

Operations Research

Multi-variate Nonlinear Programming

Ling-Chieh Kung

Department of Information Management
National Taiwan University

Road map

- ▶ **Multi-variate convex analysis.**
- ▶ Solving constrained NLPs.
- ▶ Applications.

Convex analysis

- ▶ We have learned how to solve single-variate NLPs.
 - ▶ An optimal solution either satisfies the **FOC** or is a boundary point.
 - ▶ If the NLP is a **CP**, a feasible point satisfying the FOC is optimal.
- ▶ The above facts actually apply to **multi-variate NLPs**.
- ▶ We need to be able to determine whether a multi-variate function is convex, concave, or neither.
- ▶ We will still focus on **twice differentiable** functions.
 - ▶ Let's extend the notion of derivatives first.

Partial derivatives

- ▶ For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, its i th **partial derivative** is $\frac{\partial f(x)}{\partial x_i}$.
 - ▶ E.g., the partial derivatives for

$$f(x_1, x_2, x_3) = x_1^2 + x_2x_3 + x_3^3$$

are

$$\frac{\partial f(x)}{\partial x_1} = 2x_1, \quad \frac{\partial f(x)}{\partial x_2} = x_3 \quad \text{and} \quad \frac{\partial f(x)}{\partial x_3} = x_2 + 3x_3^2.$$

- ▶ It also has **second-order partial derivatives**:
 - ▶ For the same f , we have

$$\frac{\partial^2 f(x)}{\partial x_1^2} = 2, \quad \frac{\partial^2 f(x)}{\partial x_2^2} = 0, \quad \frac{\partial^2 f(x)}{\partial x_3^2} = 6x_3,$$

$$\frac{\partial^2 f(x)}{\partial x_1x_2} = \frac{\partial^2 f(x)}{\partial x_2x_1} = 0, \quad \frac{\partial^2 f(x)}{\partial x_1x_3} = \frac{\partial^2 f(x)}{\partial x_3x_1} = 0, \quad \frac{\partial^2 f(x)}{\partial x_2x_3} = \frac{\partial^2 f(x)}{\partial x_3x_2} = 1.$$

Symmetry of second-order derivatives

- ▶ For a second-order derivatives, we have the following fact:

Proposition 1

For a twice differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, if its second-order derivatives are all continuous, then

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$$

for all $i = 1, \dots, n$, $j = 1, \dots, n$.

- ▶ For all functions we will see in this course, the above property holds.

Multi-variate convex functions

- ▶ For $f : \mathbb{R} \rightarrow \mathbb{R}$, f is convex if and only if $f''(x) \geq 0$ for all x .
- ▶ For $f : \mathbb{R}^n \rightarrow \mathbb{R}$, is it true that f is convex if and only if $\frac{\partial^2 f(x)}{\partial x_i^2} \geq 0$ for all x_i , $i = 1, \dots, n$?
- ▶ Consider $f(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^2 + x_1 + x_2$. Is it convex at $(0, 0)$?
 - ▶ We have

$$\frac{\partial f(0,0)}{\partial x_1} = (2x_1 + 4x_2 + 1) \Big|_{(x_1, x_2) = (0,0)} = 1 \quad \text{and} \quad \frac{\partial^2 f(0,0)}{\partial x_1^2} = 2 > 0.$$

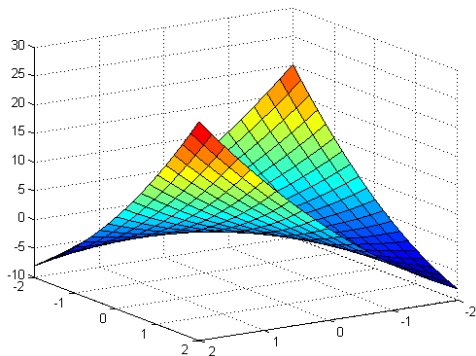
- ▶ We also have

$$\frac{\partial f(0,0)}{\partial x_2} = (2x_2 + 4x_1 + 1) \Big|_{(x_1, x_2) = (0,0)} = 1 \quad \text{and} \quad \frac{\partial^2 f(0,0)}{\partial x_1^2} = 2 > 0$$

- ▶ Is f convex at $(0, 0)$?

Multi-variate convex functions

- ▶ This is necessary but **insufficient!**
- ▶ $\frac{\partial^2}{\partial x_1^2} f(0, 0) \geq 0$ and $\frac{\partial^2}{\partial x_2^2} f(0, 0) \geq 0$ only imply that f is convex **along the two axes!**
 - ▶ Along $(1, -1)$, e.g., f is not convex.
- ▶ We need to test whether f is convex **in all directions.**



$$f(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^2 + x_1 + x_2.$$

Gradients and Hessians

- ▶ For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, collecting its first- and second-order partial derivatives generates its **gradient** and **Hessian**:

Definition 1 (Gradients and Hessians)

For a multi-variate twice differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, its gradient and Hessian are

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \quad \text{and} \quad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \ddots & \\ \vdots & & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

- ▶ In this course, all Hessians are **symmetric**.

Example

- For $f(x_1, x_2, x_3) = x_1^2 + x_2x_3 + x_3^3$, the gradient is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \frac{\partial f(x)}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ x_3 \\ x_2 + 3x_3^2 \end{bmatrix}.$$

- The Hessian is

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f(x)}{\partial x_3 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_3 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_3^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 6x_3 \end{bmatrix}.$$

- What are $\nabla f(3, 2, 1)$ and $\nabla^2 f(3, 2, 1)$?

Convexity of twice differentiable functions

- ▶ Recall the following theorem for single-variate functions:

Proposition 2

For a single-variate twice differentiable function $f(x)$:

- ▶ *f is convex in $[a, b]$ if $f''(x) \geq 0$ for all $x \in [a, b]$.*
 - ▶ *\bar{x} is an interior local min only if $f'(\bar{x}) = 0$.*
 - ▶ *If f is convex, x^* is a global min if and only if $f'(x^*) = 0$.*
- ▶ We have an analogous theorem for multi-variate functions:

Proposition 3

For a multi-variate twice differentiable function $f(x)$:

- ▶ *f is convex in F if $\nabla^2 f(x)$ is positive semi-definite for all $x \in F$.*
 - ▶ *\bar{x} is an interior local min only if $\nabla f(x) = 0$.*
 - ▶ *If f is convex, x^* is a global min if and only if $\nabla f(x^*) = 0$.*
- ▶ What is **positive semi-definiteness** (PSD)?

Positive semi-definite matrices

- ▶ Positive semi-definite Hessians in \mathbb{R}^n are **generalizations** of nonnegative second-order derivatives in \mathbb{R} .

Definition 2 (Positive semi-definite matrices)

A symmetric matrix A is positive semi-definite if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$.

- ▶ Example 1: For $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, we have

$$x^T Ax = 2x_1^2 + 2x_1x_2 + 2x_2^2 = (x_1 + x_2)^2 + x_1^2 + x_2^2 \geq 0 \quad \forall x \in \mathbb{R}^2.$$

- ▶ Example 2: For $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, we have $x^T Ax = x_1^2 + 4x_1x_2 + x_2^2$, which is negative when $x_1 = 1$ and $x_2 = -1$.

Positive semi-definite matrices

- ▶ Given a function f , when is its Hessian $\nabla^2 f$ PSD?

Proposition 4

For a symmetric matrix A , the following statements are equivalent:

- ▶ *A is positive semi-definite.*
 - ▶ *A 's eigenvalues are all nonnegative.*
 - ▶ *A 's leading principal minors are all nonnegative.*
-
- ▶ A 's eigenvalues λ and eigenvectors x satisfy $Ax = \lambda x$.
 - ▶ A 's k th leading principal minors is the determinant of the upper-left k by k submatrix.
- ▶ Given a function f , we will:
 - ▶ Find its Hessian.
 - ▶ Find its eigenvalues or leading principal minors.
 - ▶ Determine over what region the Hessian is PSD.
 - ▶ Over that region, the function is convex.

An example

- Consider the NLP

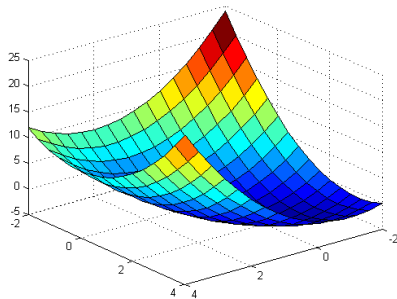
$$\min_{x \in \mathbb{R}^2} f(x_1, x_2),$$

where

$$f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2 - 2x_1 - 4x_2.$$

- Its gradient and Hessian are

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + x_2 - 2 \\ x_1 + 2x_2 - 4 \end{bmatrix} \quad \text{and} \quad \nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$



An example

$$f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2 - 2x_1 - 4x_2.$$

- ▶ To find the eigenvalues of $\nabla^2 f(x_1, x_2)$, recall that

$$Ax = \lambda x \Leftrightarrow (A - \lambda I)x = 0 \Leftrightarrow \det(A - \lambda I) = 0.$$

- ▶ For our $\nabla^2 f(x_1, x_2)$, we have

$$\begin{vmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{vmatrix} = 0 \Leftrightarrow 3 - 4\lambda + \lambda^2 = 0 \Leftrightarrow \lambda = 1 \text{ or } 3.$$

- ▶ Or by leading principal minors:

$$|2| = 2 \quad \text{and} \quad \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3.$$

- ▶ So $\nabla^2 f(x_1, x_2)$ is PSD and thus $\min_{x \in \mathbb{R}^2} f(x_1, x_2)$ is a CP. The FOC requires $2x_1^* + x_2^* - 2 = 0$ and $x_1^* + 2x_2^* - 4 = 0$, i.e., $(x_1^*, x_2^*) = (0, 2)$.

Another example

- ▶ Consider $f(x_1, x_2) = x_1^3 + 4x_1x_2 + x_2^2 + x_1 + x_2$. When is it convex?
- ▶ Its Hessian is

$$\begin{bmatrix} 6x_1 & 4 \\ 4 & 2 \end{bmatrix}.$$

- ▶ When is the Hessian positive semi-definite?
 - ▶ We need the first leading principal minor $6x_1 \geq 0$.
 - ▶ We need the second leading principal minor $6x_1 - 16 \geq 0$.
- ▶ Therefore, the function is convex if and only if $x_1 \geq \frac{8}{3}$.

Road map

- ▶ Multi-variate convex analysis.
- ▶ **Solving constrained NLPs.**
- ▶ Applications.

Solving constrained NLPs

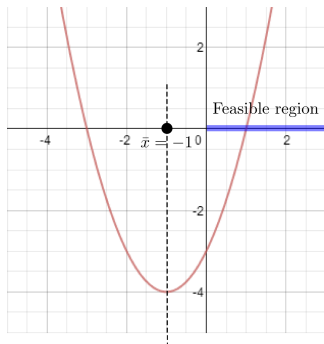
- ▶ For **unconstrained NLPs**, we have enough tools:
 - ▶ We may determine whether the objective function is convex.
 - ▶ We may use the FOC to find all local minima.
- ▶ How about **constrained NLPs**?
- ▶ We may always try the following strategy:
 - ▶ Ignore all the constraints.
 - ▶ Find a global 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.



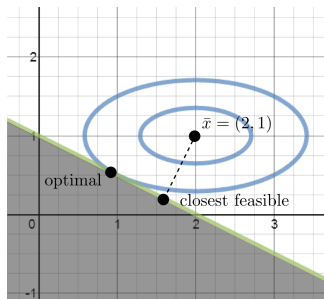
$$f(x) = x^2 + 2x - 3.$$

Solving multi-variate constrained NLPs

- ▶ Let's solve

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

- ▶ 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

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

- ▶ 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.

Lagrangian relaxation

- ▶ For an original NLP

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

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

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

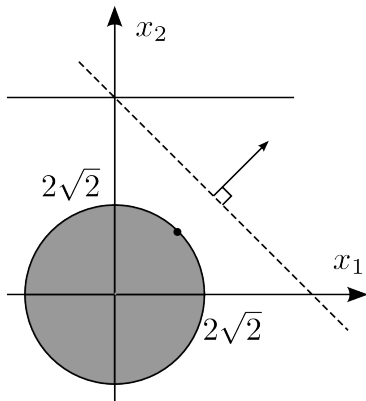
- ▶ Let's assume that λ_i s are given for a while.
- ▶ To help solve the NLP, we should have $\lambda_i \geq 0$. This **rewards feasibility** and **penalize infeasibility**.
- ▶ $\mathcal{L}(x|\lambda) = f(x) + \sum_{i=1}^m \lambda_i [b_i - g_i(x)]$ is the **Lagrangian** given λ .
- ▶ λ_i s are the **Lagrange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

An example

- ▶ Consider the following example

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

- ▶ For this original NLP, the optimal solution is $x^* = (2, 2)$. $z^* = 4$.
- ▶ What are the Lagrangian and Lagrangian relaxation?



An example

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

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

- ▶ The Lagrangian relaxation is

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

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

Lagrangian relaxation provides a bound

- ▶ The Lagrangian relaxation provides a **bound** for the original NLP.

Proposition 5

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

Proof. We have

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

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

Lagrangian duality

- ▶ Given a constrained original NLP, solving its Lagrangian 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 **Lagrangian dual program**.

- ▶ Lagrange multipliers are **dual variables** in NLP.
- ▶ LP duality is a special case of Lagrangian duality: The Lagrangian relaxation of an LP is the dual LP.
- ▶ Lagrangian duality possesses several properties (beyond the scope).
 - ▶ Just intuitively treat λ_i as the dual variable for constraint i .

The KKT condition

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

Proposition 6 (KKT condition)

For a “regular” NLP

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

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 $i = 1, \dots, m$,
- ▶ $\lambda \geq 0$ 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 $i = 1, \dots, m$.

- ▶ All NLPs in this course (and most in the world) are “regular”.
- ▶ The condition is necessary for general 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 $i = 1, \dots, m$.
 - ▶ 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** $\mathcal{L}(\bar{x}|\lambda)$:

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

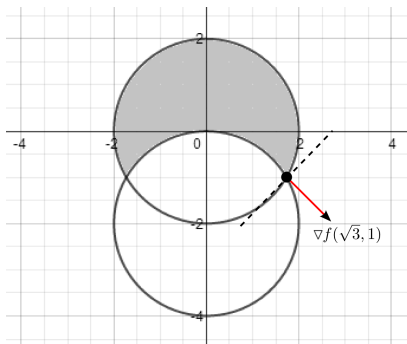
- ▶ **Complementary slackness:** $\lambda_i [b_i - g_i(\bar{x})] = 0$ for all $i = 1, \dots, m$.
 - ▶ Dual variable \times primal slack = 0.
 - ▶ If a constraint is **nonbinding**, the Lagrange multiplier is 0.
- ▶ Let's visualize the KKT condition.

Visualizing the KKT condition

- ▶ Consider

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

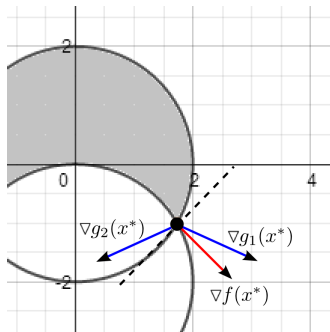
- ▶ Graphically, $x^* = (\sqrt{3}, 1)$ is optimal.
- ▶ What happens to ∇f , ∇g_1 , and ∇g_2 at x^* ?



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.



Applying the KKT condition

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

- ▶ 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

$$x_1^2 + x_2^2 \leq 4, -x_1^2 - (x_2 + 2)^2 \leq -4$$

$$\lambda_1 \geq 0, \lambda_2 \geq 0$$

$$1 - 2(\lambda_1 - \lambda_2)x_1 = 0, -1 - 2(\lambda_1 - \lambda_2)x_2 + 4\lambda_2 = 0$$

$$\lambda_1(4 - x_1^2 - x_2^2) = 0, \lambda_2(-4 + x_1^2 + (x_2 + 2)^2) = 0.$$

The KKT condition for analysis

- ▶ 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.
 - ▶ Nonlinear equations are hard to solve (even numerically).
 - ▶ Too time consuming in general.
- ▶ Nevertheless, we will see that the KKT condition is useful for analyzing many problems in business and economics.

Road map

- ▶ Multi-variate convex analysis.
- ▶ Solving constrained NLPs.
- ▶ **Applications.**

Multi-product EOQ problem

- ▶ Recall that we have solved the EOQ problem

$$\min_{q \geq 0} \frac{hq}{2} + \frac{KD}{q},$$

where h is the unit holding cost per year, K is the ordering cost per order, and D is the annual demand. The EOQ is $q^* = \sqrt{\frac{2KD}{h}}$.

- ▶ What if we procure two products? We solve

$$\min_{q_1 \geq 0, q_2 \geq 0} \frac{h_1 q_1}{2} + \frac{K_1 D_1}{q_1} + \frac{h_2 q_2}{2} + \frac{K_2 D_2}{q_2}.$$

The problem is separable; the optimal quantities are the two EOQs.

Multi-product EOQ problem

- ▶ What if we have only a limited space for these two products?
- ▶ We solve

$$\begin{aligned} \min_{q_1 \geq 0, q_2 \geq 0} \quad & \frac{h_1 q_1}{2} + \frac{K_1 D_1}{q_1} + \frac{h_2 q_2}{2} + \frac{K_2 D_2}{q_2} \\ \text{s.t.} \quad & v_1 q_1 + v_2 q_2 \leq W, \end{aligned}$$

where W is the total space and v_i is the volume of product i .

- ▶ Assumptions:
 - ▶ We assume that products can be “in any shape”.
 - ▶ This constraint can also be modeling budgets or something else.
 - ▶ We do not try to “synchronize” the procurement processes (so we assume the orders for the two products may arrive at the same time).
- ▶ How to solve this problem?
- ▶ To simplify the derivation, assume that $v_1 = v_2 = 1$ and $h_1 = h_2 = h$.

Convexity of the problem

- ▶ Our (simplified) two-product EOQ problem

$$\begin{aligned} \min_{q_1 \geq 0, q_2 \geq 0} \quad & \frac{hq_1}{2} + \frac{K_1 D_1}{q_1} + \frac{hq_2}{2} + \frac{K_2 D_2}{q_2} \\ \text{s.t.} \quad & q_1 + q_2 \leq W, \end{aligned}$$

is a CP:

- ▶ The objective function is convex; the Hessian matrix

$$\begin{bmatrix} \frac{2K_1 D_1}{q_1^3} & 0 \\ 0 & \frac{2K_2 D_2}{q_2^3} \end{bmatrix}$$

is positive semi-definite.

- ▶ The feasible region is convex.
- ▶ A local minimum is a global minimum.

The FOC for the Lagrangian

- ▶ The Lagrangian is

$$\mathcal{L}(q|\lambda) = \frac{hq_1}{2} + \frac{K_1 D_1}{q_1} + \frac{hq_2}{2} + \frac{K_2 D_2}{q_2} + \lambda(W - q_1 - q_2).$$

- ▶ The FOC for the Lagrangian is

$$\begin{aligned} \frac{\partial}{\partial q_1} \mathcal{L}(q|\lambda) &= \frac{h}{2} - \frac{K_1 D_1}{q_1^2} - \lambda = 0 \text{ and} \\ \frac{\partial}{\partial q_2} \mathcal{L}(q|\lambda) &= \frac{h}{2} - \frac{K_2 D_2}{q_2^2} - \lambda = 0. \end{aligned}$$

Note that this must be satisfied by **any optimal solution!**

- ▶ Therefore, we have

$$\frac{K_1 D_1}{q_1^2} = \frac{K_2 D_2}{q_2^2} \quad \Leftrightarrow \quad \frac{q_1}{q_2} = \sqrt{\frac{K_1 D_1}{K_2 D_2}}.$$

Solving the multi-product EOQ problem

- ▶ Now we are ready to solve our two-product EOQ problem

$$\min_{q_1 \geq 0, q_2 \geq 0} \left\{ \frac{hq_1}{2} + \frac{K_1 D_1}{q_1} + \frac{hq_2}{2} + \frac{K_2 D_2}{q_2} \mid q_1 + q_2 \leq W \right\}.$$

- ▶ If the unconstrained optimal solution $(\bar{q}_1, \bar{q}_2) = \left(\sqrt{\frac{2K_1 D_1}{h}}, \sqrt{\frac{2K_2 D_2}{h}} \right)$ satisfies $\bar{q}_1 + \bar{q}_2 \leq W$, it is optimal.
- ▶ Otherwise, the capacity constraint must be binding. The solution to the two equalities

$$q_1 + q_2 = W \quad \text{and} \quad \frac{q_1}{q_2} = \sqrt{\frac{K_1 D_1}{K_2 D_2}}$$

is optimal; i.e., $(\tilde{q}_1, \tilde{q}_2) = \left(\frac{W}{1 + \sqrt{\frac{K_2 D_2}{K_1 D_1}}}, \frac{W}{1 + \sqrt{\frac{K_1 D_1}{K_2 D_2}}} \right)$ is optimal.

Solving the multi-product EOQ problem

- Collectively, the optimal solution is

$$(q_1^*, q_2^*) = \begin{cases} \left(\sqrt{\frac{2K_1 D_1}{h}}, \sqrt{\frac{2K_2 D_2}{h}} \right) & \text{if } \sqrt{\frac{2K_1 D_1}{h}} + \sqrt{\frac{2K_2 D_2}{h}} \leq W \\ \left(\frac{W}{1 + \sqrt{\frac{K_2 D_2}{K_1 D_1}}}, \frac{W}{1 + \sqrt{\frac{K_1 D_1}{K_2 D_2}}} \right) & \text{otherwise.} \end{cases}$$

