最佳化理論-簡述

黃志勝 (Tommy Huang)
義隆電子 人工智慧研發部
國立陽明交通大學 AI學院 合聘助理教授
國立台北科技大學 電資學院合聘助理教授

Introduction

- Optimization(最佳化): 做最佳(Best)的判斷
- 什麼是"最佳" → measure "goodness" by a function f (cost function or objective function)
- 通常是希望要去最小化 f (Smaller = better)

Introduction

- Optimization problem:

$$\begin{aligned} &minimize\ f(x) \\ &subject\ to\ x \in \Omega \end{aligned}$$

- 如果你要找的問題是Maximization，直接把目標函數取負號即可。

- 如何解最佳化問題:
- 1. Analytically
- 2. Numerically

Introduction

Example1:

科技公司預計下半年推出A和B兩款筆電，企劃規劃

- A筆電BOM cost是5000元，組裝成本是5000元。
- B筆電競筆電BOM cost是20000元，組裝成本是1000元。

公司開發BOM採購預算上限是4000萬，組裝預算上限為600萬。

但A和B賣出去，每台獲利皆是2000元，假設推出的筆電都能賣出去，請問A和B各需要賣出幾台才能獲得最大利潤?

Introduction

Example1 :

假設A筆電賣出x_1台，B筆電賣出x_2台(單位:萬元)

- 依照題目

$$0.5x_1 + 2x_2 \leq 4000$$
$$0.5x_1 + 0.1x_2 \leq 600$$

目標函數就是最大化利潤

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2\}$$

	BOM	組裝
A	0.5	0.5
B	2	0.1
總成本	4000	600

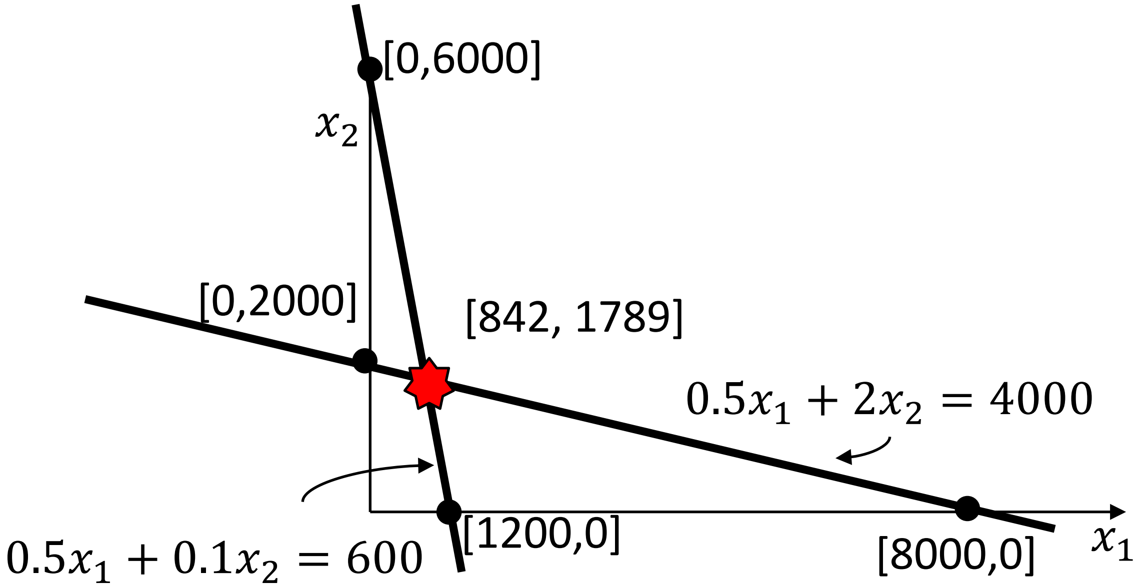
Introduction

Example1 :

$$0.5x_1 + 2x_2 \leq 4000$$
$$0.5x_1 + 0.1x_2 \leq 600$$

目標函數就是最大化利潤

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2\}$$



$$\begin{cases} 0.5x_1 + 2x_2 = 4000 \\ 0.5x_1 + 0.1x_2 = 600 \end{cases} \Rightarrow \begin{cases} x_1 + 4x_2 = 8000 \\ 5x_1 + x_2 = 6000 \end{cases} \Rightarrow$$

$$\begin{cases} 19x_1 = 16000 \\ 19x_2 = 34000 \end{cases} \Rightarrow \begin{cases} x_1 = 842.11 \\ x_2 = 1789.47 \end{cases}$$

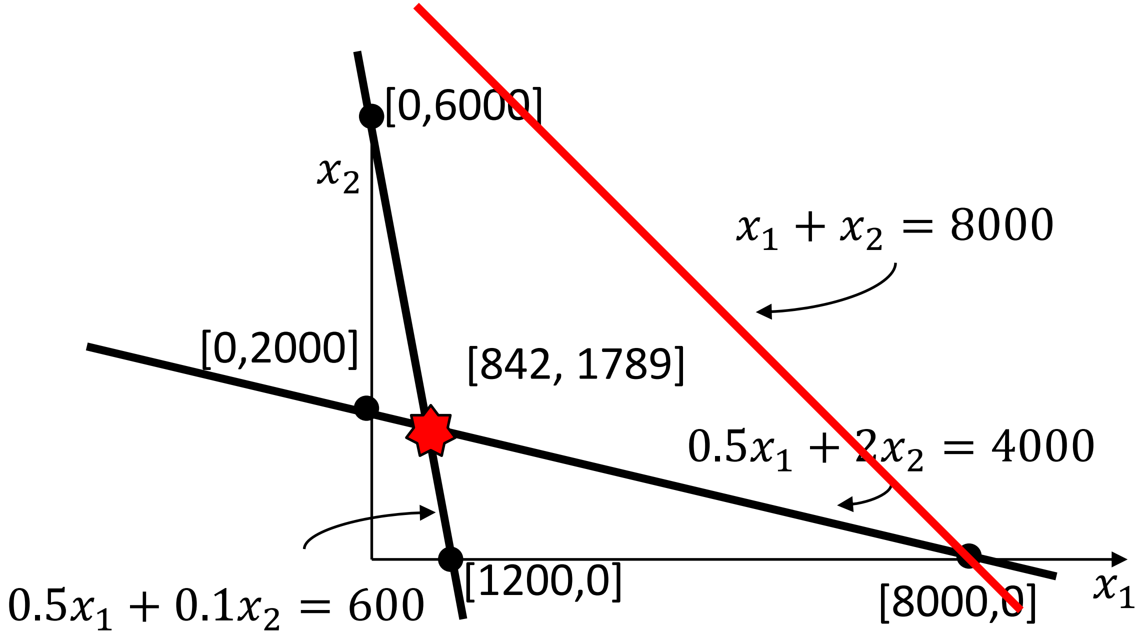
Introduction

Example1 :

$$0.5x_1 + 2x_2 \leq 4000$$
$$0.5x_1 + 0.1x_2 \leq 600$$

目標函數就是最大化利潤

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2\}$$



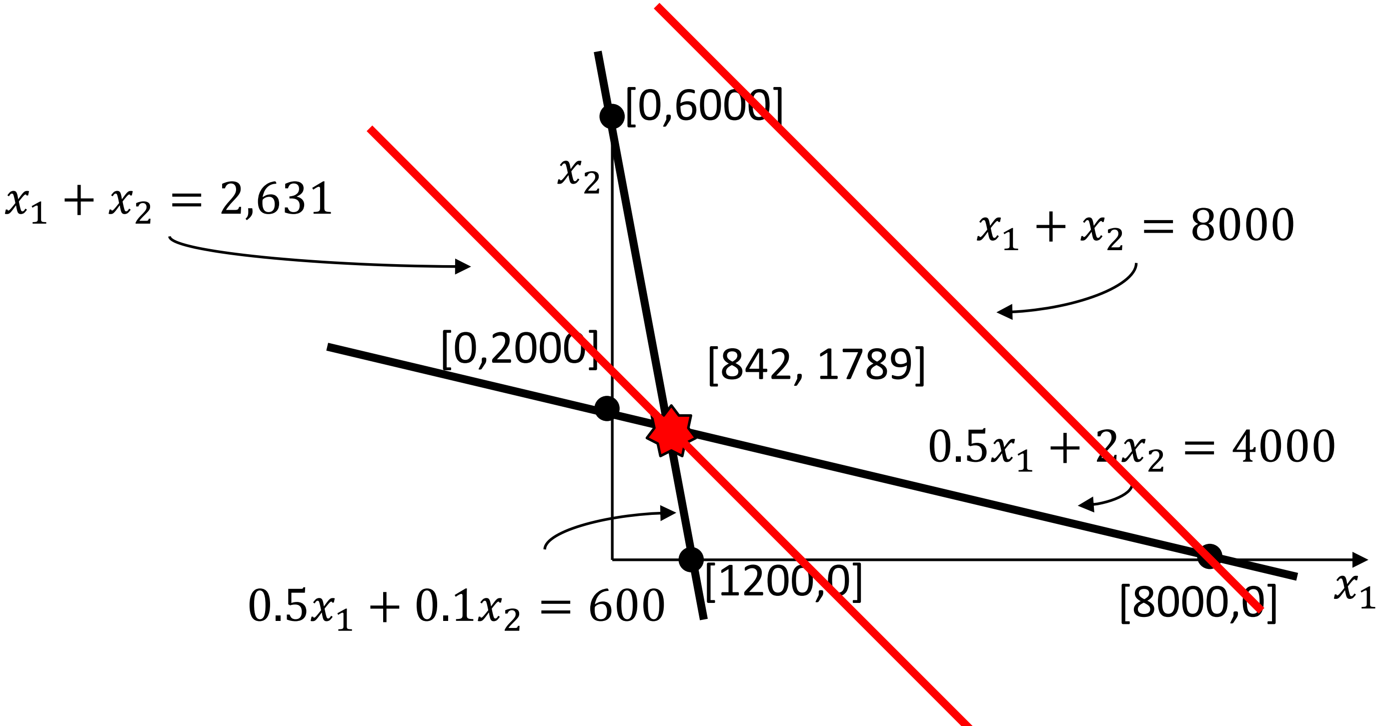
Introduction

Example1：

$$0.5x_1 + 2x_2 \leq 4000$$
$$0.5x_1 + 0.1x_2 \leq 600$$

目標函數就是最大化利潤

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2\}$$



公司開發BOM採購預算上限是4000萬，組裝預算上限為600萬。

A筆電賣出$x_1 = 842$台，B筆電賣出$x_2 = 1789$台會有最大利潤

$0.2x_1 + 0.2x_2 = 526.2$萬

Introduction

- Example2 :

假設銀行每個月存款利率是r, 我們每個月固定將錢存進銀行存n個月，儲存的錢最多不超過D元，如何在n個月之後，最大化儲存的金錢。

$$\max f(x) = (1+r)^n x_1 + (1+r)^{n-1} x_2 + \cdots + (1+r) x_n$$

$$subject\ to \quad x_1 + x_2 + \cdots + x_n \le D$$

$$x_1, x_2, \ldots, x_n \ge 0$$

答案: 把錢D全部存在第一個月。

Introduction

- Example2 :

假設銀行每個月存款利率是*0.1*, 我們每個月固定將錢存進銀行存2個月，儲存的錢最多不超過*100*元，如何在2個月之後，最大化儲存的金錢。

$$\begin{cases} \max f(x) = (1+0.1)^2 x_1 + (1+0.1)^1 x_2 \\ subject\ to \quad x_1 + x_2 \leq 100 \\ \qquad\qquad x_1, x_2 \geq 0 \end{cases} \Rightarrow \begin{cases} min -f(x) = -1.21x_1 - 1.1x_2 \\ subject\ to \quad -x_1 - x_2 + 100 \geq 0 \\ \qquad\qquad x_1, x_2 \geq 0 \end{cases}$$

$$L = -1.21x_1 - 1.1x_2 + \alpha(-x_1 - x_2 + 100) + \beta_1 x_1 + \beta_2 x_2$$

$$\frac{\partial L}{\partial x_1} = -1.21 - \alpha + \beta_1 = 0 \Rightarrow \alpha = 1.21 - \beta_1$$

$$\frac{\partial L}{\partial x_2} = -1.1 - \alpha + \beta_2 = 0 \Rightarrow \alpha = 1.1 - \beta_2$$

$$\frac{\partial L}{\partial \alpha} = x_1 + x_2 = 100 \Rightarrow x_1 = 100 - x_2$$

$$\frac{\partial L}{\partial \beta_1} = x_1 = 0, \frac{\partial L}{\partial \beta_2} = x_2 = 0$$

$$x_1 = 0 \Rightarrow x_2 = 100, f(x) = 1.21 * 0 + 1.1 * 100 = 110$$
$$x_2 = 0 \Rightarrow x_1 = 100, f(x) = 1.21 * 100 + 1.1 * 0 = 121$$

Introduction

Example1 :

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2\}$$
$$subject\ to$$
$$0.5x_1 + 2x_2 \leq 4000$$
$$0.5x_1 + 0.1x_2 \leq 600$$

Lagrange Dual Problem

$$L(\boldsymbol{x}, \alpha_1, \alpha_2) = -0.2x_1 - 0.2x_2 + \alpha_1(-0.5x_1 - 2x_2 + 4000) + \alpha_2(-0.5x_1 - 0.1x_2 + 600)$$

$$\frac{\partial}{\partial x_1} L(\boldsymbol{x}, \alpha_1, \alpha_2) = -0.2 + 0.5\alpha_1 + 0.5\alpha_2 = 0$$
$$\frac{\partial}{\partial x_2} L(\boldsymbol{x}, \alpha_1, \alpha_2) = -0.2 + 2\alpha_1 + 0.1\alpha_2 = 0$$
$$\frac{\partial}{\partial \alpha_1} L(\boldsymbol{x}, \alpha_1, \alpha_2) = 0.5x_1 + 2x_2 - 4000 = 0$$
$$\frac{\partial}{\partial \alpha_2} L(\boldsymbol{x}, \alpha_1, \alpha_2) = 0.5x_1 + 0.1x_2 - 600 = 0$$

$$0.5\alpha_1 + 0.5\alpha_2 = 0.2$$
$$2\alpha_1 + 0.1\alpha_2 = 0.2$$
$$0.5x_1 + 2x_2 = 4000$$
$$0.5x_1 + 0.1x_2 = 600$$

$$0.5\alpha_1 = 0.8 \Rightarrow \alpha_1 = 1.6$$
$$1.9\alpha_2 = 0.6 \Rightarrow \alpha_2 = 0.315789$$
$$\begin{cases} x_1 = 842.11 \\ x_2 = 1789.47 \end{cases}$$

Introduction

Example3:

科技公司預計下半年推出A和B兩款筆電，企劃規劃

- A筆電BOM cost是5000元，組裝成本是5000元。
- B電競筆電BOM cost是20000元，組裝成本是1000元。
- C電競筆電BOM cost是30000元，組裝成本是4000元。

公司開發BOM採購預算上限是4000萬，組裝、行銷預算上限為600萬。

但A、B和C賣出去，每台獲利皆是2000元，假設推出的筆電都能賣出去，請問A、B和C各需要賣出幾台才能獲得最大利潤?

Introduction

	BOM	組裝
A	0.5	0.5
B	2	0.1
C	3	0.4
總成本	4000	600

Example3 : (單位:萬元)

假設A筆電賣出x_1台，B筆電賣出x_2台，C筆電賣出x_3台

依照題目

$$0.5x_1 + 2x_2 + 3x_3 \leq 4000$$
$$0.5x_1 + 0.1x_2 + 0.4x_3 \leq 600$$

目標函數就是最大化利潤

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2 + 0.2x_3\}$$

畫圖解?

Introduction

Example3 :

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2 + 0.2x_3\}$$
$$subject\ to$$
$$0.5x_1 + 2x_2 + 3x_3 \leq 4000$$
$$0.5x_1 + 0.1x_2 + 0.4x_3 \leq 600$$

Lagrange Dual Problem

$$L(\boldsymbol{x}, \alpha_1, \alpha_2) = -0.2x_1 - 0.2x_2 - 0.2x_3 + \alpha_1\,(-0.5x_1 - 2x_2 - 3x_3 + 4000) + \alpha_2(-0.5x_1 - 0.1x_2 - 0.4x_3 + 600)$$

$$\frac{\partial}{\partial x_1} L(\boldsymbol{x}, \alpha_1, \alpha_2) = -0.2 + 0.5\alpha_1 + 0.5\alpha_2 = 0$$

$$\frac{\partial}{\partial x_2} L(\boldsymbol{x}, \alpha_1, \alpha_2) = -0.2 + 2\alpha_1 + 0.1\alpha_2 = 0$$

$$\frac{\partial}{\partial x_2} L(\boldsymbol{x}, \alpha_1, \alpha_2) = -0.2 + 3\alpha_1 + 0.4\alpha_2 = 0$$

$$\frac{\partial}{\partial \alpha_1} L(\boldsymbol{x}, \alpha_1, \alpha_2) = 0.5x_1 + 2x_2 + 3x_3 - 4000 = 0$$

$$\frac{\partial}{\partial \alpha_2} L(\boldsymbol{x}, \alpha_1, \alpha_2) = 0.5x_1 + 0.1x_2 + 0.4x_3 - 600 = 0$$

$$0.5\alpha_1 + 0.5\alpha_2 = 0.2$$
$$2\alpha_1 + 0.1\alpha_2 = 0.2$$
$$3\alpha_1 + 0.4\alpha_2 = 0.2$$

?

$$0.5x_1 + 2x_2 + 3x_3 - 4000 = 0$$
$$0.5x_1 + 0.1x_2 + 0.4x_3 - 600 = 0$$

Introduction

Example3 :

$$\max\{f(\boldsymbol{x}) = 0.2x_1 + 0.2x_2 + 0.2x_3\}$$
$$\textit{subject to}$$
$$0.5x_1 + 2x_2 + 3x_3 \leq 4000$$
$$0.5x_1 + 0.1x_2 + 0.4x_3 \leq 600$$

Lagrange Dual Problem

$$L(\boldsymbol{x}, \alpha_1, \boldsymbol{\alpha_2}) = -0.2x_1 - 0.2x_2 - 0.2x_3 + \ \alpha_1\,(-0.5x_1 - 2x_2 - 3x_3 + 4000) + \alpha_2(-0.5x_1 - 0.1x_2 - 0.4x_3 + 600)$$

$\alpha_1, \boldsymbol{\alpha_2}$給定值，假設都是1

$$L(\boldsymbol{x}\ ) = -1.2x_1 - 2.3x_2 - 3.6x_3 + 4600$$

梯度下降法請看範例程式

$$\frac{\partial}{\partial x} L(x\ ) = \begin{bmatrix} -1.2 \\ -2.3 \\ -3.6 \end{bmatrix}$$

Example3 :

$$\min\{-f(\boldsymbol{x}) = -0.2x_1 - 0.2x_2 - 0.2x_3\}$$

$subject\ to$

$$0.5x_1 + 2x_2 + 3x_3 \le 4000$$

$$0.5x_1 + 0.1x_2 + 0.4x_3 \le 600$$

----
iter:1
x:[0 0 0]
f(x):-0.0
con1(x):4000.0
con2(x):600.0
----
iter:501
x:[60. 115. 180.]
f(x):-71.00000000000037
con1(x):3199.9999999999936
con2(x):486.4999999999995
----
iter:1501
x:[180. 345. 540.]
f(x):-213.00000000000423
con1(x):1599.9999999999482
con2(x):259.49999999999176
----
iter:2001
x:[240. 460. 720.]
f(x):-284.00000000000784
con1(x):799.9999999999081
con2(x):145.99999999998698

---------梯度找到的最佳解-----
number of devices:[300. 575. 900.]
$f(\boldsymbol{x})$ = **355.0000000000115**
4000.0000000001314
567.5000000000177

--------調整梯度找的最佳解------
number of devices:[299. 574. 899.]
$f(\boldsymbol{x})$ = 354.4000000000115
3994.5000000001314
566.5000000000177
--------調整梯度找的最佳解------
number of devices:[301. 576. 901.]
$f(\boldsymbol{x})$ = 355.6000000000115
4005.5000000001314
568.5000000000177

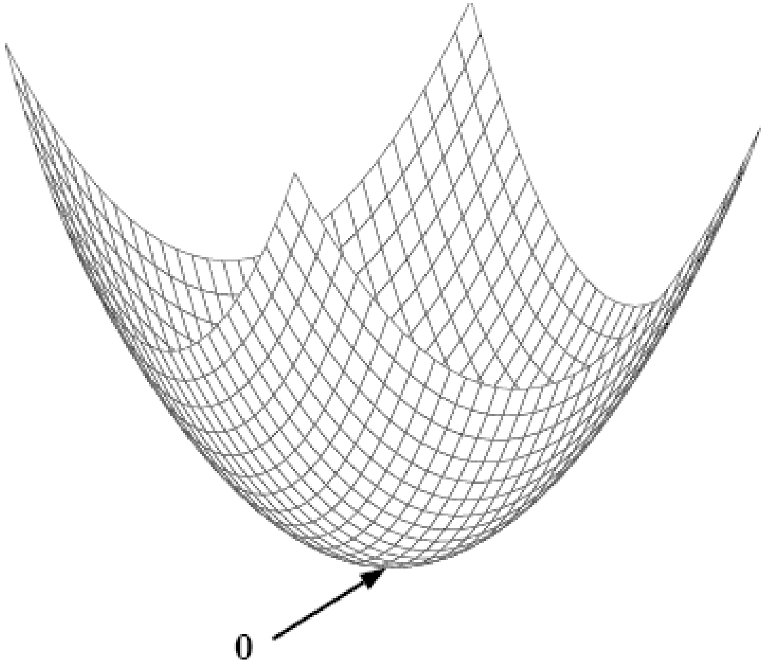
Quadratic function

$$f(\boldsymbol{x}) = \boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} + c$$

Q: Symmetric

If Q is positive definite, then f is parabolic “bowl”.

- Quadratics are useful in the study in Optimization.



$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in R^{n*1}, b \in R^{n*1}, c \in R$$

$$Q = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix} \in R^{n*n}$$

Derivatives

$$f: R \to R$$

The derivative of f is a function $f': R \to R$

$$f'(x) = \frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$f: R^n \to R$$

The gradient of f is a function, $\nabla f: R^n \to R^n$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

Jacobian

$$f: R^n \to R$$

The gradient of f is a function,
$\nabla f: R^n \to R^n$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

函數的微分

The derivative of $\boldsymbol{f}$ is a function

$$D\boldsymbol{f}: R^n \to R^{m*n}$$

$$\boldsymbol{f} = [f_1 \quad \dots \quad f_m]^T: R^n \to R^m$$

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

Jacobian Matrix

函數組成的向量的微分

$$f: R^n \to R \Rightarrow \nabla f(\boldsymbol{x}) = D\boldsymbol{f}(\boldsymbol{x})^T$$

Hessian

如果∇f可微分存在則稱f為twice differentiable。

The second derivative of f is Hessian of f :

$$F = D^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

Example

EX1:

$$f: R \to R \Rightarrow f(x) = x^2 + x + 1 \Rightarrow f'(x) = 2x + 1$$

EX2:

$$f: R^n \to R \Rightarrow f(\boldsymbol{x}) = x_1^2 + 2x_2^2 + 3x_1 + 4x_2 + 1$$

$$\Rightarrow f(\boldsymbol{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 1$$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + 3 \\ 4x_2 + 4 \end{bmatrix}$$

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \partial \dfrac{\frac{\partial f(\boldsymbol{x})}{\partial x_1}}{\partial x_1} & \partial \dfrac{\frac{\partial f(\boldsymbol{x})}{\partial x_1}}{\partial x_2} \\ \partial \dfrac{\frac{\partial f(\boldsymbol{x})}{\partial x_2}}{\partial x_1} & \partial \dfrac{\frac{\partial f(\boldsymbol{x})}{\partial x_2}}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \partial \dfrac{2x_1 + 3}{\partial x_1} & \partial \dfrac{2x_1 + 3}{\partial x_2} \\ \partial \dfrac{4x_2 + 4}{\partial x_1} & \partial \dfrac{4x_2 + 4}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

Example

EX2:

$$f: R^n \to R \Rightarrow f(\boldsymbol{x}) = x_1^2 + 2x_2^2 + 3x_1 + 4x_2 + 1$$

$$\Rightarrow f(\boldsymbol{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 1$$

$$f(\boldsymbol{x}) = \boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} + c$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, Q = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$\nabla f(\boldsymbol{x}) = 2Q\boldsymbol{x} + \boldsymbol{b} = 2 \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3 \\ 4x_2 + 4 \end{bmatrix}$$

$$D\boldsymbol{f}(\boldsymbol{x}) = 2Q = 2 \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

Chain rule

$$f: R^n \longrightarrow R$$

$$g: R \longrightarrow R^n$$

$$F: R \longrightarrow R, F(t) = f\big(g(t)\big)$$

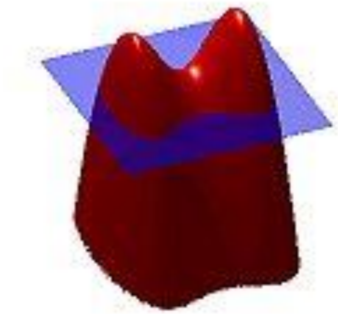
$$F'^{(t)} = Df\big(g(t)\big) * Dg(t)$$

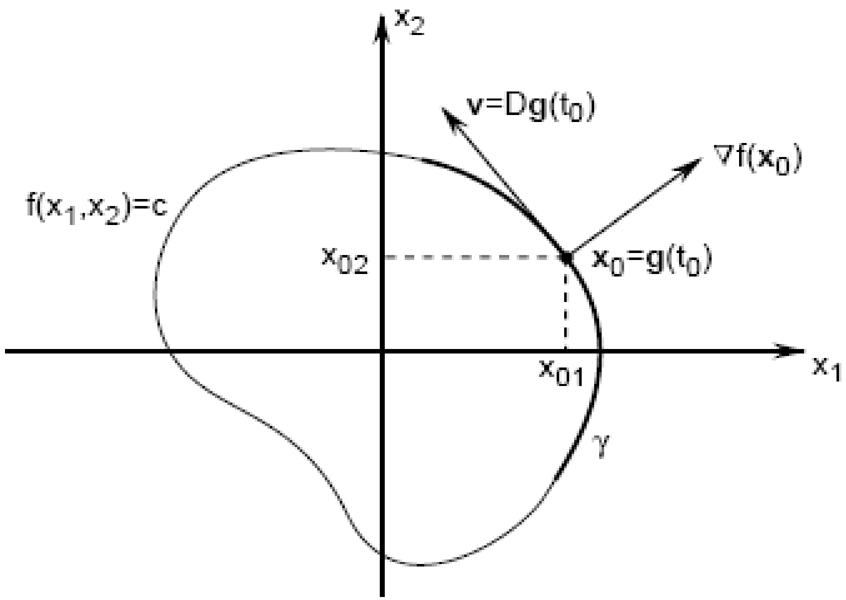
$$= \nabla f\big(g(t)\big)^{\mathrm{T}} g'(t)$$

$$= g'(t)^T \nabla f\big(g(t)\big)$$

Gradient and level sets

Given $f: R^n \longrightarrow R, \nabla f(x_0)$ is orthogonal to the level set at x_0





$g(t)$ *is the position of the particle at time* t, $g(0) = x_0$

$f(g(t))$: constant for all t

$$F(t) = f\big(g(t)\big)$$

$$F'(0) = 0$$

$$\Rightarrow F'^{(0)} = g'(0)^T \nabla f\big(g(0)\big) = 0$$

$g'(0)$和$\nabla f(x_0)$: *orthogonal*

https://zh.m.wikipedia.org/zh-tw/%E6%B0%B4%E5%B9%B3%E9%9B%86%E6%96%B9%E6%B3%95

Taylor's formula

$f: R \longrightarrow R$ is in C^1,

is a term such that $o(h)/h \to 0$ as $h \to 0$

$f: R \longrightarrow R$ is in C^2,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + o((x - x_0)^2)$$

Taylor's formula

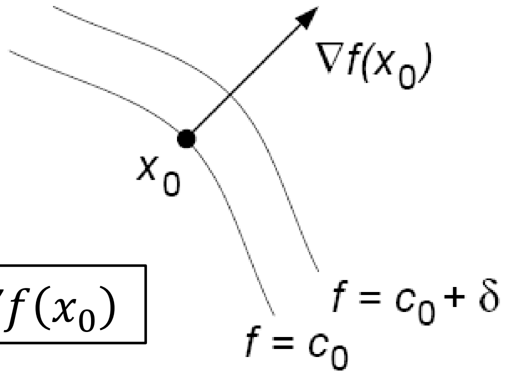
$f\colon R^n \longrightarrow R$ is in C^1,

is a term such that $o(h)/h \to 0$ as $h \to 0$

$f\colon R^n \longrightarrow R$ is in C^2,

$$
\begin{aligned}
& f(x) \\
& = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T F(x_0)(x - x_0) \\
& + o(\|x - x_0\|^2)
\end{aligned}
$$

Taylor’s formula vs Gradient



$$x_\alpha = x_0 + \alpha\nabla f(x_0), \alpha > 0$$

$$\boxed{x_\alpha - x_0 = x_0 + \alpha\nabla f(x_0) - x_0 = \alpha\nabla f(x_0)}$$

$$f(x_\alpha) = f(x_0) + \nabla f(x_0)^T \boxed{(x_\alpha - x_0)} + o(\|x_\alpha - x_0\|)$$

$$f(x_\alpha) = f(x_0) + \alpha\nabla\|f(x_0)\|^2 + o(\alpha)$$

$$\Rightarrow f(x_\alpha) > f(x_0)$$

所以從Taylor’s formula 可以證明如果要找最小值，必須往梯度的反方向走。

Gradient Descent vs Newton's Method

Taylor's formula 可以證明如果要找最小值，必須往梯度的反方向走

單變量

$$x_{t+1} = x_t - \alpha f'(x_t), \alpha > 0$$

$$x_{t+1} = x_t - \frac{f'(x_t)}{f''(x_t)}$$

多變量

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \alpha \nabla f(\boldsymbol{x}_t), \alpha > 0$$

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \nabla f(\boldsymbol{x}_t) F^{-1}$$

$$\boldsymbol{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_d \end{bmatrix}$$

下周梯度下降法，我們在介紹使用起來的差異。

Lagrange Dual Problem

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}), \boldsymbol{x} \in \boldsymbol{R}^d$$
$$s.t. \quad f_i(\boldsymbol{x}), i = 1,2,\dots,m$$

$\downarrow$ Lagrange Dual Problem

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x})$$

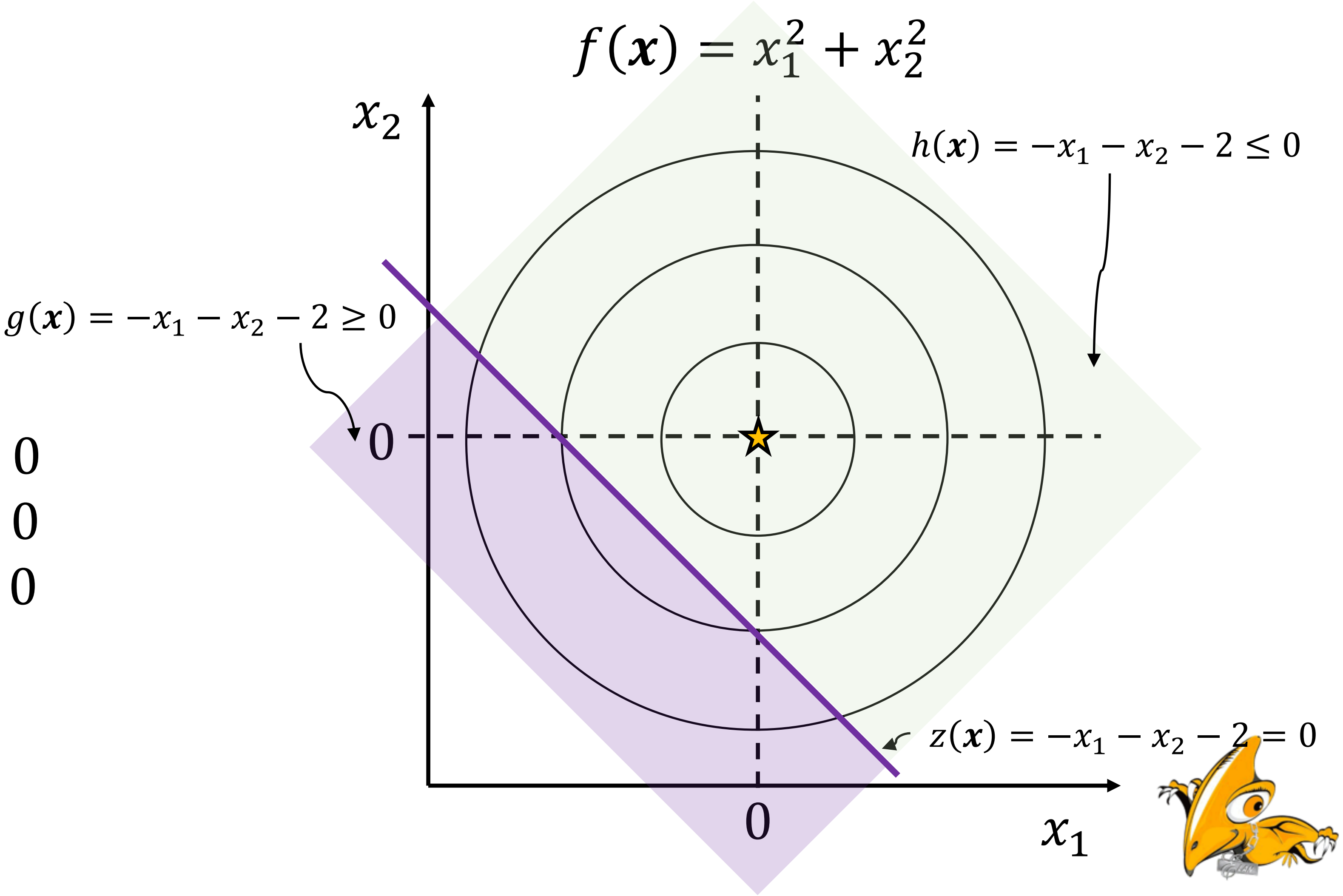
Optimal problem with Constraint

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x})$$

$$s.t. \quad \begin{aligned} g(\boldsymbol{x}) &\geq 0 \\ h(\boldsymbol{x}) &\leq 0 \\ z(\boldsymbol{x}) &= 0 \end{aligned}$$

EX:

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$
$$h(\boldsymbol{x}) = -x_1 - x_2 - 2 \leq 0$$
$$z(\boldsymbol{x}) = -x_1 - x_2 - 2 = 0$$



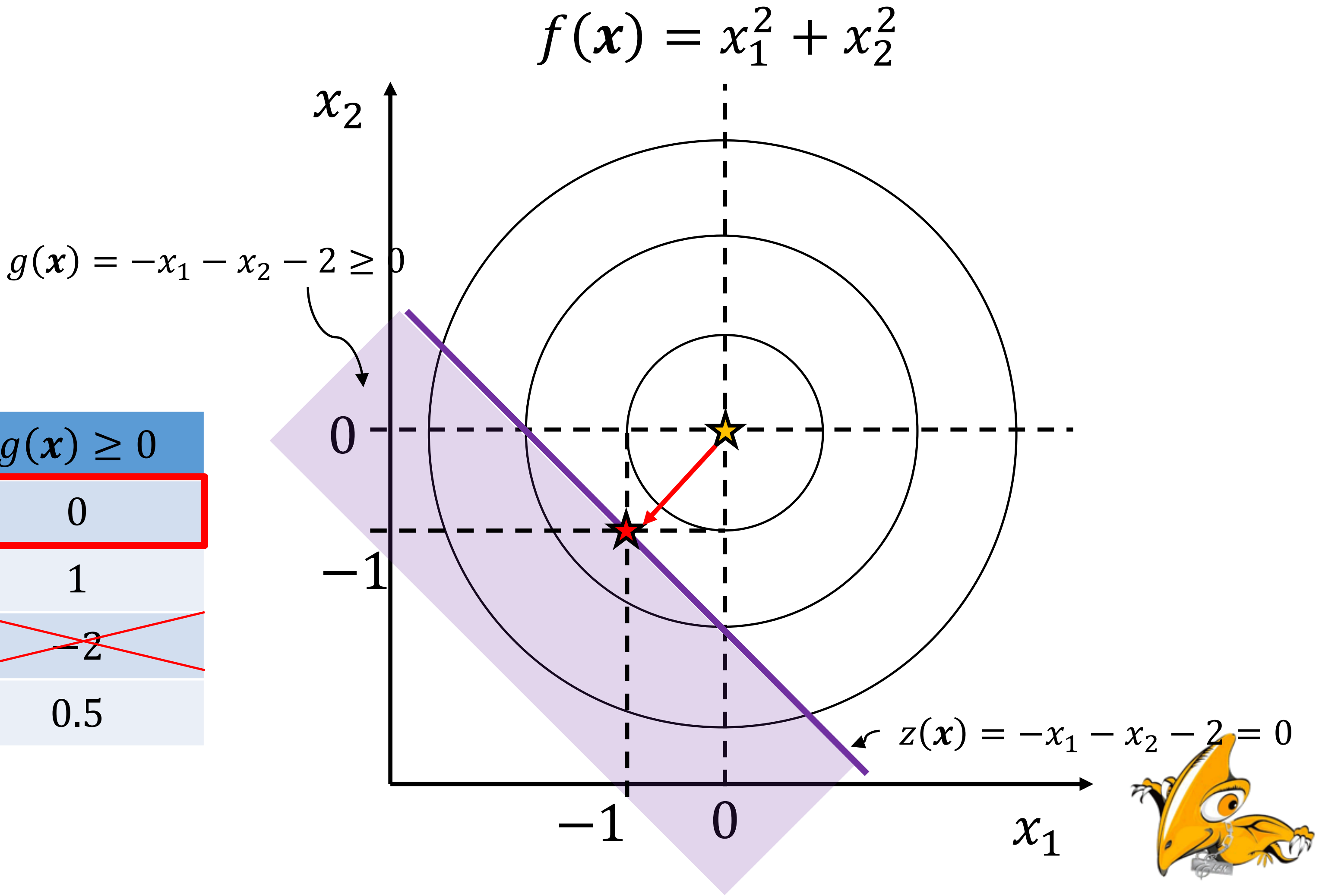
Optimal problem with Constraint

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x})$$
$$s.t. \quad g(\boldsymbol{x}) \geq 0$$

EX:

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$

x_1	x_2	$f(\boldsymbol{x})$	$g(\boldsymbol{x}) \geq 0$
-1	-1	2	0
-1.5	-1.5	4.5	1
0	0	0	$\cancel{-2}$
-1.25	-1.25	3.125	0.5



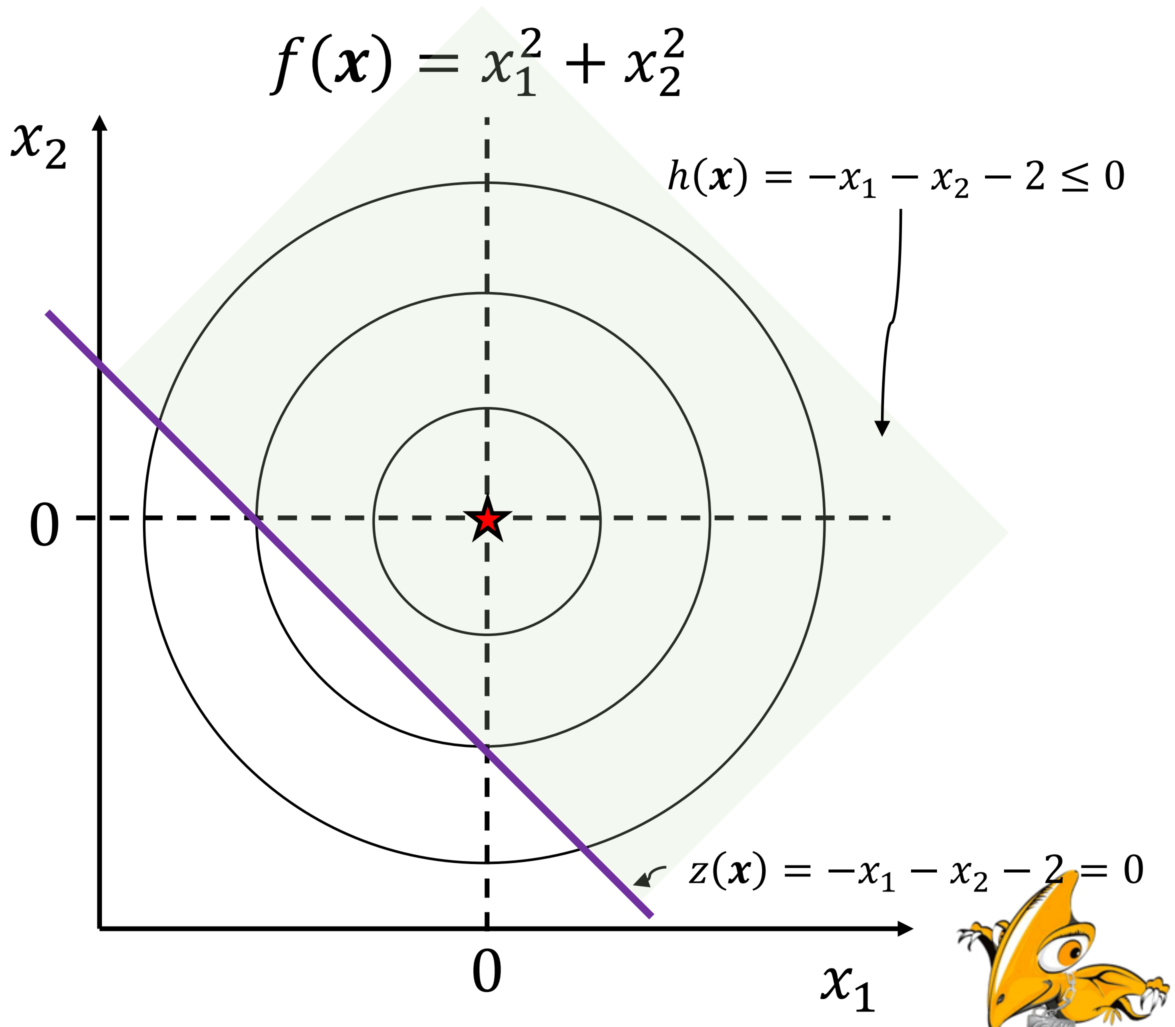
Optimal problem with Constraint

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x})$$
$$s.t. \quad h(\boldsymbol{x}) \le 0$$

EX:

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$h(\boldsymbol{x}) = -x_1 - x_2 - 2 \le 0$$

x_1	x_2	$f(\boldsymbol{x})$	$h(\boldsymbol{x}) \le 0$
0	0	0	-2
-1	-1	2	0
0.5	0.5	0.5	-3
-1.25	-1.25	3.125	0.5



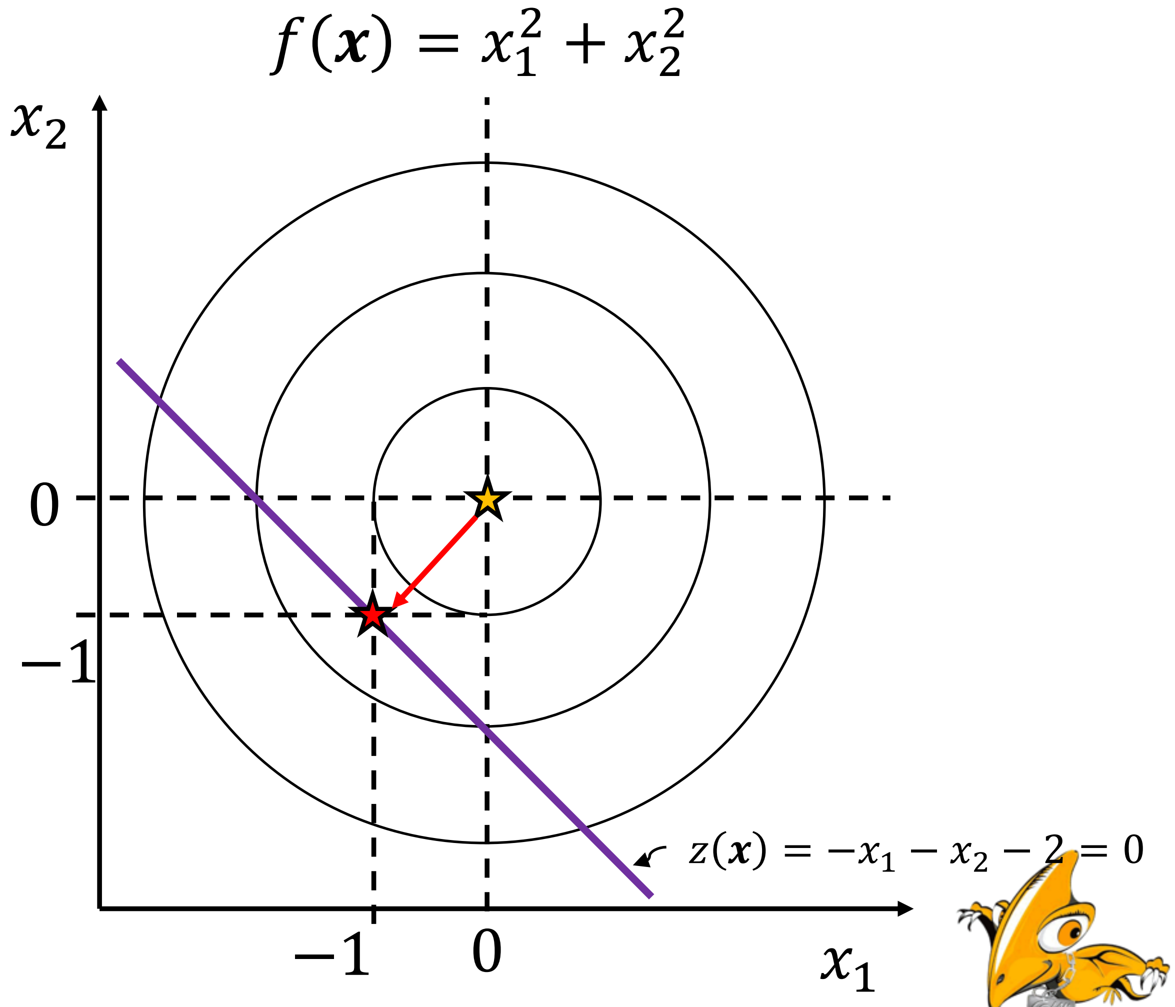
Optimal problem with Constraint

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x})$$
$$s.t. \quad z(\boldsymbol{x}) = 0$$

EX:

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$z(\boldsymbol{x}) = -x_1 - x_2 - 2 = 0$$

x_1	x_2	$f(\boldsymbol{x})$	$\mathbf{z}(\boldsymbol{x}) = 0$
-1	-1	2	0
-1.5	-1.5	4.5	1
0	0	0	$\cancel{-2}$
-1.25	-1.25	3.125	0.5



Lagrange Dual Problem

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ s.t. & g(\boldsymbol{x}) \geq 0 \end{array} \Longrightarrow L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

Example:

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$
$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x}) = x_1^2 + x_2^2 + \lambda(-x_1 - x_2 - 2)$$

$$\boxed{\begin{cases} \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \dfrac{\partial L(\boldsymbol{x}, \lambda)}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial L(\boldsymbol{x}, \lambda)}{\partial x_1} \\ \dfrac{\partial L(\boldsymbol{x}, \lambda)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \lambda/2 \\ \lambda/2 \end{bmatrix} \\ \nabla_{\lambda} f(\boldsymbol{x}) = \dfrac{\partial L(\boldsymbol{x}, \lambda)}{\partial \lambda} = -x_1 - x_2 - 2 = 0 \Rightarrow x_1 = -x_2 - 2 \end{cases}}$$

微分等於0

$$x_2 = \lambda/2 = x_1 = -x_2 - 2 \Rightarrow 2x_2 = -2 \Rightarrow x_2 = -1$$
$$x_2 = -1 = \lambda/2 \Rightarrow \lambda = -2$$

ANS: $x_1 = x_2 = -1, \lambda = -2$

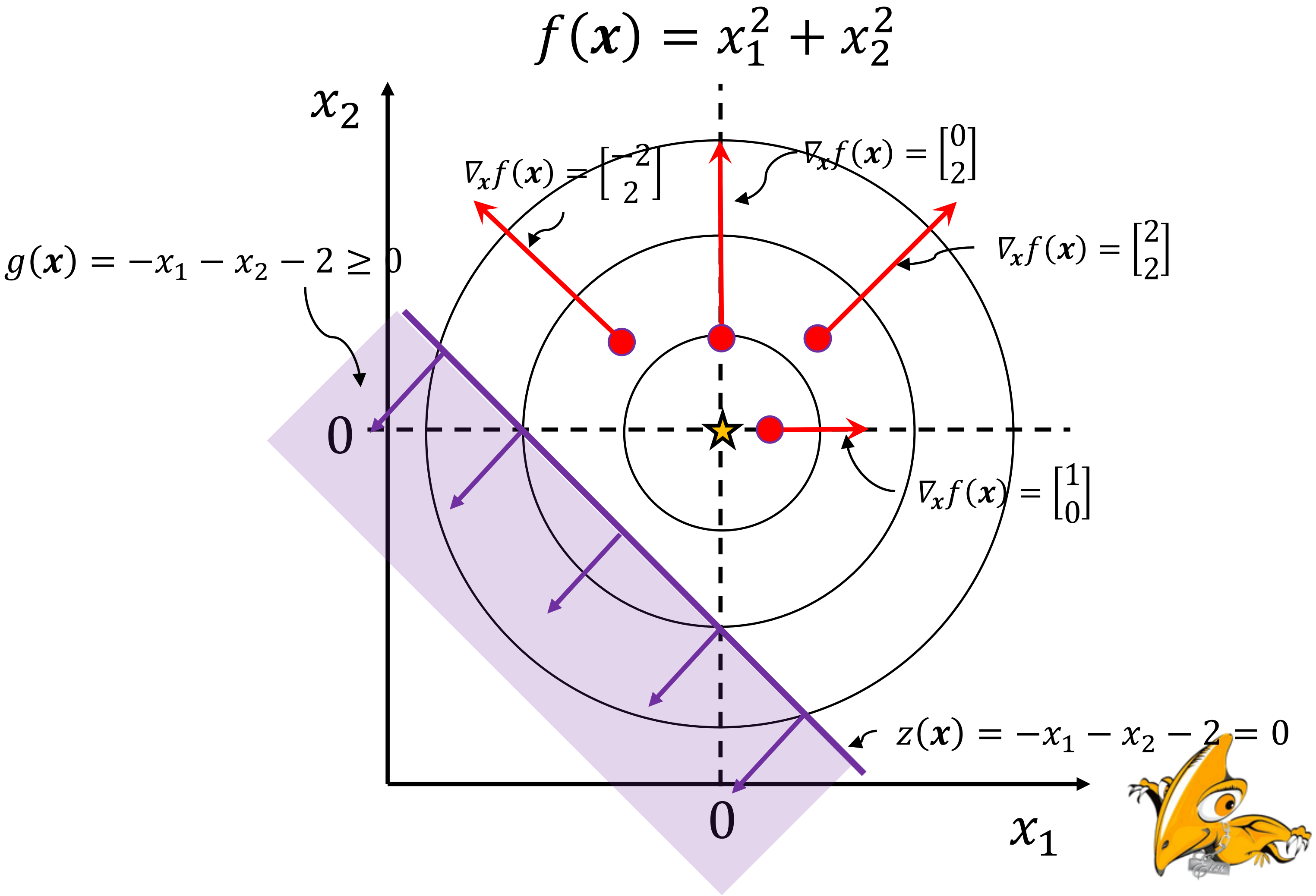
Optimal problem with Constraint

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = x_1^2 + x_2^2$$

$$s.t. \quad g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$

$$\nabla_x f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$\nabla_x g(\boldsymbol{x}) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$



假設 $\boldsymbol{x}^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \nabla_x f\left(\boldsymbol{x}^{(0)}\right) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$

假設 $\boldsymbol{x}^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \nabla_x f\left(\boldsymbol{x}^{(0)}\right) = \begin{bmatrix} 2 \\ 2 \end{bmatrix},$

假設 $\boldsymbol{x}^{(0)} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}, \nabla_x f\left(\boldsymbol{x}^{(0)}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

假設 $\boldsymbol{x}^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \nabla_x f\left(\boldsymbol{x}^{(0)}\right) = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$

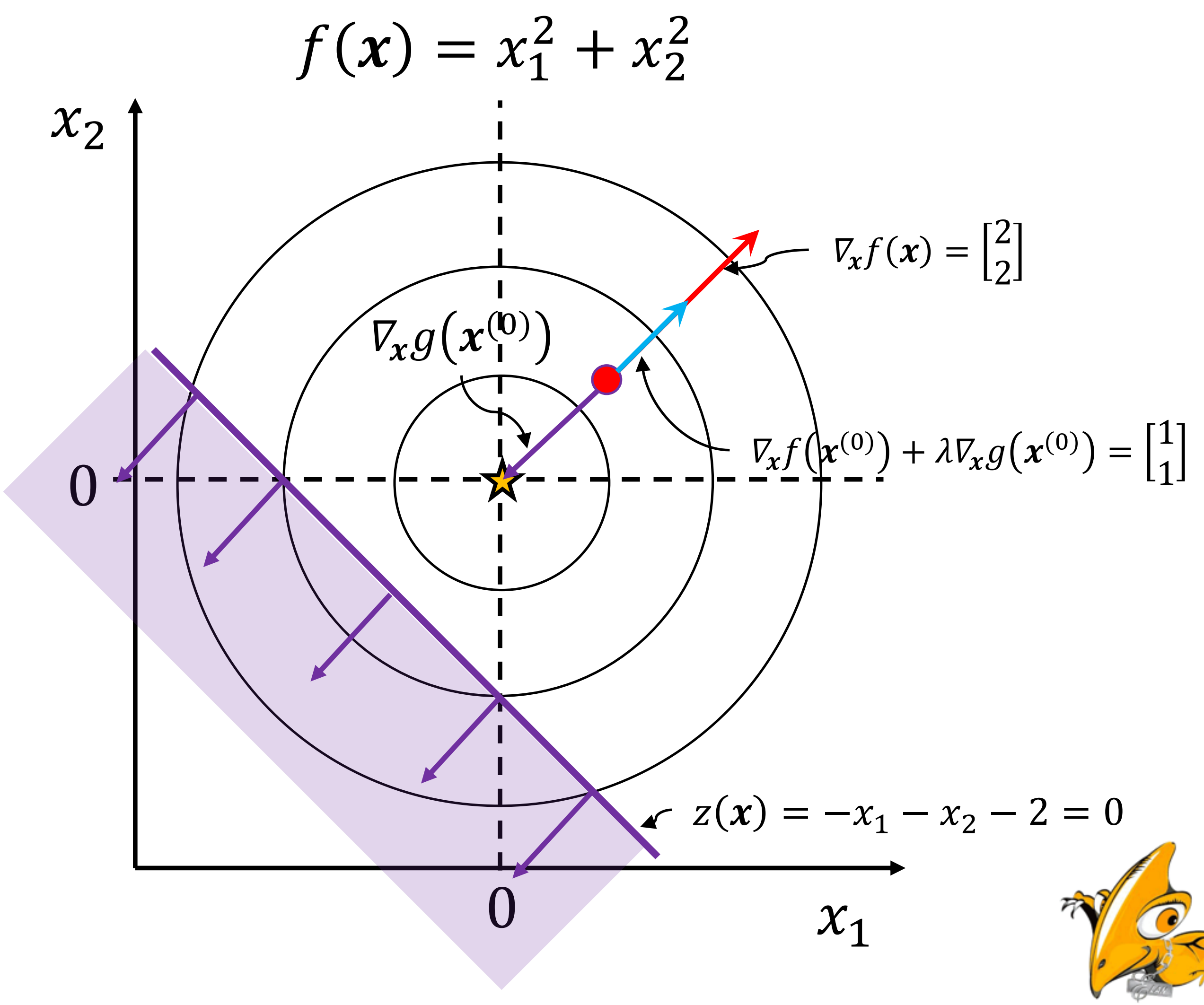
Gradient Descent in Lagrange Dual Problem

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$s.t. \quad g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$
$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}, \nabla_{\boldsymbol{x}} g(\boldsymbol{x}) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$
$$= x_1^2 + x_2^2 + \lambda(-x_1 - x_2 - 2)$$
$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) + \lambda \nabla_{\boldsymbol{x}} g(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} + \begin{bmatrix} -\lambda \\ -\lambda \end{bmatrix}$$

假設 $\boldsymbol{x}^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \lambda = 1$

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(0)}) + \lambda \nabla_{\boldsymbol{x}} g(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$



Optimal problem with Constraint

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = x_1^2 + x_2^2$$

$$s.t. \quad g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

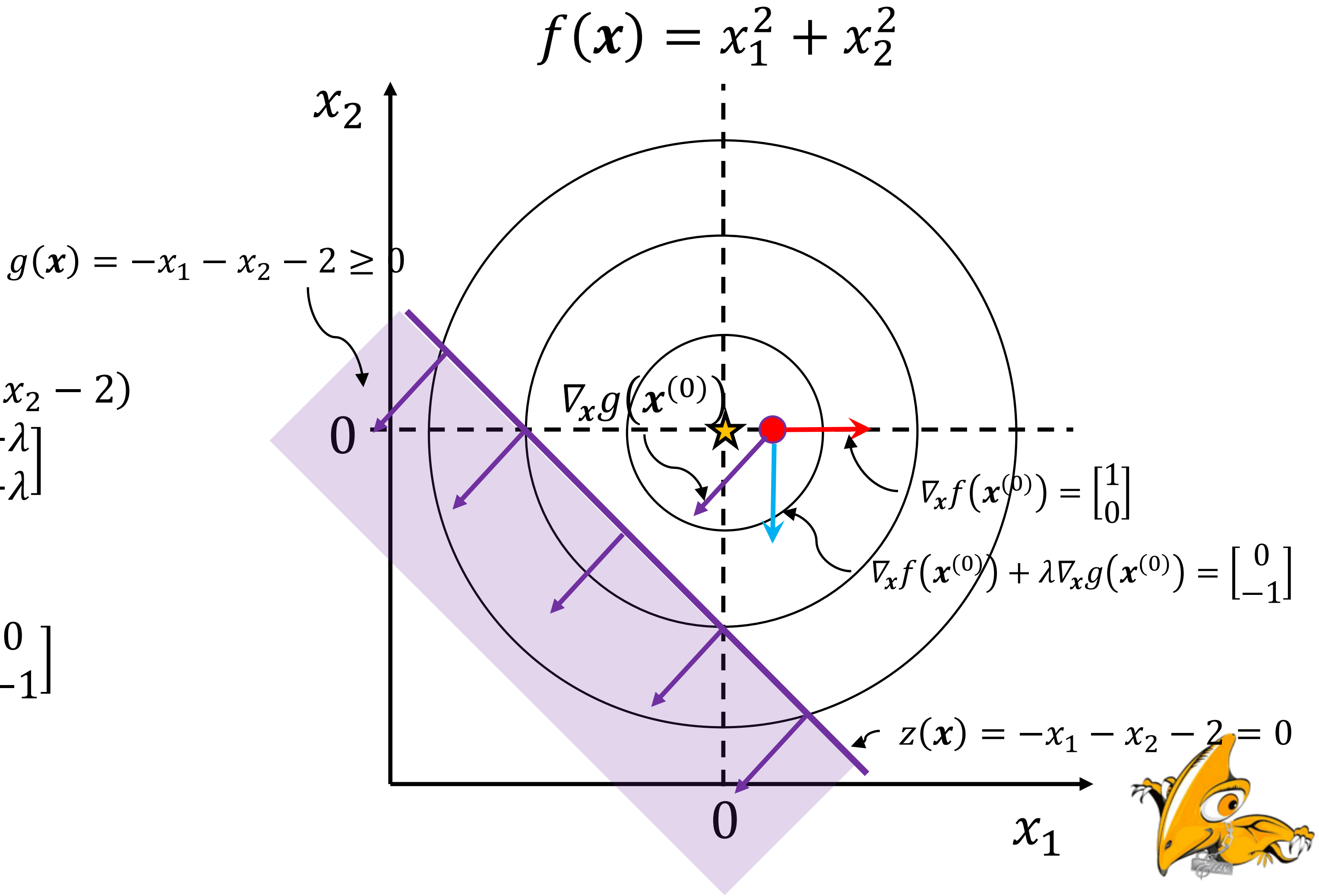
$$\nabla_{\boldsymbol{x}} g(\boldsymbol{x}) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x}) = x_1^2 + x_2^2 + \lambda(-x_1 - x_2 - 2)$$

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) + \lambda \nabla_{\boldsymbol{x}} g(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} + \begin{bmatrix} -\lambda \\ -\lambda \end{bmatrix}$$

假設 $\boldsymbol{x}^{(0)} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$,$\lambda = 1$

$$\nabla_{\boldsymbol{x}} f\left(\boldsymbol{x}^{(0)}\right) + \lambda \nabla_{\boldsymbol{x}} g\left(\boldsymbol{x}^{(0)}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$



Lagrange Dual Problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g(\boldsymbol{x}) \geq 0 \end{aligned} \Rightarrow L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

EX:

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$

ANS: $x_1 = x_2 = -1, \lambda = -2$

In deep learning, λ: parameter for **weight decay**. 這個參數是不可以設定為負數的。

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g(\boldsymbol{x}) \geq 0 \Rightarrow -g(\boldsymbol{x}) \boxed{\leq 0} \end{aligned} \Rightarrow L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda(-g(\boldsymbol{x}))$$

Lagrange Dual Problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g(\boldsymbol{x}) \geq 0 \Longrightarrow -g(\boldsymbol{x}) \leq 0 \end{aligned} \Longrightarrow L(\boldsymbol{x},\lambda) = f(\boldsymbol{x}) + \lambda(-g(\boldsymbol{x}))$$

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$

$$g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$

$$L(\boldsymbol{x},\lambda) = f(\boldsymbol{x}) + \lambda(-g(\boldsymbol{x})) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 + 2)$$

$$\begin{cases} \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \dfrac{\partial L(\boldsymbol{x},\lambda)}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial L(\boldsymbol{x},\lambda)}{\partial x_1} \\ \dfrac{\partial L(\boldsymbol{x},\lambda)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + \lambda \\ 2x_2 + \lambda \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} {}^{-\lambda}/_{2} \\ {}^{-\lambda}/_{2} \end{bmatrix} \\ \nabla_{\lambda} f(\boldsymbol{x}) = \dfrac{\partial L(\boldsymbol{x},\lambda)}{\partial \lambda} = x_1 + x_2 + 2 = 0 \Rightarrow x_1 = -x_2 - 2 \end{cases}$$

$$x_2 = {}^{-\lambda}/_{2} = x_1 = -x_2 - 2 \Rightarrow 2x_2 = -2 \Rightarrow x_2 = -1$$

$$x_2 = -1 = {}^{-\lambda}/_{2} \Rightarrow \lambda = 2$$

ANS: $x_1 = x_2 = -1, \lambda = 2$

Lagrange Dual Problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g(\boldsymbol{x}) \geq 0 \end{aligned} \Rightarrow L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

EX:

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$

ANS: $x_1 = x_2 = -1, \lambda = -2$

In deep learning, λ: parameter for **weight decay**. 這個參數是不可以設定為負數的。

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g(\boldsymbol{x}) \geq 0 \Rightarrow -g(\boldsymbol{x}) \leq 0 \end{aligned} \Rightarrow L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda(-g(\boldsymbol{x}))$$

ANS: $x_1 = x_2 = -1, \lambda = 2$

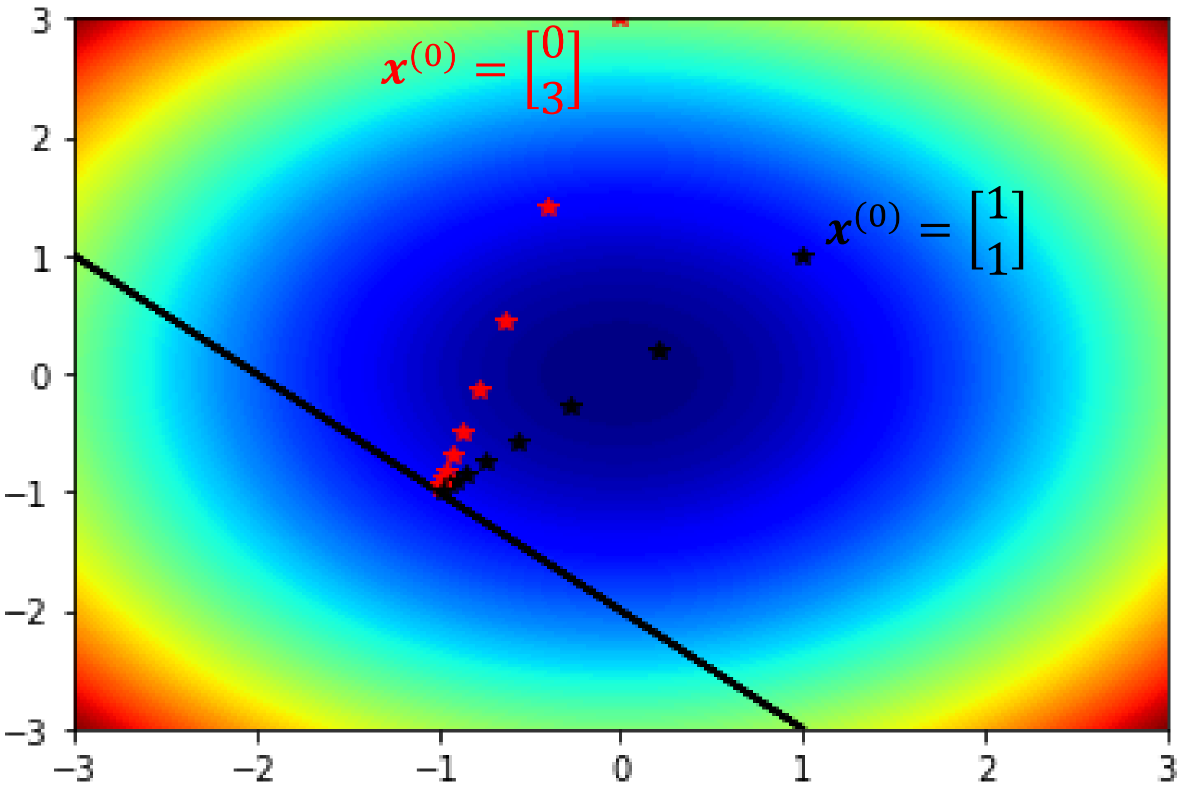
Gradient Descent in Lagrange Dual Problem

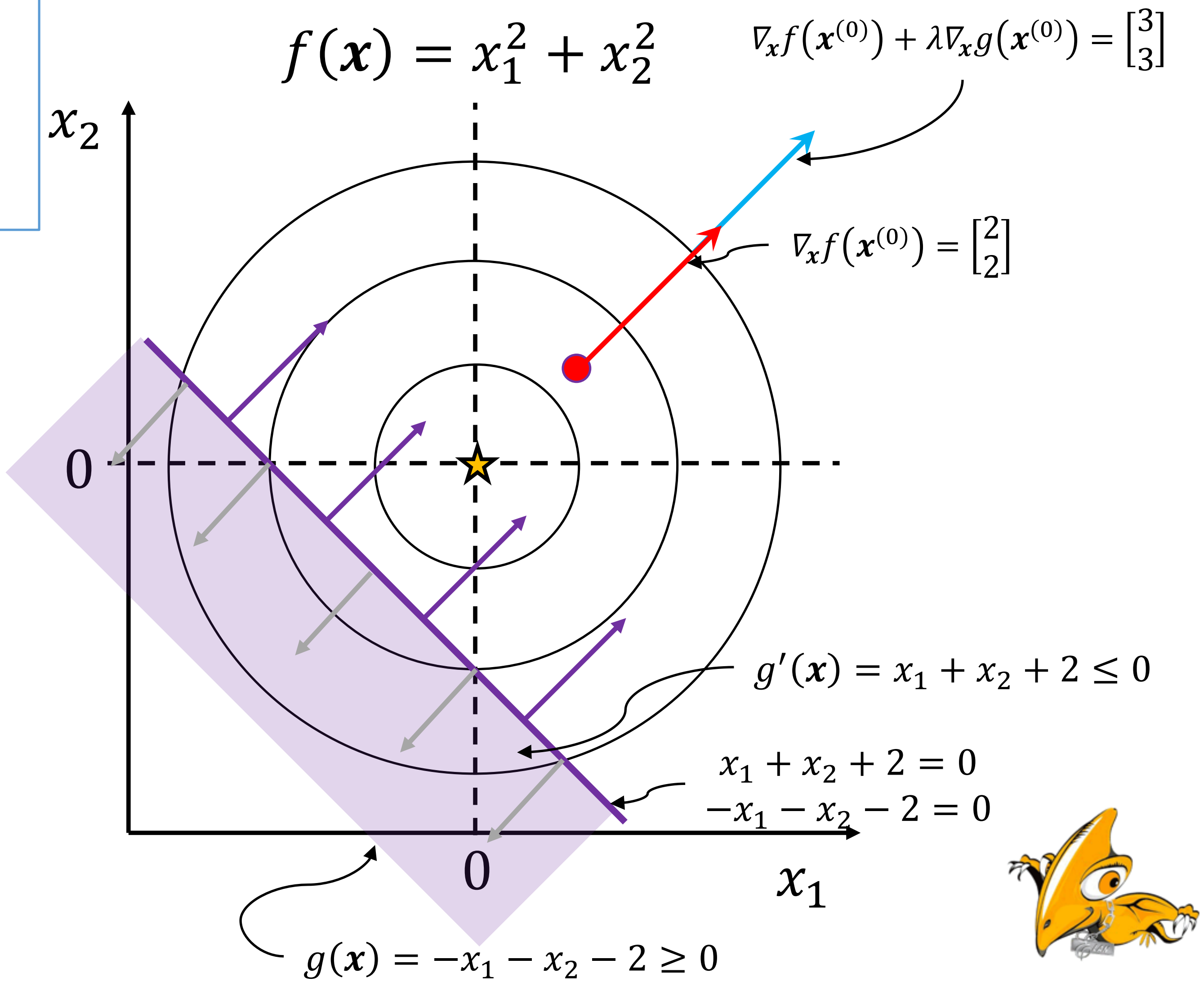
$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = x_1^2 + x_2^2$$

$$s.t. \quad g(\boldsymbol{x}) = -x_1 - x_2 - 2 \geq 0$$

$$\Rightarrow g'(\boldsymbol{x}) = -g(\boldsymbol{x}) = x_1 + x_2 + 2 \leq 0$$

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}, \nabla_{\boldsymbol{x}} g'(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$



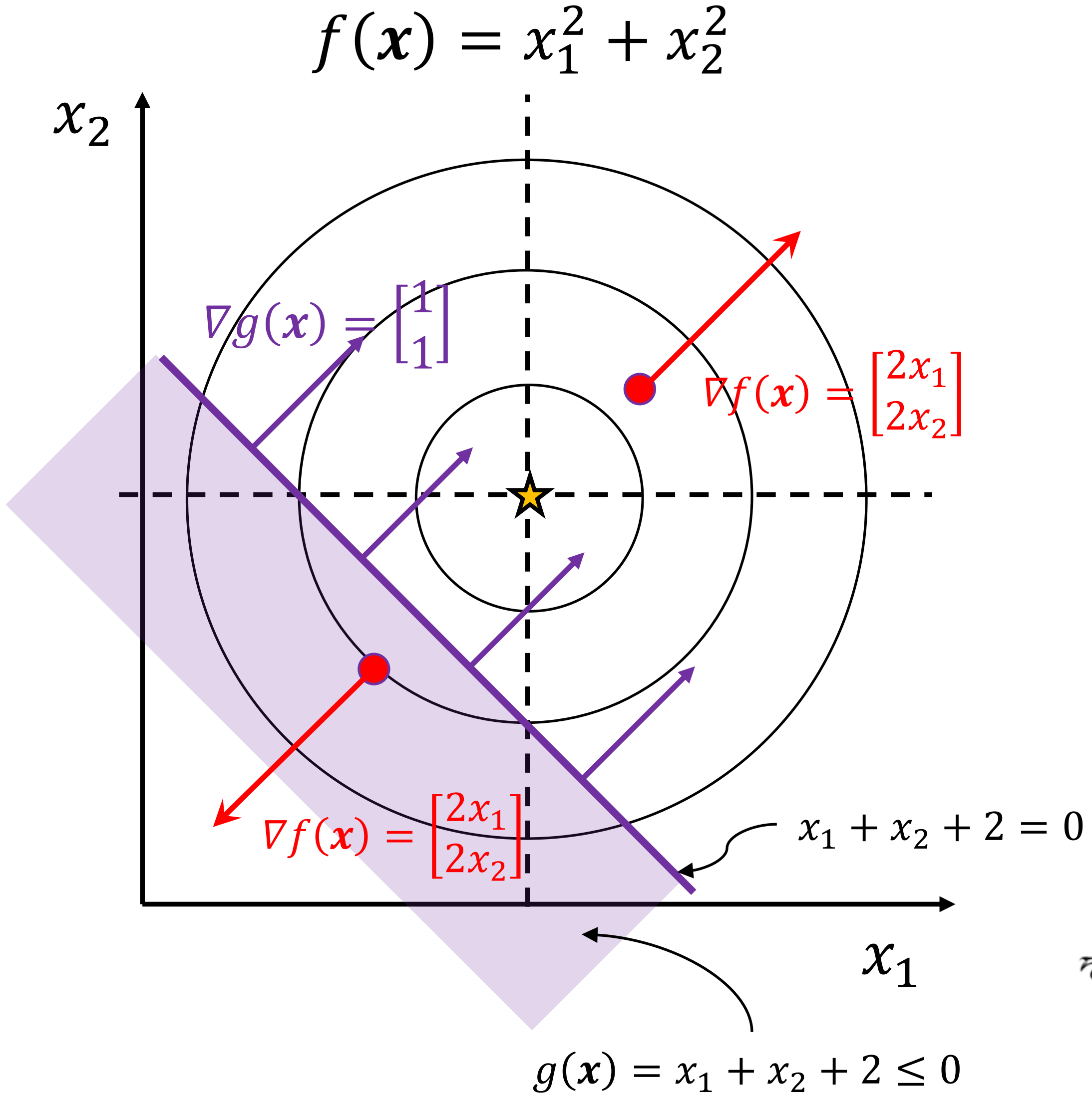


Lagrange Dual Problem

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x})$$
$$s.t. \quad g(\boldsymbol{x}) \le 0 \quad \Rightarrow L(\boldsymbol{x}, \lambda)$$
$$= f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

求解

$$\nabla L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$$
$$\Rightarrow \nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = 0$$
$$\Rightarrow -\nabla f(\boldsymbol{x}) = \lambda \nabla g(\boldsymbol{x})$$



Lagrange Dual Problem with multi-Constraints

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g(\boldsymbol{x}) \leq 0 \end{aligned} \Longrightarrow L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

$$\downarrow$$

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g_1(\boldsymbol{x}) \leq 0 \\ & g_2(\boldsymbol{x}) \leq 0 \\ & \ldots \\ & g_{\mathrm{m}}(\boldsymbol{x}) \leq 0 \end{aligned} \Longrightarrow L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \lambda_1 g_1(\boldsymbol{x}) + \lambda_2 g_2(\boldsymbol{x}) + \ldots + \lambda_{\mathrm{m}} g_{\mathrm{m}}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{\mathrm{m}} \lambda_i g_i(\boldsymbol{x})$$

求解

$$\nabla L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$$

$$\Rightarrow \nabla f(\boldsymbol{x}) + \sum_{i=1}^{\mathrm{m}} \lambda_i \nabla g_i(\boldsymbol{x}) = 0$$

$$\Rightarrow -\nabla f(\boldsymbol{x}) = \sum_{i=1}^{\mathrm{m}} \lambda_i \nabla g_i(\boldsymbol{x})$$

Lagrange Dual Problem with multi-Constraints

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\
s.t. \quad & g_1(\boldsymbol{x}) \le 0 \\
& g_2(\boldsymbol{x}) \le 0 \quad \Rightarrow L(\boldsymbol{x}, \boldsymbol{\lambda}) \\
& \cdots \\
& g_{\mathrm{m}}(\boldsymbol{x}) \le 0
\end{aligned}
$$

$$= f(\boldsymbol{x}) + \sum_{i=1}^{\mathrm{m}} \lambda_i g_i(\boldsymbol{x})$$

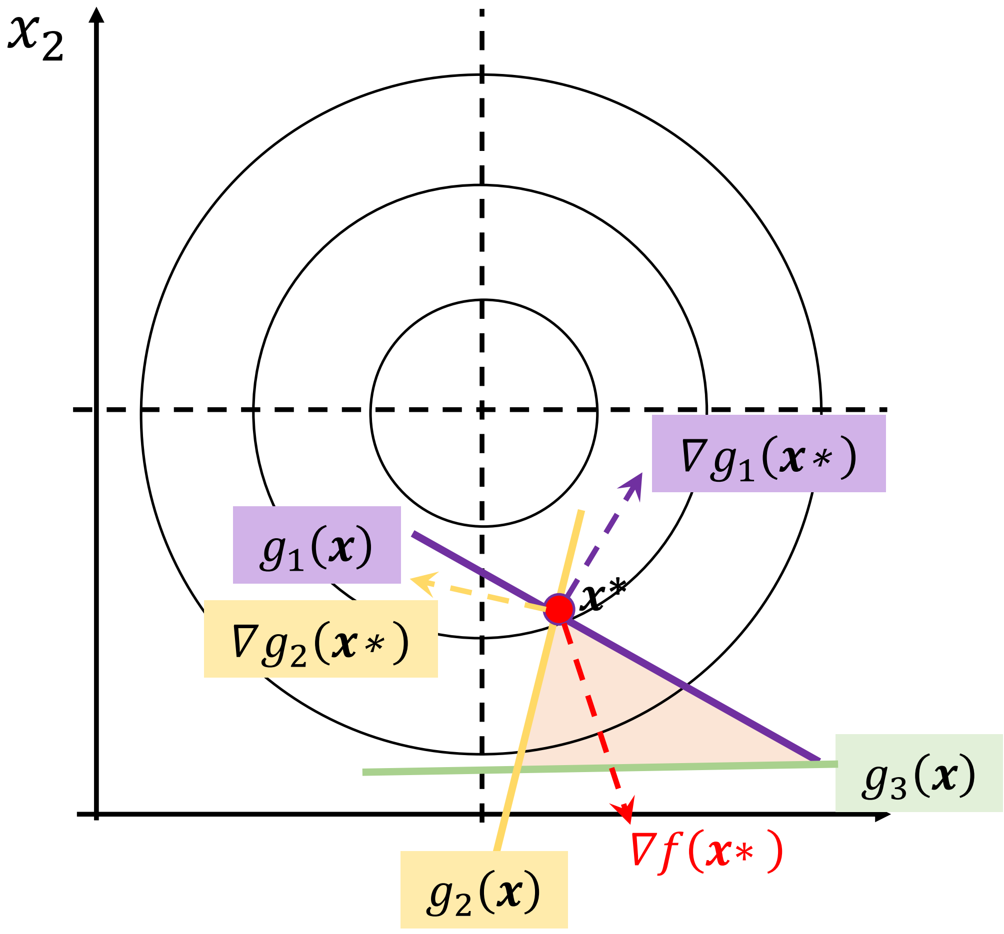
求解

$$\nabla L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$$

$$\Rightarrow \nabla f(\boldsymbol{x}) + \sum_{i=1}^{\mathrm{m}} \lambda_i \nabla g_i(\boldsymbol{x}) = 0$$

$$\Rightarrow -\nabla f(\boldsymbol{x}) = \sum_{i=1}^{\mathrm{m}} \lambda_i \nabla g_i(\boldsymbol{x})$$

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$



$$
\begin{aligned}
& -\nabla f(\boldsymbol{x}) \\
& = \lambda_1 \nabla g_1(\boldsymbol{x}) + \lambda_2 \nabla g_2(\boldsymbol{x}) \\
& + \lambda_3 \nabla g_3(\boldsymbol{x})
\end{aligned}
$$

$$g_1(\boldsymbol{x}^*) = \mathbf{0}$$

$$g_2(\boldsymbol{x}^*) = \mathbf{0}$$

$$g_3(\boldsymbol{x}^*) \neq \mathbf{0} < \mathbf{0}$$

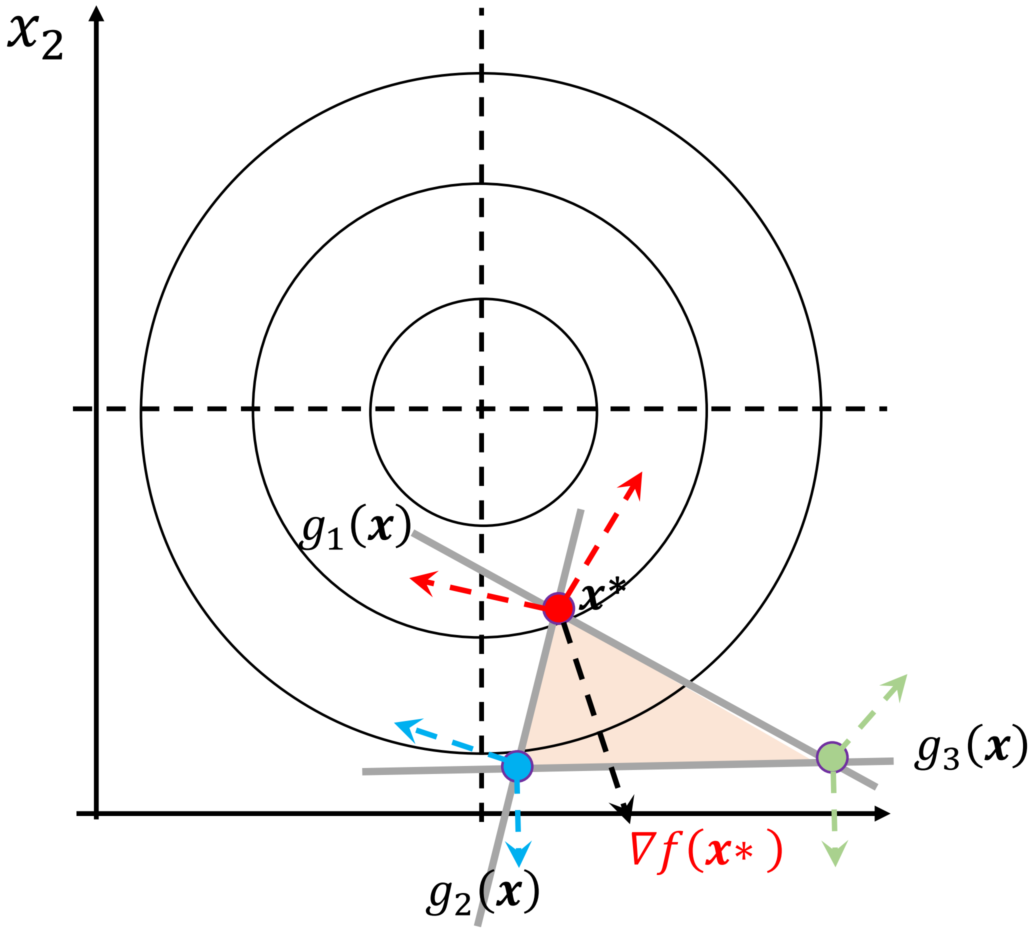
$$\Rightarrow \lambda_3 \nabla g_3(\boldsymbol{x}^*) = \mathbf{0}$$

($\nabla g_3(\boldsymbol{x}^*)$應該不為0)

$$\Rightarrow \lambda_3 = 0$$

Lagrange Dual Problem with multi-Constraints

$f(\boldsymbol{x}) = x_1^2 + x_2^2$



藍色點的梯度(∇g_2和∇g_3)和(兩條藍色的虛線向量和)
綠色點的梯度(∇g_1和∇g_3)和(兩條綠色的虛線向量和)
都不會和黑色梯度($\nabla f(\boldsymbol{x})$，目標函數的梯度)剛好反方向→達到約束條件得功用。

只有當λ<0，才會把藍色點的梯度方向/綠色點的梯度和目標函數的梯度反方向。

但我們想要得到的解應該是紅色點。
所以

$$\lambda_i \geq 0$$

$\lambda_i = 0$: 約束條件$g_i(x)$是鬆弛的。
$\lambda_i > 0$: 約束條件$g_i(x)$是緊緻的。