Convex Functions

Guy Lebanon

October 11, 2006

In this note we define convex functions and describe some important properties. We describe in some detail the concept of convex conjugacy as it plays a major role in computational statistics.

Definition 1. *The graph of a function $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$. is the set $\{(x, y) : x \in A, f(x) = y\} \subset \mathbb{R}^{n+1}$ and the epi-graph of f is the set $\{(x, t) : x \in A, f(x) \leq t\} \subset \mathbb{R}^{n+1}$.*

Definition 2. *Let $A \subset \mathbb{R}^n$ be a convex set. A function $f : A \to \mathbb{R}$ is convex if*

$$\forall \theta \in [0, 1], \qquad \forall x, y \in A, \qquad f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

If we have strict inequality above we say that f is strictly convex. If $-f$ is convex, we say that f is concave.

Note that for the above definition to make sense the requirement that A is convex is necessary. Geometrically, the above definition is equivalent to saying the every convex combination of points on the graph of the function is above or on the graph itself, i.e. in its epi-graph. This is also equivalent to saying that a function is convex iff its epi-graph is a convex set.

Example: All affine functions are both concave and convex (but not strictly).

If f is convex with domain A, we define its extension $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ as $\tilde{f}(x) = f(x)$ for $x \in A$ and $f(x) = +\infty$ otherwise. Thus we can convert a convex function on a convex set A to a convex function on $\mathbb{R}^n$ that is equivalent to f in some sense. When dealing with extended infinite values of extended real-valued functions we need to use the appropriate arithmetics e.g. $\infty \geq \infty > c \in \mathbb{R}$ and $\infty + \infty = \infty$.

Proposition 1 (First order differentiability condition)**.** *A differentiable function f on a convex domain is convex iff $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$ i.e. the graph is above the second order Taylor approximation plane.*

A consequence of the above result is that for a convex function f, $\nabla f(x) = 0$ implies that x is a global minimum.

Proposition 2 (Second order differentiability condition)**.** *If f is twice differentiable on a convex domain A, then it is convex iff the Hessian matrix $H(x)$ is positive semi-definite for all $x \in A$.*

The second order condition above makes it relatively easy to check whether a differentiable function is convex. Examples: the following functions are convex: exponential, logarithm, norm functions, cumulant generating function of exponential family distributions (log of the normalization term), Kullback-Leibler divergence. Similarly, it can be shown that the entropy and the log of the determinant are concave functions.

Proposition 3 (Jensen's Inequality)**.** *For a convex function f and a RV X, $f(\mathsf{E}(X)) \leq \mathsf{E} f(X)$.*

The following operations preserve convexity of functions.

- A weighted combination with positive weights of convex functions is convex. If $w_i > 0$ and $f_1, \ldots, f_n$ are convex then $\sum w_i f_i$ is convex (with a similar result for integration rather than summation). This can be seen by the second order condition for convexity.
- The point-wise maximum or supremum of convex functions is convex (this is a consequence of the fact that the intersection of convex epi-graphs is a convex epi-graph).
- If f is convex in (x, y) and C is a convex set then $\inf_{y \in C} f(x, y)$ is convex in x.

Definition 3. *The convex or Fenchel conjugate of* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *is* $f^*(y) = \sup_x (y^\top x - f(x))$.

Since the f^* is a point-wise supremum of convex functions in y, it is convex (even if f is not convex). Also, f^* is always lower semi-continuous since it is the point-wise supremum of lsc affine functions in y.

Example: For $f(x) = \alpha^\top x + \beta$ we have

$$f^*(y) = \sup_x (y^\top x - f(x)) = \sup_x (y^\top x - \alpha^\top x - \beta) = \sup_x (x^\top (y - \alpha) - \beta) \Rightarrow f^*(y) = \begin{cases} -\beta & y = \alpha \\ \infty & \text{otherwise} \end{cases}.$$

Example: For $f(x) = -\log x$, we have $f^*(y) = \sup_x (yx + \log x)$. If $y \geq 0$ $f^*(y) = \infty$. If $y < 0$, the supremum is achieved at $x = -1/y$ and then $f^*(y) = -1 + \log(-1/y) = -\log(-y) - 1$.

Example: For $f(x) = e^x$, $f^*(y) = \sup_x (yx - e^x)$. Again, for $y < 0$, $f^*(y) = \infty$. On the other hand, $f^*(0) = \sup -e^x = 0$ and for $y > 0$, the supremum is attained at $\log y$ and therefore $f^*(y) = y \log y - y$ for $y > 0$.

Example: For $f(x) = x \log x$, $f^*(y) = \sup_x (yx - x \log x)$. For all y the supremum is attained at $x = e^{y-1}$ and we get $f^*(y) = ye^{y-1} - e^{y-1}(y-1) = e^{y-1}$.

Example: For $f(x) = \frac{1}{2} x^\top Q x$ with Q symmetric positive definite, $f^*(y) = \sup_x y^\top x - \frac{1}{2} x^\top Q x$ and for all y, the supremum is attained at $0 = y^\top - x^\top Q$ or $x = Q^{-1} y$ hence $f^*(y) = y^\top Q^{-1} y - \frac{1}{2} y^\top Q^{-1} Q Q^{-1} y = \frac{1}{2} y^\top Q^{-1} y$.

Example: For I_S the indicator function of a set S ($I_S(x) = 0$ if $x \in S$ and ∞ otherwise), we have $I_S^*(y) = \sup_x y^\top x - I_S(x) = \sup_{x \in S} y^\top x$. In particular, $I_v^*(\lambda) = v^\top \lambda$, $I^*_{\{v: -c \leq v_i \leq c\}} = \sum_i c|\lambda_i| = c\|\lambda\|_1$, and $I^*_{\{v: \|v\|_2 \leq c\}} = \lambda^\top c \frac{\lambda}{\|\lambda\|_2} = \beta \|\lambda\|_2$.

In the case that f is differentiable the convex conjugate is also called the Legendre transform. Although not necessary, in the following we assume that f is also convex and defined on all $\mathbb{R}^n$ (possibly taking ∞ values). In this case, the supremum in $f^*(y)$ is attained at x' for which $0 = y - \nabla f(x')$. If the supremum is attained $f^*(y) = (\nabla f(x'))^\top x' - f(x')$. In other words, for an arbitrary z such that $y = \nabla f(z)$ we have $f^*(y) = z^\top \nabla f(z) - f(z) = (\nabla f(z), -1)^\top (z, f(z))$.

This leads to the following geometric interpretation of the Legendre transform in the space $\mathbb{R}^{n+1}$ that contains the epi-graph. First observe that the hyperplane $\alpha^\top (y, t) + \beta$ characterized by $\alpha = \begin{pmatrix} \nabla f(z) \\ -1 \end{pmatrix}$ and $\beta = -\begin{pmatrix} \nabla f(z) \\ -1 \end{pmatrix}^\top \begin{pmatrix} z \\ f(z) \end{pmatrix}$ is supporting the graph of f at $(z, f(z))$ since

$$\begin{aligned} (y,t) \in \operatorname{epi}(f) \quad &\Rightarrow \quad t \geq f(y) \geq f(z) + \nabla f(z)^\top (y - z) \quad \Rightarrow \quad \begin{pmatrix} \nabla f(z) \\ -1 \end{pmatrix}^\top \left(\begin{pmatrix} y \\ t \end{pmatrix} - \begin{pmatrix} z \\ f(z) \end{pmatrix} \right) \leq 0 \\ &\Rightarrow \alpha^\top \begin{pmatrix} y \\ t \end{pmatrix} + \beta \leq 0. \end{aligned}$$

where above we used the first order condition for convexity. The vertical distance of the hyperplane $\alpha^\top (y,t) + \beta$ from the origin is β which happens to be minus the Legendre transform $f^*(\nabla f(z))$. This leads to the interpretation of the Legendre transform $f^*(y)$ as the vertical offset of the supporting hyperplane to the graph of f at $(z, f(z)$ where $y = \nabla f(z)$. Intuitively, the Legendre transform maps a differentiable convex function to vertical offsets of supporting hyperplanes.

For convex lower semi-continuous functions f, $(f^*)^* = f$ and hence the convex conjugate is invertible (its inverse is itself).