School of Mathematics

Duality in Optimization

Jacek Gondzio
Email: J.Gondzio@ed.ac.uk
URL: http://www.maths.ed.ac.uk/~gondzio

Outline

- **Convexity**
  - convex sets, convex functions
  - local optimum, global optimum
- **Duality**
  - Lagrange duality
  - Wolfe duality
  - primal-dual pairs of LPs and QPs
  - geometric interpretation of duality

Reading:

Bertsekas, D., *Nonlinear Programming*,
Athena Scientific, Massachusetts, 1995. ISBN 1-886529-14-0.

Optimization

Consider the general optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \end{array}$$

where $x \in \mathcal{R}^n$, and $f : \mathcal{R}^n \mapsto \mathcal{R}$ and $g : \mathcal{R}^n \mapsto \mathcal{R}^m$ are convex, twice differentiable.

Basic Assumptions:

f and g are convex
$\Rightarrow$ If there exists a **local** minimum then it is a **global** one.

f and g are twice differentiable
$\Rightarrow$ We can use the **second order Taylor approximations** of them.

Glossary

LP: Linear Programming
both f and g are linear.

QP: Quadratic Programming
f is quadratic and g is linear.

NLP: Nonlinear Programming
f or g is nonlinear.

SOCP: Second-Order Cone Programming
f, g are conic (quadratic) functions.

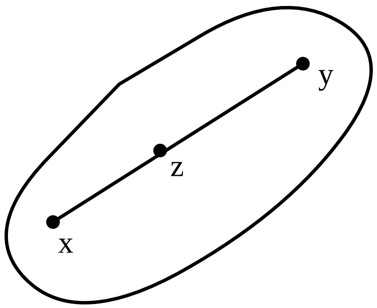
SDP: Semidefinite Programming
f, g are functions of positive definite matrices.

IPMs: Interior Point Methods.

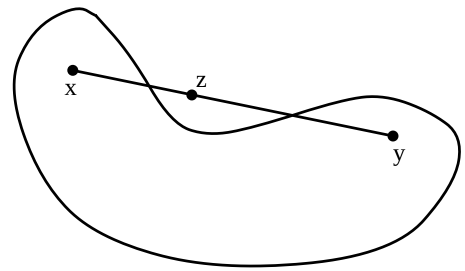
Convexity

Convexity is a key property in optimization.

Def. A set $C \subset \mathcal{R}^n$ is *convex* if $\lambda x + (1-\lambda)y \in C, \ \forall x, y \in C, \ \forall \lambda \in [0,1]$.

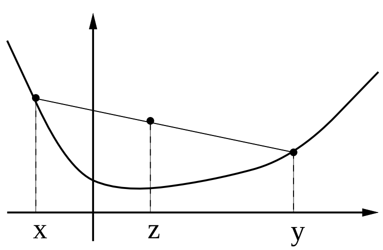


Convex set

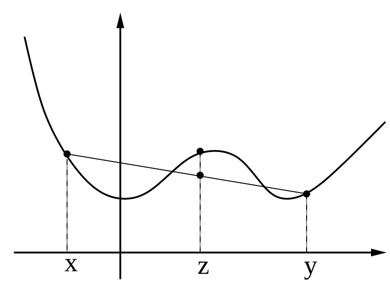


Nonconvex set

Def. Let C be a convex subset of $\mathcal{R}^n$. A function $f : C \mapsto \mathcal{R}$ is *convex* if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \ \forall x, y \in C, \ \forall \lambda \in [0,1]$.



Convex function



Nonconvex function

Convexity (cont'd)

Def. Let C be a convex subset of $\mathcal{R}^n$.
A function $f : C \mapsto \mathcal{R}$ is *concave* if
$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y), \; \forall x, y \in C, \; \forall \lambda \in [0,1]$.

Remark. A function $f : C \mapsto \mathcal{R}$ is concave if and only if function $-f$ is convex.

Def. Let C be a convex subset of $\mathcal{R}^n$.
A function $f : C \mapsto \mathcal{R}$ is *strictly convex* if
$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \; \forall x, y \in C, \; \forall \lambda \in (0,1)$.

Def. Let C be a convex subset of $\mathcal{R}^n$.
A function $f : C \mapsto \mathcal{R}$ is *strictly concave* if
$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda) f(y), \; \forall x, y \in C, \; \forall \lambda \in (0,1)$.

Convexity and Optimization

Consider a problem

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \end{array}$$

where X is a set of feasible solutions
and $f : X \to \mathcal{R}$ is an objective function.

Def. A vector $\hat{x}$ is a **local** minimum of f if

$$\exists \epsilon > 0 \text{ such that } f(\hat{x}) \leq f(x), \ \forall x \mid \|x - \hat{x}\| < \epsilon.$$

Def. A vector $\hat{x}$ is a **global** minimum of f if

$$f(\hat{x}) \leq f(x), \ \forall x \in X.$$

Lemma. If X is a convex set and $f : X \mapsto \mathcal{R}$ is a convex function, then a **local** minimum is a **global** minimum.

Proof.
Suppose that x is a local minimum, but not a global one. Then $\exists y \neq x$ such that $f(y) < f(x)$.
From convexity of f, for all $\lambda \in [0,1]$, we have

$$\begin{aligned} f((1-\lambda)x+\lambda y) &\leq (1-\lambda)f(x)+\lambda f(y) \\ &< (1-\lambda)f(x)+\lambda f(x) = f(x). \end{aligned}$$

In particular, for a sufficiently small λ, the point $z = (1-\lambda)x+\lambda y$ lies in the ϵ-neighbourhood of x and $f(z) < f(x)$. This contradicts the assumption that x is a local minimum.

Useful properties

1. For any collection $\{C_i \mid i \in I\}$ of convex sets, the intersection $\bigcap_{i \in I} C_i$ is convex.

2. The vector sum $\{x_1 + x_2 \mid x_1 \in C_1, x_2 \in C_2\}$ of two convex sets C_1 and C_2 is convex.

3. The image of a convex set under a linear transformation is convex.

4. If C is a convex set and $f : C \mapsto \mathcal{R}$ is a convex function, the level sets $\{x \in C \mid f(x) \leq \alpha\}$ and $\{x \in C \mid f(x) < \alpha\}$ are convex for all scalars α.

5. For any collection $\{f_i : C \mapsto \mathcal{R} \mid i \in I\}$ of convex functions, the weighted sum, with weights $w_i > 0$, $i \in I$, i.e. the function $f = \sum_{i \in I} w_i f_i : C \mapsto \mathcal{R}$, is convex.

6. If I is an index set, $C \in \mathcal{R}^n$ is a convex set, and $f_i : C \mapsto \mathcal{R}$ is convex $\forall i \in I$, then the function $h : C \mapsto \mathcal{R}$ defined by

$$h(x) = \sup_{i \in I} f_i(x)$$

is also convex.

7. Let $C \in \mathcal{R}^n$ be a convex set and $f : C \mapsto \mathcal{R}$ be differentiable over C.
(a) The function f is convex if and only if

$$f(y) \geq f(x) + \nabla^T f(x)(y - x), \quad \forall x, y \in C.$$

(b) If the inequality is strict for $x \neq y$, then f is strictly convex.

8. Let $C \in \mathcal{R}^n$ be a convex set and $f : C \mapsto \mathcal{R}$ be twice continuously differentiable over C.
(a) If $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$, then f is convex.
(b) If $\nabla^2 f(x)$ is positive definite for all $x \in C$, then f is strictly convex.
(c) If f is convex, then $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$.

9. Let $C \in \mathcal{R}^n$ be a convex set and Q a square matrix. Let $f(x) = x^T Q\, x$ be a quadratic function $f : C \mapsto \mathcal{R}$.
(a) f is convex iff Q is positive semidefinite.
(b) f is strictly convex iff Q is positive definite.

Duality

Consider a general optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \\ & x \in X \subseteq \mathcal{R}^n, \end{array} \tag{1}$$

where $f : \mathcal{R}^n \mapsto \mathcal{R}$ and $g : \mathcal{R}^n \mapsto \mathcal{R}^m$.
The set X is arbitrary; it may include, for example, an integrality constraint.

Let $\hat{x}$ be an optimal solution of (1) and define

$$\hat{f} = f(\hat{x}).$$

Introduce Lagrange multiplier $y_i \geq 0$ for every inequality constraint $g_i(x) \leq 0$.
Define $y = (y_1, \ldots, y_m)^T$ and the **Lagrangian**

$$L(x, y) = f(x) + y^T g(x),$$

y are also called *dual* variables.

Consider the problem

$$L_D(y) \;=\; \min_x L(x,y) \quad s.t. \quad x \in X \subseteq \mathcal{R}^n.$$

Its optimal solution x depends on y and so does the optimal objective $L_D(y)$.

Lemma. For any $y \geq 0$, $L_D(y)$ is a lower bound on $\hat{f}$ (the optimal solution of (1)), i.e.,

$$\hat{f} \geq L_D(y) \qquad \forall y \geq 0.$$

Proof.

$$\begin{aligned} \hat{f} = & \;\min \;\{f(x) \mid g(x) \leq 0, x \in X\} \\ \geq & \;\min \;\left\{f(x) + y^T g(x) \mid g(x) \leq 0, y \geq 0, x \in X\right\} \\ \geq & \;\min \;\left\{f(x) + y^T g(x) \mid y \geq 0, x \in X\right\} \\ = & \;L_D(y). \end{aligned}$$

Corollary.

$$\hat{f} \geq \max_{y \geq 0} L_D(y), \quad \text{i.e.,} \quad \hat{f} \geq \max_{y \geq 0} \min_{x \in X} L(x,y).$$

Lagrangian Duality

If $\exists i \; g_i(x) > 0$, then

$$\max_{y \geq 0} L(x,y) = +\infty$$

(we let the corresponding y_i grow to $+\infty$).
If $\forall i \; g_i(x) \leq 0$, then

$$\max_{y \geq 0} L(x,y) = f(x),$$

because $\forall i \; y_i g_i(x) \leq 0$ and the maximum is attained when

$$y_i g_i(x) = 0, \qquad \forall i = 1, 2, ..., m.$$

Hence the problem (1) is equivalent to the following **MinMax** problem

$$\min_{x \in X} \max_{y \geq 0} L(x,y),$$

which could also be written as follows:

$$\hat{f} = \min_{x \in X} \max_{y \geq 0} L(x,y).$$

Consider the following problem

$$\min \; \{ f(x) \mid g(x) \leq 0, x \in X \} \,,$$

where f, g and X are arbitrary.
With this problem we associate the **Lagrangian**

$$L(x, y) = f(x) + y^T g(x),$$

y are *dual* variables (Lagrange multipliers).
The **weak duality** always holds:

$$\min_{x \in X} \max_{y \geq 0} L(x, y) \;\geq\; \max_{y \geq 0} \min_{x \in X} L(x, y).$$

We have not made **any** assumption about functions f and g and set X.

If f and g are convex, X is convex and certain regularity conditions are satisfied, then

$$\min_{x \in X} \max_{y \geq 0} L(x, y) \;=\; \max_{y \geq 0} \min_{x \in X} L(x, y).$$

This is called the **strong duality**.

Notation: Consider again the problem

$$\begin{array}{cc} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \\ & x \in X \subseteq \mathcal{R}^n, \end{array}$$

where $f : \mathcal{R}^n \mapsto \mathcal{R}$ and $g : \mathcal{R}^n \mapsto \mathcal{R}^m$.
Take $x \in X \subseteq \mathcal{R}^n$ and $y \in Y = \{y \in \mathcal{R}^m, y \geq 0\}$ and write the Lagrangian

$$L(x,y) = f(x) + y^T g(x).$$

Define the **primal function**

$$L_P(x) = \begin{cases} f(x) & \text{if } \forall i \; g_i(x) \leq 0 \\ +\infty & \text{if } \exists i \; g_i(x) > 0. \end{cases}$$

Observe that

$$L_P(x) = \max_{y \geq 0} L(x,y). \tag{2}$$

Define the **dual function**

$$L_D(y) = \min_{x \in X} L(x,y). \tag{3}$$

Primal & Dual Problems

The problem (1) can be formulated as looking for $\hat{x} \in X \subseteq \mathcal{R}^n$ such that

$$L_P(\hat{x}) = \min_{x \in X} L_P(x).$$

It is called the **primal problem**.

The problem

$$L_D(\hat{y}) = \max_{y \geq 0} L_D(y).$$

is called the **dual problem**.

The **weak duality** can be rewritten as:

$$L_P(\hat{x}) \geq L_D(\hat{y}).$$

Primal & Dual Feasibility Sets

Def. *Primal feasible set.*

$$X_P = \{x : x \in X, g_i(x) \leq 0, i = 1, 2, \ldots, m\}.$$

Def. *Dual feasible set.*
A tuple $(x, y) \in \mathcal{R}^{n+m}$ is feasible for the dual problem if

$$(x,y) \in Y_D = \{(x,y) : x \in X, y \in Y, L_D(y) = L(x,y)\}.$$

Def. *Dual optimal solution.*
A tuple $(\hat{x}, \hat{y}) \in \mathcal{R}^{n+m}$ is called dual optimal if $(\hat{x}, \hat{y}) \in Y_D$ and $\hat{y}$ maximizes $L_D(y)$.

Primal & Dual Bounds

Lemma. If $x^1 \in X_P$ and $(x^2, y^2) \in Y_D$ (i.e., x^1 is primal feasible and (x^2, y^2) is dual feasible), then

$$L_P(x^1) \geq L_D(y^2).$$

Proof. Since $x^1 \in X_P$ we get $L_P(x^1) = f(x^1)$. For any $y \in Y$, from definition (2) we have $L_P(x^1) \geq L(x^1, y)$. In particular, for $y = y^2$:

$$L_P(x^1) \geq L(x^1, y^2). \tag{4}$$

On the other hand, $(x^2, y^2) \in Y_D$ hence for any $x \in X$ from (3) we have $L(x, y^2) \geq L_D(y^2)$ and, in particular, for $x = x^1$:

$$L(x^1, y^2) \geq L_D(y^2). \tag{5}$$

From (4) and (5) we get

$$f(x^1) = L_P(x^1) \geq L(x^1, y^2) \geq L_D(y^2),$$

which completes the proof.

Any *primal* feasible solution provides an **upper** bound for the *dual* problem, and any *dual* feasible solution provides a **lower** bound for the *primal* problem.

Duality and Convexity

The weak duality holds regardless of the form of functions f, g and set X:

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \;\geq\; \max_{y \geq 0} \min_{x \in X} L(x,y).$$

What do we need for the *inequality* in the weak duality to become an *equation*?
If

- $X \subseteq \mathcal{R}^n$ is open and convex;
- f and g are convex;
- optimal solution is finite;
- some mysterious *regularity conditions* hold,

then strong duality holds. That is

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \;=\; \max_{y \geq 0} \min_{x \in X} L(x,y).$$

An example of regularity conditions:
$\exists x \in int(X)$ such that $g(x) < 0$.

Lagrange duality does not need differentiability.
Suppose f and g are convex and differentiable. Suppose X is convex.
The **dual function**

$$L_D(y) = \min_{x \in X} L(x, y).$$

requires minimization with respect to x.
Instead of **minimization** with respect to x,
we ask for a **stationarity** with respect to x:

$$\nabla_x L(x, y) = 0.$$

Lagrange dual problem:

$$\max_{y \geq 0} L_D(y) \qquad \left(\text{i.e., } \max_{y \geq 0} \min_{x \in X} L(x, y) \right).$$

Wolfe dual problem:

$$\begin{array}{rc} \max & L(x, y) \\ \text{s.t.} & \nabla_x L(x, y) = 0 \\ & y \geq 0. \end{array}$$

Duality: Example

Consider the nonlinear program:

$$\min_{x_1,x_2} \; f(x) = x_1^2 + x_2^2$$
$$\text{s.t.} \quad x_1 + x_2 \geq 1.$$

$f(x) = x_1^2 + x_2^2$ and $g(x) = 1 - x_1 - x_2$ are convex.
Observe that $\hat{x} = (0,0)$ is the only stationary point of f and since f is convex it is the unique unconstrained minimizer of f. However this point is infeasible and since there are no other possible unconstrained local optima, the constrained optimum must lie on the boundary of the feasible region, and so satisfies $x_1 + x_2 = 1$.
Using this to eliminate x_2 gives $f(x_1, 1-x_1) = x_1^2 + (1-x_1)^2$, which has a minimum at $x_1 = \frac{1}{2}$. Hence constrained minimizer is at $\hat{x} = (\frac{1}{2}, \frac{1}{2})$, with minimum $\hat{f} = 0.5$.

Duality: Example (continued)

Lagrangian:

$$L(x,y) = x_1^2 + x_2^2 + y(1 - x_1 - x_2).$$

The Lagrangian dual function:

$$L_D(y) = \min_x \, [x_1^2 + x_2^2 + y(1 - x_1 - x_2)].$$

For any y the Lagrangian $L(x,y)$ is convex in x. We can use the stationarity condition to replace the minimization. We write:

$$\nabla_x L(x,y) = \begin{bmatrix} 2x_1 - y \\ 2x_2 - y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which gives $x_1 = 0.5y$ and $x_2 = 0.5y$.

Example (continued)

Having substituted $x_1 = 0.5y$ and $x_2 = 0.5y$, we obtain:

$$L_D(y) = y - \frac{1}{2}y^2.$$

The dual problem

$$\max_{y \geq 0} L_D(y),$$

thus becomes

$$\max_{y \geq 0} [y - \frac{1}{2}y^2].$$

It has the obvious solution $\hat{y} = 1$.
We observe that $L_D(\hat{y}) = \frac{1}{2} = \hat{f} = f(\hat{x})$, so strong duality holds.
We have calculated these optimal solutions $\hat{y}$ and $\hat{x}$, but even if we did not already know they were optimal, the Corollary 3 would confirm they were optimal.
As observed, this is a convex program so it was to be expected that strong duality would hold and the duality gap would be zero.

Duality and Convexity

The weak duality holds regardless of the form of functions f, g and set X:

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \;\geq\; \max_{y \geq 0} \min_{x \in X} L(x,y).$$

What do we need for the *inequality* in the weak duality to become an *equation*?
If

- $X \subseteq \mathcal{R}^n$ is open and convex;
- f and g are convex;
- optimal solution is finite;
- some mysterious *regularity conditions* hold,

then strong duality holds. That is

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \;=\; \max_{y \geq 0} \min_{x \in X} L(x,y).$$

An example of regularity conditions:
$\exists x \in int(X)$ such that $g(x) < 0$.

Lagrange duality does not need differentiability.
Suppose f and g are convex and differentiable. Suppose X is convex.
The **dual function**

$$L_D(y) = \min_{x \in X} L(x, y).$$

requires minimization with respect to x.
Instead of **minimization** with respect to x,
we ask for a **stationarity** with respect to x:

$$\nabla_x L(x, y) = 0.$$

Lagrange dual problem:

$$\max_{y \geq 0} L_D(y) \qquad \left(\text{i.e., } \max_{y \geq 0} \min_{x \in X} L(x, y) \right).$$

Wolfe dual problem:

$$\begin{array}{rc} \max & L(x, y) \\ \text{s.t.} & \nabla_x L(x, y) = 0 \\ & y \geq 0. \end{array}$$

Lagrange Duality and Wolfe Duality

If f and g are convex and differentiable and $X = \mathcal{R}^n$, then $\min_x L(x, y)$ occurs where $\nabla_x L(x, y) = 0$.

Hence the Wolfe Dual Problem is equivalent to the Langrangian Dual Problem.

However even then, the Wolfe Dual Problem is not necessarily a convex problem.

Lagrangian duality is very general:
no assumptions on f, g and X are made.

Wolfe duality requires differentiability of f and g.

Dual Linear Program

Consider a linear program

$$\begin{array}{lrcl} \min & c^T x & & \\ \text{s.t.} & Ax & = & b, \\ & x \geq 0, & & \end{array}$$

where $c, x \in \mathcal{R}^n, b \in \mathcal{R}^m$, $A \in \mathcal{R}^{m \times n}$.
We associate Lagrange multipliers $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$ $(s \geq 0)$ with the constraints $Ax = b$ and $x \geq 0$, and write the **Lagrangian**

$$L(x, y, s) \; = \; c^T x - y^T (Ax - b) - s^T x.$$

To determine the *Lagrangian dual*

$$L_D(y, s) = \min_{x \in X} L(x, y, s)$$

we need stationarity with respect to x:

$$\nabla_x L(x, y, s) \; = \; c - A^T y - s \; = \; 0.$$

Hence

$$
\begin{array}{rcl}
L_D(y,s) & = & c^T x - y^T(Ax-b) - s^T x \\
 & = & b^T y + x^T(c - A^T y - s) = b^T y.
\end{array}
$$

and the **dual LP** has a form:

$$
\begin{array}{rrrrl}
\max & b^T y & & & \\
\text{s.t.} & A^T y & + & s & = c, \\
 & y \text{ free}, & & s \geq 0, &
\end{array}
$$

where $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$.

Primal Problem	**Dual Problem**
$\min \ c^T x$	$\max \ b^T y$
$\text{s.t.} \ Ax = b,$	$\text{s.t.} \ A^T y + s = c,$
$x \geq 0;$	$s \geq 0.$

Dual Quadratic Program

Consider a quadratic program

$$\begin{array}{ll} \min & c^T x + \frac{1}{2} x^T Q\, x \\ \text{s.t.} & Ax = b, \\ & x \geq 0, \end{array}$$

where $c, x \in \mathcal{R}^n, b \in \mathcal{R}^m$, $A \in \mathcal{R}^{m \times n}$, $Q \in \mathcal{R}^{n \times n}$.
We associate Lagrange multipliers $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$ $(s \geq 0)$ with the constraints $Ax = b$ and $x \geq 0$, and write the **Lagrangian**

$$L(x, y, s) \;=\; c^T x + \frac{1}{2} x^T Q\, x - y^T (Ax - b) - s^T x.$$

To determine the *Lagrangian dual*

$$L_D(y, s) = \min_{x \in X} L(x, y, s)$$

we need stationarity with respect to x:

$$\nabla_x L(x, y, s) \;=\; c + Qx - A^T y - s \;=\; 0.$$

Hence

$$\begin{aligned} L_D(y,s) &= c^T x + \tfrac{1}{2} x^T Q\, x - y^T (Ax - b) - s^T x \\ &= b^T y + x^T (c + Qx - A^T y - s) - \tfrac{1}{2} x^T Q\, x \\ &= b^T y \; - \; \tfrac{1}{2} x^T Q\, x, \end{aligned}$$

and the **dual QP** has the form:

$$\begin{array}{rc} \max & b^T y - \frac{1}{2} x^T Q\, x \\ \text{s.t.} & A^T y + s - Qx = c, \\ & x, s \geq 0, \end{array}$$

where $y \in \mathcal{R}^m$ and $x, s \in \mathcal{R}^n$.

Primal Problem	**Dual Problem**
$\min \; c^T x + \frac{1}{2} x^T Q x$	$\max \; b^T y - \frac{1}{2} x^T Q x$
$\text{s.t.} \; Ax = b,$	$\text{s.t.} \; A^T y + s - Qx = c,$
$x \geq 0;$	$s \geq 0.$

Geometric View of Duality

Consider a mapping which for any $x \in X$ defines a point in $\mathcal{R}^{m+1}$ of the form $(g(x), f(x))$. We write $x \mapsto (g, f)$. Let H be the image of X.
Example $n = 2,\ m = 1$. Hence: $x \in X \subseteq \mathcal{R}^2$ and $f : \mathcal{R}^2 \mapsto \mathcal{R}$ and $g : \mathcal{R}^2 \mapsto \mathcal{R}$.
Lagrange multiplier: $y \in \mathcal{R}$ $(y \geq 0)$.

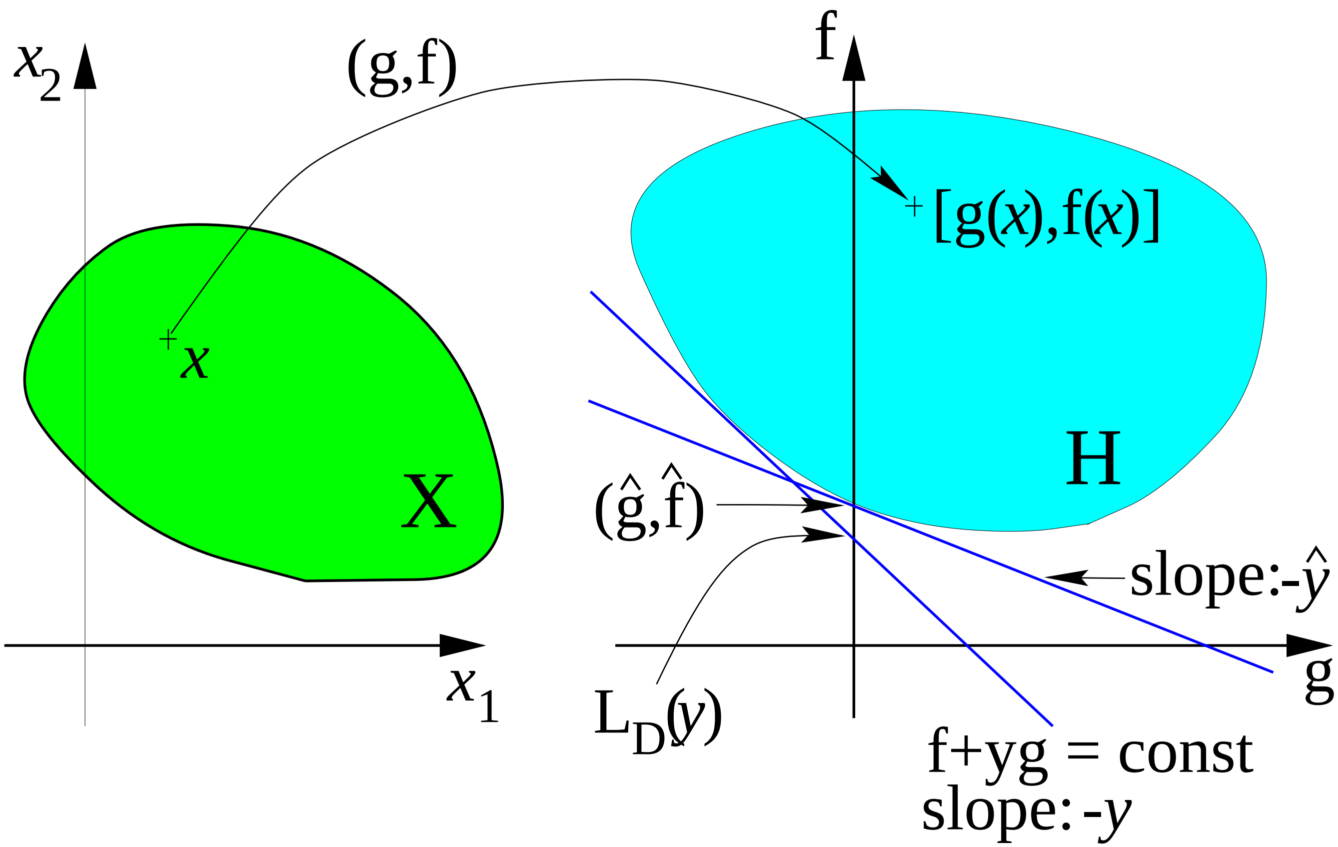


Figure Interpretation

Primal problem:

We look for a point $(g, f) \in H$ such that
$g \leq 0$ and f attains its minimum.
This is the point $(\hat{g}, \hat{f})$ in the Figure.

Figure Interpretation

Dual problem:

Take $y \geq 0$. To find $L_D(y)$, we need to minimize $f(x) + yg(x)$ with respect to $x \in X$. This corresponds to the minimization of the linear form $f + yg$ in the set H.
For a given $y \geq 0$, the linear form $f + yg$ has a fixed slope (equal to $-y$) and the minimum is attained when the line $f + yg$ touches the bottom of H. We say that "the hyperplane $f + yg$ supports the set H".
The intersection of the supporting plane and the f line determines the value of $L_D(y)$.

The dual problem consists in finding such a slope y that $L_D(y)$ is maximized, i.e., the intersection of the supporting plane and the f axis is the highest possible.

There are two supporting hyperplanes in the Figure. The one corresponding to $\hat{y}$ corresponds to the maximum of $L_D(y)$.

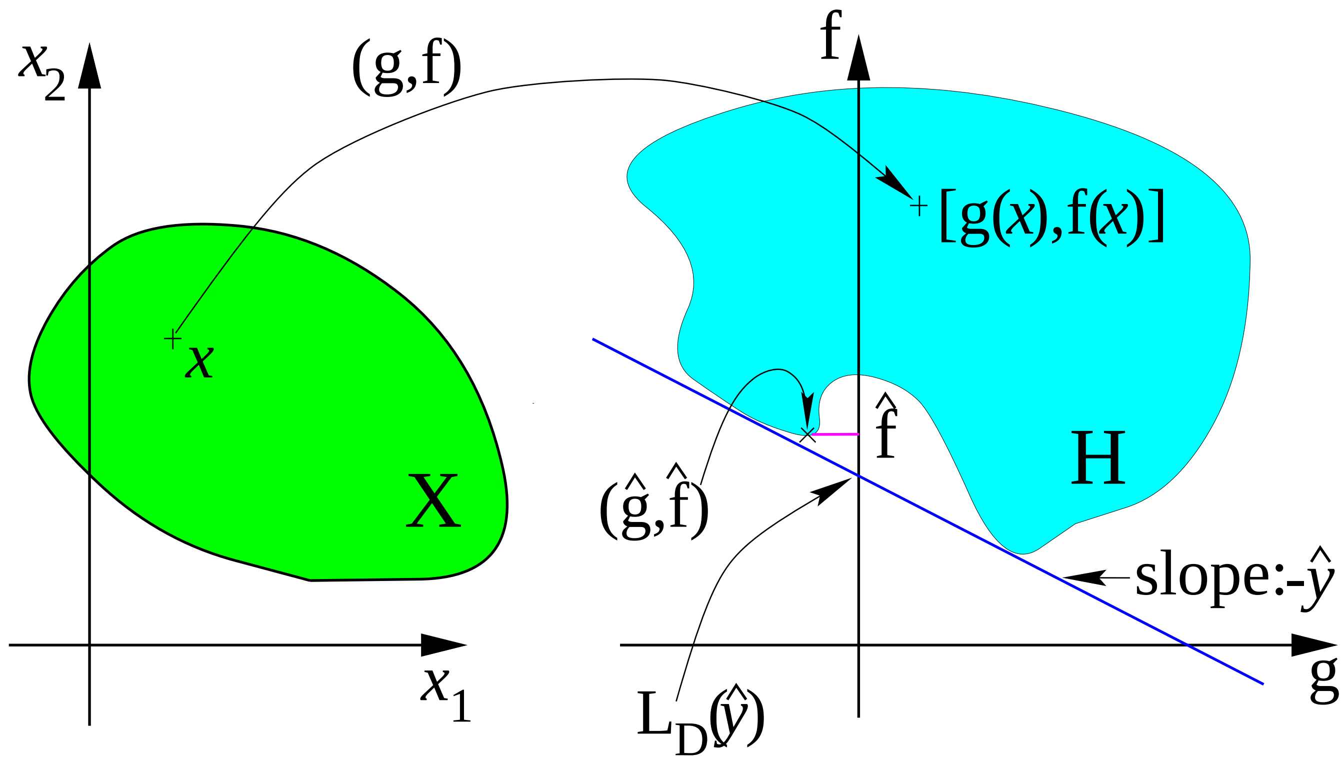
Nonzero Duality Gap

When sufficient conditions for strong duality are not satisfied, we may observe a nonzero duality gap:

$$\min_{x \in X} \max_{y \geq 0} L(x,y) - \max_{y \geq 0} \min_{x \in X} L(x,y) \;>\; 0.$$

In the Figure below:

$$\hat{f} - L_D(\hat{y}) \;>\; 0.$$



Treatment of Equality Constraints

Let $h : \mathcal{R}^n \mapsto \mathcal{R}^k$ define an *equality* constraint $h(x) = 0$ (understood as $h_j(x) = 0,\ j = 1, ..., k$). Replace $h_j(x) = 0$ with two inequalities:

$$h_j(x) \leq 0 \quad \text{and} \quad -h_j(x) \leq 0.$$

Then the optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \\ & h(x) = 0, \\ & x \in X \subseteq \mathcal{R}^n, \end{array}$$

where $f : \mathcal{R}^n \mapsto \mathcal{R}$, $g : \mathcal{R}^n \mapsto \mathcal{R}^m$ and $h : \mathcal{R}^n \mapsto \mathcal{R}^k$, becomes:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \\ & h(x) \leq 0, \\ & -h(x) \leq 0, \\ & x \in X \subseteq \mathcal{R}^n. \end{array}$$

Equality Constraints (continued)

Use nonnegative Lagrange multipliers $y \in \mathcal{R}^m$ for g constraints.
Use a pair of Lagrange multipliers $u_j^+ \geq 0$ and $u_j^- \geq 0$ for inequalities $h_j(x) \leq 0$ and $-h_j(x) \leq 0$, respectively. In other words, use two vectors $u^+ \geq 0$ and $u^- \geq 0$, both in $\mathcal{R}^k$ and write the Lagrangian

$$\begin{aligned} L(x, y, u^+, u^-) &= f(x) + y^T g(x) + (u^+)^T h(x) - (u^-)^T h(x) \\ &= f(x) + y^T g(x) + (u^+ - u^-)^T h(x) \\ &= f(x) + y^T g(x) + u^T h(x), \end{aligned}$$

where the vector $u = u^+ - u^- \in \mathcal{R}^k$ has no sign restriction.
The Lagrangian becomes:

$$L(x, y, u) = f(x) + y^T g(x) + u^T h(x),$$

and all theoretical results derived earlier can be replicated for this new problem formulation.