Operations Research III: Theory

Lagrange Duality and the KKT Condition

Ling-Chieh Kung

Department of Information Management
National Taiwan University

Road map

- **Lagrange relaxation**.
- The KKT condition.
- More about Lagrange duality.

Solving constrained NLPs

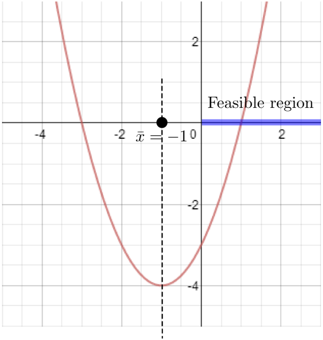
- For **unconstrained NLPs**, we have enough tools:
  - We may determine whether the objective function is convex.
  - We may use the FOC to find all local minima.
- How about **constrained NLPs**?
- We may always try the following strategy:
  - Ignore all the constraints.
  - Find a global minimum.
  - If it is feasible, it is optimal.
- If an unconstrained global minimum is infeasible, what should we do?

Solving single-variate constrained NLPs

- Let's solve

$$\min_{x \geq 0} \; f(x) = x^2 + 2x - 3.$$

- We have $f'(x) = 2x + 2$ and $f''(x) = 2$.
- f is convex and the solution satisfying the FOC is $\bar{x} = -1$. However, it is infeasible!
- For a single-variate NLP, the feasible solution that is **closest** to the FOC-solution is optimal.



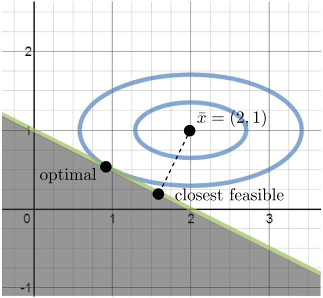
$$f(x) = x^2 + 2x - 3.$$

Solving multi-variate constrained NLPs

- Let's solve

$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^2} & f(x) = (x_1 - 2)^2 + 4(x_2 - 1)^2 \\ \text{s.t.} & x_1 + 2x_2 \leq 2. \end{array}$$

- For this CP, the FOC-solution $\bar{x} = (2, 1)$ is infeasible.
- The closest feasible point is **not** optimal!
- We need a way to deal with constraints.



$$f(x) = x^2 + 2x - 3.$$

Relaxation with rewards

- Recall our strategy: First ignore all constraints, and then ...
- Ignoring all constraints is "too much"!
  - An infeasible solution should be bad.
  - But this cannot be revealed in the relaxed NLP.
  - While we allow one to violate constraints, we **encourage** feasibility.
- Consider an original NLP

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq b_i \quad \forall i = 1, ..., m. \end{aligned}$$

- How to allow one to violate constraints but encourage feasibility?
  - For constraint i, let's associate a unit **reward** $\lambda_i \geq 0$ to it.
  - If a solution $\bar{x}$ satisfies constraint i (so $b_i - g_i(\bar{x}) \geq 0$), "reward" the solution by $\lambda_i[b_i - g_i(\bar{x})]$. Let's add this into the relaxed NLP.

Lagrange relaxation

- For an original NLP

$$z^* = \max_{x \in \mathbb{R}^n} \left\{ f(x) \middle| g_i(x) \leq b_i \ \forall i = 1, ..., m \right\}, \tag{1}$$

  we relax the constraints and add **rewards for feasibility** into the objective function:

$$z^L(\lambda) = \max_{x \in \mathbb{R}^n} \ f(x) + \sum_{i=1}^{m} \lambda_i \Big[b_i - g_i(x) \Big]. \tag{2}$$

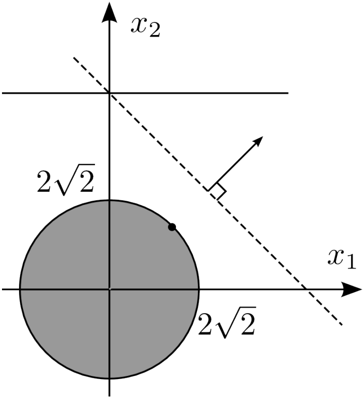
  - Let's assume that λ_is are given for a while.
  - To help solve the NLP, we should have $\lambda_i \geq 0$. This **rewards feasibility** and **penalize infeasibility**.
- $\mathcal{L}(x|\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i [b_i - g_i(x)]$ is the **Lagrangian** given λ.
- λ_is are the **Lagrange multipliers**.

An example

- Consider the following example

$$
\begin{array}{rccccl}
z^* = & \max & x_1 & + & x_2 & \\
& \text{s.t.} & x_1^2 & + & x_2^2 & \leq 8 \\
& & & & x_2 & \leq 6.
\end{array}
$$

- For this original NLP, the optimal solution is $x^* = (2, 2)$. $z^* = 4$.
- What may we do with Lagrange relaxation?



An example

- The original NLP is $z^* = \max_{x \in \mathbb{R}^2} \left\{ x_1 + x_2 \middle| x_1^2 + x_2^2 \leq 8, x_2 \leq 6 \right\}$.
- Given Lagrange multipliers $\lambda = (\lambda_1, \lambda_2) \geq 0$, the Lagrangian is

$$\mathcal{L}(x|\lambda) = x_1 + x_2 + \lambda_1(8 - x_1^2 - x_2^2) + \lambda_2(6 - x_2).$$

- We may solve

$$z^L(\lambda) = \max_{x \in \mathbb{R}^2} \mathcal{L}(x|\lambda)$$

  given any $\lambda \geq 0$. E.g.,
  - $z^L(0,1) = \max_{x \in \mathbb{R}^2} \; x_1 + 6 = \infty$.
  - $z^L(1,2) = \max_{x \in \mathbb{R}^2} \; -x_1^2 + x_1 - x_2^2 - x_2 + 20 = 20.5$.
  - $z^L(1,0) = \max_{x \in \mathbb{R}^2} \; -x_1^2 + x_1 - x_2^2 - x_2 + 8 = 8.5$.
- All the $z^L(\lambda)$ above is greater than $z^* = 4$! Will this always be true?

Lagrange relaxation provides a bound

- Lagrange relaxation provides a **bound** for the original NLP.

Proposition 1 (Weak duality of Lagrange relaxation)

For the two NLPs defined in (1) and (2), $z^L(\lambda) \geq z^*$ for all $\lambda \geq 0$.

Proof. We have

$$\begin{aligned} z^* &= \max_{x \in \mathbb{R}^n} \left\{ f(x) \middle| g_i(x) \leq b_i \ \forall i = 1, ..., m \right\} \\ &\leq \max_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^m \lambda_i [b_i - g_i(x)] \middle| g_i(x) \leq b_i \ \forall i = 1, ..., m \right\} \\ &\leq \max_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^m \lambda_i [b_i - g_i(x)] \right\} = z^L(\lambda), \end{aligned}$$

where the first inequality relies on $\lambda \geq 0$. □

Lagrange duality

- Given a constrained original NLP, solving its Lagrange relaxation gives us some information.
- A similar situation happened to LP!
  - Any feasible dual solution gives a bound to the primal LP.
  - We look for an dual optimal solution that gives a tight bound.
- Given that $z^L(\lambda) \geq z^*$ for all $\lambda \geq 0$, it is natural to define

$$\min_{\lambda \geq 0} \; z^L(\lambda)$$

  as the **Lagrange dual program**.
  - Lagrange multipliers are **dual variables** in NLP.

Road map

- Lagrange relaxation.
- **The KKT condition**.
- More about Lagrange duality.

The KKT condition

- Now we present the most useful optimality condition for general NLPs:

Proposition 2 (KKT condition)

For a "regular" NLP

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & f(x) \\ s.t. \quad & g_i(x) \leq b_i \quad \forall i = 1, ..., m, \end{aligned}$$

if $\bar{x}$ is a local max, then there exists $\lambda \in \mathbb{R}^m$ such that

- $g_i(\bar{x}) \leq b_i$ *for all* $i = 1, ..., m$,
- $\lambda \geq 0$ *and* $\nabla f(\bar{x}) = \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x})$, *and*
- $\lambda_i [b_i - g_i(\bar{x})] = 0$ *for all* $i = 1, ..., m$.

- All NLPs in this course (and most in the world) are "regular".
- The condition is necessary for all NLPs but also sufficient for CPs.

The KKT condition

- There are three conditions for $\bar{x}$ to be a local maximum.
- **Primal feasibility**: $g_i(\bar{x}) \leq b_i$ for all $i = 1, ..., m$.
  - It must be feasible.
- **Dual feasibility**: $\lambda \geq 0$ and $\nabla f(\bar{x}) = \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x})$.
  - The equality is the **FOC for the Lagrangian** $\mathcal{L}(\bar{x}|\lambda)$:

$$\nabla \left\{ f(x) + \sum_{i=1}^{m} \lambda_i [b_i - g_i(x)] \right\} = 0 \quad \Leftrightarrow \quad \nabla f(\bar{x}) - \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0.$$

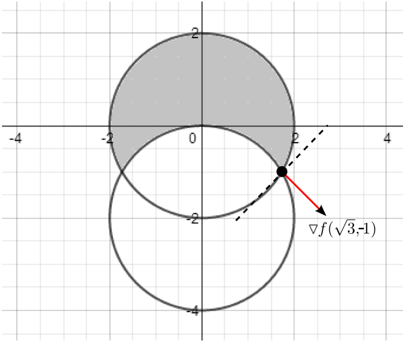
- **Complementary slackness**: $\lambda_i [b_i - g_i(\bar{x})] = 0$ for all $i = 1, ..., m$.
  - Dual variable $\times$ primal slack $= 0$.
  - If a constraint is **nonbinding**, the Lagrange multiplier is 0.
- Let's visualize the KKT condition.

Visualizing the KKT condition

- Consider

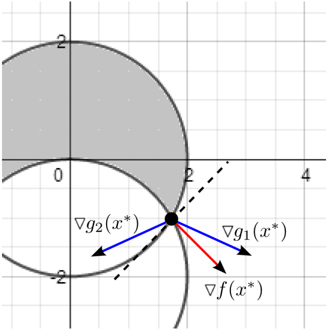
$$\begin{aligned} \max \quad & x_1 - x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \leq 4 \\ & -x_1^2 - (x_2+2)^2 \leq -4. \end{aligned}$$

- Graphically, $x^* = (\sqrt{3}, -1)$ is optimal.
- What happens to ∇f, ∇g_1, and ∇g_2 at x^*?



Visualizing the KKT condition

$$\begin{aligned} \max \quad & f(x) = x_1 - x_2 \\ \text{s.t.} \quad & g_1(x) = x_1^2 + x_2^2 \leq 4 \\ & g_2(x) = -x_1^2 - (x_2+2)^2 \leq -4. \end{aligned}$$



- We have $\nabla f(x) = (1, -1)$, $\nabla g_1(x) = (2x_1, 2x_2)$, and $\nabla g_2(x) = (-2x_1, -2(x_2+2))$.
- Therefore, $\nabla f(x^*) = (1, -1)$, $\nabla g_1(x^*) = (2\sqrt{3}, -2)$, and $\nabla g_2(x^*) = (-2\sqrt{3}, -2)$.
- The existence of $\lambda \geq 0$ such that $\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*)$ simply means that ∇f is "**in between**" ∇g_1 and ∇g_2 at x^*.
  - Otherwise there is a feasible improving direction.
  - Complementary slackness $\lambda_i[b_i - g_i(x^*)]$ says that only constraints binding at x^* matter.

Example 1

- A retailer sells products 1 and 2 at supply quantities q_1 and q_2. For product i, the market-clearing price is

$$p_i = a_i - b_i q_i, \quad i = 1, 2,$$

where $a_i > 0$ and $b_i > 0$ are known parameters for $i = 1, 2$. The retailer sets q_1 and q_2 to maximize its total profit while ensuring that the total supply does not exceed $K > 0$.

  - Formulate the retailer's problem.
  - Is this a convex program?
  - Solve the retailer's problem.
  - How do the optimal quantities change with K? Does that make sense?

Example 1: formulation

- The formulation is

$$\begin{aligned} \max_{q_1 \geq 0, q_2 \geq 0} \quad & q_1(a_1 - b_1 q_1) + q_2(a_2 - b_2 q_2) \\ \text{s.t.} \quad & q_1 + q_2 \leq K. \end{aligned}$$

- Let $f(q_1, q_2) = -\Big[q_1(a_1 - b_1 q_1) + q_2(a_2 - b_2 q_2)\Big]$, we have $\nabla^2 f(q_1, q_2) = \begin{bmatrix} 2b_1 & 0 \\ 0 & 2b_2 \end{bmatrix}$, which is positive semi-definite because $b_i > 0$. This implies that $f(q_1, q_2)$ is convex, i.e., the objective function $-f(q_1, q_2)$ is concave. As we are maximizing a concave function subject to linear constrains, this is a convex program.
- However, the first-order solution $(q_1, q_2) = (\frac{a_1}{2b_1}, \frac{a_2}{2b_2})$ may be infeasible.

Example 1: KKT condition

- The Lagrangian is

$$\mathcal{L}(x|\lambda) = q_1(a_1 - b_1 q_1) + q_2(a_2 - b_2 q_2) + \lambda(K - q_1 - q_2).$$

$\nabla \mathcal{L} = 0$ requires

$$\frac{\partial}{\partial q_1}\mathcal{L} = a_1 - 2b_1 q_1 - \lambda = 0$$
$$\frac{\partial}{\partial q_2}\mathcal{L} = a_2 - 2b_2 q_2 - \lambda = 0.$$

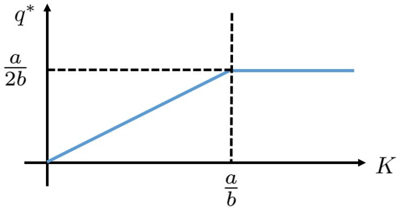
  - If $\lambda = 0$, we have $(q_1, q_2) = (\frac{a_1}{2b_1}, \frac{a_2}{2b_2})$. This is optimal if $\frac{a_1}{2b_1} + \frac{a_2}{2b_2} \leq K$.
  - If $\lambda > 0$, we have $q_1 + q_2 = K$. Solving the three equations results in $(q_1, q_2) = (\frac{2b_2 K + a_1 - a_2}{2(b_1 + b_2)}, \frac{2b_1 K + a_2 - a_1}{2(b_1 + b_2)})$. This is optimal if $\frac{a_1}{2b_1} + \frac{a_2}{2b_2} > K$.

Example 1: solution and interpretation

- The optimal solution is

$$(q_1^*, q_2^*) = \begin{cases} \left(\dfrac{a_1}{2b_1}, \dfrac{a_2}{2b_2}\right) & \text{if } \dfrac{a_1}{2b_1} + \dfrac{a_2}{2b_2} \leq K \\ \left(\dfrac{2b_2K + a_1 - a_2}{2(b_1+b_2)}, \dfrac{2b_1K + a_2 - a_1}{2(b_1+b_2)}\right) & \text{otherwise.} \end{cases}$$

- If K increases, q_1 and q_2 will weakly increase. If $a_1 = a_2$ and $b_1 = b_2$:

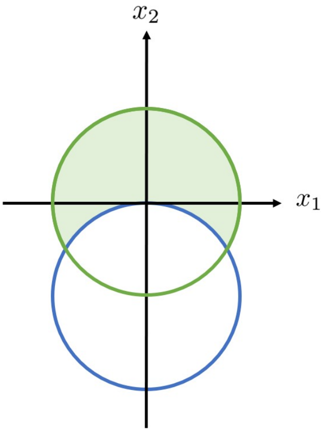


Example 2

- Let's solve

$$\begin{aligned} \max \quad & x_1 - x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \leq 4 \\ & -x_1^2 - (x_2+2)^2 \leq -4. \end{aligned}$$

- Note that this nonlinear program is **nonconvex**.
  - A typical numerical algorithm does not work.



Example 2: KKT condition

- The Lagrangian is

$$\mathcal{L}(x|\lambda) = x_1 - x_2 + \lambda_1(4 - x_1^2 - x_2^2) + \lambda_2(-4 + x_1^2 + (x_2 + 2)^2).$$

- $\frac{\partial \mathcal{L}(x|\lambda)}{\partial x_1} = 1 - 2(\lambda_1 - \lambda_2)x_1$ and $\frac{\partial \mathcal{L}(x|\lambda)}{\partial x_2} = -1 - 2(\lambda_1 - \lambda_2)x_2 + 4\lambda_2$.
- A solution $\bar{x}$ is a local maximum only if there exists λ such that

$$\begin{aligned}
&x_1^2 + x_2^2 \leq 4, -x_1^2 - (x_2 + 2)^2 \leq -4 && \text{(PF-1, PF-2)}\\
&\lambda_1 \geq 0, \lambda_2 \geq 0 && \text{(DFS-1, DFS-2)}\\
&1 - 2(\lambda_1 - \lambda_2)x_1 = 0, -1 - 2(\lambda_1 - \lambda_2)x_2 + 4\lambda_2 = 0 && \text{(DFF-1, DFF-2)}\\
&\lambda_1(4 - x_1^2 - x_2^2) = 0, \lambda_2(-4 + x_1^2 + (x_2 + 2)^2) = 0, && \text{(CS-1, CS-2)}
\end{aligned}$$

Example 2: analysis

- To find all solutions that satisfy the KKT condition, we **examine all four cases**.
- **Case 1.** $\lambda_1 > 0$, $\lambda_2 > 0$: By (CS-1) and (CS-2), we have $x_1^2 + x_2^2 = 4$ and $x_1^2 + (x_2 + 2)^2 = 4$. Solving the two equations results in $(x_1, x_2) = (\sqrt{3}, -1)$ and $(-\sqrt{3}, -1)$. They certainly satisfy (PF-1) and (PF-2).
  - Plugging $(\sqrt{3}, -1)$ into (DFF-1) and (DFF-2) result in $\lambda_1 = \frac{1}{4} + \frac{1}{4\sqrt{3}}$ and $\lambda_2 = \frac{1}{4} - \frac{1}{4\sqrt{3}}$. As this satisfies (DFS-1) and (DFS-2), $(\sqrt{3}, -1)$ is a KKT point.
  - Plugging $(-\sqrt{3}, -1)$ into (DFF-1) and (DFF-2) result in $\lambda_1 = \frac{1}{4} - \frac{1}{4\sqrt{3}}$ and $\lambda_2 = \frac{1}{4} + \frac{1}{4\sqrt{3}}$. As this satisfies (DFS-1) and (DFS-2), $(-\sqrt{3}, -1)$ is also a KKT point.

Example 2: analysis

- **Case 2.** $\lambda_1 > 0$, $\lambda_2 = 0$: Plugging $\lambda_2 = 0$ into (DFF-1) and (DFF-2) leads to

$$1 - 2\lambda_1 x_1 = 0 \quad \text{and} \quad -1 - 2\lambda_1 x_2 = 0,$$

which imply $x_1 = -x_2$. As $\lambda_1 > 0$ implies

$$x_1^2 + x_2^2 = 4$$

by (CS-1), we have two candidate solutions $(x_1, x_2) = (-\sqrt{2}, \sqrt{2})$ and $(\sqrt{2}, -\sqrt{2})$. Note that each of them requires $\lambda_1 = -\frac{1}{2\sqrt{2}}$ and $\lambda_1 = \frac{1}{2\sqrt{2}}$ due to (DFF-1) and (DFF-2), respectively. As the former violates $\lambda_1 > 0$, it cannot be a KKT point. For the latter, though $\lambda_1 > 0$ is good, it violates the second primal constraint $-x_1^2 - (x_2 + 2)^2 \leq -4$. We thus conclude that there is no KKT point under this case.

Example 2: analysis

- **Case 3.** $\lambda_1 = 0$, $\lambda_2 > 0$: Similar to Case 2, $\lambda_2 > 0$ and (CS-2) require

$$x_1^2 + (x_2 + 2)^2 = 4,$$

and $\lambda_1 = 0$ and (DFF-1) and (DFF-2) require

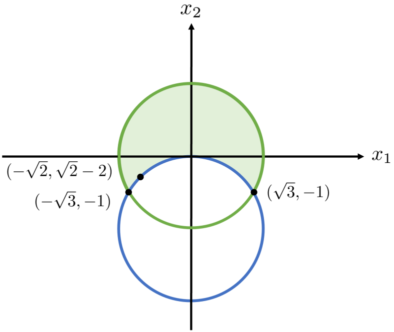
$$1 + 2\lambda_2 x_1 = 0 \quad \text{and} \quad -1 + 2\lambda_2 x_2 + 4\lambda_2 = 0.$$

Solving the three equations yields a KKT point $(x_1, x_2) = (-\sqrt{2}, \sqrt{2} - 2)$ with $\lambda_2 = \frac{1}{2\sqrt{2}}$. Note that another solution $(x_1, x_2) = (\sqrt{2}, -\sqrt{2} - 2)$ has $\lambda_2 = -\frac{1}{2\sqrt{2}}$, which violates $\lambda_2 > 0$.

- **Case 4.** $\lambda_1 = 0$, $\lambda_2 = 0$: As (DFF-1) and (DFF-2) become $1 = 0$ and $-1 = 0$ in this case, there is no KKT point in this case.

Example 2: visualization

- To summarize, we have:
  - $\lambda_1 > 0$, $\lambda_2 > 0$: Two KKT points $(\sqrt{3}, -1)$ and $(-\sqrt{3}, -1)$.
  - $\lambda_1 > 0$, $\lambda_2 = 0$: No KKT point.
  - $\lambda_1 = 0$, $\lambda_2 > 0$: A KKT point $(-\sqrt{2}, \sqrt{2} - 2)$.
  - $\lambda_1 = 0$, $\lambda_2 = 0$: No KKT point.
- These are the **only candidates of local maxima** (and thus global maxima).
  - Direct comparison shows that $(\sqrt{3}, -1)$ is global optimal.
  - Note that not all of these three points are local optimal.



The KKT condition for analysis

- The condition is named by three scholars Karush, Kuhn, and Tucker.
- In general, if there are n variables and m constraints.
  - There are n primal variables (x) and m dual variables (λ).
  - There are n equalities for dual feasibility.
  - There are m equalities for complementary slackness.
- As those equalities are nonlinear, there may be multiple solutions satisfying those equalities.
  - Those inequalities are then used to eliminate some solutions.
- If we have all local maxima, we compare them for a global maximum.
- Finding all local optima can be time consuming in general.
  - 2^m cases to examine.
  - Nonlinear equations are hard to solve (even numerically).
- The KKT condition is still useful for analyzing many constrained nonlinear optimization problems.

Road map

- Lagrange relaxation.
- The KKT condition.
- **More about Lagrange duality**.

More about Lagrange duality

- Recall that, for the **primal NLP** defined in (1)

$$z^* = \max_{x \in \mathbb{R}^n} \left\{ f(x) \middle| g_i(x) \leq b_i \ \forall i = 1, ..., m \right\},$$

  we define its **Lagrange dual program** in (2) as

$$\min_{\lambda \geq 0} \ z^L(\lambda) = \min_{\lambda \geq 0} \left\{ \max_{x \in \mathbb{R}^n} \ f(x) + \sum_{i=1}^{m} \lambda_i \Big[b_i - g_i(x) \Big] \right\}.$$

- We have derived the **weak duality** theorem: $z^L(\lambda) \geq z^*$ for all $\lambda \geq 0$.
- Below we plan to talk more about for **Lagrange duality**:
  - The **convexity** of $z^L(\lambda)$.
  - **Strong duality** (which holds for "regular" convex programs).
  - An **example** of solving the Lagrange dual program.
  - **Linear Programming duality** is a special case of Lagrange duality.

Convexity of Lagrange relaxation

- Is it reasonable to solve the Lagrange dual program $\min_{\lambda \geq 0} z^L(\lambda)$?

Proposition 3 (Convexity of the Lagrange dual program)

The function $z^L(\lambda)$ defined in (2) is convex over $\lambda \in [0, \infty)^n$.

A nonrigorous proof. It is true that:

- For any fixed $x \in \mathbb{R}^n$, the Lagrangian

$$\mathcal{L}(x|\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i \Big[b_i - g_i(x) \Big]$$

  is a linear function of λ.
- The maximum of convex functions is a convex function, i.e., $\max_{i=1,\ldots,m}\{g_i(x)\}$ is convex if $g_i(x)$ is convex for all $i = 1, ..., m$.

Combining the two facts implies that $z^L(\lambda) = \max_{x \in \mathbb{R}^n} \mathcal{L}(x|\lambda)$ is a convex function over the region in which λ is defined, i.e., $[0, \infty)^n$. $\square$

Convexity of Lagrange relaxation

- The above proposition shows that the Lagrange dual program

$$\min_{\lambda \geq 0} z^L(\lambda)$$

  is a **convex program** for **any** primal NLP.
- It thus reasonable to ask someone to solve a Lagrange dual program.
- In most practical applications, a Lagrange dual program is solved by **numerical algorithms**.
  - In this lecture, we will give you one example in which a Lagrange dual program is solved analytically.

Strong duality

- Let

$$w^* = \min_{\lambda \geq 0} \; z^L(\lambda) \tag{3}$$

be the optimized objective value to the Lagrange dual program.

- Weak duality implies that $w^* \geq z^*$.
- Is it possible for the upper bound to be **tight**, i.e., $w^* = z^*$?

Proposition 4 (Strong duality of Lagrange relaxation)

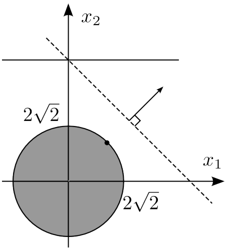
For the two NLPs defined in (1) and (2) and w^* defined in (3), $w^* = z^*$ if the primal NLP in (1) is a "regular" convex program.

- This is the **strong duality** of Lagrange relaxation.
- The proof is beyond the scope of this course. We will give you an example.
- All NLPs in this course (and most in the world) are "regular".

An example

- Recall the example

$$\begin{array}{rrcrcrcl} z^* = & \max & x_1 & + & x_2 & & \\ & \text{s.t.} & x_1^2 & + & x_2^2 & \leq & 8 \\ & & & & x_2 & \leq & 6. \end{array}$$



- For this primal NLP, the optimal solution is $x^* = (2, 2)$. $z^* = 4$.
- Lagrange relaxation leads to

$$w^* = \min_{\lambda_1 \geq 0, \lambda_2 \geq 0} z^L(\lambda) = \min_{\lambda_1 \geq 0, \lambda_2 \geq 0} \max_{x \in \mathbb{R}^2} \mathcal{L}(x|\lambda),$$

where $\mathcal{L}(x|\lambda) = x_1 + x_2 + \lambda_1(8 - x_1^2 - x_2^2) + \lambda_2(6 - x_2)$.
- Is $z^L(\lambda)$ convex? Is $w^* = z^*$?

An example: finding the dual program

- To solve the Lagrange dual program, first we solve

$$\max_{x \in \mathbb{R}^2} \mathcal{L}(x|\lambda) = \max_{x \in \mathbb{R}^2} \Big\{ x_1 + x_2 + \lambda_1(8 - x_1^2 - x_2^2) + \lambda_2(6 - x_2) \Big\}.$$

- This function is clearly jointly concave for x_1 and x_2. The first-order condition yields

$$x_1 = \frac{1}{2\lambda_1} \quad \text{and} \quad x_2 = \frac{1 - \lambda_2}{2\lambda_1}.$$

  Plugging these into $\mathcal{L}(x|\lambda)$ results in

$$z^L(\lambda) = \max_{x \in \mathbb{R}^2} \ \mathcal{L}(x|\lambda) = \frac{1 + (1 - \lambda_2)^2}{4\lambda_1} + 8\lambda_1 + 6\lambda_2.$$

- The Lagrange dual program is to look for $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ to minimize $z^L(\lambda)$.

An example: convexity of the dual program

- The Lagrange dual program

$$\min_{\lambda_1 \geq 0, \lambda_2 \geq 0} \frac{1 + (1 - \lambda_2)^2}{4\lambda_1} + 8\lambda_1 + 6\lambda_2$$

  is another constrained NLP.
  - Luckily, we know how to analytically solve it!
- $z^L(\lambda)$ is convex over $(0, \infty)^2$. To see this, note that

$$\nabla z^L(\lambda) = \begin{bmatrix} -\frac{1+(1-\lambda_2)^2}{4\lambda_1^2} + 8 \\ -\frac{1-\lambda_2}{2\lambda_1} + 6 \end{bmatrix} \text{ and } \nabla^2 z^L(\lambda) = \begin{bmatrix} \frac{1+(1-\lambda_2)^2}{2\lambda_1^3} & \frac{1-\lambda_2}{2\lambda_1^2} \\ \frac{1-\lambda_2}{2\lambda_1^2} & \frac{1}{2\lambda_1} \end{bmatrix}.$$

  As $\frac{1+(1-\lambda_2)^2}{2\lambda_1^3} > 0$, and $|\nabla^2 z^L(\lambda)| = \frac{1}{4\lambda_1^4} > 0$, the convexity is proved.
- Proposition 3 is indeed true for this example.

An example: solving the dual program

- To solve

$$\min_{\lambda_1 \geq 0, \lambda_2 \geq 0} \frac{1 + (1 - \lambda_2)^2}{4\lambda_1} + 8\lambda_1 + 6\lambda_2,$$

  let's apply the KKT condition.
  - Note that $\lambda_1 \geq 0$ cannot be binding at an optimal solution. Let's ignore it directly.
- Let $\mu \geq 0$ be the Lagrange multiplier for $\lambda_2 \geq 0$, the Lagrangian is

$$\frac{1 + (1 - \lambda_2)^2}{4\lambda_1} + 8\lambda_1 + 6\lambda_2 - \mu\lambda_2.$$

  The KKT condition requires an optimal solution to satisfy

$$-\frac{1 + (1 - \lambda_2)^2}{4\lambda_1^2} + 8 = 0, -\frac{1 - \lambda_2}{2\lambda_1} + 6 - \mu = 0, \text{ and } \mu\lambda_2 = 0.$$

An example: solving the dual program

- Suppose that $\mu > 0$.
  - This implies $\lambda_2 = 0$.
  - $-\frac{1+(1-\lambda_2)^2}{4\lambda_1^2} + 8 = 0$ requires $\lambda_1 = \frac{1}{4}$.
  - $-\frac{1-\lambda_2}{2\lambda_1} + 6 - \mu = 0$ requires $\mu = 4$, which is feasible.
- Suppose that $\mu = 0$.
  - $-\frac{1-\lambda_2}{2\lambda_1} + 6 = 0$ requires $\lambda_2 = 1 - 12\lambda_1$.
  - Plugging this into $-\frac{1+(1-\lambda_2)^2}{4\lambda_1^2} + 8 = 0$ results in $1 + 112\lambda_1^2 = 0$, which is impossible.
- The only KKT point is $(\lambda_1, \lambda_2) = (\frac{1}{4}, 0)$. Plugging this into $z^L(\lambda)$ gives us $w^* = 4$, which exactly equals z^*.
- Proposition 4 is indeed true for this example.

Lagrange duality vs. LP duality

- The last thing we want to do is to connect Linear Programming and Nonlinear Programming.
- The term "duality" is used both in LP duality and Lagrange duality. Somehow there must be some similarity.
- We will demonstrate one fact: LP duality is a **special case** of Lagrange duality.
  - Once we apply Lagrange duality on a primal LP, its Lagrange dual program will be its dual LP!

An example

- Consider an LP

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

- Let $\lambda \in \mathbb{R}^m$ be the Lagrange multipliers, the Lagrange relaxation is

$$\begin{aligned} z^L(\lambda) &= \max_{x \geq 0} \ c^T x + \lambda^T (b - Ax) \\ &= \lambda^T b + \max_{x \geq 0} \ (c^T - \lambda^T A) x \end{aligned}$$

- We set **no sign restriction** for λ. In fact, there is no multiplier with a specific sign that may reward feasibility or penalize infeasibility for equality constraints.

An example: Lagrange dual program

- The Lagrange dual program is

$$\begin{aligned} & \min_{\lambda} \; z^L(\lambda) \\ = & \min_{\lambda} \; \left\{ \lambda^T b + \max_{x \geq 0} \, (c^T - \lambda^T A) x \right\}. \end{aligned}$$

- The Lagrange dual program is to search for λ that minimizes $z^L(\lambda)$, which depends on the outcome of a maximization problem.
- The dual program is meaningful **only if $c^T \leq \lambda^T A$**. To see this, note that if $(c^T)_i > (\lambda^T A)_i$ for any i, $\max_{x \geq 0}(c^T - \lambda^T A)x$ will be unbounded because we may keep increasing x_i to infinity.
- In other words, no choice of λ that violates $c^T \leq \lambda^T A$ may be optimal to the Lagrange dual program.

An example: dual linear program

- The Lagrange dual program is

$$\min_{\lambda} z^*(\lambda) = \min_{\lambda : c^T \leq \lambda^T A} \left\{ \lambda^T b + \max_{x \geq 0} (c^T - \lambda^T A)x \right\}.$$

- If λ satisfies $c^T \leq \lambda^T A$, we know $\max_{x \geq 0} (c^T - \lambda^T A)x = 0$.
- The Lagrange dual becomes

$$\begin{array}{rl} \min\limits_{\lambda \text{ urs.}} & \lambda^T b \\ \text{s.t.} & \lambda^T A \geq c^T, \end{array}$$

which is exactly the dual LP of

$$\begin{array}{rl} \max\limits_{x \geq 0} & c^T x \\ \text{s.t.} & Ax = b. \end{array}$$