1: $1 < p \leq 2$. Then $f''(x) = (p-1)x^{p-2}$, hence $1/f''(x) = x^{2-p}/(p-1)$ is concave. Thus D_f is jointly convex by Theorem 3.3.(i).

Case 2: $2 < p < +\infty$. Then $f(z)/z^2 = z^{p-2}/p$ tends to $+\infty$ as z does. Consequently, by Lemma 4.1, D_f is not separately convex. $\square$

Lemma 4.3. Suppose $\inf I = 0$, $\lim_{x\to 0^+} f(x) = 0$, and D_f is separately convex. Then $0 \leq \lim_{x\to 0^+} f(x)/(x\ln(x) - x) < +\infty$.

Proof. Let $L := \lim_{x\to 0^+} f(x)/(x\ln(x) - x)$. We must show that L exists and is a nonnegative finite real number. As in the proof of Lemma 4.1, we employ space-saving notation such as $f(0) := \lim_{x\to 0^+} f(x) = 0$. Consider

$$q(x) := \frac{f'(x)}{\ln(x)}$$

for small $x > 0$.

It suffice to show that $q(0)$ exists in $[0, \infty)$: indeed, this implies $q(0) = L$ by L'Hospital's Rule, which is what we want.

Since f is convex, f' is increasing. Hence $f'(0)$ exists in $[-\infty, +\infty)$. If $f'(0) > -\infty$, then $q(0) = 0$ and we are done. Hence assume $f'(0) = -\infty$. Since D_f is separately convex, Theorem 3.3.(ii) yields again $f''(y) + f'''(y)(y - x) \geq 0$. Letting tend x to 0 from the right shows that $f''(y) + yf'''(y) \geq 0$, for all $y \in I$. Hence $y \mapsto yf''(y)$ is increasing. It follows that

$$\lim_{x\to 0^+} xf''(x) \in [0, +\infty).$$

By L'Hospital Rule, $q(0) \in [0, +\infty)$ as is L. $\square$

Example 4.4. Suppose $f(x) = -\frac{1}{p}x^p$ on $I = (0, +\infty)$ for $p \in (-\infty, 1) \setminus \{0\}$. L'Hospital's Rule applied twice easily yields $\lim_{x\to 0^+} f(x)/(x\ln(x) - x) = +\infty$; thus, by Lemma 4.3, D_f is not separately convex.

5. RELAXING THE ASSUMPTIONS

RELAXING $f'' > 0$ TO $f'' \geq 0$

The function $x \mapsto \frac{1}{4}x^4$ is strictly convex, but its second derivative has a zero; consequently, none of the above results is applicable. The following technique allows to handle such function within our framework. For $\epsilon > 0$, let

$$f_\epsilon := f + \epsilon \tfrac{1}{2}|\cdot|^2.$$

Then we obtain readily the following.

Observation 5.1. D_f is jointly (resp. separately) convex if and only if each D_{f_ϵ} is.

We only give one example to show how this Observation can be used rather than writing down several slightly more general results.

Example 5.2. Suppose $f(x) = \frac{1}{4}x^4$ on $I = (-1, 1)$. Since $f''(0) = 0$, we cannot use our previous results. Now $f_\epsilon''(x) = f''(x) + \epsilon = 3x^2 + \epsilon$ and $f_\epsilon'''(x) = 6x$. Pick $0 < \epsilon < \frac{1}{2}$, let $x = 2\sqrt{\epsilon}$ and $y = \frac{1}{3}x$. With this assignment, $f_\epsilon''(y) + f_\epsilon'''(y)(y - x) = -3\epsilon < 0$. Thus, by Theorem 3.3.(ii), D_{f_ϵ} is not separately convex. Hence, by Observation 5.1, D_f is not separately convex.

RELAXING $f \in C^3(I)$ TO $f \in C^2(I)$

The assumption on differentiability on f can be weakened and several slightly more general result could be obtained — we illustrate this through the following variant of Theorem 3.3.(i).

Theorem 5.3. Suppose $f \in C^2(I)$ and $f'' > 0$. Then D_f is jointly convex if and only if $1/f''$ is concave.

Proof. D_f is jointly convex $\Leftrightarrow$ $\nabla^2 D_f$ exists almost everywhere in $I \times I$, and whenever it exists,

$$\nabla^2 D_f(x, y) = \begin{pmatrix} f''(x) & -f''(y) \\ -f''(y) & f''(y) + f'''(y)(y - x) \end{pmatrix}$$

is positive semidefinite (by Theorem 2.7) $\Leftrightarrow$ for almost every $y \in I$, $f'''(y)$ exists and $f''(y) + f'''(y)(y - x) \geq \big(f''(y)\big)^2$, for all $x \in I$ (by using a Fubini-type argument such as [34, Lemma 6.120]) $\Leftrightarrow$ $1/f''$ is concave (by applying Theorem 2.6 to $g := -1/f''$). $\square$

Example 5.4. Let $I := (-\infty, 1)$ and $f(x) := \frac{1}{2}x^2$, if $x \leq 0$; $x + (1 - x)\ln(1 - x)$, if $0 \leq x < 1$. Then $f \in C^2(I) \setminus C^3(I)$ so that Theorem 3.3 does not apply directly. However: since $1/f''(x) = \min\{1, 1 - x\}$ is concave, Theorem 5.3 yields the joint convexity of D_f.

6. JOINT AND SEPARATE CONVEXITY IN GENERAL

The results above concern a function defined on an interval. This allows us to handle functions on $\mathbb{R}^N$ that are separable. The general case is a little more involved; however, the main patterns observed so far generalize quite beautifully.

Throughout this section, we assume that

$$\boxed{U \text{ is a convex nonempty open set in } \mathbb{R}^N}$$

and that

$$\boxed{f \in C^3(U) \text{ with } f'' \text{ positive definite on } U.}$$

Clearly, f is strictly convex. Let

$$\boxed{H := H_f := \nabla^2 f = f''}$$

denote the Hessian of f. $H(x)$ is a real symmetric positive semidefinite matrix, for every $x \in U$. Recall that the real symmetric matrices can be partially ordered by the *Loewner ordering*

$$\boxed{H_1 \succeq H_2 \quad :\Leftrightarrow \quad H_1 - H_2 \text{ is positive semidefinite,}}$$

and they form a Euclidean space with the inner product

$$\boxed{\langle H_1, H_2 \rangle := \operatorname{trace}(H_1 H_2).}$$

Further information can be found in [18, Section 7.7], [29, Section 16.E], and in the very recent [5]. We will be using results from these sources throughout.

We start by generalizing Theorem 3.3.

Theorem 6.1. (i) D_f is jointly convex if and only if

$$H(y) + \big(\nabla H(y)\big)(y - x) \succeq H(y)H^{-1}(x)H(y), \quad \text{for all } x, y \in U. \tag{J}$$

(ii) D_f is separately convex if and only if

$$H(y) + \big(\nabla H(y)\big)(y - x) \succeq 0, \quad \text{for all } x, y \in U. \tag{S}$$

Proof. "(i)": The Hessian of the function $U \times U \to [0,+\infty) : (x,y) \mapsto D_f(x,y)$ is (compare to the proof of Theorem 3.3) the block matrix

$$\nabla^2 D_f(x,y) = \begin{pmatrix} H(x) & -H(y) \\ -H(y) & H(y) + \big(\nabla H(y)\big)(y-x) \end{pmatrix}.$$

Using standard criteria for positive semidefiniteness of block matrices (see [18, Section 7.7]) and remembering that H is positive definite, we obtain that $\nabla^2 D_f(x,y)$ is positive semidefinite for all x, y if and only if (J) holds. "(ii)": For fixed $x \in U$, similarly discuss positive semidefiniteness of the Hessian of the map $y \mapsto D_f(x,y)$. $\square$

Corollary 6.2. The following are equivalent:

(i) D_f is jointly convex.

(ii) $\nabla H^{-1}(y)(x-y) \succeq H^{-1}(x) - H^{-1}(y)$, for all $x, y \in U$.

(iii) H^{-1} is *(matrix) concave*, i.e.,

$$H^{-1}(\lambda x + \mu y) \succeq \lambda H^{-1}(x) + \mu H^{-1}(y),$$

for all $x, y \in U$ and $\lambda, \mu \in [0,1]$ with $\lambda + \mu = 1$.

(iv) $x \mapsto \langle P, H^{-1}(x)\rangle$ is concave, for every $P \succeq 0$.

Proof. Consider the mapping $U \to \mathbb{R}^{N\times N} : Y \mapsto H(y)H^{-1}(y)$. It is constant, namely the identity matrix. Take the derivative with respect to y. Using an appropriate product rule (see, for instance, [24, page 469f in Section 17.3]) yields $0 = H(y)\big(\big(\nabla H^{-1}(y)\big)(z)\big) + \big(\big(\nabla H(y)\big)(z)\big)H^{-1}(y)$, for every $z \in \mathbb{R}^N$. In particular, after setting $z = x-y$, multiplying by $H^{-1}(y)$ from the left, and re-arranging, we obtain

$$H^{-1}(y)\big(\big(\nabla H(y)\big)(y-x)\big)H^{-1}(y) = \big(\nabla H^{-1}(y)\big)(x-y).$$

"(i)$\Leftrightarrow$(ii)": The equivalence follows readily from the last displayed equation and Theorem 6.1.(i). "(ii)$\Leftrightarrow$(iii)": The proof of [30, Theorem 42.A] works without change in the present positive semidefinite setting. "(iii)$\Leftrightarrow$(iv)": is clear, since the cone of positive semidefinite matrices is self-dual. $\square$

Example 6.3. Suppose Q is an N-by-N real symmetric and positive definite matrix. Let $U = \mathbb{R}^N$ and $f(x) = \frac{1}{2}\langle x, Qx\rangle$. We assume that Q is not a diagonal matrix. Then f is not separable, hence none of the results in previous chapters are applicable. Now $f''(x) = Q = H(x)$, for all $x \in \mathbb{R}^N$. Hence $H^{-1}(x) = Q^{-1}$ is constant, and thus trivially matrix-concave. By Corollary 6.2, the Bregman distance D_f is jointly convex.

SPECTRAL FUNCTIONS

In this last subsection, we discuss convex functions defined on the N-by-N real symmetric matrices, denoted by $\mathcal{S}$ and equipped with the inner product

$$\boxed{\langle S_1, S_2\rangle := \operatorname{trace}(S_1 S_2).}$$

Suppose further

$$\boxed{U \subseteq \mathbb{R}^N \text{ is convex, nonempty, and open,}}$$

and

$$\boxed{f \in C^3(U) \text{ is convex, symmetric, with } f'' > 0.}$$

Then f induces a *spectral function*

$$\boxed{F := f \circ \lambda,}$$

where $\lambda : \mathcal{S} \to \mathbb{R}^N$ is the *eigenvalue map* (ordered decreasingly). Then the function F is *orthogonally invariant*: $F(X) = F(U^T XU)$, for all $X \in \mathcal{S}$ and every orthogonal N-by-N matrix U, and $F \circ \Lambda = f \circ \lambda$. For further information, we refer the reader to [25] and [5]; in [1, Section 7.2], we discussed the Bregman distance D_F in detail.

Our one-dimensional results showed the special status of the separable energy and the separable Boltzmann-Shannon entropy: if $f(x) = \sum_{n=1}^N \frac{1}{2}x_n^2$ or $f(x) = \sum_{n=1}^N x_n \ln(x_n) - x_n$, then D_f is jointly convex. It is natural to inquire about D_F, the Bregman distance induced by the corresponding spectral function $F = f \circ \lambda$. Now if f is the separable energy, then F is the energy on $\mathcal{S}$ and thus D_F is jointly convex. This is rather easy.

The same is true for the Boltzmann-Shannon entropy — this is essentially known as *Lindblad's Theorem* [28]:

Theorem 6.4. If $f(x) = \sum_{n=1}^N x_n \ln(x_n) - x_n$, then D_F is jointly convex.

Proof. For $X \in \mathcal{S}$ positive definite with eigenvectors $\lambda(X)$ ordered decreasingly, denote the diagonal matrix with $\lambda(X)$ on the diagonal by $\Lambda(X)$ so that $X = U\Lambda(X)U^T$ for some orthogonal matrix U. Recall that (see [3]) if g is a function from real interval to $\mathbb{R}$, then g acting on a diagonal matrix is defined as the diagonal matrix obtained by applying g to each diagonal entry; moreover, this is used to define $g(X) := Ug(\Lambda(X))U^T$. Also, we diagonalize a positive definite Y by $Y = V\Lambda(Y)V^T$, for some orthogonal matrix V. Then

$$\begin{aligned} D_F(X,Y) &= F(X) - F(Y) - \langle \nabla F(Y), X - Y\rangle \\ &= F(U\Lambda(X)U^T) - F(V\Lambda(Y)V^T) - \langle \nabla F(Y), X - Y\rangle \\ &= f(\lambda(X)) - f(\lambda(Y)) - \langle \nabla F(Y), X - Y\rangle. \end{aligned}$$

On the other hand, using [25, Corollary 3.3],

$$\begin{aligned}\langle \nabla F(Y), X\rangle &= \langle V(\nabla f)(\Lambda(Y))V^T, X\rangle = \langle V \ln(\Lambda(Y))V^T, X\rangle = \langle \ln(Y), X\rangle \\ &= \operatorname{trace}(X\ln(Y))\end{aligned}$$

and similarly

$$\begin{aligned}\langle \nabla F(Y), Y\rangle &= \langle V \ln(\Lambda(Y))V^T, Y\rangle = \operatorname{trace}(V\ln(\Lambda(Y))V^TY) \\ &= \operatorname{trace}(\ln(\Lambda(Y))V^TYV) = \operatorname{trace}(\ln(\Lambda(Y))\Lambda(Y)) \\ &= \textstyle\sum_{n=1}^N \lambda_n(Y)\ln(\lambda_n(Y)) = f(\lambda(Y)) + \operatorname{trace}(Y).\end{aligned}$$

Altogether,

$$\begin{aligned}D_F(X,Y) &= f(\lambda(X)) - f(\lambda(Y)) - \langle \nabla F(Y), X\rangle + \langle \nabla F(Y), Y\rangle \\ &= f(\lambda(X)) - \operatorname{trace}(X\ln(Y)) + \operatorname{trace}(Y) \\ &= \textstyle\sum_{n=1}^N \big(\lambda_n(X)\ln(\lambda_n(X)) - \lambda_n(X)\big) - \operatorname{trace}(X\ln(Y)) + \operatorname{trace}(Y) \\ &= \operatorname{trace}(\Lambda(X)\ln(\Lambda(X))) - \operatorname{trace}(X) - \operatorname{trace}(X\ln(Y)) + \operatorname{trace}(Y) \\ &= \operatorname{trace}(U\Lambda(X)U^TU\ln(\Lambda(X))U^T) + \operatorname{trace}(Y-X) - \operatorname{trace}(X\ln(Y)) \\ &= \operatorname{trace}(X\ln(X)) + \operatorname{trace}(Y-X) - \operatorname{trace}(X\ln(Y)) \\ &= \operatorname{trace}(Y-X) + \operatorname{trace}(X(\ln(X)-\ln(Y))).\end{aligned}$$

Clearly, $\operatorname{trace}(Y-X)$ is jointly convex. Finally, Lindblad's Theorem [28] (see also [3, Theorem IX.6.5]) precisely states that $\operatorname{trace}(X(\ln(X)-\ln(Y)))$ is jointly convex. Therefore, $D_F(X,Y)$ is jointly convex. ☐

Remark 6.5. Let f be the (separable) Boltzmann-Shannon entropy as in Theorem 6.4, and let F be its corresponding spectral function. Then Corollary 6.2.(iii) implies that $(\nabla^2 F)^{-1}$ is Loewner concave. (See [26,27] for further information on computing second derivatives of spectral functions.)

It would be interesting to find out about the general case. We thus conclude with a question.

Open Problem 6.6. *Is D_f jointly convex if and only if D_F is?*

We believe that Lewis and Sendov's recent work on the second derivative of a spectral function F [26,27] in conjunction with our second derivative characterization of joint convexity of D_F (Corollary 6.2.(iii)) will prove useful in deciding this problem.

REFERENCES

1. H. H. Bauschke and J. M. Borwein. Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis* 4(1):27–67, 1997.
2. H. H. Bauschke, D. Noll, A. Celler, and J. M. Borwein. An EM-algorithm for dynamic SPECT tomography. *IEEE Transactions on Medical Imaging*, 18(3):252–261, 1999.
3. R. Bhatia. *Matrix Analysis.* Springer-Verlag, 1996.
4. J. M. Borwein and L. C. Hsu. On the Joint Convexity of the Bregman Distance. Unpublished manuscript, 1993.
5. J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization.* Springer-Verlag, 2000.
6. L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *U.S.S.R. Computational Mathematics and Mathematical Physics* 7(3):200–217, 1967.
7. C. Byrne and Y. Censor. Proximity function minimization and the convex feasibility problem for jointly convex Bregman distances. Preprint, 1998.
8. C. Byrne and Y. Censor. Proximity Function Minimization Using Multiple Bregman Projections, with Applications to Split Feasibility and Kullback-Leibler Distance Minimization. Preprint, 1999.
9. D. Butnariu, Y. Censor, and S. Reich. Iterative Averaging of Entropic Projections for Solving Stochastic Convex Feasibility Problems. *Computational Optimization and Applications* 8(1):21–39, 1997.
10. D. Butnariu and A. N. Iusem. On a proximal point method for convex optimization in Banach spaces. *Numerical Functional Analysis and Optimization* 18(7–8):723–744, 1997.
11. D. Butnariu and A. N. Iusem. *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization.* Kluwer, 2000.
12. D. Butnariu, A. N. Iusem, and R. S. Burachik. Iterative methods for solving stochastic convex feasibility problems. To appear in *Computational Optimization and Applications.*
13. Y. Censor and S. A. Zenios. *Parallel Optimization.* Oxford University Press, 1997.
14. Centre for Experimental and Constructive Mathematics. *Computational Convex Analysis* project at `www.cecm.sfu.ca/projects/CCA`.
15. F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory.* Springer-Verlag, 1998.
16. I. Csiszár and G. Tusnády. Information geometry and alternating minimization procedures. *Statistics and Decisions* (Supplement 1), 205–237, 1984.

17. P. P. B. Eggermont and V. N. LaRiccia. On EM-like algorithms for minimum distance estimation. Preprint, 2000.
18. R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.
19. A. N. Iusem. A short convergence proof of the EM algorithm for a specific Poisson model. *Revista Brasileira de Probabilidade e Estatistica* 6:57–67, 1992.
20. A. N. Iusem. Personal communication.
21. K. C. Kiwiel. Free-steering relaxation methods for problems with strictly convex costs and linear constraints. *Mathematics of Operations Research* 22(2):326-349, 1997.
22. K. C. Kiwiel. Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization* 35(4):1142–1168, 1997.
23. K. C. Kiwiel. Generalized Bregman projections in convex feasibility problems. *Journal of Optimization Theory and Applications* 96(1):139–157, 1998.
24. S. Lang. *Undergraduate Analysis (Second Edition).* Springer-Verlag, 1997.
25. A. S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization* 6(1):164–177, 1996.
26. A. S. Lewis and H. S. Sendov. Characterization of Twice Differentiable and Twice Continuously Differentiable Spectral Functions. Preprint, 2000.
27. A. S. Lewis and H. S. Sendov. Quadratic Expansions of Spectral Functions. Preprint, 2000.
28. G. Lindblad. Entropy, information and quantum measurements. *Communications in Mathematical Physics* 33:305–322, 1973.
29. A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, 1979.
30. A. W. Roberts and D. E. Varberg. *Convex Functions.* Academic Press, 1973.
31. R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.
32. R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer-Verlag, 1998.
33. M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point method algorithm and some new results on the theory of Bregman functions. To appear in *Mathematics of Operations Research.*
34. K. R. Stromberg. *An Introduction to Classical Real Analysis.* Wadsworth, 1981.