Why convex optimization? One of the simplest convex optimization problem is to find the optimum coefficients of a regression function $\hat{y}(x,b) = Ab$, where the functional basis is the column vector of A that spans the space of the functional $\hat{y}(x)$ in it's x domain. When the optimum coeffients vector b is found, it approximates the true $y(x) = \hat{y}(x,b) + e(x)$ with minimum error. The minimum error is a convex function defined as

$$\epsilon(b) = ||y(x) - \hat{y}(x,b)||_2 = ||y - Ab||_2, \tag{1}$$

the 2-norm is taken over either discrete sum or continuous integral of x. The p-norm is a convex function, hence composing over a affine set gives convexity to $\epsilon(b)$. Now the optimization problem is to find the $b = [b_1, \cdots, b_n]$ that minimizes $\epsilon(b) \in \mathcal{R}_+$.

0.1 Perspective function preserves convex sets

$f(x,t) = \frac{x}{t}$, where $x \in \mathcal{R}^n$ and $t \in \mathcal{R}_{++}$, and (x,t) is convex. Given any two sets (x_1,t_1) and (x_2,t_2), the function $f(\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2) = \frac{\lambda x_1 + (1-\lambda)x_2}{\lambda t_1 + (1-\lambda)t_2} = \frac{\lambda t_1}{\lambda t_1 + (1-\lambda)t_2}\frac{x_1}{t_1} + \frac{(1-\lambda)t_2}{\lambda t_1 + (1-\lambda)t_2}\frac{x_2}{t_2} = \nu P(x_1,t_1) + (1-\nu)P(x_2,t_2)$.

The "Perspective of a Function", $f(x,t) = tg(x/t)$, preserving convexity of the convex function g, is proven by using the "Perspective function P" on the epigraph of f and g as: epi$f = \{(x,t,v) \mid v \geq f = tg(x/t)\}$ and epi$g = \{(x/t, v/t) \mid v/t \geq g(x/t)\}$, therefore under the same condition $v \geq tg(x/t)$, the perspective function maps epi$f$ to epi$g$, $P(x,t,v) = \frac{(x,v)}{t}$. Since the perspective function preserves convex sets and epi$g$ is convex by g being convex, hence, epi$f$ must be convex, which indicates a convex f. ■

0.2 Operations that preserve convexity

- Nonnegative Weighted Sum of convex functions (viewed as conic set of convex functions (vectors))
- Composition with an affine mapping
- Pointwise Maximimum and supremum
- Composition (with differentiability conditions or without (extended-value extension))
- Perspective Function

0.2.1 Nonneg Weighted Sum

$f = w_1 f_1 + w_2 f_2 + \cdots$, where $w_i \geq 0$ and f_i's are convex functions. Suppose $f(x_2, x_2) = w_1 f_1(x_1) + w_2 f_2(x_2)$ is a nonnegative weighted sum of two convex functions, then the domain of f_1 and f_2 doesn't have to be the same.

0.3 Conjugate function

0.3.1 Definition

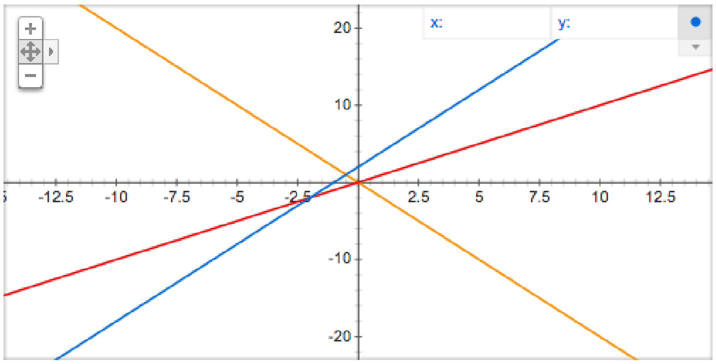
Let $f : R^n \longrightarrow R$. The conjugate function $f^* : R^n \longrightarrow R$ of f is

$$f^*(y) = \sup_{x \in \mathrm{dom} f} \left(y^T x - f(x) \right). \tag{2}$$

0.3.2 Examples (convex functions)

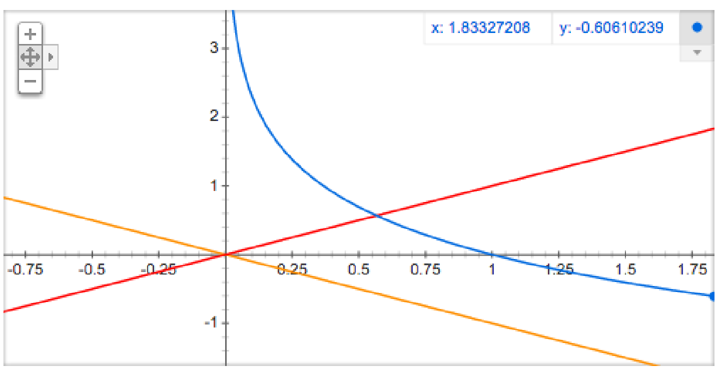
1. Affine function $f(x) = ax + b$. $yx - ax - b$ is bounded iff $y = a$ (i.e., there exists a supremum over the domain for any given y), since the domain is $[-\infty\infty]$, which gives $f^*(y) = -b$.

Graph for 2*x+2, x, -(2*x)



2. Negative logarithm $f(x) = -\log x$ (log is the natural ln). $yx + \log x$ is unbounded if $y \geq 0$, otherwise $(y < 0)$ $f^*(y)$ reaches maximum at $x = -1/y$ (set $\frac{\partial f^*(y)}{\partial x} = 0$) , which is $f^*(y) = -\log(-y) - 1$. In the figure, $f(y = 1) = x$ (red) and $f(y = -1) = -x$ (orange)

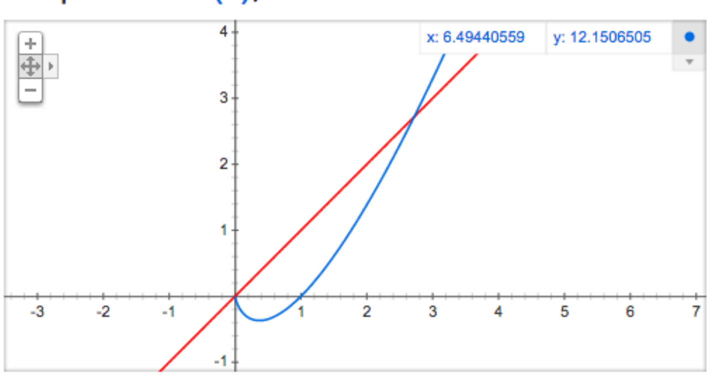
Graph for -ln(x), x, -x



3. Exponential $f(x) = e^x$. $yx - e^x$ is unbounded if $y < 0$, hence when $y > 0$ the maximum is reached at $x = \log y$ with $f^*(y) = y \log y - y$. For $y = 0$, $f^*(y) = 0$.

4. Negative entropy $f(x) = x \log x$ with $\mathrm{dom} f = R_+$.

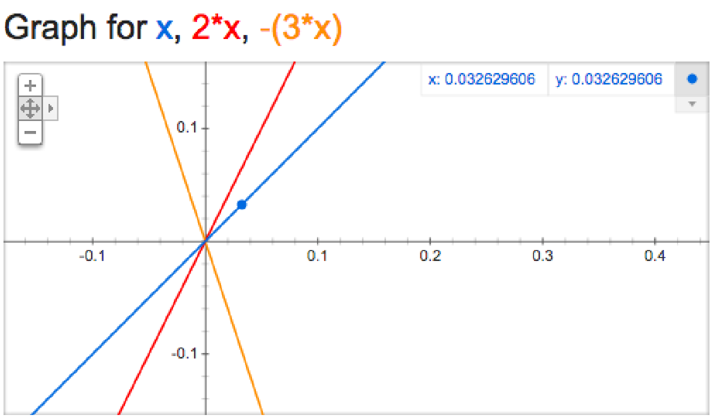
Graph for x*ln(x), x



5. Inverse $f(x) = 1/x$ on R_{++}. $f^*(y \leq 0) = -2(-y)^{1/2}$, notice at $y = 0$, the maximum $f^* = 0$ is attained as $x \longrightarrow \infty$,

6. Strictly convex quadratic function $f(x) = \frac{1}{2}x^TQx$ with $Q \in S^n_{++}$, has the conjugate $f^* = y^Tx - \frac{1}{2}x^TQx$ which has a maximum at $x = Q^{-1}y$ for all $y \longrightarrow f^*(y) = \frac{1}{2}y^TQ^{-1}y$, where Q^{-1} is also S^n_{++}.
Remarks: The conjugate of a convex quadratic function is also a convex quadratic function.

7. Indicator function $I_S(y) = 0$, $\text{dom}I_S = S$. The conjugate $I^*_S(x) = \sup(y^Tx)$, $y \in C$, is the support function (convex), which gives the pointwise maximum over linear functions of x at any given x (i.e., different y gives different linear functions y^Tx of x, so at a point x_i, $\sup(y^Tx_i)$ finds the maximum linear function). The figure shows three linear functions of x at three different $y = 1, 2, -3$, so the pointwise maximum function for $x > 0$ is at $y = 2$.



0.3.3 Basic Properties

Fenchel's inequality
Conjugate of the conjugate
Differentiable functions
Find the slope of $f(x)$ at x, which gives the maximum x^Ty (i.e., $x^T\nabla f(x)$) to pointwise maximize $f^*(y)$!!
Conjugate of a differentiable function $f \Rightarrow$ *Legendre transform* of f. Suppose $\text{dom}f = \mathcal{R}^n$, any maximizer x^* of $y^Tx - f(x)$ satisfies $y = \nabla f(x^*)$ (set $\nabla_x f^*(y) = 0$). Therefore, given an arbitrary x satisfying $y = \nabla f(x)$, we can find

$$f^*(y) = x^T\nabla f(x) - f(x).$$

Therefore, the maximized y at x is the slope of both $f(x)$ and the affine function x^Ty (passing through the origin).
Scaling and composition with affine transformation
Given $g(x) = af(x) + b$, the conjugate

$$g^*(y) = x^Ty - g(x) = x^Ty - af(x) - b = a(x^T(y/a) - f(x)) - b = af^*(y/a) - b.$$

Given $g(x) = f(Ax + b)$, where $A \in \mathcal{R}^{n\times n}$ and $b \in \mathcal{R}^n$, and set $z = Ax + b => x = A^{-1}(z - b)$

$$g^*(y) = x^Ty - g(x) = x^Ty - f(Ax+b) = (A^{-1}(z-b))^Ty - f(z) = z^TA^{-T}y - b^TA^{-T}y - f(z) = f^*(A^{-T}y) - b^TA^{-T}y,$$

$$\text{dom}g^* = y = A^TA^{-T}y = A^T\text{dom}f^*.$$

0.4 Quasiconvex Functions

0.4.1 Definition

f is *quasiconvex* if its domain and all its sublevel sets $S_\alpha = \{x \in \text{dom}f | f(x) \leq \alpha\}$ for $\alpha \in \mathcal{R}$ are convex.
f is *quasiconcave* if $-f$ is quasiconvex.
f is *quasilinear* if its both quasiconvex and quasiconcave, and all its level sets $S_\alpha = \{x | f(x) = \alpha\}$ are convex.
Remarks: Quasiconvex functions can be non-convex, with convex domain and sublevel sets.

0.4.2 Examples

1. $\log x$ (Concave function) Both quasiconvex and quasiconcave $\Rightarrow$ quasilinear.
2. Ceiling function (Discontinuous function) $f(x) := \{z \in \mathcal{Z} | z \geq x, x \in \mathcal{R}\}$ is quasilinear.
3. $f(x_1, x_2) = x_1 x_2$(Neither convex nor concave function, Hessian is indefinite; one positive and one negative eigenvalue) dom$f = \mathcal{R}^2_+$, is a quisiconcave function, its superlevel sets $\{x | x_1 x_2 \geq \alpha\}$ are convex.

0.4.3 Basic Properties

Jensen's inequality for quasiconvex function:
A function f is quasiconvex $\Leftrightarrow$ dom$f$ is convex and for any $x, y \in$dom$f$ and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\}. \tag{3}$$

The value of the function on a segment doesn't exceed the maximum of its value at the two endpoints, which could be a non-convex function. (see figure 3.10)
Quasiconvexity by a line intersection:
f is quasiconvex $\Leftrightarrow$ f restricted to a line intersection on its domain is quasiconvex.
Remarks:
A continuous quasiconvex function on $\mathcal{R}$ has a point c in the domain that is a **global minimizer**, and f is nonincreasing and nondecreasing at the left and right of c, respectively.

0.5 Differntiable Quasiconvex Functions

0.5.1 1-order conditions

Suppose f is differentiable. f is quasiconvex $\Leftrightarrow$ dom$f$ is convex and for all x, $y \in$ dom$f$

$$f(y) \leq f(x) \Rightarrow \nabla f(x)^T (y - x) \leq 0. \tag{4}$$

The definition of quasiconvex function shows that, if $y > x$ ($y < x$) and $f(y) \leq f(x)$, f is nonincreasing (nondecreasing) which will always give negative (positive) $\nabla f(x)$. An analogy to convex function which satisfies $f(y) \geq f(x) + \nabla f(x)^T (y - x)$. $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{y | f(y) \leq f(x)\}$

0.5.2 2-order conditions

Suppose f is twice differentiable. If f is quasiconvex for all $x \in$dom$f$ and all $y \in \mathcal{R}^n$, then if

$$y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y \geq 0. \tag{5}$$

For a quasifunction on $\mathcal{R}$, this reduces to

$$f'(x) = 0 \Rightarrow f''(x) \geq 0.$$

i.e, whenever the slope is zero, a quasiconvex function has a positive curvature. For the $\mathcal{R}^n$ case, when $\nabla f(x) \neq 0$ (y is in the $(n-1)$-dim subspace $\nabla f(x)^\perp$ orthogonal (thus independent) to $\nabla f(x)$), $\nabla^2 f(x)$ is positive semidefinite on $\nabla f(x)^\perp$, and have at most one negative eigenvalue in the 1-dim subspace of $\nabla f(x)$. (proof ommitted)

0.6 Operations that Preserve Quasiconvexity

- Nonneg weighted maximum
- Composition
- Minimization

0.6.1 Nonneg weighted maximum

A nonneg weighted, $w_i \geq 0$ maximum of quasiconvex functions f_i,

$$f(x) = \max\{w_1 f_1(x), \ldots, w_m f_m(x)\} \tag{6}$$

is quasiconvex. This can be extended to the pointwise supremum of quasiconvex functions $g(x, y)$ for all y (y is like the integer index $i = 1, \ldots, m$, but just in real number space),

$$f(x) = \sup_{y \in C} \left(w(y) g(x, y) \right). \tag{7}$$

0.6.2 Composition

- If $h \in \mathcal{R}$ nondecreasing, and $g \in \mathcal{R}^n \to \mathcal{R}$ quasiconvex, then $f = h \circ g$ quasiconvex.
- If f quasiconvex, then operating on $\mathrm{dom} f$ with affine or linear-fractional transformation on x preserves quasiconvexity. e.g, $f(Ax + b)$ or $f((Ax + b)/(cx + d))$.

0.6.3 Minimization

Suppose $f(x, y)$ is quasiconvex jointly in x and y, and C is convex, then prove that

$$g(x) = \inf_{y \in C} f(x, y) \tag{8}$$

is quasiconvex. This boils down to proving that the sublevel set $K = \{x | g(x) \leq \alpha\}$ is convex for an arbitrary α. We know that $g(x) \leq \alpha \Leftrightarrow f(x, y) \leq \alpha + \epsilon$, where $\epsilon > 0$. Given two points (x_1, y_1) and (x_2, y_2) in the convex sublevel set $\{(x, y) | f(x, y) \leq \alpha + \epsilon\}$, with quasiconvexity of f,

$$f(\theta x_1 + (1 - \theta) x_2, \theta y_1 + (1 - \theta) y_2) \leq \alpha + \epsilon. \tag{9}$$

By the iff condition,

$$g(\theta x_1 + (1 - \theta) x_2) \leq \alpha, \tag{10}$$

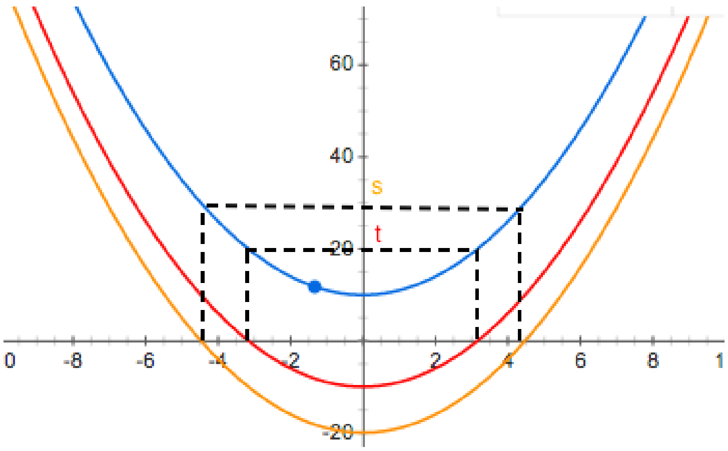
which implies that $\theta x_1 + (1 - \theta) x_2 \in K$, hence $g(x)$ is quasiconvex.

0.7 Representation via family of convex functions

The sublevel sets of a quasiconvex function f can be represented by a family of sublevel sets of convex functions,

$$f(x) \leq t \Leftrightarrow \phi_t(x) \leq 0. \tag{11}$$

i.e., the t-sublevel set of f equals the 0-sublevel set of convex function ϕ labeled by t. The figure shows that the blue f given two sublevel s and t, where $s \geq t$, then the orange $\phi_s(x)$ is below the red $\phi_t(x)$, $\phi_s(x) \leq \phi_t(x)$. i.e., due to the s-sublevel set of f contains the t-sublevel set of f, with both sets being convex.



0.7.1 Example: Convex over Concave function

Purpose: A convex over concave function f is quasiconvex, and the t**-sublevel sets** could be represented by the 0-sublevel sets of a family of convex functions (indexed by t).
Suppose $p(x)$ is convex and $q(x)$ is concave, and $p > 0$ and $q > 0$ on a convex set C, then $f(x) = p(x)/q(x)$ is quasiconvex with the relationship

$$f(x) \leq t \Leftrightarrow p(x) - tq(x) \leq 0 \tag{12}$$

which shows that $\phi_t = p(x) - tq(x)$ is convex for $t \leq 0$ (i.e., $-q(x)$ is convex).

1 Log-concave and Log-convex functions

1.1 Definition

- If $f > 0$ for all $x \in \mathrm{dom} f$, and $\log f(x)$ is concave, then f is a log-concave function.
- f is log-convex if $\log f$ is convex and $f > 0$.
- Therefore, f is log-convex iff $1/f$ is log-concave ($-\log f = \log(1/f)$ is concave and $1/f > 0$).
- Without logarithms, f is log-concave iff $f > 0$, $\mathrm{dom} f$ is convex, and for all $x, y \in \mathrm{dom} f$ and $0 \leq \theta \leq 1$ we have

$$f(\theta x + (1-\theta)y) \geq f(x)^{\theta} f(y)^{1-\theta}. \tag{13}$$

  If we take $\theta = 1/2$, then f at the mean of x and y is at least the geometric mean [1] of f at the two points.

[1] geometric mean of two values is $\sqrt{x_1 x_2}$, of three values is $(x_1 x_2 x_3)^{1/3}$. Measures the central tendency of a set of numbers. The geomtric mean of x_1 and x_2 gives the edge of a square, which has the same area as a rectangle with the edge of x_1 and x_2.

1.2 Properties

1.2.1 Convexity (concavity) by twice differentiable condition

Suppose dom$f$ is convex, f is log-convex iff

$$\nabla^2 \log f \succeq 0, \tag{14}$$

which gives

$$\frac{1}{f}\nabla^2 f - \frac{1}{f^2}\nabla f \nabla f^T \succeq 0 \Rightarrow f\nabla^2 f \succeq \nabla f \nabla f^T \tag{15}$$

2 Function Convexity wrt Generalized Inequality

2.1 Monotonicity wrt generalized inequality

2.1.1 *K-non-decreasing*

f is called K-*non-decreasing* if $K \subset \mathcal{R}^n$ is a proper cone (convex, solid, pointed, closed), and

$$x \succeq_K y \Rightarrow f(x) \leq f(y). \tag{16}$$