Relative Convexity

J. A. Palmer

August 20, 2003

According to [1], the foundations of the theory of convex functions were laid by Jensen in 1905-6, though it is noted that Hölder had earlier considered inequalities essentially similar to those defined by Jensen as convex. The basic inequality defining convex functions is the following,

$$\varphi(EX) \;\le\; E\,\varphi(X)$$

which, for invertible φ can also be expressed as,

$$EX \;\le\; \varphi^{-1}\big(E\,\varphi(X)\big)$$

The term $\varphi^{-1}\big(E\,\varphi(X)\big)$ can be seen as a generalized mean value [2],[1, ch.3]. The L_p norms $(E|X|^p)^{1/p}$ for example fit this framework. The notion of convexity was generalized by B. Jessen in [2] to compare two functions in terms of the means defined by them. An increasing function φ is defined to be *convex with respect to* another increasing function ψ if,

$$\psi^{-1}\big(E\,\psi(X)\big) \;\le\; \varphi^{-1}\big(E\,\varphi(X)\big)$$

This can be written,

$$\varphi\circ\psi^{-1}\big(E\,\psi(X)\big) \;\le\; E\,\varphi\circ\psi^{-1}(\psi(X))$$

which shows that φ is convex with respect to ψ if and only if $\varphi\circ\psi^{-1}$ is convex. The latter formulation, used in [2, 1, 3, 4], does not require φ to be invertible. This formulation is not antisymmetric and thus does not define a partial ordering. Here we shall use an independently derived formulation that is similar, but is anti-symmetric and defines a partial ordering, while not requiring invertibility of either function in the relation.

We now give an informal derivation of generalized convexity from a different perspective. Let $f : \mathbb{R} \to \mathbb{R}$ be increasing on the interval (a, b). The basic criterion for convexity of f is $f(\alpha x + \bar{\alpha} y) \le \alpha x + \bar{\alpha} y$ for all $x, y \in (a, b)$, $0 \le \alpha \le 1$. Intuitively, this criterion asserts that for any two points x and y in (a, b), the function value at all intervening points is less than the value of the linear function defined to match the value of f at the points x and y. In the intervals

(a, x) and (y, b), the value of the convex function f will be greater than that of the linear function [5]. Analytically, we have f convex on (a, b) if,

$$f(y) \le f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(y - x_0) \quad \forall\, x_0 \le y \le x_1$$

The inequality is reversed for y in (a, x_0), or (x_1, b). Letting $x_1 \to x_0 \equiv x$, we have the ordinary definition of convexity for differentiable f,

$$f(y) \ge f(x) + f'(x)(y - x) \quad \forall\, x, y \in (a, b)$$

Thus convexity of a function on an interval can be seen as a relationship between the function and a linear model of the function based on the function value and first derivative.

Let $g : \mathbb{R} \to \mathbb{R}$ be strictly increasing on (a, b). From considerations similar to those given in the linear case, we can define f to be convex with respect to g on (a, b) if a model of g using an affine transform of f, given by $\alpha f + \beta$, defined so that f and g are equal at two given points, behaves in a manner similar to the line in the convex case (see Figure 1.)

For any three points $x_0 < y < x_1$ in the interval (a, b), we then have,

$$f(y) \le f(x_0) + \frac{f(x_1) - f(x_0)}{g(x_1) - g(x_0)}\big(g(y) - g(x_0)\big) \quad \forall y \in (x_0, x_1)$$

with the inequality reversed in (a, b) outside of (x_0, x_1). Again letting $x_1 \to x_0 \equiv x$, we have f convex relative to g on (a, b) if,

$$f(y) \ge f(x) + \frac{f'(x)}{g'(x)}\big(g(y) - g(x)\big) \quad \forall\, x, y \in (a, b) \tag{0.1}$$

It can be verified that (0.1) is equivalent to the differential definition of the convexity of the composite function $f \circ g^{-1}$ on the interval $(g(a), g(b))$. Thus this formulation agrees with the generalized mean value formulation.

If f and g are twice differentiable on (a, b), we can use the second derivative criterion for convexity to derive a simple criterion for f convex with respect to g. It can be verified that the condition,

$$D^2[f \circ g^{-1}](x) \ge 0 \quad \forall\, x \in (g(a), g(b))$$

for f and g increasing, is equivalent to,

$$\frac{f''(x)}{f'(x)} \ge \frac{g''(x)}{g'(x)} \quad \forall\, x \in (a, b) \tag{0.2}$$

These facts were rediscovered in the course of our research, but have been known for some time. The second differential criterion can be found in [6, 7].

Various other related generalizations of convexity have been proposed. Another line of development starts from the theory of convexity of higher order

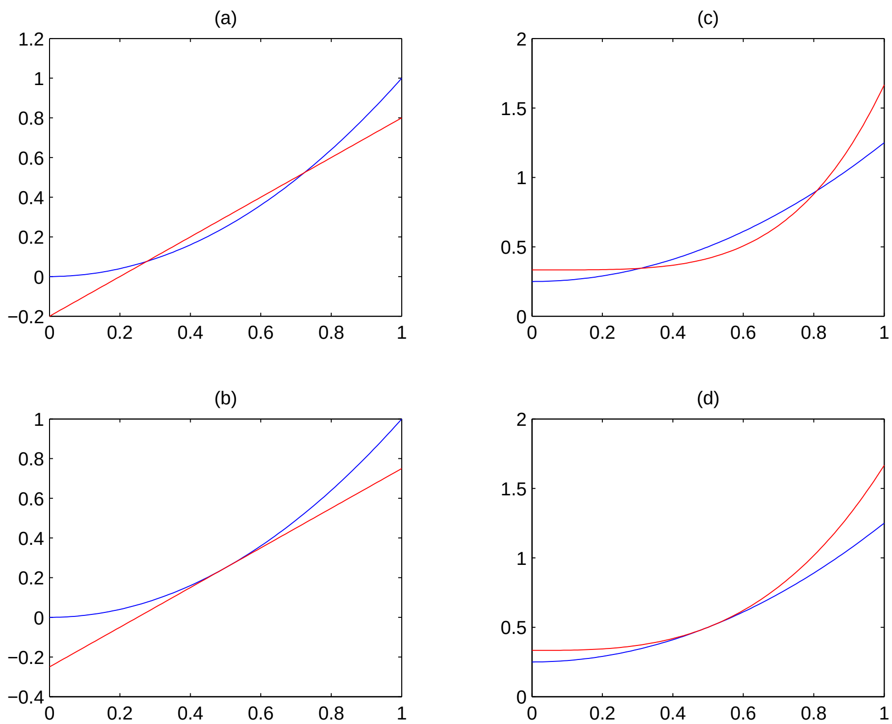

Figure 1: (a) A convex function "bends upward" with respect to a linear function when defined to have the same value at two distinct points. (b) When defined to have the same value and the same slope at a point, a convex function lies completely above a linear function. (c) x^4 (red) is convex with respect to x^2 (blue), and thus $\alpha x^4 + \beta$ "bends more" than $\gamma x^2 + \delta$ when the functions are equated at two distinct points. (c) When the two functions are set to have the same value and the same slope at a point, the quartic lies entirely above the quadratic.

considered in Nörlund [8] and E. Hopf [9] and generalized by Popoviciu [10]. This line was continued by Karlin [11, 12]. The emphasis in these developments is on systems of functions especially for use in approximation theory. The generalization of convexity given in [2] can be considered as a special case of this theory. Beckenbach [13] considers a further generalization of convexity as a relationship between functions.

Relative convexity

For invertible g, f is defined to be *convex with respect to* g if $f \circ g^{-1}$ is convex on $(g(a), g(b))$ [2, 1, 3, 4]. As mentioned, this does not in general imply that g is concave with respect to f. A slightly modified definition allows us to define a reciprocal relation that induces a partial ordering, and is not limited to invertible g. We refer to the modified definition as relative convexity.

Definition 1. *f is **convex relative to** g if there exists a function h that is convex and strictly increasing on the range of g and such that $f = h(g)$. f is **concave relative to** g if g is convex relative to f.*

It is immediate that f is *concave relative to* g iff h is strictly increasing and concave. Since $f = h(g)$ with h strictly increasing, f and g must be increasing and decreasing on the same intervals. We shall say that such functions are *isotonic*, and are members of the same isotonic class. It is thus necessary for f and g to be isotonic for f to be convex or concave relative to g. If f and g are affinely related with positive multiplier, i.e. $\exists\, \alpha > 0, \beta \in \mathbb{R}$ such that $g = \alpha f + \beta$, then we shall say that f and g are of the same *type*. A function type defines an equivalence class. Functions with the same type are both convex and concave with respect to each other.

We now show that this relation induces a partial ordering on isotonic classes. Let f convex with respect to g be denoted $f \succ g$, with f concave relative to g denoted $f \prec g$.

Theorem 1 (Partial Ordering)**.** *The relative convexity relation induces a partial ordering of function types on isotonic classes.*

Proof. We show that the relation is reflexive, antisymmetric, and transitive.
Reflexivity: $\forall\, f$ we have $f = h(f)$ where h is the identity function, which is strictly increasing and convex. Thus $f \prec f$.
Antisymmetry: Suppose $f \succ g$ and $f \prec g$. Then $f = h_1(g)$ with h_1 strictly increasing increasing and convex, and $f = h_2(g)$ with h_2 strictly increasing and concave. Thus $h_1 = h_2$ must be linear (with strictly positive slope), so that f and g are of the same type.
Transitivity: Suppose $f \succ g$ and $g \succ w$. Then $f = h_1(g)$ and $g = h_2(w)$ with h_1 and h_2 strictly increasing and convex on their respective domains. Thus $f = h_1(h_2(w))$, and since composition of strictly increasing convex functions

produces another strictly increasing convex function, $h_1 \circ h_2$ is increasing and convex and $f \succ w$. □

Relative convexity is transitive, but comparability under the relation is not. Thus even though we may have $f \prec h$ and $g \prec h$, this does not imply that f can be ordered with respect to g. If, however, $f \prec h$ and $h \prec g$, then $f \prec g$ by transitivity.

The following theorem gives an equivalent criterion that is useful in proving properties of the relation without reference to the function h.

Theorem 2. *f is convex relative to g on (a, b) iff $\forall\, x \in (a, b)\, \exists\, \lambda \in [0, \infty]$ such that,*

$$f(y) - f(x) \geq \lambda\big(g(y) - g(x)\big) \quad \forall\, y \in (a, b) \tag{0.3}$$

Proof. If $f = h(g)$ and h is strictly increasing and convex on the range of g, then from the ordinary definition of convexity, we have $\forall\, x \in (a, b)\, \exists\, \lambda > 0$ such that,

$$f(y) - f(x) = h(g(y)) - h(g(x)) \geq \lambda\big(g(y) - g(x)\big) \quad \forall\, y \in (a, b) \tag{0.4}$$

Conversely, suppose (0.3) holds and define the function h that maps the range of g to the range of f by mapping $g(x)$ to $f(x)$ for each $x \in (a, b)$. To show that this defines a single valued mapping, i.e. a function, we show that (0.3) implies that if $g(x) = g(y)$, then $f(x) = f(y)$. If $g(x) = g(y)$, then we have $f(y) - f(x) \geq \lambda(x) \cdot 0 = 0$, and $f(x) - f(y) \geq \lambda(y) \cdot 0 = 0$, so that $f(x) = f(y)$. Thus all x values that map to the same g also map to the same f, so that there is a unique value in range of f associated with each value in the range of g. Then (0.3) shows that h is convex on the range of g. And since (0.3) also implies that if $g(y) > g(x)$, then $f(y) > f(x)$ or $h(g(y)) > h(g(x))$, we also have that h is strictly increasing. □

If f and g are differentiable and $f \succ g$, then $\lambda = h'(g(x))$. Since $f(x) = h(g(x))$, we have $f'(x) = h'(g(x))g'(x)$, or $h'(g(x)) = f'(x)/g'(x)$. Thus for $f \succ g$, we have,

$$f(y) - f(x) \geq \frac{f'(x)}{g'(x)}\big(g(y) - g(x)\big) \quad \forall\, y \in (a, b) \tag{0.5}$$

For $f \prec g$,

$$f(y) - f(x) \leq \frac{f'(x)}{g'(x)}\big(g(y) - g(x)\big) \quad \forall\, y \in (a, b) \tag{0.6}$$

A definition similar to (0.3) is given in [2], with λ allowed to be negative. We require λ to be positive to ensure that f is convex relative to g if and only if g is concave relative to f. This also has the effect of ensuring that the subdifferential of the interior of range of g contains only strictly positive λ, which allows us to define the partial ordering. It is immediate from Theorem 2 that $f \prec g$ if and only if $-f \succ -g$.

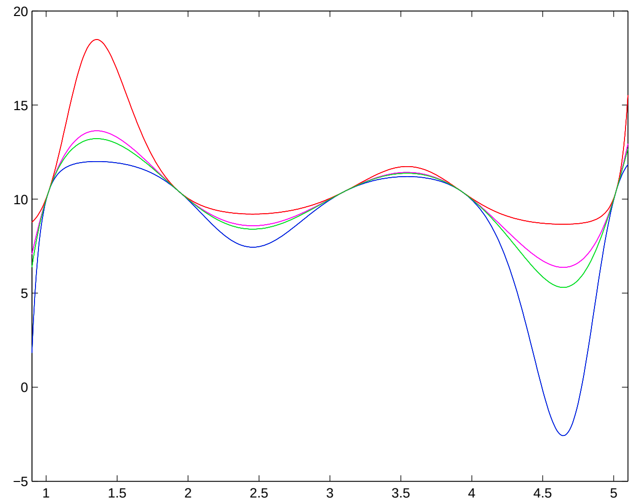

Figure 2: A chain of functions in a partial ordering, $-e^{-f} \prec \log(f) \prec f \prec e^f$ on the interval shown, where f is a fifth degree polynomial.

As a simple example of the use of Theorem 2, we prove the following lemma on addition.

Lemma 1. (a) *If* $f, g \prec h$, *then* $f + g \prec h$. *Likewise if* $f, g \succ h$, *then* $f + g \succ h$. (b) *If* $f \prec g$, *then* $f \prec f + g \prec g$.

Proof. (a) From the definition, we have $f(y) - f(x) \leq \lambda(h(y) - h(x))$ and $g(y) - g(x) \leq \xi(h(y) - h(x))$ for all y, with $\lambda, \xi > 0$. Adding these inequalities, we get $f(y) + g(y) - (f(x) + g(x)) \leq (\lambda + \xi)(h(y) - h(x))$, with $\lambda + \xi > 0$. Thus, $f + g \prec h$. Similarly, $f, g \succ h$ implies $f + g \succ h$. (b) We have $f(y) - f(x) \leq \lambda(g(y) - g(x))$ and $f(y) - f(x) \leq f(y) - f(x)$. Thus $f(y) + g(y) - (f(x) + g(x)) \leq (1 + \lambda)(g(y) - g(x))$, and $f + g \prec g$. Similarly, with $f(y) - f(x) \geq f(y) - f(x)$, we have $f(y) + g(y) - (f(x) + g(x)) \geq (1 + \lambda^{-1})(f(y) - f(x))$, or $f(y) - f(x) \leq \frac{\lambda}{1+\lambda}\big(f(y) + g(y) - (f(x) + g(x))\big)$, and $f \prec f + g$. □

Lemma 1a shows that the set of functions convex (or concave) relative to h is a convex cone. The f and g in the definition need not be comparable. Lemma 1b shows that if f and g are comparable, then adding f and g in a sense mixes the convexity of the functions. We will make this more explicit later. Lemma 1 can easily be extended to integrals over a parameterized set of functions.

The following theorem is important in the application of the theory to estimation.

Theorem 3. *If* $f, g \succ h$, *then* $\log(e^f + e^g) \succ h$.

Proof. We have $f(y) \geq f(x) + \lambda(h(y) - h(x))$ and $g(y) \geq g(x) + \xi(h(y) - h(x))$, and thus,

$$e^{f(y)} \geq e^{f(x)} e^{\lambda(h(y)-h(x))} \quad \text{and} \quad e^{g(y)} \geq e^{g(x)} e^{\xi(h(y)-h(x))}$$

Adding these inequalities and dividing by $\exp f(x) + \exp g(x)$, we get,

$$\begin{aligned} \frac{e^{f(y)} + e^{g(y)}}{e^{f(x)} + e^{g(x)}} &\geq \left(\frac{e^{f(x)}}{e^{f(x)} + e^{g(x)}}\right) e^{\lambda(h(y)-h(x))} + \left(\frac{e^{g(x)}}{e^{f(x)} + e^{g(x)}}\right) e^{\xi(h(y)-h(x))} \\ &\geq e^{\gamma(h(y)-h(x))} \end{aligned}$$

where $\gamma = [\lambda \exp f(x) + \xi \exp g(x)]/[\exp f(x) + \exp g(x)] > 0$, and the second inequality follows from the convexity of the exponential function. Taking the logarithm of both sides gives the desired result. $\square$

If f and g are twice differentiable, then a simple criterion can be derived for $f \succ g$.

Theorem 4. *If f and g are twice differentiable on (a, b), then,*

$$f \succ g \qquad \textit{iff} \qquad \frac{f''}{|f'|} \geq \frac{g''}{|g'|}$$

Proof. (a) Let the function h be defined by $f = h(g)$. We have $f \succ g$ if and only if h'' is non-negative on the range of g. Since,

$$h''(g(x)) = \frac{|f'(x)|}{g'(x)^2} \left(\frac{f''(x)}{|f'(x)|} - \frac{g''(x)}{|g'(x)|}\right)$$

we have h'' non-negative on the range of g if and only if $\frac{f''(x)}{|f'(x)|} \geq \frac{g''(x)}{|g'(x)|} \quad \forall x \in (a, b)$. $\square$

Theorem 4 provides a measure that can be used to determine the relative convexity ordering of functions. We have $f \succ g$ on (a, b) if $f''/|f'| \geq g''/|g'|$ uniformly on (a, b). Thus we can compare several functions simultaneously rather than pairwise. The measure can be seen as a sort of curvature measure that is invariant to affine scaling of the function, or invariant over types. We shall refer to $f''/|f'|$ as the *function curvature* of f. The function curvature can be related to the geometric curvature of the level curves of f for multivariate functions.

The following theorem gives a relationship between the convexity of convex and concave functions to their Fenchel-Legendre conjugates.

Theorem 5. *$f \succ x^2$ if and only if $f^* \prec \phi^2$. Also, $f \prec \log$ if and only if $f^* \succ \log$.*

Proof. For simplicity suppose f is twice differentiable, though the result holds in the non-differentiable case as well. We have $f \succ x^2$ if and only if $f''(x)/f'(x) \geq 1/x$ for $x \in (0, \infty)$. Since under our assumptions f^* is twice differentiable, we have $f \succ x^2$ if and only if $f''(x)/f'(x) \geq 1/x$ for $x \in (0, \infty)$. But this holds if and only if,

$$\frac{f^{*\prime\prime}(\phi)}{f^{*\prime}(\phi)} = \frac{1}{x(\phi) f''(x(\phi))} \leq \frac{1}{f'(x(\phi))} = \frac{1}{\phi}$$

that is, if and only if $f^* \prec \phi^2$.

Similarly, $f \prec \log$ if and only if $f''(x)/f'(x) \leq -1/x$ for $x \in (0, \infty)$. But this holds if and only if,

$$\frac{f^{*\prime\prime}(\phi)}{f^{*\prime}(\phi)} = \frac{1}{x(\phi) f''(x(\phi))} \geq -\frac{1}{f'(x(\phi))} = -\frac{1}{\phi}$$

that is, if and only if $f^* \succ \log$. $\square$

Thus the two self dual functions $\frac{1}{2}x^2$ and $e \log x$ are central in a sense among convex and concave functions respectively. This is similar to the result that an increasing function f is convex if and only if f^{-1} is concave. These topics are further developed in [14].

The symmetry about $\frac{1}{2}x^2$ implies that if $p(x)$ is strongly sub-gaussian, then the dual problem will be equivalent to estimation with a strongly super-gaussian density, and thus amenable to the results given here.

References

[1] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge, 1934,1959.

[2] B. Jessen. Bemaerkinger om konvekse functioner og uligheder imellem middelvaerdier. I. *Matematisk tidsskrift. B.*, pages 17–28, 1931.

[3] A. W. Roberts and D. E. Varberg. *Convex Functions.* Academic Press, New York, 1973.

[4] J. E. Pecaric, F. Proschan, and Y. L. Tong. *Convex Functions, Partial Orderings, and Statistical Applications.* Academic Press, Inc., 1992.

[5] J. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms I.* Sringer-Verlag, 1993.

[6] G. T. Cargo. Comparable means and generalized convexity. *J. Math. Anal. Appl.*, 12:387–392, 1965.

[7] J. G. Mikusinski. Sur les moyenned de la forme $\psi^{-1}\left[\sum q\psi(x)\right]$. *Studia Math.*, 10:90–96, 1949.

[8] N. E. Nörlund. *Vorlesungen über differenzenrechnung.* Chelsea Publishing Company, New York, 1954,1923.

[9] E. Hopf. Über die zusammenhänge zwischen gewissen höheren differenzenquotienten reeler funktionen einer reellen variablen und deren differenzierbarkeitseigenschaften, 1926.

[10] T. Popoviciu. Notes sur les fonctions convexes d'ordre superieur. *Mathematica*, 12:81–92, 1936.

[11] S. Karlin and A. Novikoff. Generalized convex inequalities. *Pacific J. Math.*, 13:1251–1279, 1963.

[12] S. Karlin and W. J. Studden. *Tchebycheff Systems: with applications in analysis and statistics*. Interscience, New York, 1966.

[13] E. F. Beckenbach. Generalized convex functions. *Bull. Am. Math. Soc.*, 43:363–371, 1937.

[14] J. A. Palmer. Function curvature, relative convexity, and conjugate curvature. Technical report, ECE Dept., UCSD, 2002.