CS257 Linear and Convex Optimization

Lecture 14

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

December 7, 2020

Recap: Lagrange Condition

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = 0, \ i = 1, 2, \ldots, k \end{aligned}$$

Regular point. $\boldsymbol{x}$ is a regular point of $\boldsymbol{h}$ if $\nabla h_1(\boldsymbol{x}), \ldots, \nabla h_k(\boldsymbol{x})$ are linearly independent.

First-order necessary condition. If $\boldsymbol{x}^*$ is a local minimum and also a regular point, then there exist Lagrange multipliers $\lambda_1^*, \ldots, \lambda_k^*$ s.t. the following Lagrange condition holds

$$\begin{cases} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^k \lambda_i^* \nabla h_i(\boldsymbol{x}^*) = \mathbf{0} \\ h_i(\boldsymbol{x}^*) = 0, \quad i = 1, 2, \ldots, k \end{cases}$$

or

$$\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}, \quad \text{where } \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^k \lambda_i h_i(\boldsymbol{x})$$

Note. For $\boldsymbol{x} \in \mathbb{R}^n$, we assumed $k < n$, but also true for $k = n$ (why?)

Recap: Lagrange Condition for Convex Problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i = 0, \ i = 1, 2, \ldots, k \end{aligned}$$

First-order optimality condition for convex problem. If f is convex and $h_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i$ are affine, then $\boldsymbol{x}^*$ is a global minimum if and only if there exist Lagrange multipliers $\lambda_1^*, \ldots, \lambda_k^*$ s.t.

$$\begin{cases} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^k \lambda_i^* \boldsymbol{a}_i = \boldsymbol{0} \\ h_i(\boldsymbol{x}^*) = \boldsymbol{a}_i^T \boldsymbol{x}^* - b_i = 0, \quad i = 1, 2, \ldots, k \end{cases}$$

Note. Regularity is not needed in the convex case. We assumed regularity, i.e. $\operatorname{rank} \boldsymbol{A} = k$ for $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$, for simplicity, but it is not necessary for Lagrange condition to hold. If $\operatorname{rank} \boldsymbol{A} = k$, $\boldsymbol{\lambda}^*$ is unique; if $\operatorname{rank} \boldsymbol{A} < k$, either infeasible or $\boldsymbol{\lambda}^*$ is not unique.

Recap: KKT Conditions

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = 0, \; i = 1, 2, \ldots, k \\ & g_j(\boldsymbol{x}) \le 0, \; j = 1, 2, \ldots, m \end{aligned}$$

A constraint is active at $\boldsymbol{x}_0$ if it holds with equality at $\boldsymbol{x}_0$.

A point $\boldsymbol{x}_0$ is regular if the gradients of all active constraints at $\boldsymbol{x}_0$ are linearly independent.

KKT conditions.

1. (dual feasibility) $\mu_j^* \ge 0, j = 1, 2, \ldots, m$
2. (stationarity) $\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^k \lambda_i^* \nabla h_i(\boldsymbol{x}^*) + \sum_{j=1}^m \mu_j^* \nabla g_j(\boldsymbol{x}^*) = \mathbf{0}$
3. (complementary slackness) $\mu_j^* g_j(\boldsymbol{x}^*) = 0, j = 1, 2, \ldots, m$
4. (primal feasibility) $h_i(\boldsymbol{x}^*) = 0, i = 1, \ldots, k; \; g_j(\boldsymbol{x}^*) \le 0, j = 1, \ldots, m$

- KKT is necessary for a regular point $\boldsymbol{x}^*$ to be a local minimum.
- For a convex problem, KKT is also sufficient for $\boldsymbol{x}^*$ to be a global minimum ($\boldsymbol{x}^*$ need not be regular).

Sufficiency of KKT Conditions for Convex Problems

Theorem. For a convex problem, i.e. f and g_j are convex, and h_i are affine, if there exist $\lambda_1^*, \ldots, \lambda_k^*$ and $\mu_1^*, \ldots, \mu_m^*$ s.t. the KKT conditions are satisfied at a feasible $\boldsymbol{x}^* \in X$, then $\boldsymbol{x}^*$ is a global minimum.

Note. The previous necessary conditions assume $\boldsymbol{x}^*$ is regular point. The sufficient conditions here assume convexity but not regularity.

Proof. We show $\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \forall \boldsymbol{x} \in X$.

1. By the KKT conditions,

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = -\sum_i \lambda^* \nabla h_i(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) - \sum_{j \in J(\boldsymbol{x}^*)} \mu_j^* \nabla g_j(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*)$$

   It suffices to show $\nabla h_i(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = 0$ and $\nabla g_j(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$.

2. Since $h_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{x} - b_i$ is affine, and $h_i(\boldsymbol{x}) = h(\boldsymbol{x}^*) = 0$ by feasibility,

$$\nabla h_i(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = \boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{x}^*) = h_i(\boldsymbol{x}) - h(\boldsymbol{x}^*) = 0$$

3. For $j \in J(\boldsymbol{x}^*)$, $g_j(\boldsymbol{x}^*) = 0$ and $g_j(\boldsymbol{x}) \leq 0$. By the convexity of g_j,

$$\nabla g_j(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \leq g_j(\boldsymbol{x}) - g_j(\boldsymbol{x}^*) \leq 0$$

Example

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^3} \quad & f(\boldsymbol{x}) = x_1 + 2x_2 + x_3 \\ \text{s.t.} \quad & h(\boldsymbol{x}) = x_1 + x_2 + 2x_3 = 0 \\ & g(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 - 1 \leq 0 \end{aligned}$$

All feasible points are regular. The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \lambda, \mu) = x_1 + 2x_2 + x_3 + \lambda(x_1 + x_2 + 2x_3) + \mu(x_1^2 + x_2^2 + x_3^2 - 1)$$

The KKT conditions (including the constraints) are

$$\begin{cases} \mu \geq 0 \\ \partial_{x_1}\mathcal{L} = 1 + \lambda + 2\mu x_1 = 0 \\ \partial_{x_2}\mathcal{L} = 2 + \lambda + 2\mu x_2 = 0 \\ \partial_{x_3}\mathcal{L} = 1 + 2\lambda + 2\mu x_3 = 0 \\ \mu(x_1^2 + x_2^2 + x_3^2 - 1) = 0 \\ x_1 + x_2 + 2x_3 = 0 \\ x_1^2 + x_2^2 + x_3^2 - 1 \leq 0 \end{cases}$$

Example (cont'd)

Case I. g is inactive. Thus $\mu = 0$. But this leads to a contradiction.

$$\begin{cases} \partial_{x_1}\mathcal{L} = 1 + \lambda + 2\mu x_1 = 0 \implies \lambda = -1 \\ \partial_{x_2}\mathcal{L} = 2 + \lambda + 2\mu x_2 = 0 \implies \lambda = -2 \end{cases}$$

Case II. g is active. This essentially reduces to the example on slide 17 of Lecture 13, but we only take the solution with $\mu \geq 0$,

$$\begin{cases} x_1 = -\frac{1}{\sqrt{66}} \\ x_2 = -\frac{7}{\sqrt{66}} \\ x_3 = \frac{4}{\sqrt{66}} \\ \lambda = -\frac{5}{6} \\ \mu = \sqrt{\frac{33}{72}} \end{cases}$$

Since the problem is convex, the above gives a global minimum.

Note. By minimizing $-f$, one can verify the other solution for the example on slide 17 of Lecture 13 is a global maximum.

Example

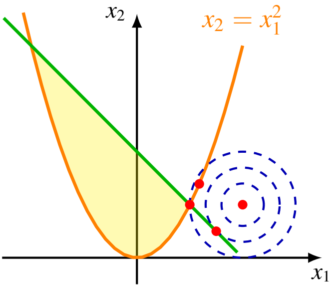
$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad f(\boldsymbol{x}) = (x_1-2)^2 + (x_2-1)^2$$
$$\text{s.t.} \quad g_1(\boldsymbol{x}) = x_1^2 - x_2 \leq 0$$
$$g_2(\boldsymbol{x}) = x_1 + x_2 - 2 \leq 0$$

All feasible points are regular. The Lagrangian is

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\mu}) = (x_1-2)^2 + (x_2-1)^2 + \mu_1(x_1^2 - x_2) + \mu_2(x_1+x_2-2)$$

The KKT conditions (including the constraints) are

$$\begin{cases} \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ \partial_{x_1}\mathcal{L} = 2(x_1-2) + 2\mu_1 x_1 + \mu_2 = 0 \\ \partial_{x_2}\mathcal{L} = 2(x_2-1) - \mu_1 + \mu_2 = 0 \\ \mu_1(x_1^2 - x_2) = 0 \\ \mu_2(x_1+x_2-2) = 0 \\ x_1^2 - x_2 \leq 0 \\ x_1 + x_2 - 2 \leq 0 \end{cases}$$



Example (cont'd)

Case I. Both g_1 and g_2 are inactive, so $\mu_1 = \mu_2 = 0$.

$$\begin{cases} \partial_{x_1}\mathcal{L} = 2(x_1 - 2) = 0 \\ \partial_{x_2}\mathcal{L} = 2(x_2 - 1) = 0 \end{cases} \implies \begin{cases} x_1 = 2 \\ x_2 = 1 \end{cases}$$

But

$$x_1^2 - x_2 = 3 > 0$$

violating $g_1 \leq 0$.

Case II. g_2 is active, but g_1 is inactive, so $\mu_1 = 0$.

$$\begin{cases} \partial_{x_1}\mathcal{L} = 2(x_1 - 2) + \mu_2 = 0 \\ \partial_{x_2}\mathcal{L} = 2(x_2 - 1) + \mu_2 = 0 \\ x_1 + x_2 - 2 = 0 \end{cases} \implies \begin{cases} x_1 = \frac{3}{2} \\ x_2 = \frac{1}{2} \\ \mu_2 = 1 \end{cases}$$

But

$$x_1^2 - x_2 = \frac{7}{4} > 0$$

violating $g_1 \leq 0$.

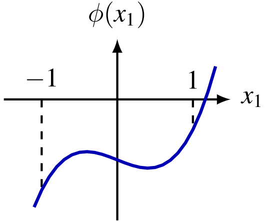
Example (cont'd)

Case III. g_1 is active, but g_2 is inactive, so $\mu_2 = 0$.

$$\begin{cases} \partial_{x_1}\mathcal{L} = 2(x_1 - 2) + 2\mu_1 x_1 = 0 \\ \partial_{x_2}\mathcal{L} = 2(x_2 - 1) - \mu_1 = 0 \\ x_1^2 - x_2 = 0 \end{cases}$$

From the last two equations,

$$x_2 = x_1^2, \quad \mu_1 = 2(x_2 - 1) = 2(x_1^2 - 1)$$



Plugging into the first equation,

$$2(x_1 - 2) + 4x_1(x_1^2 - 1) = 0 \implies \phi(x_1) \triangleq 2x_1^3 - x_1 - 2 = 0$$

Note $\mu_1 \geq 0 \implies x_1^2 \geq 1 \implies x_1 \geq 1$ or $x_1 \leq -1$.

If $x_1 \geq 1$, then $x_2 = x_1^2 \geq 1$, contradicting $x_1 + x_2 < 2$ (g_2 is inactive).
If $x_1 \leq -1$, $\phi(x_1) = 0$ has no solution since $\phi'(x_1) = 6x_1^2 - 1 > 0$ for $x_1 \leq -1$ and $\phi(-1) = -3 < 0$.

Example (cont'd)

Case IV. Both g_1 and g_2 are active.

$$\begin{cases} x_1^2 - x_2 = 0 \\ x_1 + x_2 - 2 = 0 \end{cases} \implies \begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases} \quad \text{or} \quad \begin{cases} x_1 = -2 \\ x_2 = -2 \end{cases}$$

Plugging into

$$\begin{cases} \partial_{x_1}\mathcal{L} = 2(x_1 - 2) + 2\mu_1 x_1 + \mu_2 = 0 \\ \partial_{x_2}\mathcal{L} = 2(x_2 - 1) - \mu_1 + \mu_2 = 0 \end{cases}$$

yields

$$\begin{cases} x_1 = 1 \\ x_2 = 1 \\ \mu_1 = \frac{2}{3} \\ \mu_2 = \frac{2}{3} \end{cases} \quad \text{or} \quad \begin{cases} x_1 = -2 \\ x_2 = -2 \\ \mu_1 = -\frac{2}{3} \quad \text{(violating } \mu_1 \geq 0\text{)} \\ \mu_2 = \frac{16}{3} \end{cases}$$

This is a convex problem, so $\boldsymbol{x}^* = (1, 1)^T$ is the global minimum.

Lagrange Duality

Lower Bounds in LP

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^2} \quad & f(\boldsymbol{x}) = x_1 + 2x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \geq 2 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Given a feasible solution $\boldsymbol{x}_0$, say $(1,0)^T$, can we say something about its quality as measured by $f(\boldsymbol{x}_0) - f^*$ without knowing f^*?

If we have a lower bound f_{LB} on f^*, then we can upper bound $f(\boldsymbol{x}_0) - f^*$

$$f(\boldsymbol{x}_0) - f^* \leq f(\boldsymbol{x}_0) - f_{\mathsf{LB}}$$

Note. A lower bound on f^* is the same as a lower bound on $f(\boldsymbol{x})$ for all feasible $\boldsymbol{x} \in X$.

Lower Bounds in LP (cont'd)

For any $\mu_1, \mu_2, \mu_3 \geq 0$,

$$\begin{array}{llrcrcll}
\mu_1 \times & [& 2x_1 & + & x_2 & \geq & 2 &] \\
\mu_2 \times & [& x_1 & & & \geq & 0 &] \\
\mu_3 \times & [& & & x_2 & \geq & 0 &] \\
\hline
& & (2\mu_1 + \mu_2)x_1 & + & (\mu_1 + \mu_3)x_2 & \geq & 2\mu_1 =: \psi(\boldsymbol{\mu}) &
\end{array}$$

We can set $2\mu_1 + \mu_2 = 1$ and $\mu_1 + \mu_3 = 2$ so the LHS becomes f.

Thus

$$f(\boldsymbol{x}) \geq \psi(\boldsymbol{\mu}) = 2\mu_1$$

for any $\boldsymbol{x} \in X$ and any μ_1, μ_2, μ_3 s.t.

$$2\mu_1 + \mu_2 = 1, \quad \mu_1 + \mu_3 = 2, \quad \mu_1, \mu_2, \mu_3 \geq 0$$

In particular, $f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x}) \geq \psi(\boldsymbol{\mu})$ for such $\boldsymbol{\mu}$.

Lower Bounds in LP (cont'd)

The quality of the lower bound $\psi(\boldsymbol{\mu})$ varies for different $\boldsymbol{\mu}$.

- $\psi(0,1,2) = 0$. It tells us

  $$f(1,0) - f^* \leq f(1,0) - \psi(0,1,2) = 1$$

  and

  $$0 = \psi(0,1,2) \leq f^* \leq f(1,0) = 1$$

- $\psi(\frac{1}{2}, 0, \frac{3}{2}) = 1$. It tells us

  $$f(1,0) - f^* \leq f(1,0) - \psi(\frac{1}{2}, 0, \frac{3}{2}) = 0$$

  and

  $$1 = \psi(\frac{1}{2}, 0, \frac{3}{2}) \leq f^* \leq f(1,0) = 1$$

  so $f^* = 1$ and $\boldsymbol{x}_0 = (1,0)^T$ is actually the optimal solution.

Dual LP

To get the best lower bound, we maximize over μ_1, μ_2, μ_3,

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^2} \quad & f(\boldsymbol{x}) = x_1 + 2x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \geq 2 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$	$$\begin{aligned} \max_{\boldsymbol{\mu}\in\mathbb{R}^3} \quad & \psi(\boldsymbol{\mu}) = 2\mu_1 \\ \text{s.t.} \quad & 2\mu_1 + \mu_2 = 1 \\ & \mu_1 + \mu_3 = 2 \\ & \mu_1 \geq 0 \\ & \mu_2 \geq 0 \\ & \mu_3 \geq 0 \end{aligned}$$
primal LP	dual LP

The variables μ_1, μ_2, μ_3 are called dual variables.

The number of dual variables is equal to the number of constraints in the primal problem.

The dual optimal solution is $\boldsymbol{\mu}^* = (\frac{1}{2}, 0, \frac{3}{2})^T$ and $\psi^* = 1 = f^*$.

Duality via Lagrangian

The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = x_1 + 2x_2 - \mu_1(2x_1 + x_2 - 2) - \mu_2 x_1 - \mu_3 x_2$$

If $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{x} \in X$, then

$$\begin{aligned} f(\boldsymbol{x}) &= x_1 + 2x_2 \\ &\geq x_1 + 2x_2 - \underbrace{\mu_1(2x_1 + x_2 - 2)}_{\geq 0} - \underbrace{\mu_2 x_1}_{\geq 0} - \underbrace{\mu_3 x_2}_{\geq 0} = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \end{aligned}$$

Taking the infimum over $\boldsymbol{x} \in X$ first and then relaxing the constraint,

$$f^* = \inf_{x \in X} f(\boldsymbol{x}) \geq \inf_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) \geq \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) =: \phi(\boldsymbol{\mu})$$

To maximize the lower bound, solve the dual problem

$$\begin{aligned} \max_{\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

Duality via Lagrangian (cont'd)

Rewriting the Lagrangian as

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = (1 - 2\mu_1 - \mu_2)x_1 + (2 - \mu_1 - \mu_3)x_2 + 2\mu_1$$

The dual objective

$$\phi(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = \begin{cases} 2\mu_1, & \text{if } 1 - 2\mu_1 - \mu_2 = 0, 2 - \mu_1 - \mu_3 = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

The dual problem is

$$\begin{aligned} \max_{\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\mu}) = \begin{cases} 2\mu_1, & \text{if } 1 - 2\mu_1 - \mu_2 = 0, 2 - \mu_1 - \mu_3 = 0 \\ -\infty, & \text{otherwise} \end{cases} \\ \text{s. t.} \quad & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

which is equivalent to the dual LP

$$\begin{aligned} \max_{\boldsymbol{\mu}} \quad & \psi(\boldsymbol{\mu}) = 2\mu_1 \\ \text{s. t.} \quad & 2\mu_1 + \mu_2 = 1, \quad \mu_1 + \mu_3 = 2, \quad \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

Dual of General LP

Given $\boldsymbol{c} \in \mathbb{R}^n, \boldsymbol{A} \in \mathbb{R}^{k\times n}, \boldsymbol{b} \in \mathbb{R}^k, \boldsymbol{G} \in \mathbb{R}^{m\times n}, \boldsymbol{h} \in \mathbb{R}^m$, consider

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \end{aligned}$$

For $\boldsymbol{\lambda} \in \mathbb{R}^k$, $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\boldsymbol{\mu} \geq \boldsymbol{0}$,

$$-\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{\mu}^T\boldsymbol{G}\boldsymbol{x} \geq -\boldsymbol{\lambda}^T\boldsymbol{b} - \boldsymbol{\mu}^T\boldsymbol{h} =: \psi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

If $-\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c}$, then we can lower bound f^* by $f^* \geq \psi(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

To maximize the lower bound, solve the following dual problem

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \psi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\boldsymbol{\lambda}^T\boldsymbol{b} - \boldsymbol{\mu}^T\boldsymbol{h} \\ \text{s.t.} \quad & -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

Duality via Lagrangian

The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{\mu}^T(\boldsymbol{G}\boldsymbol{x} - \boldsymbol{h}), \quad \boldsymbol{\lambda} \in \mathbb{R}^k, \boldsymbol{\mu} \in \mathbb{R}^m$$

If $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{x} \in X$, i.e. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h}$, then

$$f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} \geq \boldsymbol{c}^T\boldsymbol{x} + \underbrace{\boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})}_{=0} + \underbrace{\boldsymbol{\mu}^T(\boldsymbol{G}\boldsymbol{x} - \boldsymbol{h})}_{\leq 0} = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Taking the infimum over $\boldsymbol{x} \in X$ first and then relaxing the constraint,

$$f^* = \inf_{x \in X} f(\boldsymbol{x}) \geq \inf_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \geq \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) =: \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

To maximize the lower bound, solve the dual problem

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

Duality via Lagrangian (cont'd)

Note

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = (\boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\lambda} + \boldsymbol{G}^T\boldsymbol{\mu})^T\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{\lambda} - \boldsymbol{h}^T\boldsymbol{\mu}.$$

An affine function is bounded below iff the coefficient for $\boldsymbol{x}$ is zero[1].

The dual problem

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} -\boldsymbol{b}^T\boldsymbol{\lambda} - \boldsymbol{h}^T\boldsymbol{\mu}, & \text{if } \boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\lambda} + \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{0} \\ -\infty & \text{otherwise} \end{cases} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

which is equivalent to the dual LP

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \psi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\boldsymbol{b}^T\boldsymbol{\lambda} - \boldsymbol{h}^T\boldsymbol{\mu} \\ \text{s.t.} \quad & -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

---

[1] Consider $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + c$. If $\boldsymbol{a} = \boldsymbol{0}$, then $\inf_{\boldsymbol{x}} f(\boldsymbol{x}) = c$. If $\boldsymbol{a} \neq \boldsymbol{0}$, letting $\boldsymbol{x} = -t\boldsymbol{a}$ and $t \to +\infty$ yields $\inf_{\boldsymbol{x}} f(\boldsymbol{x}) \leq -t\|\boldsymbol{a}\|^2 + c \to -\infty$.

Lagrange Dual Function

Consider the general optimization problem (not necessarily convex),

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = 0,\ i = 1, 2, \ldots, k \\ & g_j(\boldsymbol{x}) \leq 0,\ j = 1, 2, \ldots, m \end{aligned} \tag{P}$$

The Lagrangian is

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

The (Lagrange) dual function is

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in D} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in D} \left(f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}) \right)$$

where $D = \operatorname{dom} f \cap (\bigcap_{i=1}^{k} \operatorname{dom} h_i) \cap (\bigcap_{j=1}^{m} \operatorname{dom} g_j)$ is the domain of the problem. We will downplay the role of D and focus on the case $D = \mathbb{R}^n$.

Example

Given $\boldsymbol{A} \in \mathbb{R}^{k \times n}$,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = \boldsymbol{x}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{x}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

Since $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is convex in $\boldsymbol{x}$, its minimum satisifies

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = 2\boldsymbol{x} + \boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{0} \implies \boldsymbol{x} = -\frac{1}{2} \boldsymbol{A}^T \boldsymbol{\lambda}$$

The dual function is

$$\phi(\boldsymbol{\lambda}) = \mathcal{L}\left(-\frac{1}{2}\boldsymbol{A}^T \boldsymbol{\lambda}, \boldsymbol{\lambda}\right) = -\frac{1}{4} \boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{b}^T \boldsymbol{\lambda} = -\frac{1}{4} \|\boldsymbol{A}^T \boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T \boldsymbol{\lambda}$$

Example

Given $\boldsymbol{A} \in \mathbb{R}^{k \times n}$,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{x}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{\mu}^T \boldsymbol{x}$$

Since $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is convex in $\boldsymbol{x}$, its minimum satisifies

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 2\boldsymbol{x} + \boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{\mu} = \boldsymbol{0} \implies \boldsymbol{x} = \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda})$$

The dual function is

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathcal{L}\left(\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda}), \boldsymbol{\lambda}, \boldsymbol{\mu}\right) = -\frac{1}{4}\|\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T \boldsymbol{\lambda}$$

Lower Bound for Optimal Value

For any $\boldsymbol{\lambda}$ and any $\boldsymbol{\mu} \geq \mathbf{0}$, the optimal value f^* of (P) is bounded by

$$f^* \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Proof. Let $X = \{\boldsymbol{x} : h_i(\boldsymbol{x}) = 0, \forall i; g_j(\boldsymbol{x}) \leq 0, \forall j\}$ be the feasible set.

- If $X = \emptyset$, then $f^* = +\infty$, trivially true.
- If $X \neq \emptyset$, for $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{x} \in X$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i \underbrace{h_i(\boldsymbol{x})}_{=0} + \sum_{j=1}^{m} \underbrace{\mu_j g_j(\boldsymbol{x})}_{\leq 0} = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Minimizing over $\boldsymbol{x}$,

$$f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}} f(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Concavity of Dual Function

The dual function is always concave, whether the primal problem (P) is convex or not.

Proof. Note $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is affine in $(\boldsymbol{\lambda}, \boldsymbol{\mu})$. Thus $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is the pointwise infimum of a family of affine functions indexed by $\boldsymbol{x}$, and hence concave. (Recall the pointwise supremum of convex functions is convex).

$$\begin{aligned}\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\boldsymbol{x} \in D} \left(f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}) \right) \\ &= - \underbrace{\sup_{\boldsymbol{x} \in D} \left(-f(\boldsymbol{x}) - \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) - \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}) \right)}_{\text{pointwise supremum of convex (affine) functions in } (\boldsymbol{\lambda}, \boldsymbol{\mu})}\end{aligned}$$

Example. $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\frac{1}{4}\|\boldsymbol{\mu} - \boldsymbol{A}^T\boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T\boldsymbol{\lambda}$ is concave.

Lagrange Dual Problem

To find the best lower bound given by the dual function

$$f^* \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

solve the (Lagrange) dual problem associated with the primal problem (P),

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \tag{D}$$

The dual problem (D) is always convex, whether or not (P) is convex.

$(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \mathbf{0}$ and $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$.

Note. The domain of a convex function f is $\mathrm{dom} f = \{\boldsymbol{x} : f(\boldsymbol{x}) < +\infty\}$, while the domain of a concave function f is $\mathrm{dom} f = \{\boldsymbol{x} : f(\boldsymbol{x}) > -\infty\}$. Thus the condition $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$ just means $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathrm{dom}\, \phi$.

Example

The dual problem of the following general LP

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \end{aligned}$$

is

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda},\boldsymbol{\mu}) = \begin{cases} -\boldsymbol{\lambda}^T\boldsymbol{b} - \boldsymbol{\mu}^T\boldsymbol{h}, & \text{if } -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ -\infty, & \text{otherwise} \end{cases} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

$(\boldsymbol{\lambda},\boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \boldsymbol{0}$ and $-\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c}$, which just means it is feasible for the dual LP,

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \psi(\boldsymbol{\lambda},\boldsymbol{\mu}) = -\boldsymbol{\lambda}^T\boldsymbol{b} - \boldsymbol{\mu}^T\boldsymbol{h} \\ \text{s.t.} \quad & -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

Example

The dual problem of the following problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

is

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\frac{1}{4}\|\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T \boldsymbol{\lambda} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

$(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \boldsymbol{0}$.