Nonlinear Optimization: Duality

INSEAD, Spring 2006

Jean-Philippe Vert

Ecole des Mines de Paris

`Jean-Philippe.Vert@mines.org`

Outline

- The Lagrange dual function
- Weak and strong duality
- Geometric interpretation
- Saddle-point interpretation
- Optimality conditions
- Perturbation and sensitivity analysis

The Lagrange dual function

Setting

We consider an equality and inequality constrained optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0\,, \quad i = 1, \ldots, m\,, \\ & g_j(x) \leq 0\,, \quad j = 1, \ldots, r\,, \end{aligned}$$

making *no assumption* of f, g and h.

We denote by f^* the optimal value of the decision function under the constraints, i.e., $f^* = f(x^*)$ if the minimum is reached at a global minimum x^*.

Lagrange dual function

- Remember the *Lagrangian* of this problem is the function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ defined by:

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) .$$

- We define the *Lagrange dual function* $g : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ as:

$$\begin{aligned} q(\lambda, \mu) &= \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \\ &= \inf_{x \in \mathbb{R}^n} \left(f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x) \right) . \end{aligned}$$

Properties of the dual function

When L in unbounded below in x, the dual function $q(\lambda, \mu)$ takes on the value $-\infty$. It has two important properties:

1. *q is concave in (λ, μ)*, even if the original problem is not convex.
2. The dual function yields lower bounds on the optimal value f^* of the original problem when μ is nonnegative:

$$q(\lambda, \mu) \leq f^* , \quad \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \mu \geq 0 .$$

Proof

1. For each x, the function $(\lambda, \mu) \mapsto L(x, \lambda, \mu)$ is linear, and therefore both convex and concave in (λ, μ). The pointwise minimum of concave functions is concave, therefore q *is concave*.
2. Let $\bar{x}$ be any feasible point, i.e., $h(\bar{x}) = 0$ and $g(\bar{x}) \leq 0$. Then we have, for any λ and $\mu \geq 0$:

$$\sum_{i=1}^{m} \lambda_i h_i(\bar{x}) + \sum_{i=1}^{r} \mu_i h_i(\bar{x}) \leq 0 \ ,$$

$$\implies \quad L(\bar{x}, \lambda, \mu) = f(\bar{x}) + \sum_{i=1}^{m} \lambda_i h_i(\bar{x}) + \sum_{i=1}^{r} \mu_i h_i(\bar{x}) \leq f(\bar{x}) \ ,$$

$$\implies \quad q(\lambda, \mu) = \inf_x L(x, \lambda, \mu) \leq L(\bar{x}, \lambda, \mu) \leq f(\bar{x}) \ , \quad \forall \bar{x} \ . \quad \square$$

Proof complement

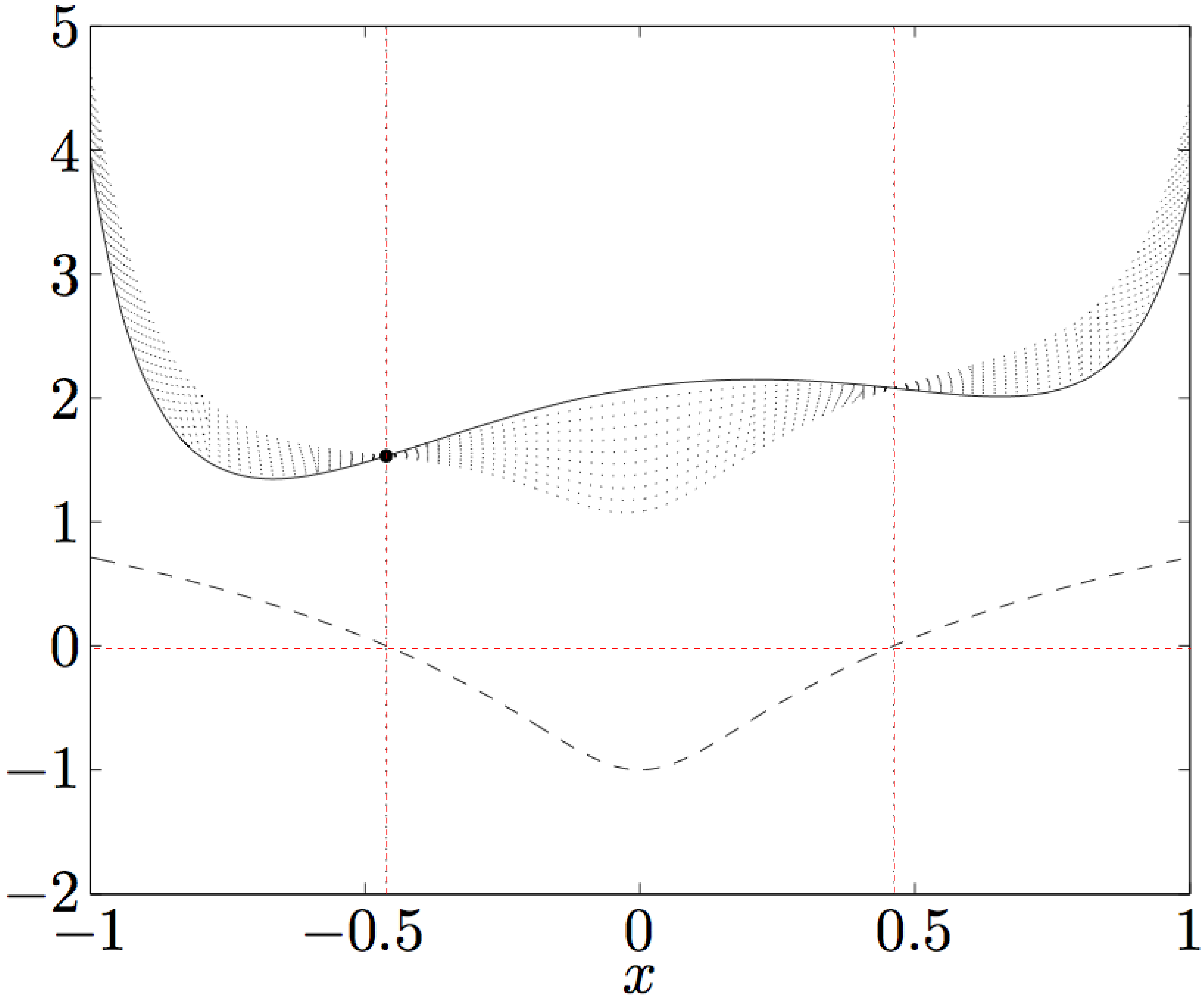
We used the fact that *the poinwise maximum (resp. minimum) of convex (resp. concave) functions is itself convex (concave).*
To prove this, suppose that for each $y \in \mathcal{A}$ the function $f(x,y)$ is convex in x, and let the function:

$$g(x) = \sup_{y \in \mathcal{A}} f(x,y) \, .$$

Then the domain of g is convex as an intersection of convex domains, and for any $\theta \in [0,1]$ and x_1, x_2 in the domain of g:

$$\begin{aligned} g\left(\theta x_1 + (1-\theta)x_2\right) &= \sup_{y \in \mathcal{A}} f(\theta x_1 + (1-\theta)x_2, y) \\ &\leq \sup_{y \in \mathcal{A}} \left(\theta f(x_1,y) + (1-\theta) f(x_2,y)\right) \\ &\leq \sup_{y \in \mathcal{A}} \left(\theta f(x_1,y)\right) + \sup_{y \in \mathcal{A}} \left((1-\theta) f(x_2,y)\right) \\ &= \theta g(x_1) + (1-\theta) g(x_2) \, . \quad \Box \end{aligned}$$

Illustration



Example 1

Least-squares solution of linear equations:

$$\begin{array}{rl} \textit{minimize} & x^\top x \\ \textit{subject to} & Ax = b\,, \end{array}$$

where $A \in \mathbb{R}^{p\times n}$. There are p equality constraints, the Lagrangian with domain $\mathbb{R}^n \times \mathbb{R}^p$ is:

$$L(x,\lambda) = x^\top x + \lambda^\top (Ax - b)\,.$$

To minimize L over x for λ fixed, we set the gradient equal to zero:

$$\nabla_x L(x,\lambda) = 2x + A^\top \lambda = 0 \qquad \Longrightarrow \qquad x = -\frac{1}{2}A^\top \lambda\,.$$

Example 1 (cont.)

Plug it in L to obtain the dual function:

$$q(\lambda) = L\left(-\frac{1}{2}A^\top\lambda, \lambda\right) = -\frac{1}{4}\lambda^\top AA^\top\lambda - b^\top\lambda$$

q is a concave function of λ, and the following lower bound holds:

$$f^* \geq -\frac{1}{4}\lambda^\top AA^\top\lambda - b^\top\lambda\,, \quad \forall\lambda \in \mathbb{R}^p\,.$$

Example 2

Standard form LP:

$$\begin{aligned} \textit{minimize} \quad & c^\top x \\ \textit{subject to} \quad & Ax = b \, , \\ & x \geq 0 \, . \end{aligned}$$

where $A \in \mathbb{R}^{p \times n}$. There are p equality and n inequality constraints, the Lagrangian with domain $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^n$ is:

$$\begin{aligned} L(x, \lambda, \mu) &= c^\top x + \lambda^\top (Ax - b) - \mu^\top x \\ &= -\lambda^\top b + \left(c + A^\top \lambda - \mu \right)^\top x \, . \end{aligned}$$

Example 2 (cont.)

L is linear in x, and its minimum can only be 0 or $-\infty$:

$$g(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) = \begin{cases} -\lambda^\top b & \textit{if } A^\top \lambda - \mu + c = 0 \\ -\infty & \textit{otherwise.} \end{cases}$$

g is linear on an affine subspace and therefore concave.
The lower bound is non-trivial when λ and μ satisfy $\mu \geq 0$ and $A^\top \lambda - \mu + c = 0$, giving the following bound:

$$f^* \geq -\lambda^\top b \quad \textit{if} \quad A^\top \lambda + c \geq 0 \,.$$

Example 3

Inequality form LP:

$$\begin{aligned} \textit{minimize} \quad & c^\top x \\ \textit{subject to} \quad & Ax \leq b \,, \end{aligned}$$

where $A \in \mathbb{R}^{p \times n}$. There are p inequality constraints, the Lagrangian with domain $\mathbb{R}^n \times \mathbb{R}^p$ is:

$$\begin{aligned} L(x,\mu) &= c^\top x + \mu^\top (Ax - b) \\ &= -\mu^\top b + \left(A^\top \mu + c\right)^\top x \,. \end{aligned}$$

Example 3 (cont.)

L is linear in x, and its minimum can only be 0 or $-\infty$:

$$g(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) = \begin{cases} -\mu^\top b & \textit{if } A^\top \mu + c = 0 \\ -\infty & \textit{otherwise.} \end{cases}$$

g is linear on an affine subspace and therefore concave. The lower bound is non-trivial when μ satisfies $\mu \geq 0$ and $A^\top \mu + c = 0$, giving the following bound:

$$f^* \geq -\mu^\top b \quad \textit{if} \quad A^\top \mu + c = 0 \quad \textit{and } \mu \geq 0 \, .$$

Example 4

Two-way partitioning:

$$\begin{aligned} \textit{minimize} \quad & x^\top W x \\ \textit{subject to} \quad & x_i^2 = 1 \,, \quad i = 1, \ldots, n \,. \end{aligned}$$

This is a *nonconvex* problem, the feasible set contains 2^n discrete points ($x_i = \pm 1$).
Interpretation : partition $(1, \ldots, n)$ in two sets, W_{ij} is the cost of assigning i, j to the same set, $-W_{ij}$ the cost of assigning them to different sets. Lagrangian with domain $\mathbb{R}^n \times \mathbb{R}^n$:

$$\begin{aligned} L(x, \lambda) &= x^\top W x + \sum_{i=1}^n \lambda_i \left(x_i^2 - 1 \right) \\ &= x^\top \left(W + diag(\lambda) \right) x - \mathbf{1}^\top \lambda \,. \end{aligned}$$

Example 4 (cont.)

For M symmetric, the minimum of $x^\top M x$ is 0 if all eigenvalues of M are nonnegative, $-\infty$ otherwise. We therefore get the following dual function:

$$q(\lambda) = \begin{cases} -\mathbf{1}^\top \lambda & \textit{if } W + diag(\lambda) \succeq 0 \,, \\ -\infty & \textit{otherwise.} \end{cases}$$

The lower bound is non-trivial for λ such that $W + diag(\lambda) \succeq 0$. This holds in particular for $\lambda = -\lambda_{\min}(W)$, resulting in:

$$f^* \geq -\mathbf{1}^\top \lambda = n\lambda_{\min}(W) \,.$$

Weak and strong duality

Dual problem

For the (primal) problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \ , \quad g(x) \leq 0 \ , \end{aligned}$$

the *Lagrange dual problem* is:

$$\begin{aligned} \textit{maximize} \quad & q(\lambda, \mu) \\ \textit{subject to} \quad & \mu \geq 0 \ , \end{aligned}$$

where q is the (concave) Lagrange dual function and λ and μ are the Lagrange multipliers associated to the constraints $h(x) = 0$ and $g(x) \leq 0$.

Weak duality

Let d^* the optimal value of the Lagrange dual problem. Each $q(\lambda, \mu)$ is an lower bound for f^* and by definition d^* is the best lower bound that is obtained. The following *weak duality inequality* therefore *always hold:*

$$d^* \leq f^* \; .$$

This inequality holds when d^* or f^* are infinite. The difference $d^* - f^*$ is called the *optimal duality gap* of the original problem.

Application of weak duality

For *any* optimization problem, we always have:

- the dual problem is a *convex* optimization problem (="easy to solve")
- *weak duality* holds.

Hence solving the dual problem can provide *useful lower bounds for the original problem, no matter how difficult it is*. For example, solving the following SDP problem (using classical optimization toolbox) provides a non-trivial lower bound for the optimal two-way partitioning problem:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^\top \lambda \\ \text{subject to} & W + diag(\lambda) \succeq 0 \end{array}$$

Strong duality

We say that *strong duality* holds if the optimal duality gap is zero, i.e.:

$$d^* = f^* \; .$$

- If strong duality holds, then the best lower bound that can be obtained from the Lagrange dual function is *tight*
- Strong duality does *not hold* for general nonlinear problems.
- It usually holds for *convex problems*.
- Conditions that ensure strong duality for convex problems are called *constraint qualification*.

Slater's constraint qualification

Strong duality holds for a *convex* problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0 \,, \quad j = 1, \ldots, r \,, \\ & Ax = b \,, \end{aligned}$$

if it is *strictly feasible*, i.e., there exists at least one *feasible point* that satisfies:

$$g_j(x) < 0 \,, \quad j = 1, \ldots, r \,, \quad Ax = b \,.$$

Remarks

- Slater's conditions also ensure that the maximum d^* (if $> -\infty$) is *attained*, i.e., there exists a point (λ^*, μ^*) with

$$q(\lambda^*, \mu^*) = d^* = f^*$$

- They can be sharpened. For example, *strict feasibility is not required for affine constraints*.
- There exist many other types of constraint qualifications

Example 1

Least-squares solution of linear equations:

$$\begin{aligned} \text{minimize} \quad & x^\top x \\ \text{subject to} \quad & Ax = b \,, \end{aligned}$$

where $A \in \mathbb{R}^{p\times n}$. The dual problem is:

$$\textit{maximize} \quad -\frac{1}{4}\lambda^\top A A^\top \lambda - b^\top \lambda \,.$$

- Slater's conditions holds if the primal is feasible, i.e., $b \in Im(A)$. In that case *strong duality holds*.
- In fact strong duality also holds if $f^* = +\infty$: there exists z with $A^\top z = 0$ and $b^\top z \neq 0$, so the dual is unbounded above and $d^* = +\infty = f^*$.

Example 2

Inequality form LP:

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax \leq b\,, \end{aligned}$$

Remember the dual function:

$$g(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu) = \begin{cases} -\mu^\top b & \text{if } A^\top \mu + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Example 2 (cont.)

The dual problem is therefore equivalent to the following *standard form LP*:

$$\begin{aligned} \text{minimize} \quad & b^\top \mu \\ \text{subject to} \quad & A^\top \mu + c = 0, \quad \mu \geq 0 \,. \end{aligned}$$

- From the weaker form of Slater's conditions, *strong duality holds for any LP* provided the primal problem is feasible.
- In fact, $f^* = d^*$ except when both the primal LP and the dual LP are infeasible.

Example 3

Quadratic program (QP):

$$\begin{aligned} \text{minimize} \quad & x^\top P x \\ \text{subject to} \quad & Ax \leq b \,, \end{aligned}$$

where we assume $P \succ 0$. There are p inequality constraints, the Lagrangian is:

$$L(x, \mu) = x^\top P x + \mu^\top (Ax - b) \ .$$

This is a strictly convex function of x minimized for

$$x^*(\mu) = -\frac{1}{2} P^{-1} A^\top \mu.$$

Example 3

The dual function is therefore

$$q(\mu) = -\frac{1}{4}\mu^\top A P^{-1} A^\top \mu - b^\top \mu.$$

and the dual problem:

$$\begin{aligned} &\textit{maximize} && -\frac{1}{4}\mu^\top A P^{-1} A^\top \mu - b^\top \mu \\ &\textit{subject to} && \mu \geq 0\ . \end{aligned}$$

- By the weak form of Slater's conditions, *strong duality holds* if the primal problem is feasible: $f^* = d^*$.
- In fact, *strong duality always holds*, even if the primal is not feasible (in which case $f^* = d^* = +\infty$), cf LP case.

Example 4

The following QCQP problem is *not convex* if A symmetric but not positive semidefinite:

$$\begin{aligned} \textit{minimize} \quad & x^\top A x + 2b^\top x \\ \textit{subject to} \quad & x^\top x \leq 1 \,. \end{aligned}$$

Its dual problem is the following SDP (left as exercice):

$$\begin{aligned} \textit{maximize} \quad & -t-\mu \\ \textit{subject to} \quad & \begin{pmatrix} A+\mu I & b \\ b^\top & t \end{pmatrix} \succ 0 \,. \end{aligned}$$

In fact, *strong duality holds* in this case (more generally for quadratic objective and one quadratic inequality constraint, provided Slater's condition holds, see Annex B.1 in B&V).

Geometric interpretation

Setting

We consider the simple problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leq 0 \ , \end{aligned}$$

where $f, g : \mathbb{R}^n \to \mathbb{R}$. We will give a geometric interpretation of the weak and strong duality.

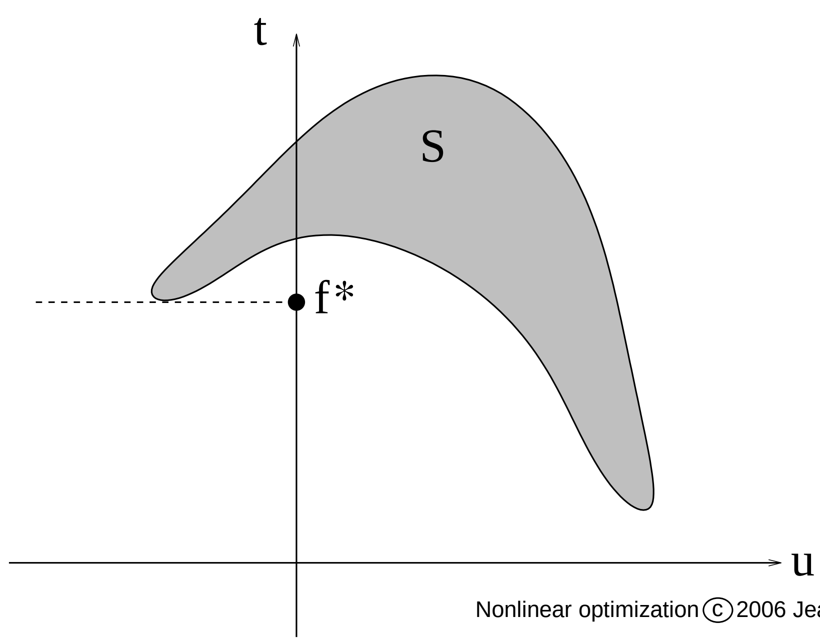
Optimal value f^*

We consider the subset of $\mathbb{R}^2$ defined by:

$$S = \{(g(x), f(x) \mid x \in \mathbb{R}^n)\} \ .$$

The optimal value f^* is determined by:

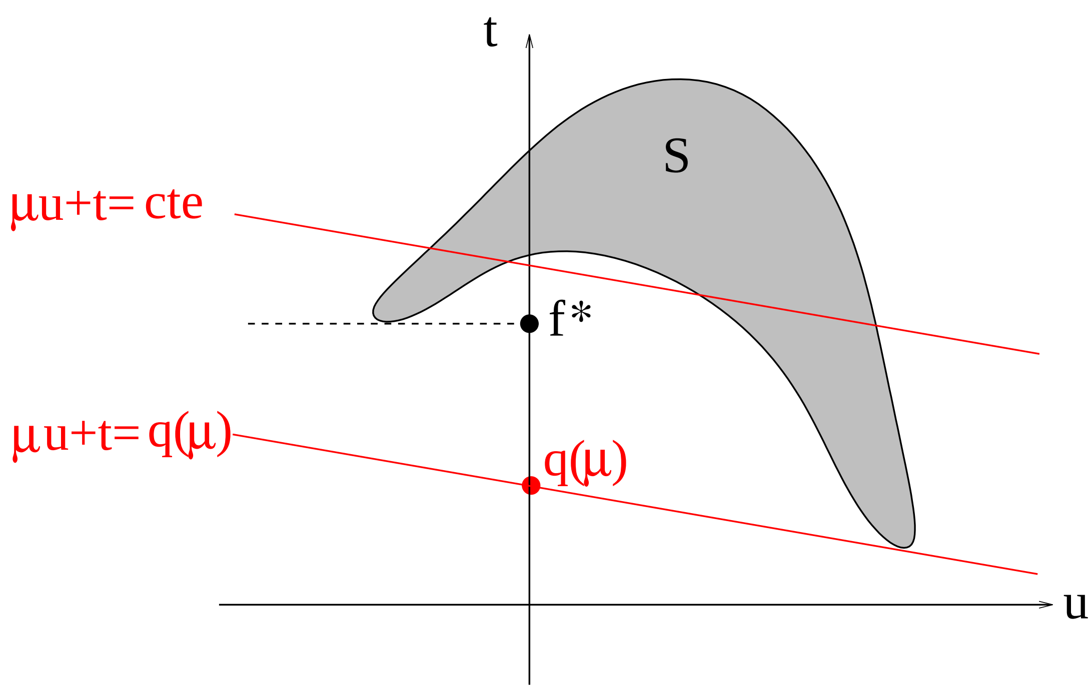
$$f^* = \inf \{t \mid (u,t) \in S, \ u \leq 0\} \ .$$



Dual function $q(\mu)$

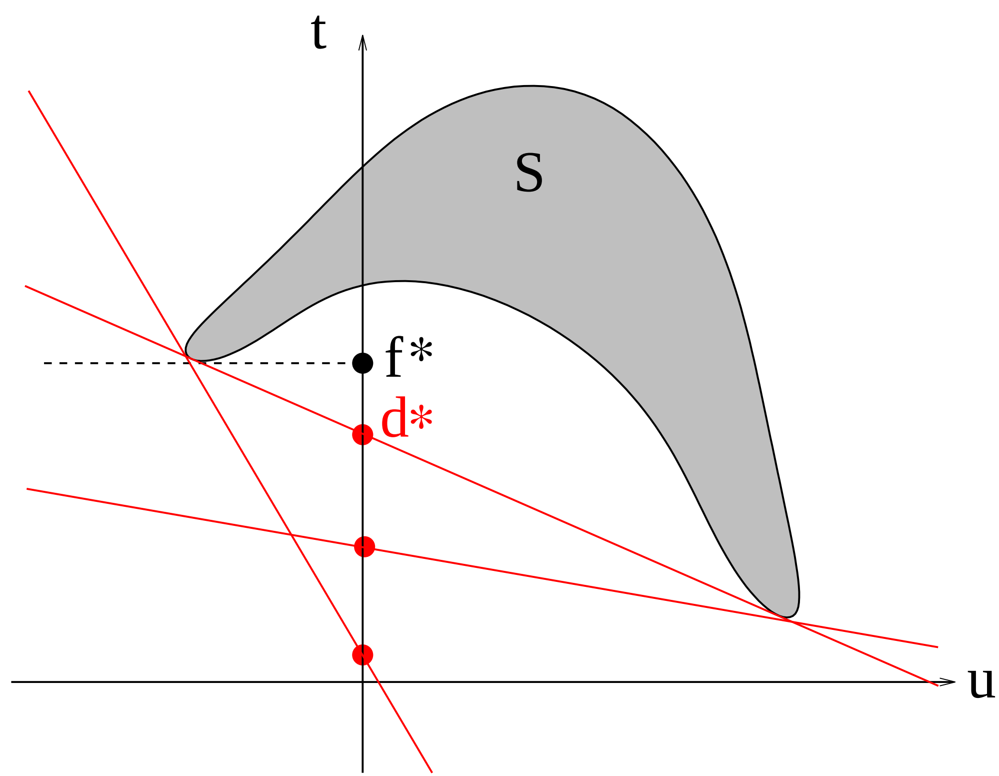
The dual function for $\mu \geq 0$ is:

$$q(\mu) = \inf_{x \in \mathbb{R}^n} \{f(x) + \mu g(x)\}$$
$$= \inf_{(u,t) \in S} \{\mu u + t\} \ .$$



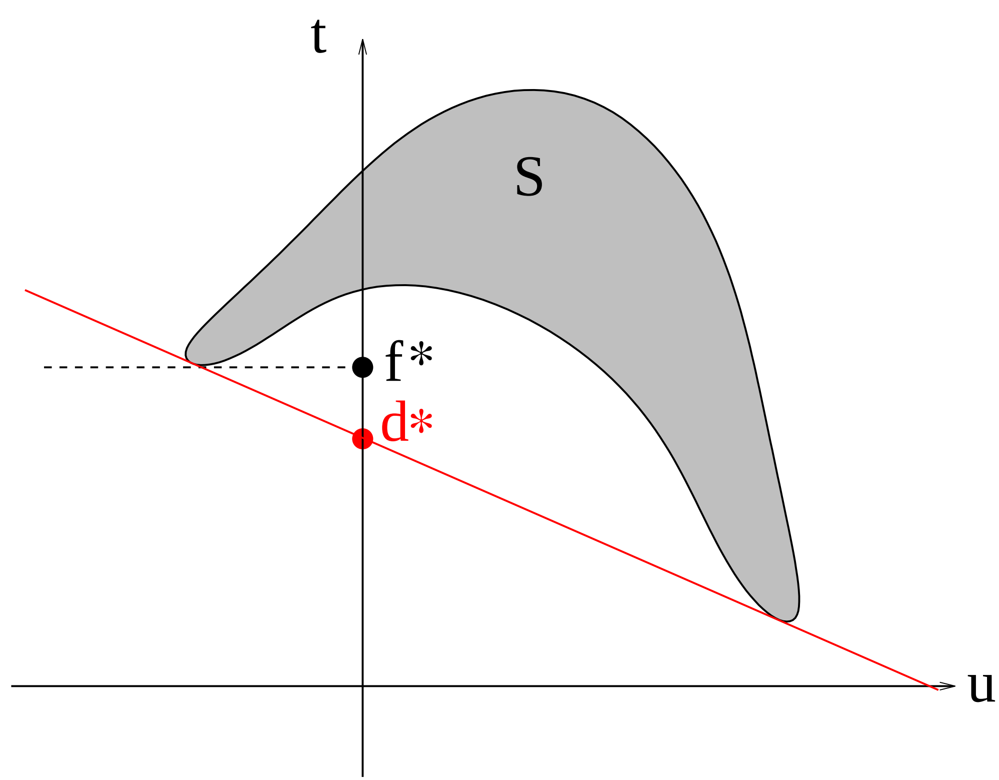
Dual optimal d^*

$$
\begin{aligned}
d^* &= \sup_{\mu \geq 0} q(\mu) \\
&= \sup_{\mu \geq 0} \inf_{(u,t) \in S} \{\mu u + t\} \; .
\end{aligned}
$$



Weak duality

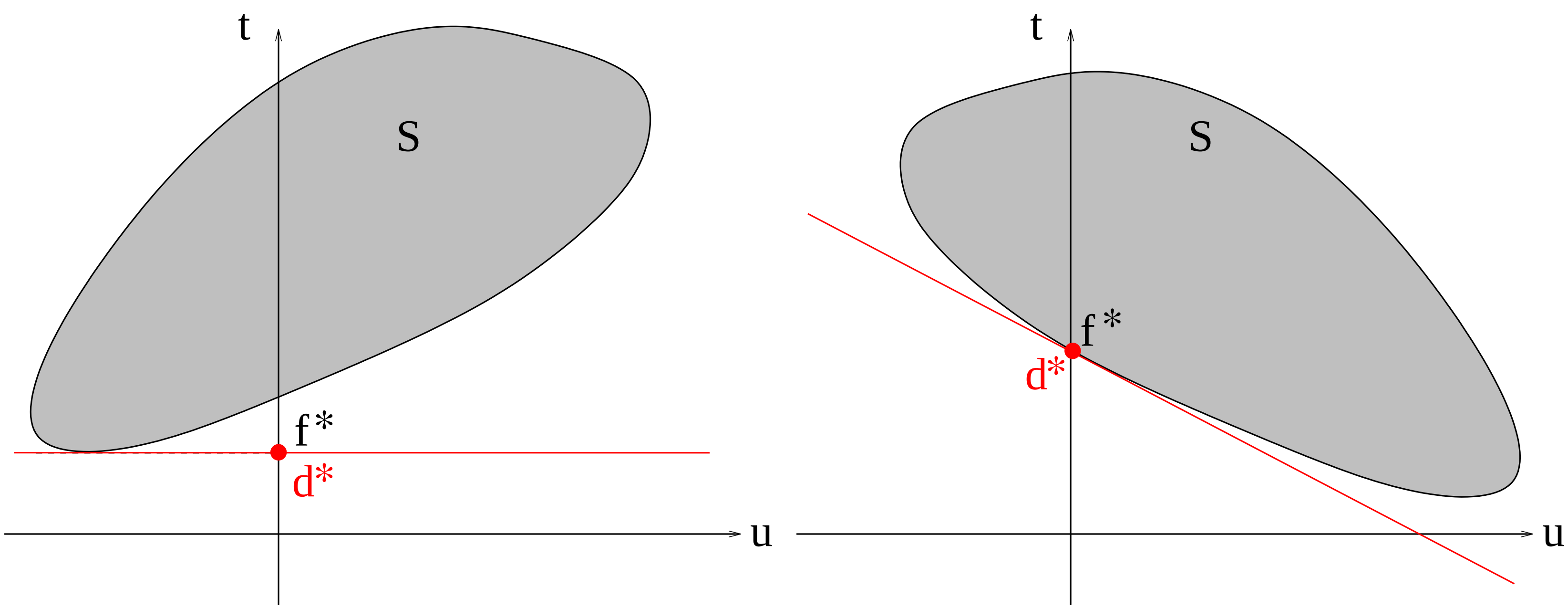
$$d^* \leq f^*$$



Strong duality

For convex problems with strictly feasible points:

$$d^* = f^*$$



Saddle-point interpretations

Setting

We consider a general optimization problem with inequality constraints:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0 \;, \quad j = 1, \ldots, r \;. \end{aligned}$$

Its Lagrangian is

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x) \;.$$

Inf-sup form of f^*

We note that, for any $x \in \mathbb{R}^n$:

$$\sup_{\mu \geq 0} L(x, \mu) = \sup_{\mu \geq 0} \left\{ f(x) + \sum_{j=1}^{r} \mu_j g_j(x) \right\}$$

$$= \begin{cases} f(x) & \text{if } g_j(x) \leq 0 \,, \quad j = 1, \ldots, r \,, \\ +\infty & \text{otherwise.} \end{cases}$$

Therefore:

$$f^* = \inf_{x \in \mathbb{R}^n} \sup_{\mu \geq 0} L(x, \mu) \,.$$

Duality

By definition we also have

$$d^* = \sup_{\mu \geq 0} \inf_{x \in \mathbb{R}^n} L(x, \mu) \ .$$

The *weak duality* can thus be rewritten:

$$\sup_{\mu \geq 0} \inf_{x \in \mathbb{R}^n} L(x, \mu) \leq \inf_{x \in \mathbb{R}^n} \sup_{\mu \geq 0} L(x, \mu) \ .$$

and the *strong duality* as the equality:

$$\sup_{\mu \geq 0} \inf_{x \in \mathbb{R}^n} L(x, \mu) = \inf_{x \in \mathbb{R}^n} \sup_{\mu \geq 0} L(x, \mu) \ .$$

Max-min inequality

In fact the weak duality does not depend on any property of L, it is just an instance of the general *max-min inequality* that states that

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) \leq \inf_{w \in W} \sup_{z \in Z} f(w,z) ,$$

for any $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $W \subset \mathbb{R}^n$ and $Z \subset \mathbb{R}^m$. When equality holds, i.e.,

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) = \inf_{w \in W} \sup_{z \in Z} f(w,z) ,$$

we say that f satisfies the *strong max-min property*. This holds only in special cases.

Proof of Max-min inequality

For any $(w_0, z_0) \in W \times Z$ we have by definition of the infimum in w:

$$\inf_{w \in W} f(w, z_0) \leq f(w_0, z_0) \, .$$

For w_0 fixed, this holds for any choice of z_0 so we can take the supremum in z_0 on both sides to obtain:

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \sup_{z \in Z} f(w_0, z) \, .$$

The left-hand side is a constant, and the right-hand side is a function of w_0. The inequality is valid for any w_0, so we can take the infimum to obtain the result:

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z) \, .$$

Saddle-point interpretation

A pair $(w^*, z^*) \in W \times Z$ is called a *saddle-point* for f if

$$f(w^*, z) \leq f(w^*, z^*) \leq f(w, z^*) \,, \quad \forall w \in W \,, z \in Z \,.$$

If a saddle-point exists then *strong max-min property holds* because:

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \geq \inf_{w \in W} f(w, z^*) = f(w^*, z^*)$$
$$= \sup_{z \in Z} f(w^*, z) \geq \inf_{w \in W} \sup_{z \in Z} f(w^*, z) \,.$$

Hence if strong duality holds, (x^*, μ^*) *form a saddle-point of the Lagrangian*. Conversily, if the Lagrangian has a saddle-point then strong duality holds.

Game interpretation

Consider a game with two players:

1. Player 1 chooses $w \in W$;
2. then Player 2 chooses $z \in Z$;
3. then Player 1 pays $f(w, z)$ to Player 2.

Player 1 wants to minimize f, while Player 2 wants to maximize it. If Player 1 chooses w, then Player 2 will choose $z \in Z$ to obtain the maximum payoff $\sup_{z \in Z} f(w, z)$. Knowing this, Player must chose w to make this payoff minimum, equal to:

$$\inf_{w \in W} \sup_{z \in Z} f(w, z) \ .$$

Game interpretation (cont.)

If Player 2 plays first, following a similar argument, the payoff will be:

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \; .$$

The general max-min inequality states that *it is better for a player to know his or her opponent's choice before choosing*. The optimal duality gap is the advantage afforded to the player who plays second. *If there is a saddle-point, then there is no advantage to the players of knowing their opponent's choice*.

Optimality conditions

Setting

We consider an equality and inequality constrained optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0\,, \quad i = 1, \ldots, m\,, \\ & g_j(x) \leq 0\,, \quad j = 1, \ldots, r\,, \end{aligned}$$

making *no assumption* of f, g and h.

We will revisit the optimality conditions at the light of duality.

Dual optimal pairs

Suppose that strong duality holds, x^* is primal optimal, (λ^*, μ^*) is dual optimal. Then we have:

$$
\begin{aligned}
f(x^*) &= q(\lambda^*, \mu^*) \\
&= \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i^* h_i(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \right\} \\
&\leq f(x^*) + \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{r} \mu_j g_j(x^*) \\
&\leq f(x^*)
\end{aligned}
$$

Hence both inequalities are in fact *equalities*.

Complimentary slackness

The first equality shows that:

$$L\left(x^*, \lambda^*, \mu^*\right) = \inf_{x \in \mathbb{R}^n} L\left(x, \lambda^*, \mu^*\right) \ ,$$

showing that x^* *minimizes the Lagrangian at* (λ^*, μ^*)*.* The second equality shows that:

$$\mu_j g_j(x^*) = 0 \ , \quad j = 1, \ldots, r \ .$$

This property is called *complementary slackness*:

the ith optimal Lagrange multiplier is zero unless the ith constraint is active at the optimum.

KKT conditions

If the functions f, g, h are differentiable and there is no duality gap, then we have seen that x^* minimizes $L(x, \lambda^*, \mu^*)$, therefore:

$$\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \nabla \lambda_i^* h_i(x^*) + \sum_{j=1}^{r} \nabla \mu_j^* g_i(x^*) = 0 \ .$$

Combined with the complimentary slackness and feasibility conditions, we recover the KKT optimality conditions that x^* must fulfill. λ^* and μ^* now have the interpretation of dual optimal.

KKT conditions for convex problems

Suppose now that the problem is convex, i.e., f and g are convex functions, h is affine, and let x^* and (λ^*, μ^*) satisfy the KKT conditions:

$$
\begin{aligned}
h_i(x^*) &= 0 \qquad i = 1, \ldots, m \\
g_j(x^*) &\leq 0 \qquad j = 1, \ldots, r \\
\mu_j^* &\geq 0 \qquad j = 1, \ldots, r \\
\mu_j^* g_j(x^*) &= 0 \qquad j = 1, \ldots, r
\end{aligned}
$$

$$
\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_i(x^*) = 0 \,,
$$

then x^* *and* (λ^*, μ^*) *are primal and dual optimal, with zero duality gap* (the KKT conditions are *sufficient* in this case).

Proof

The first 2 conditions show that x^* is feasible. Because $\mu^* \geq 0$, the Lagrangian $L(x, \lambda^*, \mu^*)$ is convex in x. Therefore, the last equality shows that x^* minimized it, therefore:

$$\begin{aligned}
q\left(\lambda^*, \mu^*\right) &= L\left(x^*, \lambda^*, \mu^*\right) \\
&= f(x^*) + \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{r} \mu_j g_j(x^*) \\
&= f(x^*)\ ,
\end{aligned}$$

showing that x^* and (λ^*, μ^*) have zero duality gap, and are therefore primal and dual optimal.

Summary

For *any* problem with *differentiable* objective and constraints:

- If x and (λ, μ) satisfy $f(x) = q(\lambda, \mu)$ (which implies in particular that x is optimal), then (x, λ, μ) satisfy KKT.
- For a *convex* problem the converse is true: x and (λ, μ) satisfy $f(x) = q(\lambda, \mu)$ if and only if they satisfy KKT.
- For a convex problem where *Slater's condition* holds, we know that strong duality holds and that the dual optimal is attained, so x is optimal if and only if there are (λ, μ) that together with x satisfy the KKT conditions.
- We showed previously without convexity assumption, if x is optimal and regular, then there exists (λ, μ) that together with x satisfy KKT. In that case, however, we do note have in general $f(x) = q(\lambda, \mu)$ (otherwise strong duality would hold).

Example 1

Equality constrained quadratic minimization:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^\top P x + q^\top x + r \\ \text{subject to} \quad & Ax = b \,, \end{aligned}$$

where $P \succeq 0$. This problem is convex with no inequality constraint, so the KKT conditions are necessary and sufficient:

$$\begin{pmatrix} P & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

This is a set of $m + n$ equations with $m + n$ variables.

Example 2

Water-filling problem (assuming $\alpha_i \geq 0$):

$$\textit{minimize} \quad -\sum_{i=1}^{n} \log\left(\alpha_i + x_i\right)$$

$$\textit{subject to} \quad x \geq 0\,, \quad \mathbf{1}^\top x = 1\,.$$

By the KKT conditions for this convex problem that satisfies Slater's conditions, x is optimal iff $x \geq 0$, $\mathbf{1}^\top x = 1$, and there exists $\lambda \in \mathbb{R}$ and $\mu \in \mathbb{R}^n$ s.t.

$$\mu \geq 0\,, \quad \mu_i x_i = 0\,, \quad \frac{1}{x_i + \alpha_i} + \mu_i = \lambda\,.$$

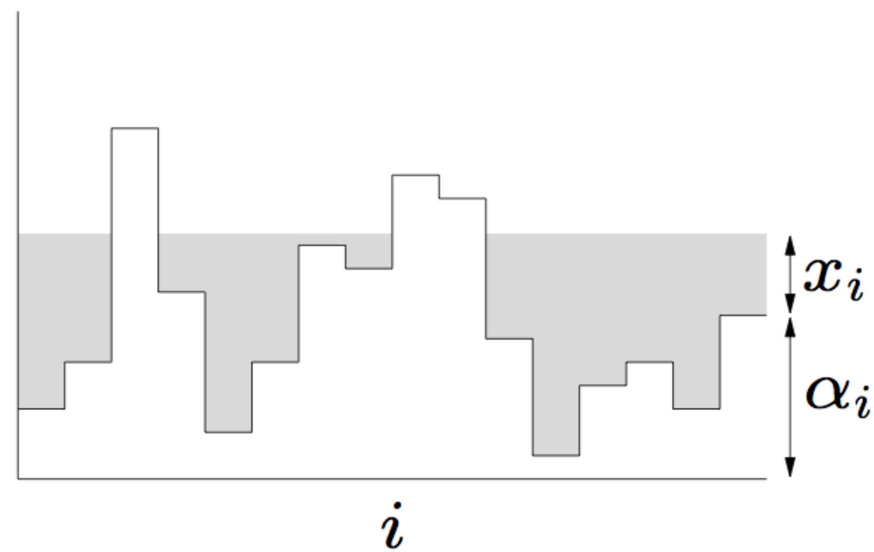
Example 2

This problem is easily solved directly:

- If $\lambda < 1/\alpha_i$: $\mu_i = 0$ and $x_i = 1/\lambda - \alpha_i$
- If $\lambda \geq 1/\alpha_i$: $\mu_i = \lambda - 1/\alpha_i$ and $x_i = 0$
- determine λ from $\mathbf{1}^\top x = \sum_{i=1}^n \max\{0, 1/\lambda - \alpha_i\} = 1$

Interpretation:

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\lambda^*$



Perturbation and sensitivity analysis

Unperturbed optimization problem

We consider the general problem:

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & h_i(x) = 0\,, \quad i = 1, \ldots, m\,, \\
& g_j(x) \leq 0\,, \quad j = 1, \ldots, r\,,
\end{aligned}
$$

with optimal value f^*, and its dual:

$$
\begin{aligned}
\text{maximize} \quad & q(\lambda, \mu) \\
\text{subject to} \quad & \mu \geq 0\,.
\end{aligned}
$$

Perturbed problem and its dual

The *perturbed problem* is

$$\begin{aligned} \textit{minimize} \quad & f(x) \\ \textit{subject to} \quad & h_i(x) = u_i \;, \quad i = 1, \ldots, m \;, \\ & g_j(x) \leq v_j \;, \quad j = 1, \ldots, r \;, \end{aligned}$$

with optimal value $f^*(u, v)$, and its dual:

$$\begin{aligned} \textit{maximize} \quad & q(\lambda, \mu) - u^\top \lambda - v^\top \mu \;. \\ \textit{subject to} \quad & \mu \geq 0 \;. \end{aligned}$$

Interpretation

- When $u = v = 0$, this coincides with the original problem: $f^*(0,0) = f^*$.
- When $v_j > 0$, we have *relaxed* the jth inequality constraint.
- When $v_j < 0$, we have *tightened* the jth inequality constraing.
- We are interested in *informations about* $f^*(u,v)$ that can be obtained from the solution of the unperturbed problem and its dual.

A global sensitivity result

Now we assume that:

- Strong duality holds, i.e., $f^* = d^*$.
- The dual optimum is attained, i.e., *there exist* (λ^*, μ^*) *such that* $d^* = q(\lambda^*, \mu^*)$*.*

Applying weak duality to the perturbed problem we obtain

$$\begin{aligned} f^*(u, v) &\geq q(\lambda^*, \mu^*) - u^\top \lambda^* - v^\top \mu^* \\ &= f^* - u^\top \lambda^* - v^\top \mu^* \,. \end{aligned}$$

Global sensitivity interpretation

$$f^*(u,v) \geq f^* - u^\top \lambda^* - v^\top \mu^* \, .$$

- If μ_j^* is large: f^* increases greatly if we tighten the jth inequality constraint ($v_j < 0$)
- If μ_j^* is small: f^* does not decrease much if we loosen the jth inequality constraint ($v_j > 0$)
- If λ_i^* is large and positive: f^* increases greatly if we decrease the ith equality constraint ($u_i < 0$)
- If λ_i^* is large and negative: f^* increases greatly if we increase the ith equality constraint ($u_i > 0$)
- If λ_i^* is small and positive: f^* does not decrease much if we increase the ith equality constraint ($u_i > 0$)
- If λ_i^* is small and negative: f^* does not decrease much if we decrease the ith equality constraint ($u_i < 0$)

Local sensitivity analysis

If (in addition) we assume that $f^*(u,v)$ *is differentiable at* $(0,0)$, then the following holds:

$$\lambda_i^* = -\frac{\partial f^*(0,0)}{\partial u_i}\,, \quad \mu_i^* = -\frac{\partial f^*(0,0)}{\partial v_i}\,,$$

In that case, the Lagrange multipliers are exactly the *local sensitivities of the optimal value with respect to constraint perturbation*. Tightening the ith inequality constraint a small amount $v_j < 0$ yields an increase in f^* of approximately $-\lambda_j^* v_j$.

Proof

For λ_i^*: from the global sensitivity result, it holds that:

$$t > 0 \implies \frac{f^*(te_i, 0) - f^*(0,0)}{t} \geq -\lambda_i^* \,,$$

and therefore

$$\frac{\partial f^*(0,0)}{\partial u_i} \geq -\lambda_i^* \,.$$

A similar analysis with $t < 0$ yields $\partial f^*(0,0)/\partial u_i \leq -\lambda_i^*$, and therefore:

$$\frac{\partial f^*(0,0)}{\partial u_i} = -\lambda_i^* \,.$$

A similar proof holds for μ_j. $\square$

Shadow price interpretation

We assume the following problem is convex and Slater's condition holds:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq v_j \;, \quad j = 1, \ldots, r \;, \end{aligned}$$

- $x \in \mathbb{R}^n$ determines how a firm operates.
- The objective f is the cost, i.e., $-f$ is the profit.
- Each constraint $g_j(x) \leq 0$ is a limit on some resource (labor, steel, warehouse space...)

Shadow price interpretation (cont.)

- $-f^*(v)$ is how much more or less profit could be made if more or less of each resource were made available to the firm.
- $\mu_j^* = -\partial f^*(0,0)/\partial v_j$ is how much more profit the firm could make for a small increase in availability of resource j.
- μ_j^* is therefore the natural or equilibrium *price* for resource j.