Some of my Favourite Convex Functions

Jon Borwein, FRSC  www.cs.dal.ca/~jborwein
Canada Research Chair, Dalhousie
Laureate Professor, Newcastle

“Harald Bohr is reported to have remarked "Most analysts spend half their time hunting through the literature for inequalities they want to use, but cannot prove."” (D.J.H. Garling)


7th joint Australia-New Zealand Mathematics Convention
NZIMA Plenary Lecture (ANZMC2008)
The first modern formalization of the concept of convex function appears in J. L. W. V. Jensen, “Om konvexe funktioner og uligheder mellem midelvaerdier.” Nyt Tidsskr. Math. B 16 (1905), pp. 49–69. Since then, at first referring to “Jensen’s convex functions,” then more openly, without needing any explicit reference, the definition of convex function becomes a standard element in calculus handbooks. (A. Guerraggio and E. Molho)\textsuperscript{1}

Convexity theory ... reaches out in all directions with useful vigor. Why is this so? Surely any answer must take account of the tremendous impetus the subject has received from outside of mathematics, from such diverse fields as economics, agriculture, military planning, and flows in networks. With the invention of high-speed computers, large-scale problems from these fields became at least potentially solvable. Whole new areas of mathematics (game theory, linear and nonlinear programming, control theory) aimed at solving these problems appeared almost overnight. And in each of them, convexity theory turned out to be at the core. The result has been a tremendous spurt in interest in convexity theory and a host of new results. (A. Wayne Roberts and Dale E. Varberg)\textsuperscript{2}
Even Three Dimensions is Subtle

AN ESSENTIALLY STRICTLY CONVEX FUNCTION WITH NONCONVEX SUBGRADIENT DOMAIN AND WHICH IS NOT STRICTLY CONVEX

max{(x-2)^2+y^2-1,-(x*y)^(1/4)}
I offer a variety of examples of convexity appearing (often unexpectedly) in my research. *(Log)* convex functions are not denatured. They are everywhere.

Each example illustrates either the power of convexity, or of modern symbolic computation, or of both …

**PUB:**

\[
f_A(x) := \sup_{A \in \mathcal{A}} \| A(x) \|
\]

**Proof.** lsc and p.w. bounded is finite hence continuous and so the linear operators are uniformly bounded.

I start with a brief advert for computer-assisted mathematics and collaborative tools.
Experimental Methodology

1. Gaining insight and intuition
2. Discovering new relationships
3. Visualizing math principles
4. Testing and especially falsifying conjectures
5. Exploring a possible result to see if it merits formal proof
6. Suggesting approaches for formal proof
7. Computing replacing lengthy hand derivations
8. Confirming analytically derived results
"Experimental mathematics is here to stay. The reader who wants to get an introduction to this exciting approach to doing mathematics can do no better than this book."

—President of the American Mathematical Society

"Let me put to the test, every mathematics library acquire a copy of this book. Every university at higher degree students require a copy on their shelf. Welcome to the rich world of computer-supported mathematics."

—Mathematical Reviews

Mathematics have always used experiments and visualisation to explore new ideas and ways to prove them. Using examples that have never been seen before, the experimental methodology this book provides the historical context of, and rationale behind, experimental mathematics. It shows how today, the use advanced computing technology provides mathematics with a new support structure, in which new ideas are discovered, and patterns are discovered. This is a perfect introduction to the history and current state of research and technology in the growing field of experimental mathematics. In fact, it is an indication of the growth that has occurred since the first edition contains over one hundred pages describing new research since the publication of the first edition.

For those interested in further examples and insights, the book Experimental Mathematics: Computational Support by Bailey is highly recommended.

Jonathan Borwein
David Bailey
A. Generalized Convexity of Volumes (Bohr-Mollerup).

B. Coupon Collecting and Convexity.

C. Convexity of Spectral Functions.

D. Madelung’s Constant for Salt.

The talk ends when I do

There are three bonus tracks!

Full details are in the four reference texts and at http://projects.cs.dal.ca/ddrive/ConvexFunctions/ with some software
The Bohrs

• One Nobel Prize
  – Nils (1885-1962)
  – Physics (1922)

• One Olympic Medal
  – Harald (1887-1951)
  – Soccer (1908)
Generalized Convexity of Volumes

A. Generalized Convexity of Gamma (Bohr-Mollerup).

\[ \Gamma(x) := \int_0^\infty e^{-t}t^{x-1} \, dt. \quad (1) \]

Theorem 1 (Bohr-Mollerup) \( \Gamma \) is the unique function \( f : (0, \infty) \rightarrow (0, \infty) \) such that:
(a) \( f(1) = 1 \); (b) \( f(x + 1) = xf(x) \);
(c) \( f \) is log-convex (\( x \rightarrow \log f(x) \) is convex).

- Application is often automatable in a computer algebra system, as I now illustrate:
Generalized Convexity of Volumes

A. Generalized Convexity of Gamma (Beta function).

The \( \beta \)-function is defined by

\[
\beta(x, y) := \int_0^1 t^{x-1}(1 - t)^{y-1} \, dt \quad (1)
\]

for \( \text{Re}(x), \text{Re}(y) > 0 \). As is often established using polar coordinates and double integrals

\[
\beta(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x + y)}. \quad (2)
\]

Proof (2) Use \( f := x \to \beta(x, y) \frac{\Gamma(x + y)}{\Gamma(y)} \). (a) and (b) are easy. For (c) we show \( f \) is log-convex via Hölder’s inequality. Thus \( f = \Gamma \) as required. QED

- \( \Gamma \) is hyper-transcendental as is \( \zeta \).
A. Convexity of Volumes (Blaschke-Santalo inequality).

For a convex body $C$ in $\mathbb{R}^n$ its polar is

$$C^\circ := \{y \in \mathbb{R}^n : \langle y, x \rangle \leq 1 \text{ for all } x \in C\}.$$ 

Denoting $n$-dimensional Euclidean volume of $S \subseteq \mathbb{R}^n$ by $V_n(S)$, **Blaschke-Santalo** says

$$V_n(C) V_n(C^\circ) \leq V_n(E) V_n(E^\circ) = V_n^2(B_n(2))$$

(1)

where maximality holds (only) for any ellipsoid $E$ and $B_n(2)$ is the Euclidean unit ball.

**Question** Explain cases of (1) as convexity estimates? Noting $B_p^\circ = B_q$ if $1/p + 1/q = 1$. 
A. Convexity of Volumes (Dirichlet Formulae).

The volume of the ball in the \( \| \cdot \|_p \)-norm, \( V_n(p) \), was first determined by Dirichlet

\[
V_n(p) = 2^n \frac{\Gamma(1 + \frac{1}{p})^n}{\Gamma(1 + \frac{n}{p})}.
\]

When \( p = 2 \),

\[
V_n = 2^n \frac{\Gamma(\frac{3}{2})^n}{\Gamma(1 + \frac{n}{2})} = \frac{\Gamma(\frac{1}{2})^n}{\Gamma(1 + \frac{n}{2})},
\]

is more concise than that usually recorded. Maple code derives this formula as an iterated integral for arbitrary \( p \) and fixed \( n \).
Generalized Convexity of Volumes

A. Convexity of Volumes (Ease of Drawing Pictures).

\[ \log \Gamma(x) \quad \log V_{a}(1/x) \text{ for } a = 4/3, 3 \]

Discover the formula for \( \sum_{n>1} V_n(2) \)
Generalized Convexity of Volumes

A. Convexity of Volumes (‘mean’ log-convexity).

Theorem 2 [(H,A) log-concavity] The function $V_\alpha(p) := 2^\alpha \Gamma(1 + \frac{1}{p})^\alpha / \Gamma(1 + \frac{\alpha}{p})$ satisfies

$$V_\alpha(p)^\lambda V_\alpha(q)^{1-\lambda} < V_\alpha \left( \frac{pq}{\lambda q + (1-\lambda)p} \right)$$  \hspace{1cm} (1)

for all $\alpha > 1$, if $p, q > 1$, $p \neq q$, and $\lambda \in (0, 1)$.

In (1) $\alpha = n$, $\frac{1}{p} + \frac{1}{q} = 1$ with $\lambda = 1 - \lambda = 1/2$ recovers the $p-$norm case of Blaschke-Santalo; and the lower bound. This extends to substitution norms. Q. How far can one take this?
"Here's where you made your mistake. "
A. Generalized Convexity of Volumes (Bohr-Mollerup).

B. Coupon Collecting and Convexity.

C. Convexity of Spectral Functions.

D. Madelung’s Constant for Salt. The talk ends when I do.
B. The origin of the problem.

Consider a network objective function $p_N$:

$$p_N(q) := \sum_{\sigma \in S_N} \left( \prod_{i=1}^{N} \frac{q_{\sigma(i)}}{\sum_{j=i}^{N} q_{\sigma(j)}} \right) \left( \sum_{i=1}^{N} \frac{1}{\sum_{j=i}^{N} q_{\sigma(j)}} \right),$$

summed over all $N!$ permutations; so a typical term is

$$\left( \prod_{i=1}^{N} \frac{q_{i}}{\sum_{j=i}^{N} q_{j}} \right) \left( \sum_{i=1}^{N} \frac{1}{\sum_{j=i}^{N} q_{j}} \right).$$

For example, with $N = 3$ this is

$$q_1q_2q_3 \left( \frac{1}{q_1 + q_2 + q_3} \right) \left( \frac{1}{q_2 + q_3} \right) \left( \frac{1}{q_3} \right) \left( \frac{1}{q_1 + q_2 + q_3} + \frac{1}{q_2 + q_3} + \frac{1}{q_3} \right).$$

This arose as the cost function in a 1999 PhD thesis on coupon collection. Ian Affleck wished to show $p_N$ was convex on the positive orthant. I hoped not!
B. Doing What is Easy.

First, we try to simplify the expression for $p_N$.

The partial fraction decomposition gives:

\[
p_1(x_1) = \frac{1}{x_1},
\]

\[
p_2(x_1, x_2) = \frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1 + x_2},
\]

\[
p_3(x_1, x_2, x_3) = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} - \frac{1}{x_1 + x_2} - \frac{1}{x_2 + x_3} - \frac{1}{x_1 + x_3} + \frac{1}{x_1 + x_2 + x_3}.
\]  

(1)

Partial fractions are an arena in which computer algebra is hugely useful. Try performing the third case in (1) by hand. It is tempting to predict the “same” pattern will hold for $N = 4$. This is easy to confirm (by computer) and so we are led to:
B. A Non-convex Integrand.

**CONJECTURE.** For each $N$, the function $p_N$ given by

$$x \mapsto \int_0^1 \left\{ 1 - \prod_{k=1}^N (1 - t^{x_k}) \right\} \, dt$$

is convex. Indeed $1/p_N$ is concave.

- Randomized numeric checks were run up to $N = 20$.
- $(N > 6)$ Computing the Hessian symbolically is impossible:
- Even just the diagonal will not fit on the largest Maple.
B. A Very Convex Integrand.  (Is there a direct proof?)

A year later, Omar Hijab suggested re-expressing $p_N$ as the joint expectation of Poisson distributions. This leads to:

If $x = (x_1, \cdots, x_n)$ is a point in the positive orthant $R^n_+$, then

$$p_N(x) = \left( \prod_{i=1}^{n} x_i \right) \int_{R^n_+} e^{-\langle x, y \rangle} \max(y_1, \cdots, y_n) \, dy$$

- $\langle x, y \rangle = x_1y_1 + \cdots + x_ny_n$ is the inner product

Now $y_i \rightarrow x_i \, y_i$ and standard techniques show $1/p_N$ is concave, since the integrand is. [We can now ignore probability if we wish!]

Q “inclusion-exclusion” convexity? **OK** for $1/g(x) > 0$, $g$ concave.
“Mathematicians are a kind of Frenchmen:

whatever you say to them they translate into their own language, and right away it is something entirely different.”

(Johann Wolfgang von Goethe)

Maximen und Reflexionen, no. 1279
Outline of Convexity Talk

A. Generalized Convexity of Volumes (Bohr-Mollerup).

B. Coupon Collecting and Convexity.

C. Convexity of Spectral Functions.

D. Madelung’s Constant for Salt.

The talk ends when I do.
C. Eigenvalues of symmetric matrices (Lewis and Davis).

\( \lambda(S) \) lists **decreasingly** the (real, resp. non-negative) eigenvalues of a (symmetric, resp. PSD) \( n \)-by-\( n \) matrix \( S \).

The **Fenchel conjugate** is the convex closed function given by

\[
 f^*(x) := \sup_y \langle y, x \rangle - f(y).
\]

**Theorem (Spectral conjugacy)** If \( f : \mathbb{R}^n \mapsto (-\infty, \infty] \) is a symmetric function, it satisfies the formula \((f \circ \lambda)^* = f^* \circ \lambda\).

**Corollary [Davis/Lewis]** Suppose \( f : \mathbb{R}^n \mapsto (-\infty, \infty] \) is symmetric. The “spectral function” \( f \circ \lambda \) is closed and convex (resp. differentiable) iff \( f \) is closed and convex (resp. differentiable). [Von Neumann for norms]
C. Three Amazing Examples (Lewis).

I. Log Determinant Let $\logb(x) := -\log(x_1 x_2 \cdots x_n)$ which is clearly symmetric and convex. The corresponding spectral function is $S \mapsto -\log \det(S)$.

II. Sum of Eigenvalues Ranging over permutations, let $f_k(x) := \max_\pi \{x_{\pi(1)} + x_{\pi(2)} + \cdots + x_{\pi(k)}\}$. This is clearly symmetric and convex. The corresponding spectral function is $\sigma_k(S) := \lambda_1(S) + \lambda_2(S) + \cdots \lambda_k(S)$. In particular the largest eigenvalue, $\sigma_1$, is a continuous convex function of $S$ and is differentiable if and only if the eigenvalue is simple.
Convexity of Spectral Functions

C. Three Amazing Examples (Lewis).

III. $k$–th Largest Eigenvalue

The $k$–th largest eigenvalue may be written as

$$\mu_k(S) = \sigma_k(S) - \sigma_{k-1}(S).$$

In particular, this represents $\mu_k$ as the difference of two convex continuous, hence locally Lipschitz, functions of $S$ and so we discover the very difficult result that for each $k$, $\mu_k(S)$ is a locally Lipschitz function of $S$.

• Hard analogues exist for singular values, hyperbolic polynomials, Lie algebras, etc.
Convexity of Barrier Functions

C. A Fourth Amazing Example (Nesterov & Nemirovskii).

IV Self-concordant Barrier Functions Let $A$ be a nonempty open convex set in $R^N$. Define, for $x \in A$,

$$F_N(x) := \lambda_N((A - x)^\circ),$$

where $\lambda_N$ is $N$-dimensional Lebesgue measure and $(A - x)^\circ$ is the polar set. Then $F_N$ is an essentially Fréchet smooth, log-convex, barrier function for $A$.

- Central to modern interior point methods.
- The orthant yields $lb(x) := - \sum_{k=1}^{N} \log x_k$.
- Hilbert space analog? (JB-JV, CUP, 2009)
"He was very big in Vienna."
Outline of Convexity Talk

A. Generalized Convexity of Volumes (Bohr-Mollerup).

B. Coupon Collecting and Convexity.

C. Convexity of Spectral Functions.

D. Madelung’s Constant for Salt.

The talk ends when I do

Full details are in the three reference texts
D. Madelung’s Constant: 
David Borwein  CMS Career Award

\[
\sum_{n,m,p} \frac{(-1)^{n+m+p}}{\sqrt{n^2 + m^2 + p^2}}
\]

This polished solid silicon bronze sculpture is inspired by the work of David Borwein, his sons and colleagues, on the conditional series above for salt, Madelung's constant. This series can be summed to uncountably many constants; one is Madelung's constant for electro-chemical stability of sodium chloride. (Convexity is hidden here too!)

This constant is a period of an elliptic curve, a real surface in four dimensions. There are uncountably many ways to imagine that surface in three dimensions; one has negative gaussian curvature and is the tangible form of this sculpture. (As described by the artist.)
Peter Borwein in front of Helaman Ferguson’s work

CMS Meeting
December 2003
SFU Harbour Centre

Ferguson uses high tech tools and micro engineering at NIST to build monumental math sculptures
D. Madelung’s Constant

\[ M_3(s) := \sum_{n,m,p} \frac{(-1)^{n+m+p}}{(n^2 + m^2 + p^2)^s} \]

\[ M_2(s) := \sum_{n,m} \frac{(-1)^{n+m}}{(n^2 + m^2)^s} \]

In many texts, the potential, \( M_3(1/2) \), is ‘added’ over increasing spheres: \( \sum_{n=1}^{\infty} (-1)^n r_3(n) / \sqrt{n} \) (\( r_3(n) \) is no. of reps. of \( n \) as sum of 3 squares).

They seem(ed) not to care \( r_3(n) / \sqrt{n} \not\to 0 \! \! \! \! \)!

The sum over increasing cubes does converge to the value chemists expect of \(-1.74756459\ldots\) (by Mellin transform methods) — needs solar-system size crystal to be realistic!
Now if $M_2$ is ‘added’ over spheres ($\ell^2$ balls) the n-th term tends to zero and the sum agrees with that over increasing squares ($\ell^\infty$) but the sum over increasing diamonds ($\ell'$) diverges—Riemann sum!

\[
M_2(s) = \sum_{n,m} (-1)^{n+m} \frac{1}{(n^2 + m^2)^s}
\]

\[
M_2(1/2) = \sum_{n=1}^{\infty} (-1)^n r_2(n) / \sqrt{n}
\]

For $C$ a closed convex symmetric body set

\[
M_C(s) := \lim_{N \to \infty} \sum_{n,m \in NC} (-1)^{n+m} \frac{1}{(n^2 + m^2)^s}
\]
\[ M_C(s) = \lim_{N \to \infty} \sum'_{n,m \in NC} \frac{(-1)^{n+m}}{(n^2 + m^2)^s} \]

**Theorem** (BBP, 1998) \( M_C(s) \) exists, is analytic and is independent of \( C \) for \( \text{Re}(s) > 1/2 \). [In \( \mathbb{R}^k \) this holds for \( \text{Re}(s) > (k - 1)/2 \).]

1. \( \text{Re}(s) > 1 \) needed for absolute convergence.
2. \( M_{\{\| \cdot \|_2 \leq 1\}}(s) = -4\zeta(s)(1 - 2^{1-s})L_4(s) \) converges precisely for \( \text{Re}(s) > 1/4 \). This relies on correctness of the wonderful exact determination of the average size of \( r_2(n) \) [Cappell and Shaneson, 2007]: the number of lattice points in a circle of radius \( \sqrt{t} \) is \( \pi t + O\left(t^{1/4 + \varepsilon}\right) \) (best possible).
A. Generalized Convexity of Volumes (Bohr-Mollerup).
B. Coupon Collecting and Convexity.
C. Convexity of Spectral Functions.
D. Madelung’s Constant for Salt.

E. Entropy and NMR.
F. Inequalities and the Maximum Principle.
G. Trefethen’s 4th Digit-Challenge Problem.
E. CONVEX CONJUGATES and NMR (MRI)

The *Hoch and Stern information measure* in complex $N$-space is $H(z) := \sum_{j=1}^{N} h(z_j/b)$ where $h$ is convex and given (for scaling $b$) by

$$h(z) := |z| \ln \left( |z| + \sqrt{1 + |z|^2} \right) - \sqrt{1 + |z|^2}$$

for quantum theoretic (NMR) reasons. Recall the *Fenchel-Legendre conjugate*

$$f^*(y) = \sup_{x} \langle x, y \rangle - f(x).$$

Our symbolic convex analysis package produced

$$h^*(z) = \cosh(|z|).$$

Compare the *Shannon entropy* $z \ln(z) - z$ whose conjugate is $\exp(z)$.

I'd never have tried by hand! Effective dual algorithms are now possible!
Knowing `Closed Forms' Helps

For example

\[(\exp \exp)^*(y) = y \ln(y) - y\{W(y) + W(y)^{-1}\}\]

where *Maple* or *Mathematica* recognize the complex *Lambert W function* given by

\[W(x)e^{W(x)} = x.\]

Thus, the conjugate's series is:

\[-1 + (\ln(y) - 1) y - \frac{1}{2} y^2 + \frac{1}{3} y^3 - \frac{3}{8} y^4 + \frac{8}{15} y^5 + O(y^6).\]

The literature is all in the last decade since W got a name!
WHAT is ENTROPY?

Despite the narrative force that the concept of entropy appears to evoke in everyday writing, in scientific writing entropy remains a thermodynamic quantity and a mathematical formula that numerically quantifies disorder. When the American scientist Claude Shannon found that the mathematical formula of Boltzmann defined a useful quantity in information theory, he hesitated to name this newly discovered quantity entropy because of its philosophical baggage. The mathematician John Von Neumann encouraged Shannon to go ahead with the name entropy, however, since “no one knows what entropy is, so in a debate you will always have the advantage.”

The American Heritage Book of English Usage, p. 158
Information Theoretic Characterizations Abound

**Theorem.** Up to a positive scalar multiple

$$H(\overrightarrow{p}) = -\sum_{k=1}^{N} p_k \log p_k$$

is the unique continuous function on finite probabilities such that [a.] **Uncertainly grows:**

$$H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$$

increases with $n$.

[b.] **Subordinate choices are respected:** for distributions $\overrightarrow{p_1}$ and $\overrightarrow{p_2}$ and $0 < p < 1$,

$$H \left( p \overrightarrow{p_1}, (1 - p) \overrightarrow{p_2} \right) = p H(\overrightarrow{p_1}) + (1 - p) H(\overrightarrow{p_2}).$$
Entropy
F. Inequalities and the Maximum Principle

- Consider the two means
  \[ \mathcal{L}^{-1}(x, y) := \frac{x - y}{\ln(x) - \ln(y)} \]
  and
  \[ \mathcal{M}(x, y) := \frac{\sqrt[3]{x^2 + y^2}}{2} \]

A conformal function estimated reduced to

\[ \mathcal{L}(\mathcal{M}(x, 1), \sqrt{x}) > \mathcal{L}(x, 1) > \mathcal{L}(\mathcal{M}(x, 1), 1) \]
for \( 0 < x < 1 \).

We first discuss showing

\[ \mathcal{E}(x) := \mathcal{L}(\mathcal{M}(x, 1), \sqrt{x}) - \mathcal{L}(x, 1) > 0. \]
I. Numeric/Symbolic Methods

- $\lim_{x \to 0^+} \mathcal{E}(x) = \infty$.

- Newton-like iteration shows that $\mathcal{E}(x) > 0$ on $[0.0, 0.9]$.

- Taylor series shows $\mathcal{E}(x)$ has 4 zeroes at 1.

  $= \frac{7}{51840} (x - 1)^4 - \frac{7}{20736} (x - 1)^5 + O((x - 1)^6)$

- Maximum Principle shows there are no more zeroes inside $C := \{z : |z - 1| = \frac{1}{4}\}$:

  $$\frac{1}{2\pi i} \int_C \mathcal{E}' = \#(\mathcal{E}^{-1}(0); C)$$
II. Graphic/Symbolic Methods

Consider the opposite (cruder) inequality

$$\Lambda := \mathcal{L}(x, 1) - \mathcal{L}(\mathcal{M}(x, 1), 1) > 0.$$ 

We may observe that it holds since:

- $\mathcal{M}$ is a mean;
- $\mathcal{L}(x, 1)$ decreases with $x$.

- There is an algorithm (Collins) for universal algebraic inequalities.
# 4. What is the global minimum of the function

\[
\exp(\sin(50x)) + \sin(60e^y) + \sin(70\sin x)
\]

\[
+ \sin(\sin(80y)) - \sin(10(x + y)) + (x^2 + y^2)/4?
\]

- no bounds are given.
This model has been numerically solved by LGO, MathOptimizer, MathOptimizer Pro, TOMLAB /LGO, and the Maple GOT (by Janos Pinter who provide the pictures).

The solution found agrees to 10 places with the announced solution (the latter was originally based (provably) on a huge grid sampling effort, interval analysis and local search).

\[ x^* \approx (-0.024627\ldots, 0.211789\ldots) \]
\[ f^* \approx -3.30687\ldots \]

Close-up picture near global solution: the problem still looks rather difficult ... Mathematica 6 can solve this by “zooming”!


“The object of mathematical rigor is to sanction and legitimize the conquests of intuition, and there was never any other object for it.”