Making sense of the Legendre transform

R. K. P. Zia
Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Edward F. Redish
Department of Physics, University of Maryland, College Park, Maryland 20742

Susan R. McKay
Department of Physics and Astronomy, University of Maine, Orono, Maine 04469

MAKING SENSE OF LEGENDRE TRANSFORM

Ben Huh ?

Gatsby Tea-talk 2013/10/18

WHAT IS L-T?

- Given a functional relation F(x), find a new different representation, G(p), that encodes the same information.

$$\{F, x\} \longleftrightarrow \{G, p\}$$

$$G(p) = xp - F(x)$$

with $p(x) = \frac{dF}{dx}$

Spring example

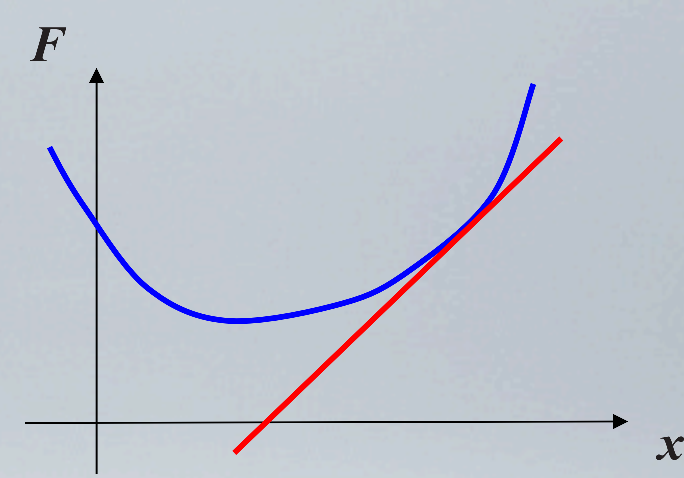
$$\mathcal{L} = K(v) - U(x) \qquad \mathcal{H} = K(p) + U(x)$$

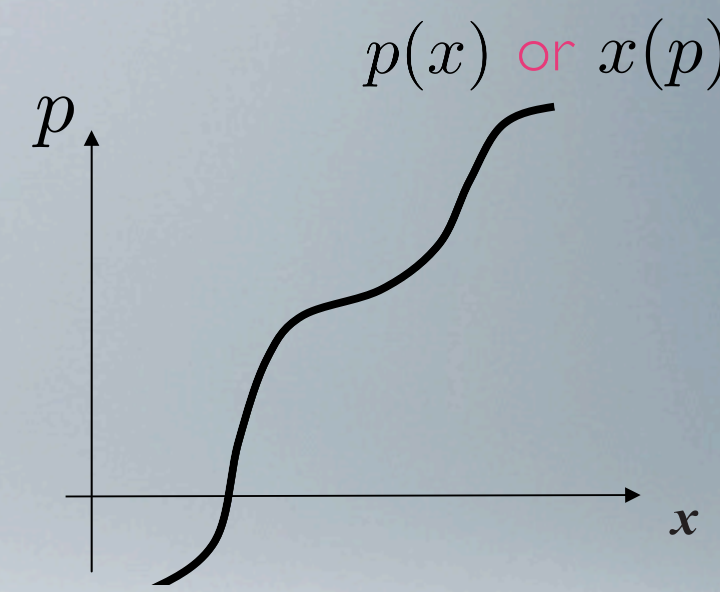
- Lagrangian $\longleftrightarrow$ Hamiltonian
- Thermodynamic potentials
  (also in info theory)
  (Entropy $\longleftrightarrow$ Helmholtz free energy)
- Dual optimization problem
  (conjugate function)

Quite confusing &
Not very well motivated

MATHEMATICS OF L-T

Given an $F(x)$, the Legendre transform provides a more convenient way of encoding the information in the function when two conditions are satisfied: (1) The function (or its negative) is strictly convex (second derivative always positive) and smooth (existence of "enough" continuous derivatives). (2) It is easier to measure, control, or think about the derivative of F with respect to x than it is to measure or think about x itself.





- F(x) is strictly convex

$$\frac{d^2F}{dx^2} > 0$$

- p(x) is strictly monotonic

$$p(x) = \frac{dF}{dx}$$

One-to-one relation between x and p

- invert! $x(p)$

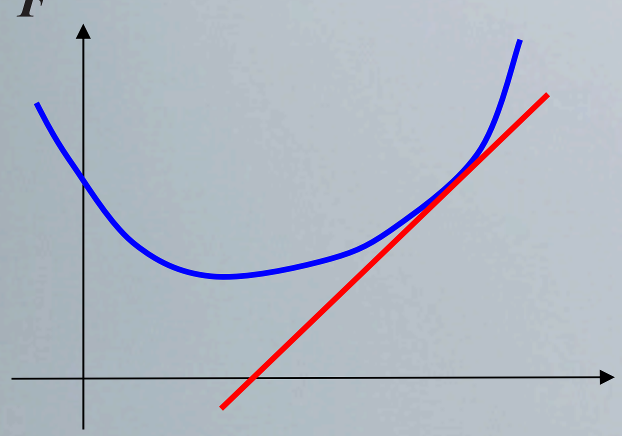
- Use p as the independent variable

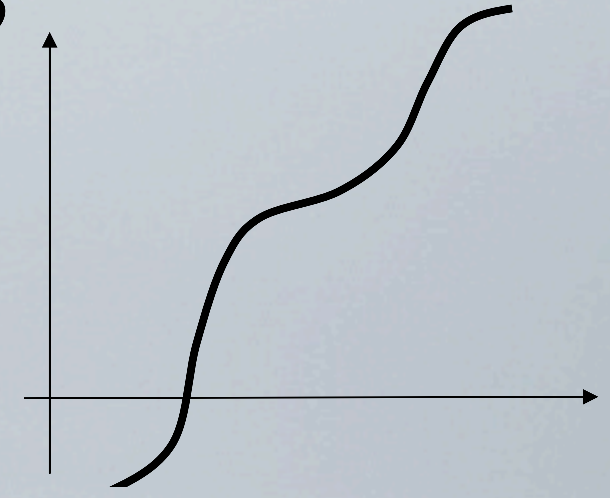
- Why not just insert it to F(x)?

why not $F(x(p))$?

why this? $G(p) \equiv px(p) - F(x(p))$

MATHEMATICS OF L-T





- Use p as the independent variable
- Why not just insert it to F(x)?

why not $F(x(p))$?

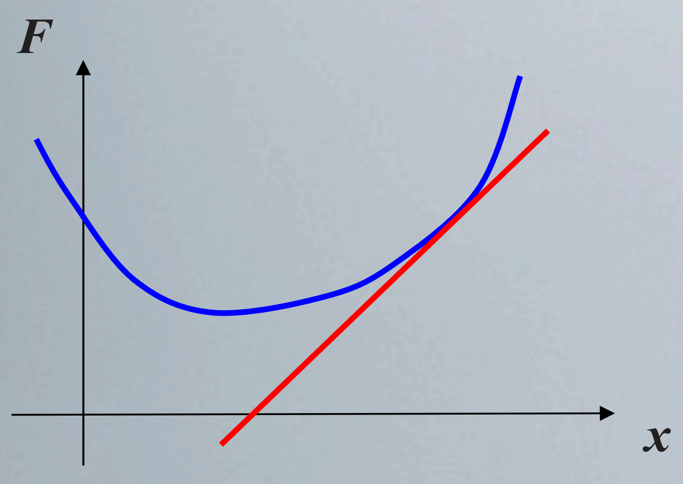
why this? $G(p) \equiv px(p) - F(x(p))$

Typically, this definition is presented with little motivation or explanation, and leaves the students to ponder: Why? Why the extra sx? Why the minus sign? Frequently, the instructor or textbook invokes another magical relation to answer such queries. Only with this peculiar definition can we have the property that "the slope of $G(s)$ is just x":

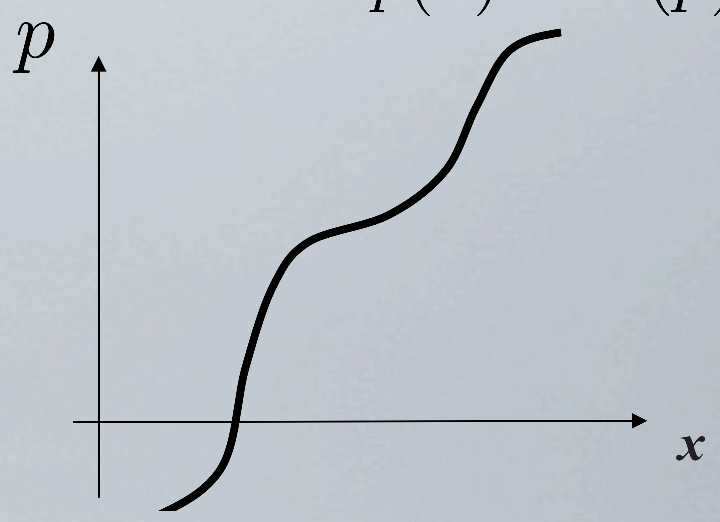
$$\frac{dG}{dp} = x(p)$$

$$G' = x + px' - \frac{dF}{dx}x'$$

GEOMETRIC INTERPRETATION



$p(x)$ or $x(p)$



Symmetry! $\{F, x\} \longleftrightarrow \{G, p\}$

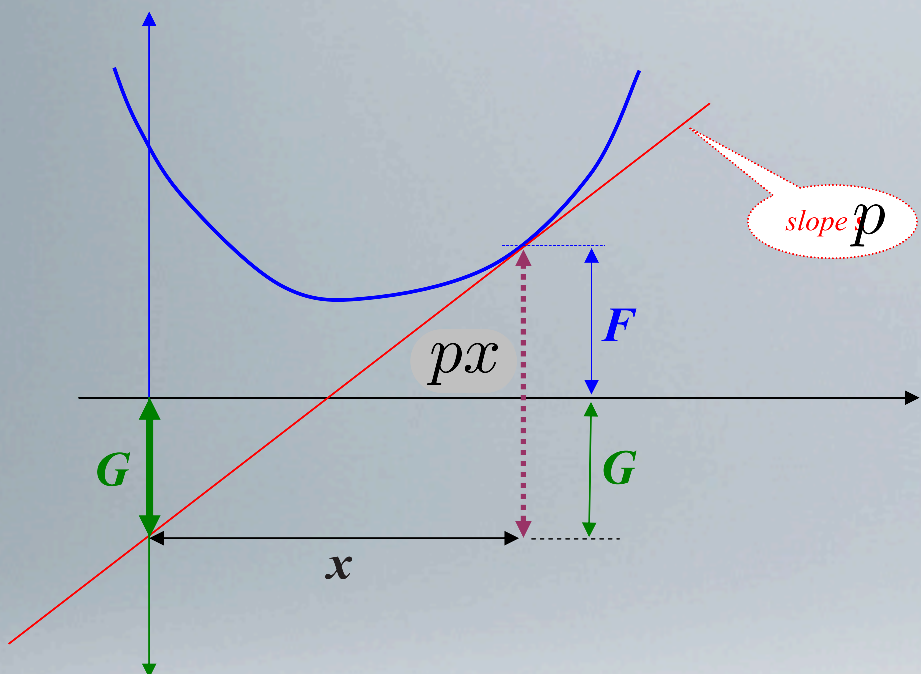
$$F(x) + G(p) = px$$

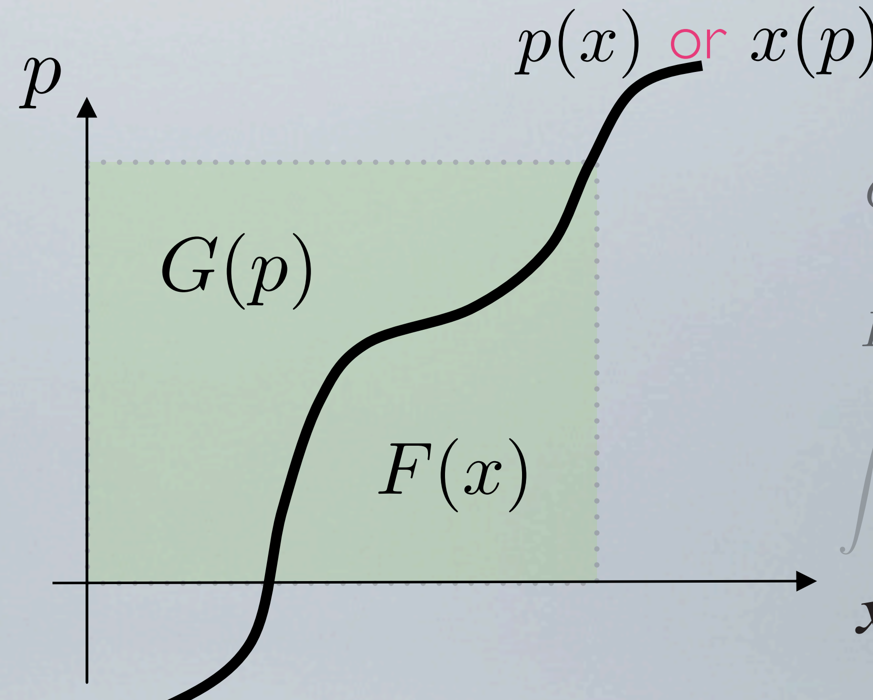
$$\frac{dF}{dx} = p(x) \qquad \frac{dG}{dp} = x(p)$$

$$G(p) \equiv px(p) - F(x(p))$$

self-inverse (involutive)

GEOMETRIC INTERPRETATION





$$G(p) = \int x(p)dp$$

$$F(x) = \int p(x)dx$$

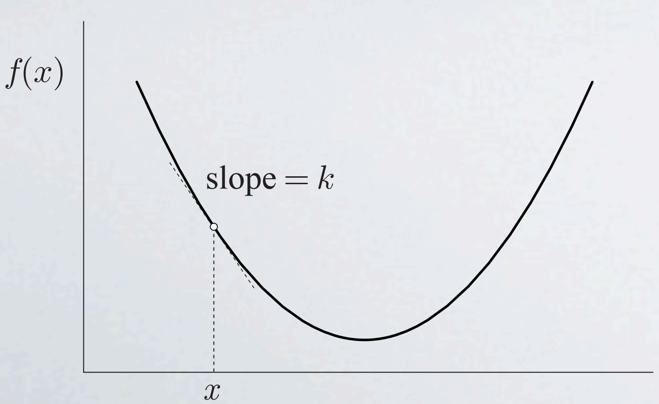
$$\int p(x)dx + \int x(p)dp = px$$

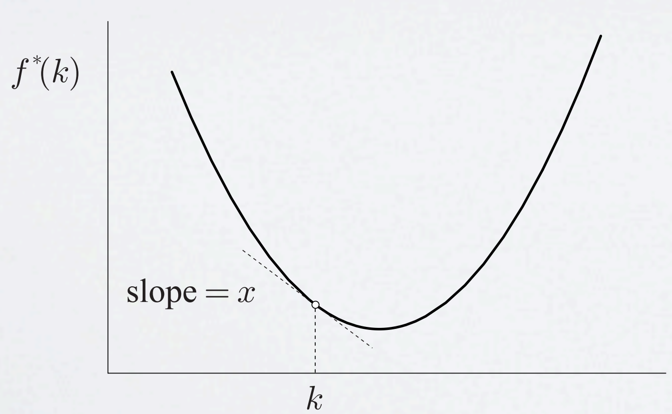
Symmetry! $\{F, x\} \longleftrightarrow \{G, p\}$

$$F(x) + G(p) = px$$

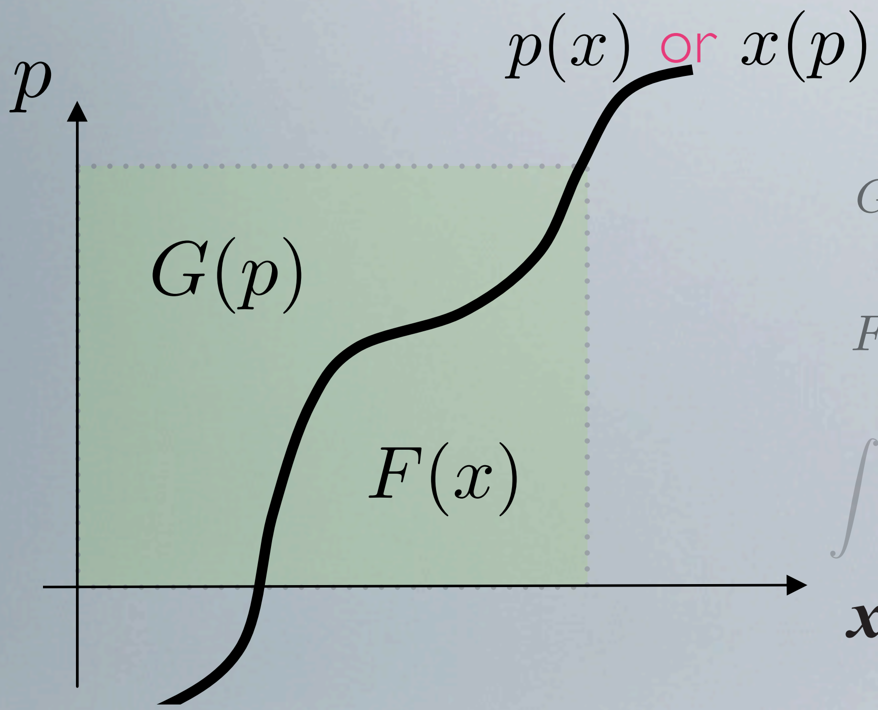
$$\frac{dF}{dx} = p(x) \qquad \frac{dG}{dp} = x(p)$$

self-inverse (involutive)





GEOMETRIC INTERPRETATION



$$G(p) = \int x(p)dp$$

$$F(x) = \int p(x)dx$$

Symmetry!

$$\{F, x\} \longleftrightarrow \{G, p\}$$

$$\int p(x)dx + \int x(p)dp = px$$

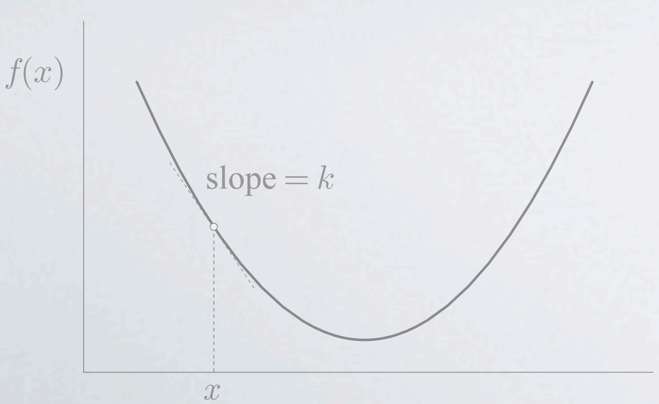
$$F(x) + G(p) = px$$

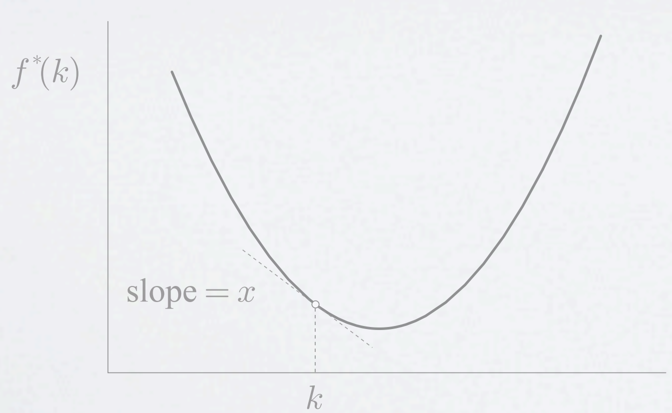
$$\frac{dF}{dx} = p(x) \qquad \frac{dG}{dp} = x(p)$$

2nd order

$$\left(\frac{d^2F}{dx^2}\right)\left(\frac{d^2G}{dp^2}\right) = 1$$

$$\frac{d^2F}{dx^2} = \frac{dp}{dx} \qquad \frac{d^2G}{dp^2} = \frac{dx}{dp}$$





$$F(x) = \frac{1}{2}\alpha x^2 \quad \longleftrightarrow \quad G(p) = \frac{1}{2\alpha}p^2$$

WHERE IS IT USEFUL?

- Lagrangian - Hamiltonian

$$\mathcal{L}(v,x) = \frac{1}{2}mv^2 - U(x) \longleftrightarrow \mathcal{H}(p,x) = pv - \mathcal{L} = \frac{1}{2m}p^2 + U(x)$$

$$p(v) = \frac{\partial \mathcal{L}}{dv} = mv \qquad v(p) = \frac{\partial \mathcal{H}}{dp} = p/m$$

- Spring (length - force)
  capacitor (charge - voltage)
  gas (volume - pressure)

$$U(x) = \frac{1}{2}kx^2 \longleftrightarrow V(f) = fx - U(x) = \frac{1}{2k}f^2$$

$$f(x) = \frac{dU}{dx} = kx \qquad x(f) = \frac{dV}{df} = f/k$$

- *Motor control*
  (movement duration - ***drive*** motivation)

- Thermodynamics
  (energy - temperature)

- Dual optimization problem

- Extensive variables - additive
  (length, charge, volume, duration, energy)

- Intensive variables
  - share common value across systems
  - easier to control
  (force, voltage, pressure, *drive*, temperature)

THERMODYNAMICS

of accessible state

$$\Omega(E) = \int_{r,p} \delta(E - \mathcal{H}(\{\vec{r}_i, \vec{p}_i\})),$$

$$\mathcal{S}(E) \equiv \ln \Omega(E)$$ Entropy

Partition function

$$Z(\beta) \equiv \int \Omega(E) e^{-\beta E} dE = \int_{r,p} e^{-\beta \mathcal{H}}$$

Laplace transform

$$\mathcal{F}(\beta) \equiv -\ln Z(\beta)$$ Helmholtz free energy

$$e^{-\mathcal{F}(\beta)} = \int e^{\mathcal{S}(E) - \beta E}$$

$$\approx e^{\max_E \{\mathcal{S}(E) - \beta E\}}$$ in thermodynamic limit (Laplace approximation)

$$\therefore \mathcal{F}(\beta) + \mathcal{S}(E) \approx \beta E$$

$$\beta = \frac{d\mathcal{S}}{dE} \qquad E = \frac{d\mathcal{F}}{d\beta}$$

inverse temperature

Geometric Interpretation

Consider the following primal problem P:

Primal Problem P

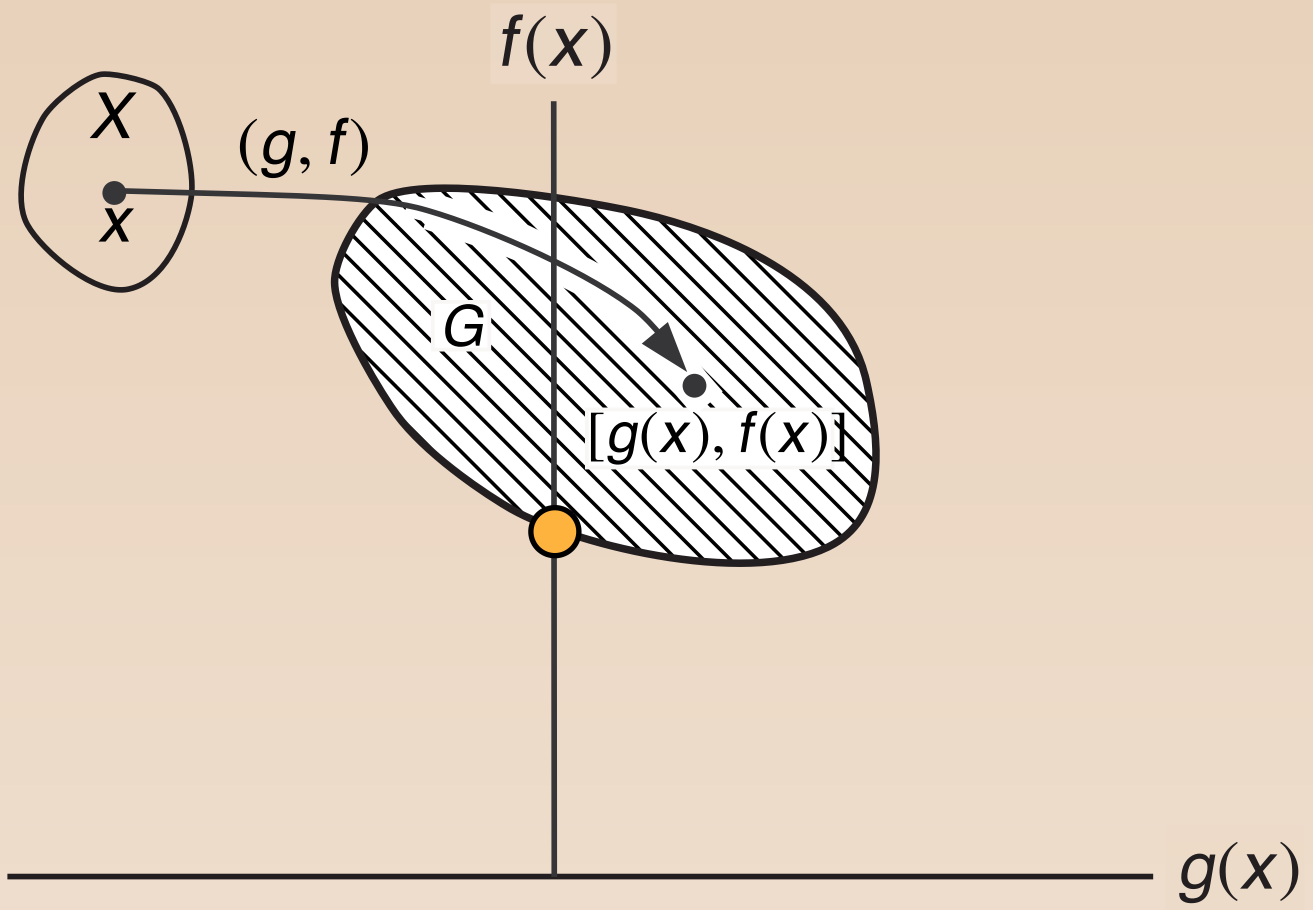
minimise $f(x)$,

subject to:

$g(x) \leq 0$,

$x \in X$,

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$.



Define the following set in $\mathbb{R}^2$:

$$G = \{(y, z) : y = g(x), z = f(x) \text{ for some } x \in X\},$$

that is, G is the image of X under the (g, f) map.

Geometric Interpretation

Lagrangian Dual Problem D

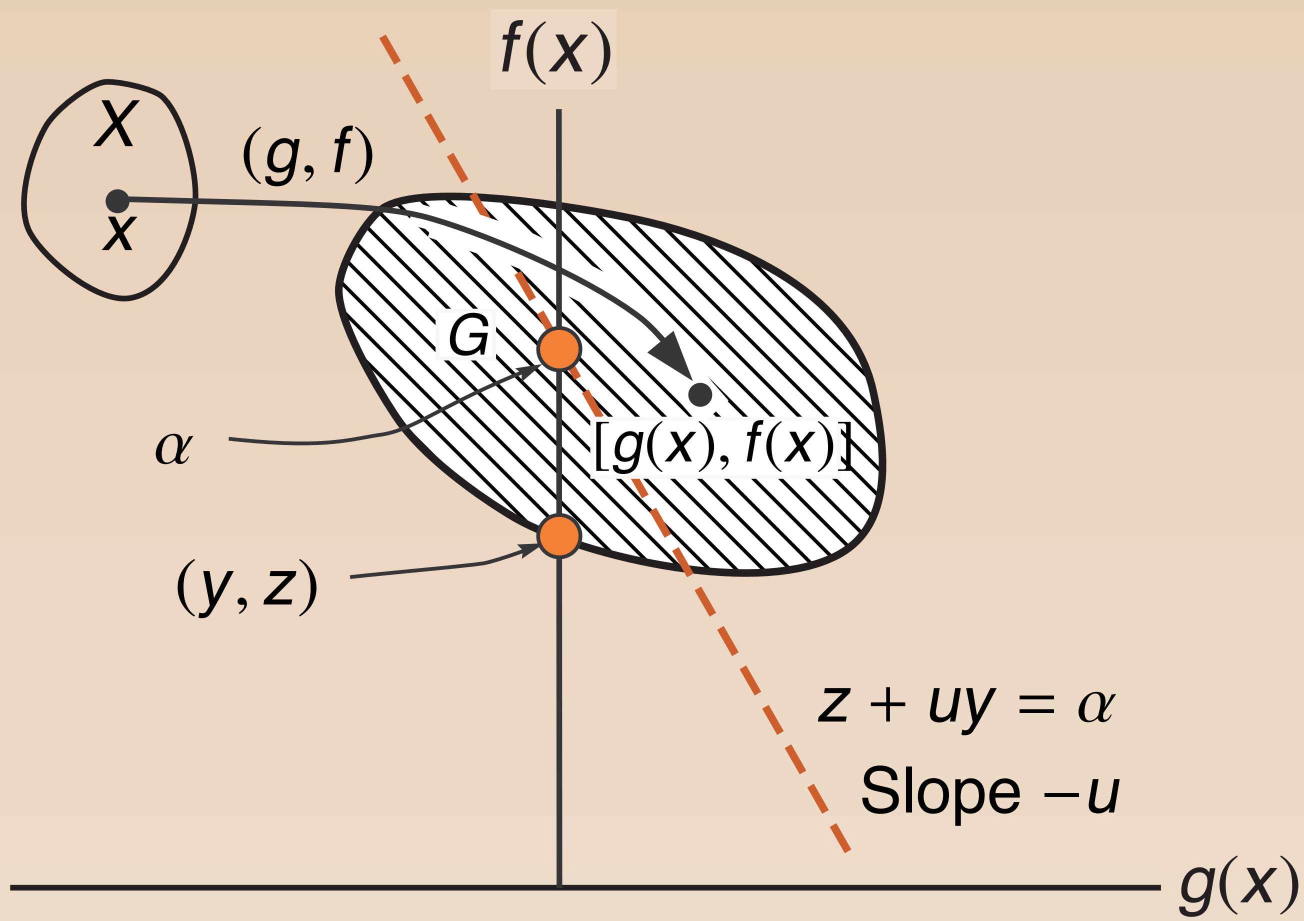
maximise $\theta(u)$,

subject to:

$u \geq 0$,

where (*Lagrangian dual subproblem*):

$$\theta(u) = \inf\{f(x)+ug(x) : x \in X\}.$$



Legendre transform (w/ wrong sign)

Given $u \geq 0$, the Lagrangian dual subproblem is equivalent to minimise $z + uy$ over points (y, z) in G. Note that $z + uy = \alpha$ is the equation of a straight line with slope $-u$ that intercepts the z-axis at α.

Geometric Interpretation

Lagrangian Dual Problem D

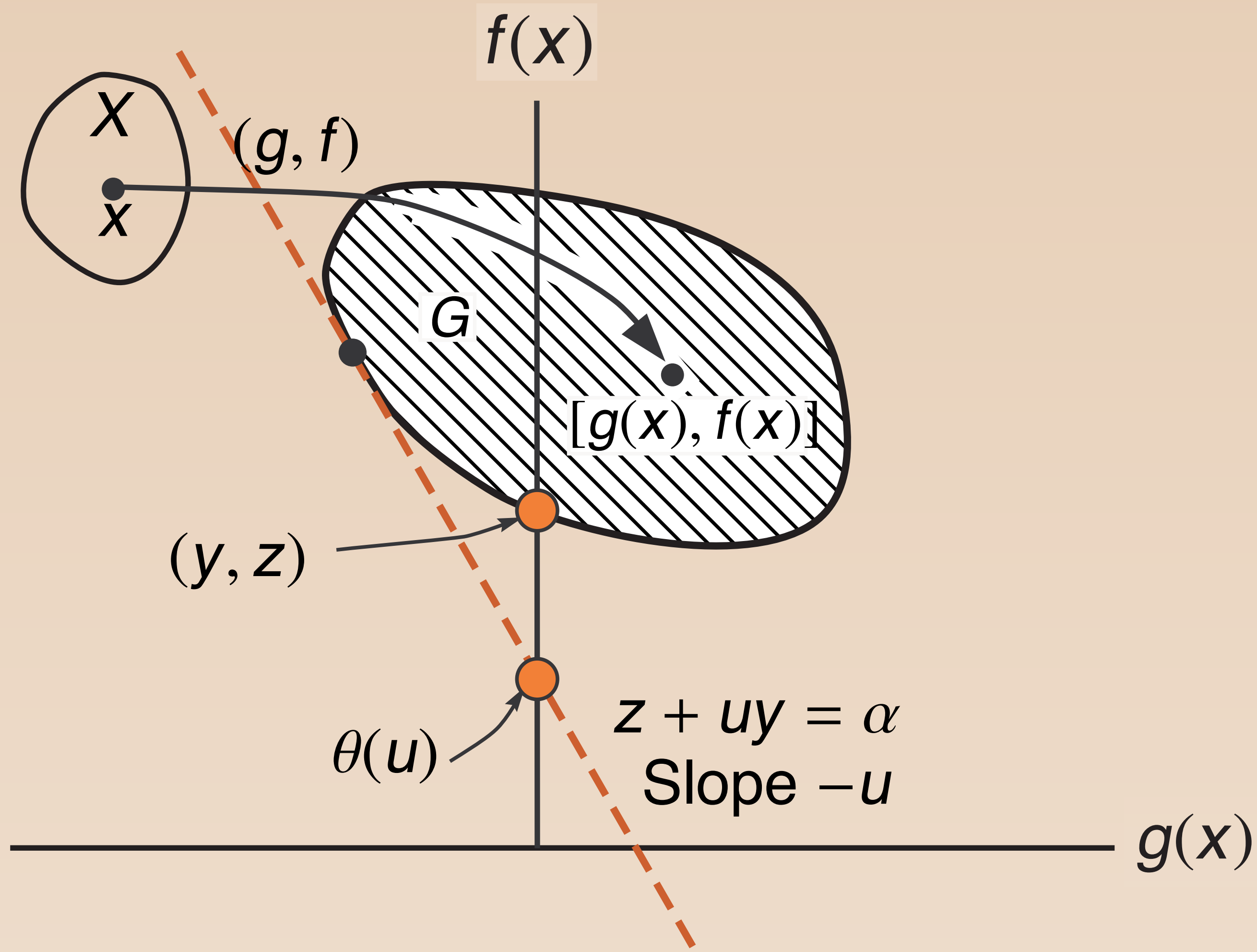
maximise $\theta(u)$,

subject to:

$u \geq 0$,

where (*Lagrangian dual subproblem*):

$$\theta(u) = \inf\{f(x)+ug(x) : x \in X\}.$$



Legendre transform (w/ wrong sign)

In order to minimise $z + uy$ over G we need to move the line $z + uy = \alpha$ parallel to itself as far down as possible, whilst it remains in contact with G. The last intercept on the z-axis thus obtained is the value of $\theta(u)$ corresponding to the given $u \geq 0$.

Geometric Interpretation

Lagrangian Dual Problem D

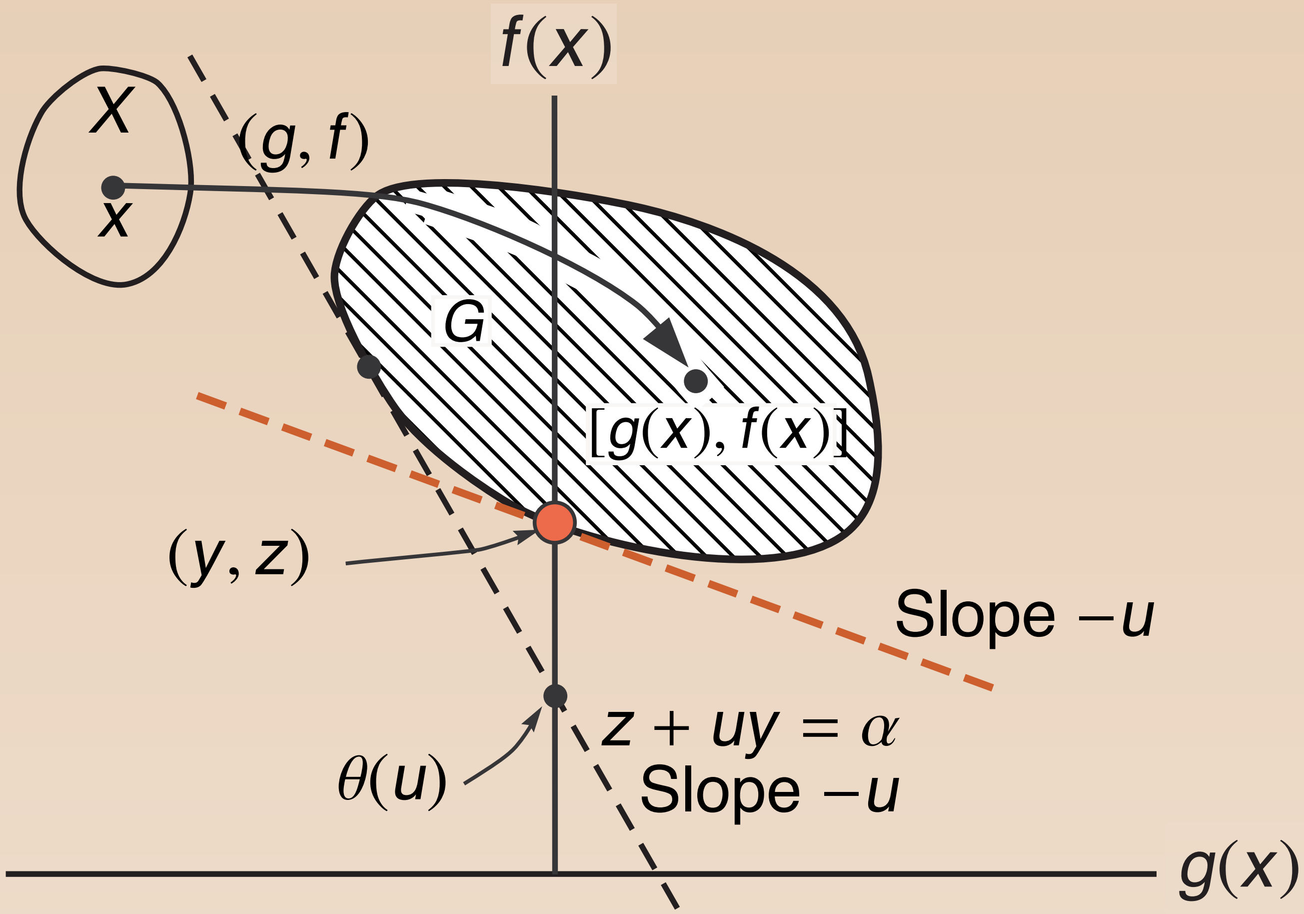
maximise $\theta(u)$,

subject to:

$u \geq 0$,

where (*Lagrangian dual subproblem*):

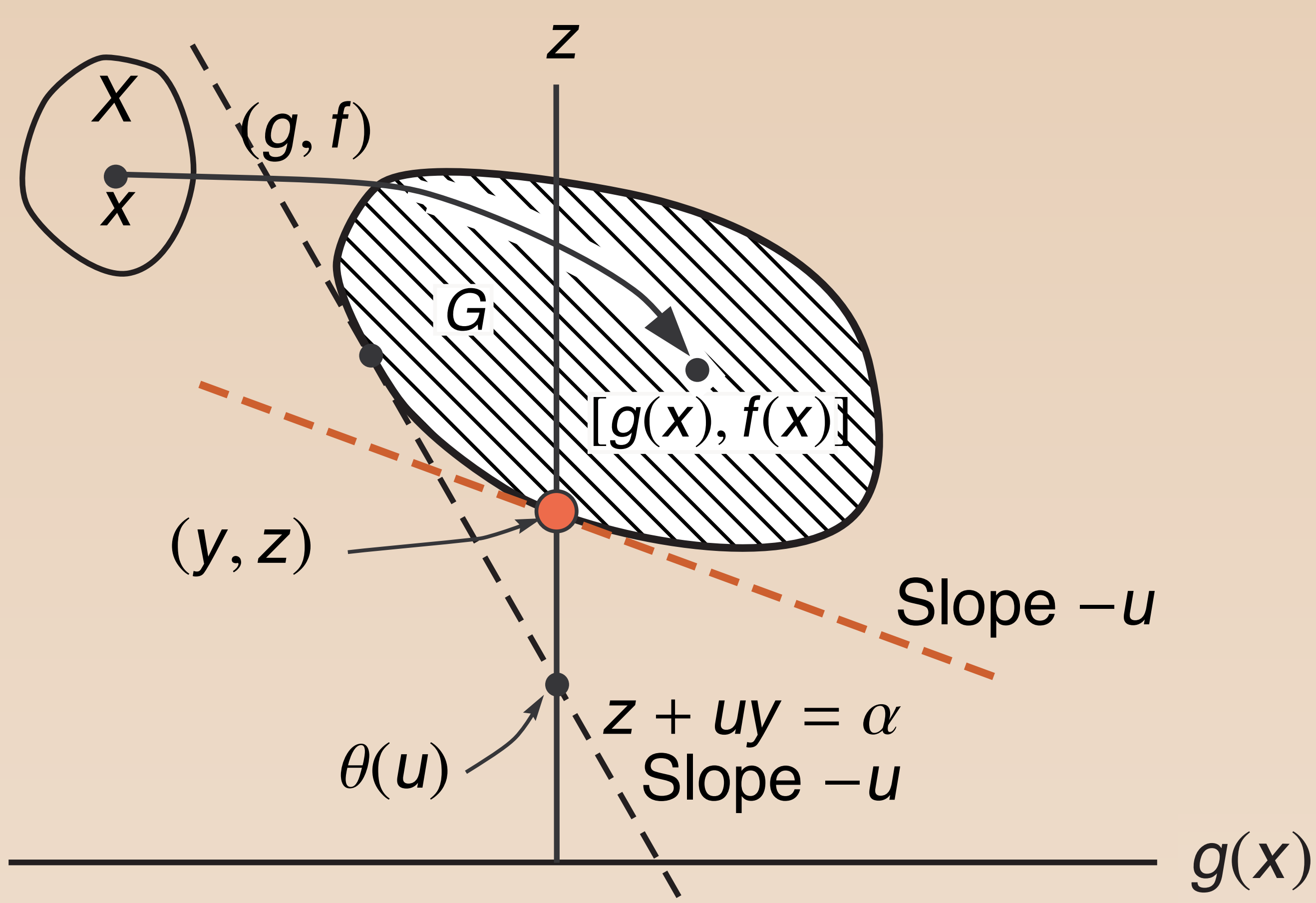
$$\theta(u) = \inf\{f(x) + ug(x) : x \in X\}.$$



Legendre transform (w/ wrong sign)

Finally, to solve the dual problem, we have to find the line with slope $-u$ ($u \geq 0$) such that the last intercept on the z-axis, $\theta(u)$, is maximal. Such a line has slope $-u$ and supports the set G at the point (y, z). Thus, the solution to the dual problem is u, and the optimal dual objective value is z.

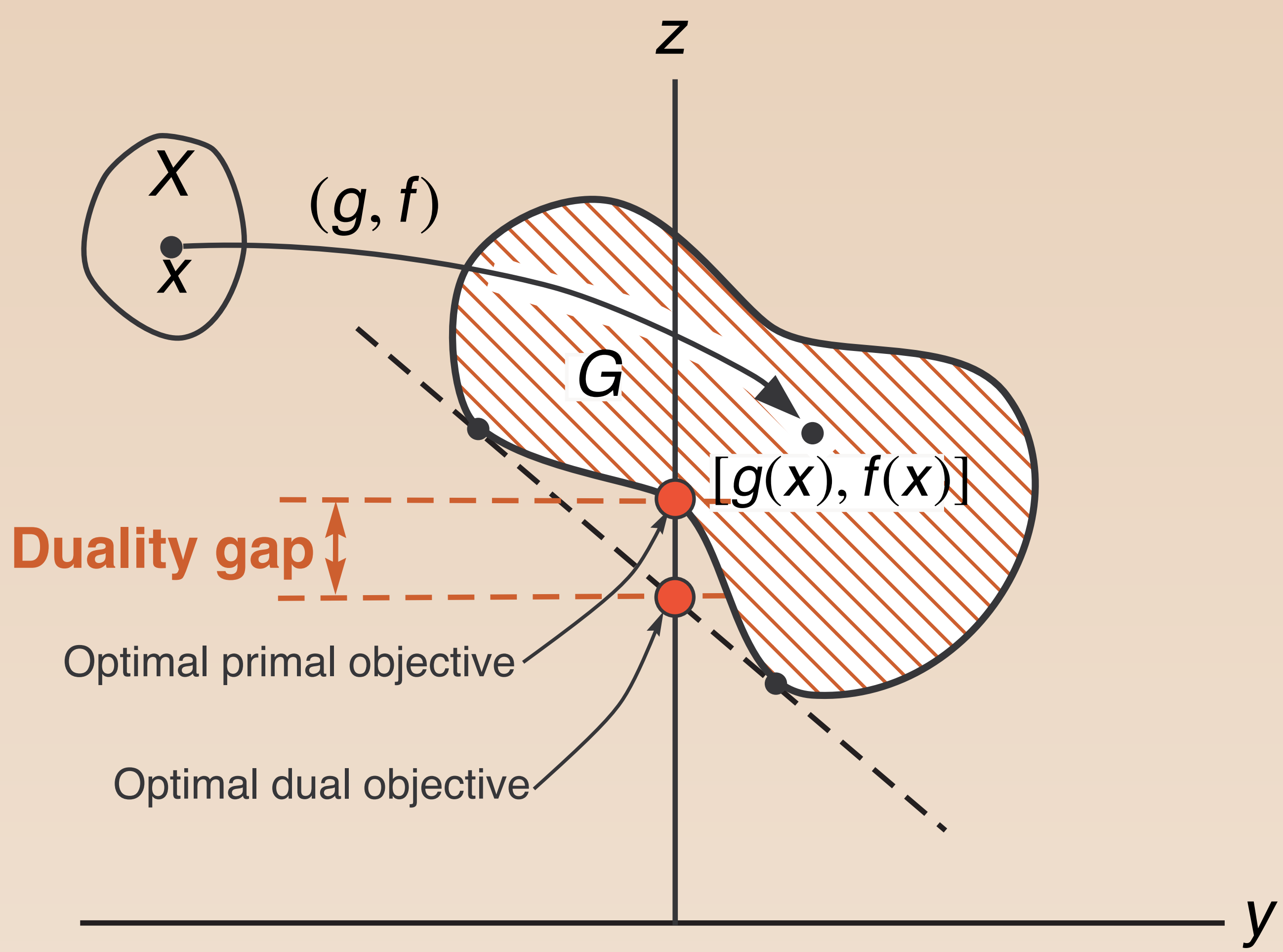
Geometric Interpretation



- The solution of the Primal problem is z, and the solution of the Dual problem is also z.
- It can be seen that, in the example illustrated, the optimal primal and dual objective values are equal. In such cases, it is said that there is no *duality gap* (strong duality).

Weak Duality

The figure shows an example of the geometric interpretation of the primal and dual problems.



Notice that, in the case shown in the figure, there exists a duality gap due to the nonconvexity of the set G.

We will see, in the **Strong Duality Theorem**, that if some suitable convexity conditions are satisfied, then there is no duality gap between the primal and dual optimisation problems.

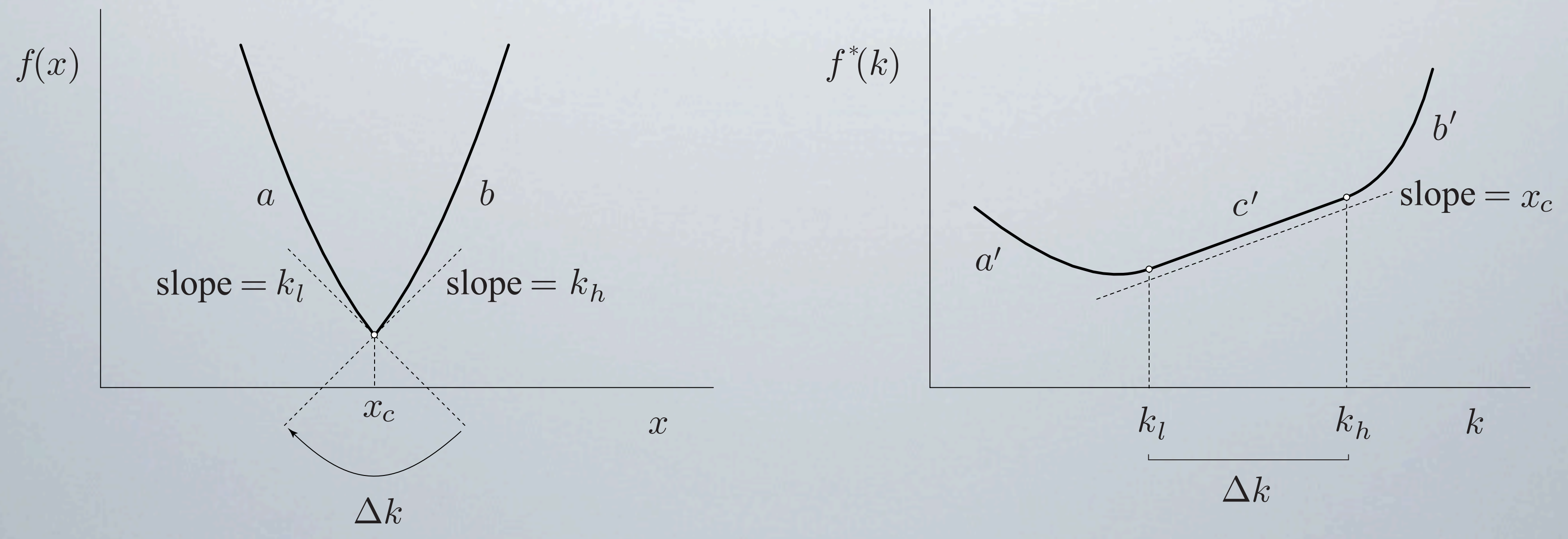


Figure 3: Function having a nondifferentiable point; its LF transform is affine.

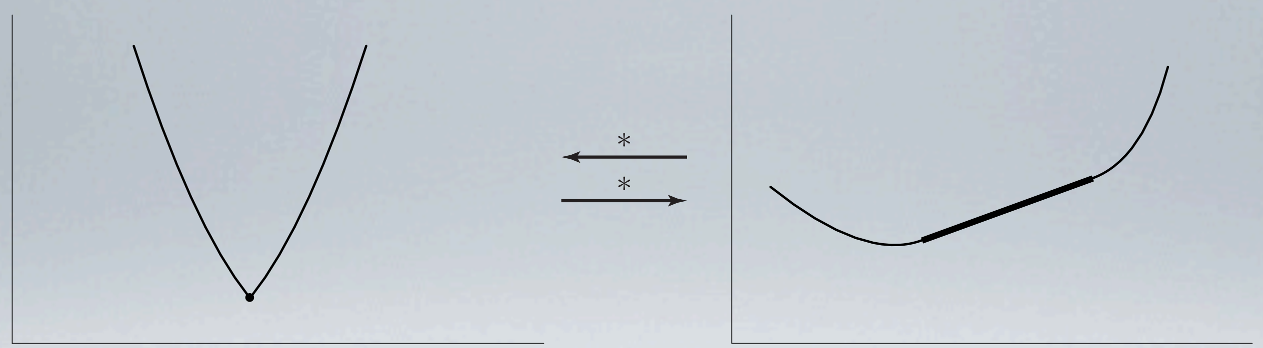


Figure 4: Nondifferentiable points are transformed into affine parts under the action of the LF transform and vice versa.

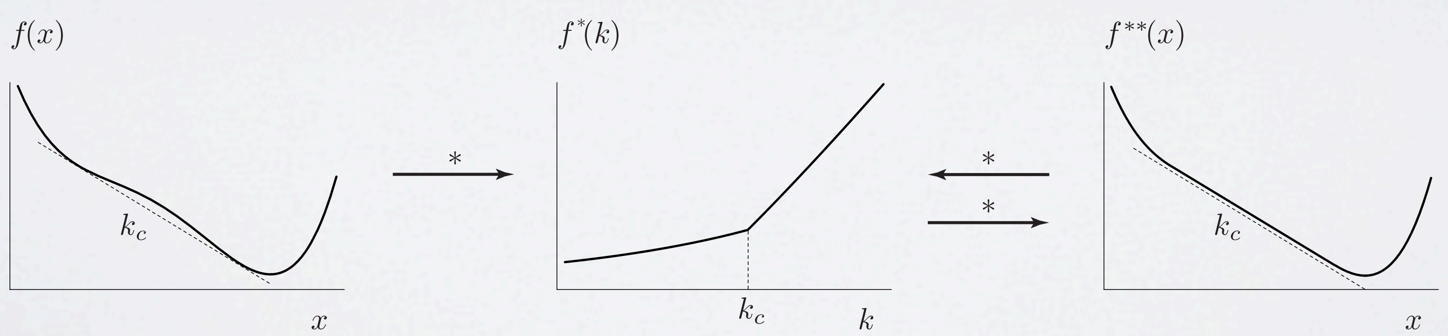


Figure 8: Structure of the LF transform for nonconvex functions.