4TE3/6TE3

Algorithms for

Continuous Optimization

(Duality in Nonlinear Optimization)

Tamás TERLAKY
Computing and Software
McMaster University

Hamilton, January 2004

terlaky@mcmaster.ca Tel: 27780

Optimality conditions for

constrained convex optimization problems

$$
\begin{array}{lll}
(CO) & \min & f(x) \\
 & \text{s.t.} & g_j(x) \leq 0, \quad j = 1, \cdots, m \\
 & & x \in \mathcal{C}.
\end{array}
$$

$$\mathcal{F} = \{x \in \mathcal{C} \mid g_j(x) \leq 0, \quad j \in J\}.$$

Let $\mathcal{C}^0$ denote the *relative interior* of the convex set $\mathcal{C}$.

Definition 1 *A vector (point) $x^0 \in \mathcal{C}^0$ is called a* Slater point *of (CO) if*

$$
\begin{array}{lll}
g_j(x^0) < 0, & \text{for all } j \text{ where} & g_j \text{ is nonlinear,} \\
g_j(x^0) \leq 0, & \text{for all } j \text{ where} & g_j \text{ is linear.}
\end{array}
$$

(CO) is Slater regular *or (CO) satisfies the Slater condition (in other words, (CO) satisfies the Slater constraint qualification).*

Ideal Slater Point

Some constraint functions $g_j(x)$ might take the value zero for all feasible points. Such constraints are called *singular* while the others are called *regular*.

$$J_s = \{j \in J \mid g_j(x) = 0 \text{ for all } x \in \mathcal{F}\},$$

$$J_r = J - J_s = \{j \in J \mid g_j(x) < 0 \text{ for some } x \in \mathcal{F}\}.$$

Remark: Note, that if (CO) is Slater regular, then all singular functions must be linear.

Definition 2 *A vector (point) $x^* \in \mathcal{C}^0$ is called an* Ideal Slater point *of the convex optimization problem (CO) if*

$$\begin{array}{ll} g_j(x^*) < 0 & \textit{for all } j \in J_r, \\ g_j(x^*) = 0 & \textit{for all } j \in J_s. \end{array}$$

Lemma 1 *If the problem (CO) is Slater regular then there exists an ideal Slater point $x^* \in \mathcal{F}$.*

Convex Farkas Lemma

Theorem 1 *Let* $\mathcal{U} \subseteq \mathbb{R}^n$ *be a convex set and a point* $w \in \mathbb{R}^n$ *with* $w \notin \mathcal{U}$ *be given. Then there is a separating hyperplane* $a^T w \leq \alpha$, *with* $a \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$ *such that* $a^T u \geq \alpha$ *for all* $u \in \mathcal{U}$ *but* $\mathcal{U}$ *is not a subset of the hyperplane* $\{x \mid a^T x = \alpha\}$.

Note that the last property says that there is a $\overline{u} \in \mathcal{U}$ such that $a^T \overline{u} > \alpha$.

Lemma 2 (Farkas) *The convex optimization problem (CO) is given and we assume that the Slater regularity condition is satisfied. The inequality system*

$$\begin{aligned} &f(x) < 0 \\ &g_j(x) \leq 0, \qquad j = 1, \cdots, m \\ &x \in \mathcal{C}. \end{aligned} \tag{1}$$

has no solution if and only if there exists a vector $y = (y_1, \cdots, y_m) \geq 0$ *such that*

$$f(x) + \sum_{j=1}^{m} y_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}. \tag{2}$$

The systems (1) and (2) are called *alternative systems*, i.e. exactly one of them has a solution.

Proof of the

Convex Farkas Lemma

If (1) is solvable, then (2) cannot hold.

To prove the other side: let us assume that (1) has no solution. With $u = (u_0, \cdots, u_m)$, we define the set $\mathcal{U} \in \mathbb{R}^{m+1}$

$$\mathcal{U} = \{u \mid \exists x \in \mathcal{C} \quad \text{with} \quad u_0 > f(x),\ u_j \geq g_j(x) \ \text{if} \ j \in J_r, \\ u_j = g_j(x) \ \text{if} \ j \in J_s\}.$$

$\mathcal{U}$ is convex (note that due to the Slater condition singular functions are linear)
due to the infeasibility of (1) it does not contain the origin.
Due to the separation Theorem 1 there exists a separating hyperplane defined by $(y_0, y_1, \cdots, y_m)$ and $\alpha = 0$ such that

$$\sum_{j=0}^{m} y_j u_j \geq 0 \quad \text{for all } u \in \mathcal{U} \tag{3}$$

and for some $\overline{u} \in \mathcal{U}$ one has

$$\sum_{j=0}^{m} y_j \overline{u}_j > 0. \tag{4}$$

The rest of the proof is divided into four parts.

I. First we prove that $y_0 \geq 0$ and $y_j \geq 0$ for all $j \in J_r$.

II. Secondly we establish that (3) holds for
$u = (f(x), g_1(x), \cdots, g_m(x))$ if $x \in \mathcal{C}$.

III. Then we prove that y_0 must be positive.

IV. Finally, it is shown by using induction that
$y_j > 0$ for all $j \in J_s$ can be reached.

Proof: Steps I and II

I. First we show that $y_0 \geq 0$ and $y_j \geq 0$ for all $j \in J_r$. Let us assume that $y_0 < 0$. Let us take an arbitrary $(u_0, u_1, \cdots, u_m) \in \mathcal{U}$. By definition $(u_0 + \lambda, u_1, \cdots, u_m) \in \mathcal{U}$ for all $\lambda \geq 0$.
Hence by (3) one has

$$\lambda y_0 + \sum_{j=0}^{m} y_j u_j \geq 0 \quad \text{for all } \lambda \geq 0.$$

For sufficiently large λ the left hand side is negative, which is a contradiction, i.e. y_0 must be nonnegative. The proof of the nonnegativity of all y_j as $j \in J_r$ goes analogously.

II. Secondly we establish that

$$y_0 f(x) + \sum_{j=1}^{m} y_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}. \qquad (5)$$

This follows from the observation that
for all $x \in \mathcal{C}$ and for all $\lambda > 0$ one has
$u = (f(x) + \lambda, g_1(x), \cdots, g_m(x)) \in \mathcal{U}$, thus

$$y_0(f(x) + \lambda) + \sum_{j=1}^{m} y_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}.$$

Taking the limit as $\lambda \longrightarrow 0$ the claim follows.

Proof: Step III

III. Thirdly we show that $y_0 > 0$. The proof is by contradiction. We already know that $y_0 \geq 0$. Let us assume to the contrary that $y_0 = 0$. Hence from (5) we have

$$\sum_{j \in J_r} y_j g_j(x) + \sum_{j \in J_s} y_j g_j(x) = \sum_{j=1}^{m} y_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}.$$

Taking an ideal Slater point $x^* \in \mathcal{C}^0$ one has

$$g_j(x^*) = 0 \ \text{ if } \ j \in J_s,$$

whence

$$\sum_{j \in J_r} y_j g_j(x^*) \geq 0.$$

Since $y_j \geq 0$ and $g_j(x^*) < 0$ for all $j \in J_r$, this implies $y_j = 0$ for all $j \in J_r$. This results in

$$\sum_{j \in J_s} y_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}. \tag{6}$$

Now, from (4), with $\bar{x} \in \mathcal{C}$ such that $\overline{u}_j = g_j(\bar{x})$ if $i \in J_s$ we have

$$\sum_{j \in J_s} y_j g_j(\bar{x}) > 0. \tag{7}$$

Because the ideal Slater point x^* is in the relative interior of $\mathcal{C}$ there exist a vector $\tilde{x} \in \mathcal{C}$ and $0 < \lambda < 1$ such that $x^* = \lambda \bar{x} + (1 - \lambda)\tilde{x}$. Using that $g_j(x^*) = 0$ for $j \in J_s$ and that the singular functions are linear one gets

Proof: Step III cntd.

$$\begin{aligned} 0 = & \textstyle\sum_{j\in J_s} y_j g_j(x^*) \\ = & \textstyle\sum_{j\in J_s} y_j g_j(\lambda\bar{x} + (1-\lambda)\tilde{x}) \\ = & \lambda \textstyle\sum_{j\in J_s} y_j g_j(\bar{x}) + (1-\lambda)\sum_{j\in J_s} y_j g_j(\tilde{x}) \\ > & (1-\lambda)\textstyle\sum_{j\in J_s} y_j g_j(\tilde{x}). \end{aligned}$$

Here the last inequality follows from (7). The inequality

$$(1-\lambda)\sum_{j\in J_s} y_j g_j(\tilde{x}) < 0$$

contradicts (6). Hence we have proved that $y_0 > 0$.

At this point we have (5) with $y_0 > 0$ and $y_j \geq 0$ for all $j \in J_r$. Dividing by $y_0 > 0$ in (5) and by defining $y_j := \frac{y_j}{y_0}$ for all $j \in J$ we obtain

$$f(x) + \sum_{j=1}^{m} y_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}. \qquad (8)$$

We finally show that y may be taken such that $y_j > 0$ for all $j \in J_s$.

Proof: Step IV

IV. To complete the proof we show by induction on the cardinality of J_s that one can make y_j non-negative for all $j \in J_s$. Observe that if $J_s = \emptyset$ then we are done. If $|J_s| = 1$ then we apply the results proved till this point to the inequality system

$$\begin{aligned} & g_s(x) < 0, \\ & g_j(x) \leq 0, \qquad j \in J_r, \\ & x \in \mathcal{C} \end{aligned} \tag{9}$$

where $\{s\} = J_s$. The system (9) has no solution, it satisfies the Slater condition, and therefore there exists a $\hat{y} \in \mathbb{R}^{m-1}$ such that

$$g_s(x) + \sum_{j \in J_r} \hat{y}_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}, \tag{10}$$

where $\hat{y}_s > 0$ and $\hat{y}_j \geq 0$ for all $j \in J_r$. Adding a sufficiently large positive multiple of (10) to (8) one obtains a positive coefficient for $g_s(x)$.

The general inductive step goes analogously.

Proof: Step IV cntd.

Assuming that the result is proved if $|J_s| = k$ then the result is proved for the case $|J_s| = k+1$. Let $s \in J_s$ then $|J_s \setminus \{s\}| = k$, and hence the inductive assumption applies to the system

$$\begin{array}{ll} g_s(x) < 0 & \\ g_j(x) \leq 0, & j \in J_s \setminus \{s\}, \\ g_j(x) \leq 0, & j \in J_r, \\ x \in \mathcal{C} & \end{array} \tag{11}$$

By construction the system (11) has no solution, it satisfies the Slater condition, and by the inductive assumption we have a $\hat{y} \in \mathbb{R}^{m-1}$ such that

$$g_s(x) + \sum_{j \in J_r \cup J_s \setminus \{s\}} \hat{y}_j g_j(x) \geq 0 \quad \text{for all } x \in \mathcal{C}. \tag{12}$$

where $\hat{y}_j > 0$ for all $j \in J_s$ and $\hat{y}_j \geq 0$ for all $j \in J_r$. Adding a sufficiently large multiple of (12) to (8) one have the desired nonnegative multipliers. □

Remark: Note, that finally we proved slightly more than was stated. We have proved that the multipliers of all the singular constraints can be made strictly positive.

!!!! Srict complementarity in LO !!!!

Goldman-Tucker Theorem

Karush–Kuhn–Tucker theory

Langrangean, Saddle point

The Lagrange function:

$$L(x,y) := f(x) + \sum_{j=1}^{m} y_j g_j(x) \tag{13}$$

where $x \in \mathcal{C}$ and $y \geq 0$. Note that for fixed y the Lagrangean is convex in x; for fixed x. it is linear in y.

Definition 3 *A vector pair* $(\bar{x}, \bar{y}) \in \mathbb{R}^{n+m}$, $\bar{x} \in \mathcal{C}$ *and* $\bar{y} \geq 0$ *is called a saddle point of the Lagrange function* L *if*

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \tag{14}$$

for all $x \in \mathcal{C}$ *and* $y \geq 0$.

A saddle point lemma

Lemma 3 *The vector* $(\bar{x}, \bar{y}) \in \mathbb{R}^{n+m}$, $\bar{x} \in \mathcal{C}$ *and* $\bar{y} \geq 0$ *is a saddle point of* $L(x, y)$ *if and only if*

$$\inf_{x\in\mathcal{C}} \sup_{y\geq 0} L(x,y) = L(\bar{x},\bar{y}) = \sup_{y\geq 0} \inf_{x\in\mathcal{C}} L(x,y). \qquad (15)$$

Proof: The saddle point inequality (14) easily follows from (15).

On the other hand, for any $(\hat{x}, \hat{y})$ one has

$$\inf_{x\in\mathcal{C}} L(x,\hat{y}) \leq L(\hat{x},\hat{y}) \leq \sup_{y\geq 0} L(\hat{x},y),$$

hence one can take the supremum of the left hand side and the infimum of the right hand side resulting in

$$\sup_{y\geq 0} \inf_{x\in\mathcal{C}} L(x,y) \leq \inf_{x\in\mathcal{C}} \sup_{y\geq 0} L(x,y). \qquad (16)$$

Using the saddle point inequality (14) one obtains

$$\inf_{x\in\mathcal{C}} \sup_{y\geq 0} L(x,y) \leq \sup_{y\geq 0} L(\bar{x},y) \leq L(\bar{x},\bar{y}) \leq \inf_{x\in\mathcal{C}} L(x,\bar{y}) \leq \sup_{y\geq 0} \inf_{x\in\mathcal{C}} L(x,y). \qquad (17)$$

Combining (17) and (16) the equality (15) follows.

□

Karush-Kuhn-Tucker Theorem

Theorem 2 *The problem (CO) is given. Assume that the Slater regularity condition is satisfied. The vector $\bar{x}$ is an optimal solution of (CO) if and only if there is a vector $\bar{y}$ such that $(\bar{x}, \bar{y})$ is a saddle point of the Lagrange function L.*

Proof: Easy: if $(\bar{x}, \bar{y})$ is a saddle point of $L(x, y)$ then $\bar{x}$ is optimal for (CO). The proof of this part does not need any regularity condition. From the saddle point inequality (14) one has

$$f(\bar{x}) + \sum_{j=1}^{m} y_j g_j(\bar{x}) \le f(\bar{x}) + \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) \le f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x)$$

for all $y \ge 0$ and for all $x \in \mathcal{C}$. From the first inequality $g_j(\bar{x}) \le 0$ for all $j = 1, \cdots, m$ follows, hence $\bar{x} \in \mathcal{F}$ is feasible for (CO). Taking the two extreme sides of the above inequality and substituting $y = 0$ we have

$$f(\bar{x}) \le f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x) \le f(x)$$

for all $x \in \mathcal{F}$, i.e. $\bar{x}$ is optimal.

KKT proof cntd.

To prove the other direction we need Slater regularity and the Convex Farkas Lemma 2. Let us take an optimal solution $\bar{x}$ of the convex optimization problem (CO). Then the inequality system

$$\begin{array}{ll} f(x) - f(\bar{x}) < 0 & \\ g_j(x) \le 0, & j = 1, \cdots, m \\ x \in \mathcal{C} & \end{array}$$

is infeasible. Applying the Convex Farkas Lemma 2 one has $\bar{y} \ge 0$ such that

$$f(x) - f(\bar{x}) + \sum_{j=1}^{m} \bar{y}_j g_j(x) \ge 0$$

for all $x \in \mathcal{C}$. Using that $\bar{x}$ is feasible one easily derive the saddle point inequality

$$f(\bar{x}) + \sum_{j=1}^{m} y_j g_j(\bar{x}) \le f(\bar{x}) \le f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x)$$

which completes the proof. □

KKT−Corollaries

Corollary 1 *Under the assumptions of Theorem 2 the vector* $\bar{x} \in \mathcal{C}$ *is an optimal solution of (CO) if and only if there exists a* $\bar{y} \geq 0$ *such that*

$$(i) \quad f(\bar{x}) = \min_{x \in \mathcal{C}} \{ f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x) \} \qquad \text{and}$$

$$(ii) \quad \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) = \max_{y \geq 0} \{ \sum_{j=1}^{m} y_j g_j(\bar{x}) \}.$$

Corollary 2 *Under the assumptions of Theorem 2 the vector* $\bar{x} \in \mathcal{F}$ *is an optimal solution of (CO) if and only if there exists a* $\bar{y} \geq 0$ *such that*

$$(i) \quad f(\bar{x}) = \min_{x \in \mathcal{C}} \{ f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x) \} \qquad \text{and}$$

$$(ii) \quad \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) = 0.$$

Corollary 3 *Let us assume that* $\mathcal{C} = \mathbb{R}^n$ *and the functions* $f, g_1, \cdots, g_m$ *are continuously differentiable functions. Under the assumptions of Theorem 2 the vector* $\bar{x} \in \mathcal{F}$ *is an optimal solution of (CO) if and only if there exists a* $\bar{y} \geq 0$ *such that*

$$(i) \quad 0 = \nabla f(\bar{x}) + \sum_{j=1}^{m} \bar{y}_j \nabla g_j(\bar{x}) \qquad \text{and}$$

$$(ii) \quad \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) = 0.$$

KKT point

Definition 4 *Let us assume that* $\mathcal{C} = \mathbb{R}^n$ *and the functions* $f, g_1, \cdots, g_m$ *are continuously differentiable functions. The vector* $(\bar{x}, \bar{y}) \in \mathbb{R}^{n+m}$ *is called a Karush–Kuhn–Tucker (KKT) point of (CO) if*

$$
\begin{array}{ll}
(i) & g_j(\bar{x}) \le 0, \text{ for all } j \in J, \\
(ii) & 0 = \nabla f(\bar{x}) + \sum_{j=1}^{m} \bar{y}_j \nabla g_j(\bar{x}) \\
(iii) & \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) = 0, \\
(iv) & \bar{y} \ge 0.
\end{array}
$$

Corollary 4 *Let us assume that* $\mathcal{C} = \mathbb{R}^n$ *and the functions* $f, g_1, \cdots, g_m$ *are continuously differentiable convex functions and the assumptions of Theorem 2 hold. Let the vector* $\bar{x} \in \mathcal{F}$ *be a KKT point, then* $\bar{x}$ *is an optimal solution of (CO).*

Lagrange dual

Definition 5 *Denote*

$$\psi(y) = \inf_{x \in \mathcal{C}} \{ f(x) + \sum_{j=1}^{m} y_j g_j(x) \}.$$

The problem

$$(LD) \quad \sup_{y \geq 0} \psi(y)$$

is called the Lagrange Dual of problem (CO).

Lemma 4 *The Lagrange Dual (LD) of (CO) is a convex optimization problem, even if the functions* $f, g_1, \cdots, g_m$ *are not convex.*

Proof: $\psi(y)$ is concave! Let $\bar{y}, \hat{y} \geq 0$ and $0 \leq \lambda \leq 1$.

$$\psi(\lambda \bar{y} + (1-\lambda)\hat{y}) = \inf_{x \in \mathcal{C}} \{ f(x) + \sum_{j=1}^{m} (\lambda \bar{y}_j + (1-\lambda)\hat{y}_j) g_j(x) \}$$

$$= \inf_{x \in \mathcal{C}} \{ \lambda [f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x)] + (1-\lambda)[f(x) + \sum_{j=1}^{m} \hat{y}_j g_j(x)] \}$$

$$\geq \inf_{x \in \mathcal{C}} \{ \lambda [f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x)] \} + \inf_{x \in \mathcal{C}} \{ (1-\lambda)[f(x) + \sum_{j=1}^{m} \hat{y}_j g_j(x)] \}$$

$$= \lambda \psi(\bar{y}) + (1-\lambda)\psi(\hat{y}).$$

□

Results on the Lagrange dual

Theorem 3 (Weak duality) *If $\bar{x}$ is a feasible solution of (CO) and $\bar{y} \geq 0$ then*

$$\psi(\bar{y}) \leq f(\bar{x})$$

and equality hold if and only if

$$\inf_{x \in \mathcal{C}} \{f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x)\} = f(\bar{x}).$$

Proof:

$$\psi(\bar{y}) = \inf_{x \in \mathcal{C}} \{f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x)\} \leq f(\bar{x}) + \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) \leq f(\bar{x}).$$

Equality holds iff $\inf_{x \in \mathcal{C}} \{f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x)\} = f(\bar{x})$ and hence $\bar{y}_j g_j(x) = 0$ for all $j \in J$. □

Corollary 5 *If $\bar{x}$ is a feasible solution of (CO), $\bar{y} \geq 0$ and $\psi(\bar{y}) = f(\bar{x})$ then the vector $\bar{x}$ is an optimal solution of (CO) and $\bar{y}$ is optimal for (LD). Further if the functions $f, g_1, \cdots, g_m$ are continuously differentiable then $(\bar{x}, \bar{y})$ is a KKT-point.*

Theorem 4 (Strong duality) *Assume that (CO) satisfies the Slater regularity condition. Let $\bar{x}$ be a feasible solution of (CO). The vector $\bar{x}$ is an optimal solution of (CO) if and only if there exists a $\bar{y} \geq 0$ such that $\bar{y}$ is an optimal solution of (LD) and*

$$\psi(\bar{y}) = f(\bar{x}).$$

Wolfe dual

Definition 6 *Assume that* $\mathcal{C} = \mathbb{R}^n$ *and the functions* $f, g_1, \cdots, g_m$ *are continuously differentiable and* **convex**. *The problem*

$$(WD) \qquad \sup\{f(x) + \sum_{j=1}^{m} y_j g_j(x)\}$$
$$\nabla f(x) + \sum_{j=1}^{m} y_j \nabla g_j(x) = 0,$$
$$y \geq 0$$

is called the Wolfe Dual of the convex optimization problem (CO).

Warning! Remember, we are only allowed to form the Wolfe dual of a nonlinear optimization problem if it is *convex*!
For nonconvex problems one has to work with the Lagrange dual.

Examples for dual problems

Linear optimization

$$\text{(LO)}\quad \min\{c^T x \mid Ax = b,\ \ x \geq 0\}.$$

$g_j(x) = (a^j)^T x - b_j$ if $j = 1, \cdots, m$;
$g_j(x) = (-a^{j-m})^T x + b_{j-m}$ if $j = m+1, \cdots, 2m$;
$g_j(x) = -x_{j-2m}$ if $j = 2m+1, \cdots, 2m+n$.

Denote Lagrange multipliers by y^-, y^+ and s, then the Wolfe dual (WD) of (LO) is:

$$\begin{array}{ll} \max & c^T x + (y^-)^T(Ax - b) + (y^+)^T(-Ax + b) + s^T(-x) \\ & c + A^T y^- - A^T y^+ - s = 0, \\ & y^- \geq 0,\ \ y^+ \geq 0,\ \ s \geq 0. \end{array}$$

Substitute $c = -A^T y^- + A^T y^+ + s$
and let $y = y^+ - y^-$ then

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

Note that the KKT conditions provide the well known complementary slackness condition $x^T s = 0$.

Quadratic optimization

$$\text{(QO)} \quad \min\{c^T x + \tfrac{1}{2}x^T Q x \mid Ax \geq b, \;\; x \geq 0\}.$$

$g_j(x) = (-a^j)^T x + b_j$ if $j = 1, \cdots, m$;
$g_j(x) = -x_{j-m}$ if $j = m+1, \cdots, m+n$.
Lagrange multipliers: y and s.

The Wolfe dual (WD) of (QO) is:

$$\begin{aligned} \max \;\; & c^T x + \tfrac{1}{2}x^T Q x + y^T(-Ax + b) + s^T(-x) \\ & c + Qx - A^T y - s = 0, \\ & y \geq 0, \;\; s \geq 0. \end{aligned}$$

Substitute $c = -Qx + A^T y + s$ in the objective.

$$\begin{aligned} \max \quad & b^T y - \tfrac{1}{2}x^T Q x \\ & -Qx + A^T y + s = c, \\ & y \geq 0, \;\; s \geq 0. \end{aligned}$$

$Q = D^T D$ (e.g. Cholesky), let $z = Dx$.
The following (QD) dual problem is obtained:

$$\begin{aligned} \max \quad & b^T y - \tfrac{1}{2}z^T z \\ & -D^T z + A^T y + s = c, \\ & y \geq 0, \;\; s \geq 0. \end{aligned}$$

Constrained maximum

likelihood estimation

Finite set of sample points x_i, $(1 \le i \le n)$.
Find the most probable density values that satisfy some linear (e.g. convexity) constraints.
Maximize the Likelihood function $\Pi_{i=1}^n x_i$ under the conditions

$$Ax \ge 0, \;\; d^T x = 1, \;\; x \ge 0.$$

The objective can equivalently replaced by

$$\min \quad -\sum_{i=1}^n \ln x_i.$$

Lagrange multipliers: $y \in \mathbb{R}^m$, $t \in \mathbb{R}$ and $s \in \mathbb{R}^n$.
The Wolfe dual (WD) is:

$$\begin{aligned} \max \quad -\sum_{i=1}^n \ln x_i + y^T(-Ax) + t(d^T x - 1) + s^T(-x) \\ -X^{-1}e - A^T y + td - s = 0, \\ y \ge 0, \;\; s \ge 0. \end{aligned}$$

Multiplying the first constraint by x^T one has

$$-x^T X^{-1} e - x^T A^T y + t x^T d - x^T s = 0.$$

Using $d^T x = 1$, $x^T X^{-1} e = n$ and the optimality conditions $y^T Ax = 0$, $x^T s = 0$ we have

$$t = n.$$

x is necessarily strictly positive,

hence the dual variable s must be zero at optimum.

Constrained maximum

likelihood estimation: 2

$$\begin{array}{ll} \max & -\sum_{i=1}^{n} \ln x_i \\ & X^{-1}e + A^T y = nd, \\ & y \geq 0. \end{array}$$

Eliminating the variables $x_i > 0$:
$x_i = \frac{1}{nd_i - a_i^T y}$ and $-\ln x_i = \ln(nd_i - a_i^T y) \; \forall \; i$.

$$\begin{array}{ll} \max & \sum_{i=1}^{n} \ln(nd_i - a_i^T y) \\ & A^T y \leq nd, \\ & y \geq 0. \end{array}$$

Example: positive duality gap

Duffin's convex optimization problem

$$
\begin{array}{lll}
\text{(CO)} & \min & e^{-x_2} \\
& \text{s.t.} & \sqrt{x_1^2 + x_2^2} - x_1 \le 0 \\
& & x \in \mathbb{R}^2.
\end{array}
$$

The feasible region is $\mathcal{F} = \{x \in \mathbb{R}^2 |\ x_1 \ge 0, x_2 = 0\}$. (CO) is not Slater regular. The optimal value of the object function is 1.

The Lagrange function is given by

$$L(x, y) = e^{-x_2} + y(\sqrt{x_1^2 + x_2^2} - x_1).$$

Now, let $\epsilon = \sqrt{x_1^2 + x_2^2} - x_1$, then

$$x_2^2 - 2\epsilon x_1 - \epsilon^2 = 0.$$

Hence, for any $\epsilon > 0$ we can find $x_1 > 0$ such that $\epsilon = \sqrt{x_1^2 + x_2^2} - x_1$ even if x_2 goes to infinity. However, when x_2 goes to infinity e^{-x_2} goes to 0. So,

$$\psi(y) = \inf_{x \in \mathbb{R}^2} e^{-x_2} + y\left(\sqrt{x_1^2 + x_2^2} - x_1\right) = 0,$$

thus the optimal value of the Lagrange dual

$$\begin{array}{lll} \text{(LD)} & \max & \psi(y) \\ & \text{s.t.} & y \geq 0 \end{array}$$

is 0. Nonzero duality gap that equals to 1!

Example: infinite duality gap

Duffin's example slightly modified

$$\begin{array}{ll} \text{min} & -x_2 \\ \text{s.t.} & \sqrt{x_1^2+x_2^2} - x_1 \leq 0. \end{array}$$

The feasible region is
$\mathcal{F} = \{x \in \mathbb{R}^2 |\ x_1 \geq 0, x_2 = 0\}$. The problem is not Slater regular.
The optimal value of the object function is 0.

The Lagrange function is given by

$$L(x,y) = -x_2 + y(\sqrt{x_1^2+x_2^2} - x_1).$$

So,

$$\psi(y) = \inf_{x\in\mathbb{R}^2} \left\{ -x_2 + y\left(\sqrt{x_1^2+x_2^2} - x_1\right)\right\} = -\infty,$$

thus **the optimal value of the Lagrange dual**

$$\begin{array}{lll} \text{(LD)} & \text{max} & \psi(y) \\ & \text{s.t.} & y \geq 0 \end{array}$$

is $-\infty$, because $\psi(y)$ is minus infinity!

Semidefinite optimization

Let $A_0, A_1, \cdots, A_n \in \mathbb{R}^{m\times m}$, let $c \in \mathbb{R}^n$, $x \in \mathbb{R}^n$.
The *primal SDO problem* is defined as

$$\begin{aligned} (PSO) \quad \text{min} \quad & c^T x \\ \text{s.t.} \quad & -A_0 + \sum_{k=1}^{n} A_k x_k \succeq 0, \end{aligned} \tag{18}$$

$F(x) = -A_0 + \sum_{k=1}^{n} A_k x_k$
The *dual SDO problem* is:

$$\begin{aligned} (DSP) \quad \text{max} \quad & \text{Tr}(A_0 Z) \\ \text{s.t.} \quad & \text{Tr}(A_k Z) = c_k, \quad \text{for all} \quad k = 1, \cdots, n, \\ & Z \succeq 0, \end{aligned} \tag{19}$$

Theorem 5 *(Weak duality) If $x \in \mathbb{R}^n$ is primal feasible and $Z \in \mathbb{R}^{m\times m}$ is dual feasible, then*

$$c^T x \geq Tr(A_0 Z)$$

with equality if and only if

$$F(x)Z = 0.$$

Proof:

$$\begin{aligned} c^T x - \text{Tr}(A_0 Z) &= \textstyle\sum_{k=1}^{n} \text{Tr}(A_k Z) x_k - \text{Tr}(A_0 Z) \\ &= \text{Tr}((\textstyle\sum_{k=1}^{n} A_k x_k - A_0) Z) \\ &= \text{Tr}(F(x) Z) \geq 0. \end{aligned}$$

□

SDO Dual as Lagrange Dual

$$
\begin{array}{lll}
(PSO') & \min & c^T x \\
& \text{s.t.} & -F(x) + S = 0 \\
& & S \succeq 0,
\end{array}
\tag{20}
$$

The Lagrange function $L(x, S, Z)$ is defined on
$\{\,(x, S, Z) \,|\, x \in \mathbb{R}^n,\ S \in \mathbb{R}^{m\times m},\ S \succeq 0,\ Z \in \mathbb{R}^{m\times m},\}$

$$L(x, S, Z) = c^T x - e^T(F(x) \circ Z)e + e^T(S \circ Z)e,$$

$X \circ Z$ is the Minkowski product, $e^T(S \circ Z)e = \text{Tr}(SZ)$.

$$L(x, S, Z) = c^T x - \sum_{k=1}^{n} x_k \text{Tr}(A_k Z) + \text{Tr}(A_0 Z) + \text{Tr}(SZ). \tag{21}$$

$$(DSDL) \qquad \max\,\{\psi(Z) \mid Z \in \mathbb{R}^{m\times m}\} \tag{22}$$

$$\psi(Z) = \min\{\, L(x, S, Z) \,|\, x \in \mathbb{R}^n,\ S \in \mathbb{R}^{m\times m},\ S \succeq 0\}. \tag{23}$$

Optimality conditions to get $\psi(Z)$:

$$\min_S \text{Tr}(SZ) = \begin{cases} 0 & \text{if } Z \succeq 0, \\ -\infty & \text{otherwise.} \end{cases} \tag{24}$$

SDO dual cntd.

If we minimize (23) in x, we need to equate the $x-$gradient of $L(x, S, Z)$ to zero:

$$c_k - \text{Tr}(A_k Z) = 0 \quad \text{for all} \quad k = 1, \cdots, n. \qquad (25)$$

Multiplying (25) by x_k and summing up

$$c^T x - \sum_{k=1}^{n} x_k \text{Tr}(A_k Z) = 0.$$

By combining the last formula and the results presented in (24) and in (25) the simplified form of the Lagrange dual (22), the Lagrange–Wolfe dual

$$\begin{array}{lll} (DSO) & \max & \text{Tr}(A_0 Z) \\ & \text{s.t.} & \text{Tr}(A_k Z) = c_k, \quad \text{for all} \quad k = 1, \cdots, n, \\ & & Z \succeq 0, \end{array}$$

follows. It is identical to (19).

Duality in cone-linear optimization

$$\begin{aligned} \min\ c^T x & \\ Ax - b &\in \mathcal{C}_1 \\ x &\in \mathcal{C}_2. \end{aligned} \tag{26}$$

Definition 7 *Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex cone. The* dual cone (polar, positive polar) *$\mathcal{C}^*$ is defined by*

$$\mathcal{C}^* := \{\, z \in \mathbb{R}^n \,|\, x^T z \geq 0 \ \text{ for all } \ x \in \mathcal{C} \}.$$

The Dual of a Cone-linear Problem

$$\begin{aligned} \min\ c^T x & \\ s - Ax + b &= 0 \\ s \in \mathcal{C}_1 \qquad & x \in \mathcal{C}_2. \end{aligned}$$

Linear equality constraints $s - Ax + b = 0$ and (s, x) must be in the convex cone

$$\mathcal{C}_1 \times \mathcal{C}_2 := \{\, (s, x) \,|\, s \in \mathcal{C}_1,\ x \in \mathcal{C}_2 \,\}.$$

The Lagrange function $L(s, x, y)$ is defined on

$$\{\, (s, x, y) \,|\, s \in \mathcal{C}_1,\ x \in \mathcal{C}_2,\ y \in \mathbb{R}^m \,\}$$

and is given by

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

Thee Lagrange dual of

the cone-linear problem

$$\max_{y \in \mathbb{R}^m} \; \psi(y)$$

where

$$\psi(y) \;=\; \min\{\, L(s,x,y) \,|\, s \in \mathcal{C}_1,\; x \in \mathcal{C}_2\}.$$

The vector s is in the cone $\mathcal{C}_1$, hence

$$\min_{s \in \mathcal{C}_1} \; s^T y = \begin{cases} 0 & \text{if } y \in \mathcal{C}_1^*, \\ -\infty & \text{otherwise.} \end{cases}$$

$$\min_{x \in \mathcal{C}_2} \; x^T(c - A^T y) = \begin{cases} 0 & \text{if } c - A^T y \in \mathcal{C}_2^*, \\ -\infty & \text{otherwise.} \end{cases}$$

By combining these: we have

$$\psi(y) = \begin{cases} b^T y & \text{if } y \in \mathcal{C}_1^* \text{ and } c - A^T y \in \mathcal{C}_2^*, \\ -\infty & \text{otherwise.} \end{cases}$$

Thus the *dual* of (26) is:

$$\begin{aligned} \max\; b^T y & \\ c - A^T y \;&\in\; \mathcal{C}_2^* \\ y \;&\in\; \mathcal{C}_1^*. \end{aligned}$$