Conjugate functions

Handout from Jeanine Wippermann

Motivation

Consider a given function $f(x)$.

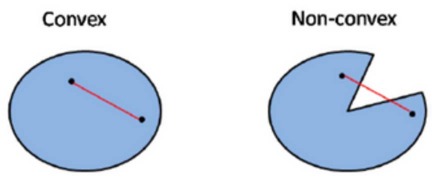
To gain deeper insights and leverage additional properties of this function, we define an associated conjugate function, denoted as $f^*(x)$.

This conjugate function $f^*(x)$ encapsulates crucial information about the convex hull of the original function $f(x)$. By utilizing the properties of the conjugate function $f^*(x)$, we can formulate and solve dual optimality problems, providing a powerful tool for addressing complex optimization scenarios.

Mathematical Background

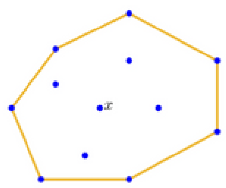
Convex set: $\forall x, y \in C \subseteq \mathbb{R}^n, \alpha \in [0,1]: \quad \alpha x + (1-\alpha)y \in C$

A convex set is defined by the property that for any two points within the set, the line segment connecting these points lies entirely within the set. This intrinsic property of convex sets ensures that they are "filled out" with no indentations or holes.



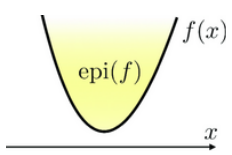
Convex hull: $\operatorname{conv}(M) = \bigcap \{C \subseteq \mathbb{R}^n | C \text{ is convex and } M \subseteq C\}$

The convex hull of a set is the smallest convex set that entirely contains the given set. It can be visualized as the shape formed by stretching a rubber band around the outermost points of the set.



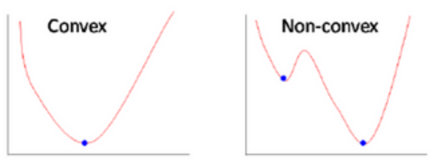
Epigraph: $\operatorname{epi}(f) = \{(x,t) \in \mathbb{R}^{n+1} \mid f(x) \leq t\}$

The epigraph is the set of all points lying on or above the graph of the function.



Convex function: $\forall x, y \in \mathbb{R}^n, \alpha \in [0,1]: \; f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$

A convex function is one where the line segment between any two points on its graph lies above or on the graph. Alternatively you can say that the epigraph of a convex function has to be convex.



Domain of a function : $\mathrm{dom}(f) = \{x \in \mathbb{R}^n | f(x) < \infty\}$

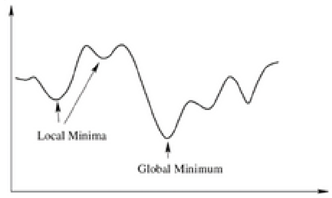
The domain of a function comprises all input values for which the function value is defined and finite. It essentially represents the set of all possible valid inputs that the function can accept and produce a meaningful output for.

Hyperplane : $H = \{x \in \mathbb{R}^n \,|\, \mathbf{a} \cdot \mathbf{x} = b\}$

Hyperplanes are flat, affine subspaces in a given space, characterized by one dimension less than the space itself. They serve to divide the entire space into two distinct regions known as halfspaces.

Optimization task :

The goal is to find the maximum or minimum of a function while considering given constraints. The global minimum is the absolute lowest value of the function over its entire domain. A local minimum, on the other hand, is the lowest value of the function within a specific region or neighborhood.

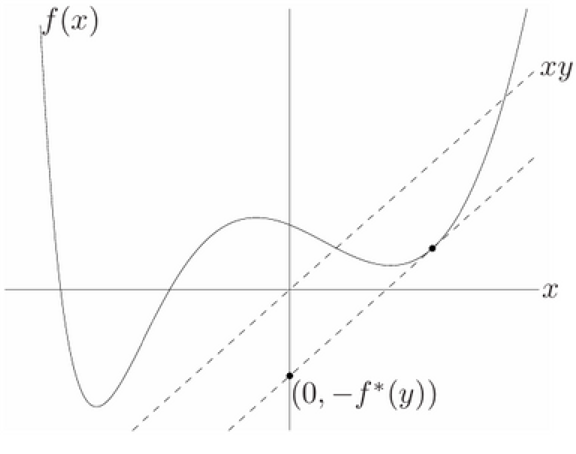


Convex functions are highly advantageous in optimization. They have the property that any local minimum is also a global minimum. This is particularly beneficial because, in general, it is often difficult to distinguish between local and global minima in non-convex functions.

Definition

We have given a function $f : \mathbf{R}^n \to \mathbf{R}$
The conjugate of the function is given by:

$$f^* : \mathbf{R}^n \to \mathbf{R}$$
$$f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x))$$



f*(y) represents the maximum gap between a linear function with slope y and f(x). If f is differentiable, this occurs at a point x where f′(x) = y.

When the maximum gab is infinity, the associated y-value is not part of the domain of f*(x).

When shifting the linear function, such that it is a tangent from below to the function, we can directly read out f*(y): the point of intersection between the tangent and the y-axis provides the value of the maximum gap.

Examples

When calculating the conjugate of a function, we might consider different cases. The supremum of yx-f(x) might be different considering different y.

Exponential function

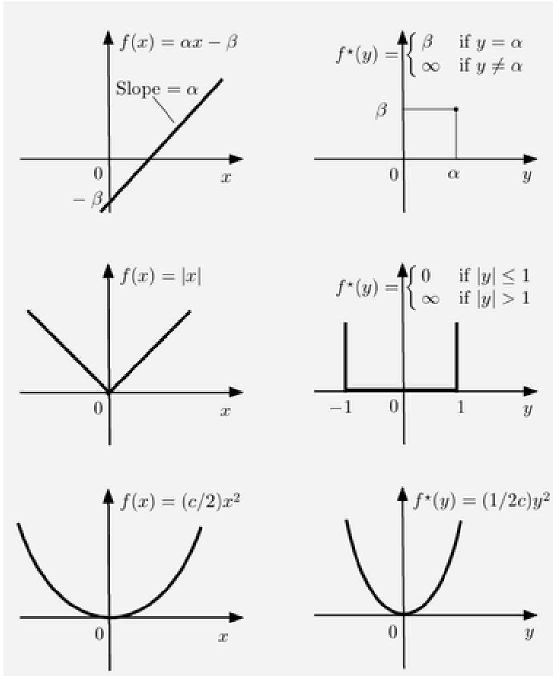
Let $f(x) = e^x$

Then $xy - e^x$ is

- $y < 0$ unlimited
- $y = 0$ limited with $\sup_{x \in dom f}(-e^x) = 0$
- $y > 0$ limited with a maximum value at $x = log\, y$

It follows $dom(f^*) = \mathbb{R}_+$ with

$$f^*(y) = y\, log(y) - y$$



Properties of the conjugate function

General properties

Convexity:

Every conjugate function f*(x) is convex, even if f(x) is not.
This is because f*(x) is the point-wise supremum of affine linear functions of y, and they are of corse convex.

Fenchels inequality: $f^*(y) + f(x) \geq y^T x \quad \forall x, y$

Fenchels inequality follow directly from the definition. It is also another way of defining a conjugate function: two functions, that satisfy Fenchels inequality are conjugate to another.

Order reversing:. $f \leq g \quad \Rightarrow \quad g^* \leq f^*$

Sums of independent functions: $f(u, v) = f_1(u) + f_2(v) \quad \Rightarrow \quad f^*(w, z) = f_1^*(w) + f_2^*(z)$

Affine Transformations:

$$g(x) = af(x) + b \quad \Rightarrow \quad g^*(y) = af^*(\tfrac{1}{a}y) - b$$

$$g(x) = f(Ax + b) \quad \Rightarrow \quad g^*(y) = f^*(A^{-T}y) - b^T A^{-T} y \qquad dom(g^*) = A^T dom(f^*)$$

Properties for convex functions

Conjugacy theorem: $(f^*)^* = f$

If f(x) is convex and closed, the conjugate of the conjugate is again f(x).
If f(x) is non-convex, the conjugate of the conjugate is a function that defines the convex hull of f(x).

Properties of differentiable functions

Simplified calculation: $f^*(y) = \nabla(x)^T x - f(x)$

If f(x) is convex and differentiable, we can find the supremum of xy-f(x) always at its stationary point. Therefore $y = \nabla f(x^*)$ must always hold. This simplifies our previous definition. If we can calculate the derivative, we can also calculate the conjugate.

Conjugate functions in optimization problems

Duality

Two different views of the same object can offer distinct perspectives, each with its own mathematical advantages depending on the context. For instance, consider the dual description of closed convex sets. One view represents these sets as a union of points, while another perspective describes them as an intersection of halfspaces. Each of these representations can be more useful than the other depending on the specific problem or application at hand.

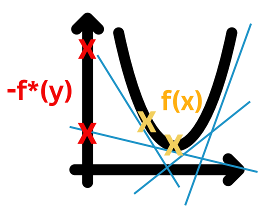
Dual description of convex functions:

The dual description of convex functions provides a different perspective from the primal description, which focuses on the values $f(x)$.

Instead, the function can be described by hyperplanes of the form:
$H = \{\mathbf{x}' \in \mathbb{R}^n | \begin{pmatrix} \mathbf{y} \\ -1 \end{pmatrix} \cdot \mathbf{x}' = f^*(\mathbf{y})\}$

These hyperplanes are associated with the points where they cross the $f(x)$-axis, leading to values of $-f^*(y)$.

Thus, while the primal description deals with the direct values of the function $f(x)$, the dual description uses the values $f^*(y)$ derived from these hyperplanes.



Fenchel primal and dual problems:

Fenchel describes two optimization problems, that are connected.

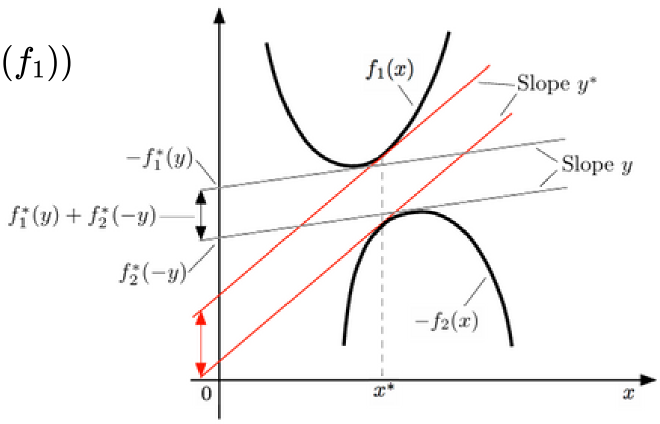
The primal optimiality problem is: $\min_x(f_1(x) + f_2(x))$
One can consider this expression as the minimal gab between the two functions f1(x) and -f2(x) in the direction of the f(x)-axis.

The dual optimality problem is: $\min_y(f_1^*(y) + f_2^*(-y))$
Imagine two tangents with same slope, one on f1(x) and one on -f(x). The expression above is the negative difference of the intersection points of the tangents with the f(x)-axis.

If f(x) and g(x) are lower semi-continuous and $0 \in core(dom(f_2) - dom(f_1))$ the solutions of the dual problems are equal.

In other cases, the dual solution provides an upper limit for the minimzer of the primal solution:

$$\min_x(f_1(x) + f_2(x)) \leq \min_y(f_1^*(y) + f_2^*(-y))$$



$$\min_x \{f_1(x) + f_2(x)\} = \max_y \{-f_1^*(y) - f_2^*(-y)\}$$