Convexity
Statistics 710
September 19, 2006

Prologue. There are interesting statistical problems in which an unknown function f, say, is to be estimated, subject to certain shape restrictions. For example, consider a regression problem

$$y_i = f(t_i) + \epsilon_i, \ i = 1, \cdots, n,$$

where $\epsilon_1, \cdots, \epsilon_n$ are random errors with means 0 and finite variances and f is known (only) to be non-decreasing, or convex, or to satisfy some other shape restriction. For example, an average response may well be increasing with time, even though individual measures may not. Global warming provides a specific example. Assuming that the error are i.i.d. for simplicity, and letting $\theta_i = f(t_i)$, least squares estimation of θ requires minimizing a function of the form

$$\sum_{i=1}^{n} [y_i - \theta_i]^2$$

with respect to $\theta = [\theta_1, \cdots, \theta_n]'$. Let $\mathcal{F}$ denote the class of allowable f. Then $\theta \in \Omega := \{[f(t_1), \cdots, f(t_n) : f \in \mathcal{F}\}$, and the latter set is often convex. For example, if f is known to be non-dreasing, then

$$\Omega = \{\theta : -\infty < \theta_1 \leq \cdots \leq \theta_n < \infty\},$$

a convex subset of $I\!R^n$.

Hilbert Spaces. Let $\mathcal{H}$ denote a (real) Hilbert space. Thus, $\mathcal{H}$ is a linear space to together with a function $\langle \cdot, \cdot \rangle \to R$ for which

$$\langle x, y \rangle = \langle y, x \rangle,$$
$$\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle$$

for all $x, y, z \in \mathcal{H}$ and $\alpha, \beta \in I\!R$, and

$$\langle x, x \rangle > 0$$

for all $x \neq 0$; and the norm in $\mathcal{H}$ is defined by

$$\|x\| = \sqrt{\langle x, x \rangle};$$

and $\mathcal{H}$ is complete in the metric $d(x, y) = \|x - y\|$.

Example 1 *a) Euclidean Space.* $\mathcal{H} = I\!R^m$ with $\langle x, y \rangle = x'y = x_1 y_1 + \cdots + x_m y_m$ or, more generally, $\langle x, y \rangle = x'Ay$, where A is a symmetric, positive definite matrix.

b) More on Euclidean Space. $\mathcal{H}$ be the set of $m \times m$ matrices with $\langle x, y \rangle = \mathrm{tr}(xy')$.

c) L^2. Let $(\Omega, \mathcal{A}, \mu)$ is a measure space; and let $L^2(\Omega, \mathcal{A}, \mu)$ be the set of (equivalence classes) of square integrable functions with

$$\langle f, g \rangle = \int fg d\mu.$$

d) Sobelov space. Let $\mathcal{S}_2$ be the set of twice differentiable functions $f : [0, 1] \to I\!R$ for which $\int_0^1 f''(x)^2 dx < \infty$, and let

$$\langle f, g \rangle = f(0)g(0) + f'(0)g'(0) + \int_0^1 f(x)g(x)dx.$$

Remarks. The primary uses of convexity will be with $\mathcal{H} = R^m$, and little is lost by focussing on this case. On the other hand many of the definitions extend to (even) more general linear spaces.

Some Properties. Let $\mathcal{H}$ is a Hilbert space. Then *Schwarz' Inequality* asserts

$$\langle x, y \rangle \leq \|x\| \times \|y\|$$

for all $x, y \in \mathcal{H}$; and the *Parallelogram Laws* assert:

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + \|y\|^2,$$
$$\|x + y\|^2 - \|x - y\|^2 = 4\langle x, y \rangle$$

for all $x, y \in \mathcal{H}$.

Linear Functional. A function $f : \mathcal{H} \to I\!R$ is said to be a linear functional if $f[\alpha x + \beta y] = \alpha f(x) + \beta f(y)$ whenever $x, y \in \mathcal{H}$ and $\alpha, \beta \in I\!R$. For example, if $y \in \mathcal{H}$, then

$$f(x) = \langle x, y \rangle, \ x \in \mathcal{H} \tag{1}$$

defines a continuous linear functional, a linear functional that is continuous in the metric. It is a theorem that *any continuous linear functional can be represented in the form (1) for some $y \in \mathcal{H}$ depending on the linear functional.*

Convex Sets. Recall that a subset $\mathcal{H}_0 \subseteq \mathcal{H}$ is called a linear subspace if it is closed under addition and scalar multiplication: that is $\alpha x + \beta y \in \mathcal{H}_0$ whenever $x, y \in \mathcal{H}_0$ and $\alpha, \beta \in \mathbb{R}$. Similarly, a subset $C \subseteq \mathcal{L}$ is said to *convex* if it contains the line joining any two of its elements; that is $\alpha x + (1-\alpha)y \in C$ whenever $x, y \in C$ and $0 \le \alpha \le 1$. A convex set C is said to be a *cone* if $\alpha x \in C$ whenever $x \in C$ and $\alpha \ge 0$ in which case $\alpha x + \beta y \in C$ whenever $x, y \in C$ and $0 \le \alpha, \beta < \infty$. Any linear subspace is a convex cone. Any ball, $B = \{x \in \mathcal{H} : \|x\| \le c\}$ is a convex set (since $\|\alpha x + (1-\alpha)y\| \le \alpha\|x\| + (1-\alpha)\|y\| \le c$ whenever $\|x\|, \|y\| \le c$, but not a convex cone.

Example 2 If $f_i,\ i \in I$ are continuous linear functionals, then

$$C = \{x \in \mathcal{H} : f_i(x) \ge 0,\ \text{for all } i \in I\}$$

is a convex cone.

Projections. *If C is a closed convex set and $z \in \mathcal{H}$, then there is a $x \in C$ for which*

$$\|x - z\| = \inf_{y \in C} \|y - z\|; \tag{2}$$

and x is the unique point in C for which

$$\langle y - x, x - z \rangle \ge 0 \tag{3}$$

for all $y \in C$. For the existence, let $\imath$ denote the infimum and let $y_n \in C$ be a sequence for which $\|y_n - z\| \to \imath$. Then

$$\|y_n - y_m\|^2 = \|y_n - z\|^2 + \|y_m - z\|^2 - 2\langle y_n - z, y_m - z\rangle$$

and

$$\begin{aligned} 2\langle y_n - z, y_m - z\rangle &= \frac{\|y_n + y_m - 2z\|^2 - \|y_n - y_m\|^2}{2} \\ &= 2\|\frac{y_n + y_m}{2} - z\|^2 - \frac{\|y_n - y_m\|^2}{2} \end{aligned}$$

by the parallelogram law. So,

$$\begin{aligned} \frac{1}{2}\|y_n - y_m\|^2 &= \|y_n - z\|^2 + \|y_m - z\|^2 - 2\|\frac{y_n + y_m}{2} - z\|^2 \\ &\le \|y_n - z\|^2 + \|y_m - z\|^2 - 2\imath^2 \\ &\to 0 \end{aligned}$$

as $m, n \to \infty$. Thus, y_n is a Cauchy sequence and, therefore, a convergent sequence. Let $x = \lim_{n\to\infty} y_n$. Then $x \in C$, since C is closed, and $\|x - z\| = \lim_{n\to\infty} \|y_n - z\| = \imath$. For (3): If $y \in C$ and $0 < t < 1$, then $x + t(y - x) = (1-t)x + ty \in C$, so that

$$0 \le \|[x + t(y - x)] - z\|^2 - \|x - z\|^2 = 2t\langle y - x, x - z\rangle + t^2\|y - x\|^2$$

and, therefore,

$$\langle y - x, x - z\rangle \geq -t\|y - x\|^2/2.$$

The inequality (3) then follows by letting $t \to 0$. Finally, it must be shown that only one $x \in C$ can satisfy (3). If x_1 and x_2 both satisfy (3), then

$$\langle x_2 - x_1, x_1 - z\rangle \geq 0,$$
$$\langle x_1 - x_2, x_2 - z\rangle \geq 0$$

by setting $y = x_2$ and $y = x_1$. Adding these two inequalities then gives

$$-\|x_1 - x_2\|^2 = \langle x_1 - x_2, x_2 - x_1\rangle \geq 0$$

and, therefore, $x_1 = x_2$. The element $x \in \mathcal{C}$ is called *the projection of z onto $\mathcal{C}$* and denoted by $\Pi_C(z)$.

If C is a convex cone, then setting $y = x/2$ and $y = 2x$ in (3) shows that $\langle x, x - z\rangle = 0$. So, x is the unique element of C for which

$$x \in C \qquad \text{and} \qquad \langle y, z - x\rangle = 0 \tag{4}$$

for all $y \in C$; and if C is a linear subspace, then $\pm y \in C$ whenever $y \in C$, so that (4) becomes

$$\langle x, z - x\rangle = 0 \qquad \text{and} \qquad \langle y, z - x\rangle \leq 0 \tag{5}$$

The Separation Theorems. For these suppose that $\mathcal{H}$ is a finite dimensional linear space–for example, $\mathcal{H} = R^m$.

Te Supporting Hyerplane Theorem: If C is a convex set and $z \notin C^o$, the interior of C, then there is an $0 \neq w \in \mathcal{H}$ for which

$$\langle w, y\rangle \geq \langle w, z\rangle \tag{6}$$

for all $y \in C$. Suppose first the $z \notin \bar{C}$, the closure of C, and let $x = \Pi_C(z)$ and $w = x - z$. Then, $w \neq 0$, and

$$\langle w, z\rangle - \langle w, y\rangle = \langle x - z, z - y\rangle = \langle x - z, z - x\rangle + \langle x - z, x - y\rangle \leq 0,$$

for all $y \in C$ by (3). If $z \in \bar{C} - C^o$, then there are $z_n \in \bar{C}'$ for which $\|z_n - z\| \to 0$ as $n \to \infty$ and $w_n \in \mathcal{H}$ for which $\|w_n\| = 1$ for all n and $\langle w_n, y\rangle \geq \langle w_n, z\rangle$ for all $y \in C$. The sequence w_n is precompact, since it is bounded; and if w denotes any limit point, then $\|w\| = 1$ and (6) holds.

The Separating Hyperplane Theorem: C_1 and C_2 are disjoint convex sets, then there is an $0 \neq w \in \mathcal{H}$ for which

$$\langle w, y_1 \rangle \geq \langle w, y_2 \rangle \tag{7}$$

for all $y_1 \in C_1$ and $y_2 \in C_2$. To see this let $C = \{y_1 - y_2 : y_1 \in C_1,\ y_2 \in C_2\}$. Then C is a convex set for which $0 \notin C$. So, by the Supporting Hyperplane Theorem, there is an $w \neq 0$ for which $\langle w, y \rangle \geq \langle w, 0 \rangle = 0$, and (7) follows by writing $y = y_1 - y_2$.

Problem 1 *In the Supporting Hyperplane Theorem, show that if $z \notin \bar{C}$, then w may be chosen so that* $\inf y \in C \langle w, y \rangle > \langle w, z \rangle$.

Convex Functions. If C is a convex subset of $\mathcal{H}$ a functon $f : C \to I\!\!R$ is said to be convex if $f[\alpha x + (1-\alpha)y] \leq \alpha f(x) + (1-\alpha) f(y)$ whenever $x, y \in C$ and $0 \leq \alpha \leq 1$; and f is said to be strictly convex if there is strict inequality whenever $x \neq y$ and $0 < \alpha < 1$.

Clearly if $f : C \to I\!\!R$ is (strictly) convex and $C_0 \subseteq C$ is a convex subset of C, then the restriction of f to C_0 is convex. Conversely, if C is convex, $f : C \to I\!\!R$, and the restriction of f to every line $L(x, y) = \{\alpha x + (1-\alpha)y : 0 \leq \alpha \leq 1\}$ is strictly convex, then f is (strictly) convex. For if $x, y \in C$ and $g(t) = f[tx + (1-t)y]$ is convex in $0 \leq t \leq 1$, then

$$f[tx + (1-t)y] = g(t) \leq tg(1) + (1-g)(0) = tf(y) + (1-t)f(x).$$

Derivatives of Convex Functions. *If $-\infty \leq a < b \leq \infty$ and $f : (a, b) \to I\!\!R$ is convex, then*

$$-\infty < f'_\ell(y) = \lim_{x \uparrow y} \frac{f(x) - f(y)}{x - y} < \infty$$

and

$$-\infty \leq f'_r(z) = \lim_{w \downarrow z} \frac{f(w) - f(z)}{w - z} < \infty$$

exist for $a < y \leq b$ and $a \leq z < b$. f'_ℓ and f'_r are both non-decreasing, and

$$-\infty < f'_\ell(x) \leq f'_r(x) < \infty$$

for $a < x < b$. If $-\infty < a < b < \infty$ and f is convex on $[a, b]$, then $-\infty \leq f'_r(a) < \infty$ and $-\infty < f'_\ell(b) \leq \infty$ exist, possibly infinite. To see this, let $a \leq x < y < z \leq b$. Then

$$y = \frac{(z-y)x + (y-x)z}{z - x}$$

so that

$$f(y) \leq \left(\frac{z-y}{z-x}\right) f(x) + \left(\frac{y-x}{z-x}\right) f(z). \tag{8}$$

Subtracting $f(x)$ from both sides of (8) leads to

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(x)}{z - x}. \tag{9}$$

Thus, the difference quotients $[f(y) - f(x)]/(y - x)$ are increasing in y, and

$$-\infty \leq f_r'(x) = \inf_{y>0} \frac{f(y) - f(x)}{y - x} < \infty$$

exists. The existence of f_ℓ' can be established similarly.

Next, subtracting $f(y)$ from both sides of (8) leads to

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y}, \tag{10}$$

and iterating (10),

$$\frac{f(x) - f(t)}{x - t} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y} \leq \frac{f(w) - f(z)}{w - z} \tag{11}$$

for $a \leq t < x < y < z < w \leq b$. That $-\infty < f_\ell'(y) \leq f_r'(y) < \infty$ for $a < y < b$ follows, as does the monotonicity of f_ℓ' and f_r'.

As a consequence: *If $a < x < y$, then*

$$f(y) - f(x) = \int_x^y f_r'(w)dw; \tag{12}$$

and if $-\infty < a < b < \infty$ and f is continuous on $[a, b]$, then (12) holds also when $x = a$ or $y = b$.

Gateaux Derivatives. If $G \subseteq \mathcal{H}$, then an $x \in G$ is called an inner point iff: for every $y \in G$, $(1 - \alpha)x + \alpha y \in G$ for sufficently small α. Any interior point is an inner point. If $f : G \to I\!R$ and x is an inner point, then f is said to have a Gateaux derivative at x if

$$df_x(y) := \lim_{t \downarrow 0} \frac{f[(1 - t)x + ty] - f(x)}{t} \tag{13}$$

exists (finite) for each $y \in G$. Then the function df_x is called the (extended) Gateaux derivative of f and x. If df_x is a continuous linear functional, then there is an element of $\mathcal{H}$, also denoted by df_x for which

$$df_x(y) = \langle df_x, y \rangle$$

for all $y \in \mathcal{H}$. For example, if G is an open subset of $I\!R^m$ and $f : G \to I\!R$ is continuously differentiable, then

$$df(x; y) = \nabla f(x)'y$$

where $\nabla f(x)$ denotes the gradient,

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_m}\right]'.$$

If f has a Gateaux derivative at every $x \in C$, then f is said to be *Gateaux differentiable on C.*

Problem 2 *Suppose that C is an convex, open subset of $\mathbb{R}^m$ and the f is twice continuously differentiable on C. Then f is convex iff*

$$\nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_j \partial x_j} : i, j = 1, \cdots, m\right]$$

is non-negative definite for all $x \in C$.

If $C \subseteq \mathcal{H}$ and $f : C \to \mathbb{R}$ convex, then f has an *extended Gateaux derivative* $df_x^e : C \to \mathbb{R}$ at every $x \in C$ in the following sense: the limit in (13) exists for all y of the form $y = z - x$, where $z \in C$. For if $z \in C$, then $g(t) = f([x + t(z - x)] = f[(1 - t)x + tz]$ is convex on $[0, 1]$ and, therefore, $g_r'(0)$ exists (possibly $-\infty$). Write

$$df_x^e(y) = \lim_{t \downarrow 0} \frac{f[x + ty] - f(x)}{t} \tag{14}$$

for $y \in C - x$. It is clear the $df_x^e(y)$ agrees with $df_x(y)$ when the latter exists.

Convex Optimization. *Let C be a convex set. Then a necessary and sufficient condition for $x^* \in C$ to minimize f is that*

$$df_{x^*}^e(y - x^*) \geq 0 \tag{15}$$

for all $y \in C$. To proof is easy. First if x^* minimizes f on C and $y \in C$, then $f[(1-t)x^*+ty] \geq f(x^*)$, so that

$$0 \leq \frac{f[x^* + t(y - x^*)] - f(x^*)}{t} \to df_x^e(y - x^*)$$

as $t \downarrow 0$. Conversely, if (15) is satisfied and $y \in C$, then $g(t) = f[(1 - t)x^* + ty]$ defines a convex function on $[0, 1]$ for which

$$g_r'(0) = df_{x^*}^e(x; y - x^*) \geq 0$$

and, therefore, $g_r(t) \geq 0$ for all $0 \leq t \leq 1$. It then follows that

$$f(y) - f(x^*) = g(1) - g(0) = \int_0^1 g_r'(t)dt \geq 0.$$

as required.

The result does not assert that the minimum is attained.

Rockafellar [2] provides a more more comprehensive account of convexity. Edwards [1] has a nice chapter on calculus in linear spaces.

References

[1] Edwards, C.H. (1973). Advanced Calculus of Several Variables. Academic Press.

[2] Rockafellar, R. (1970). *Convex Analsis*. Princeton.