Lecture 2

classical optimization

Single Variable Taylor Expansion

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{n-1}(x) + \frac{h^n}{n!} f^n(x+\theta h)$$

$$0 < \theta < 1$$

If we know <u>All</u> function derivatives at one point then we know the function values for all x!

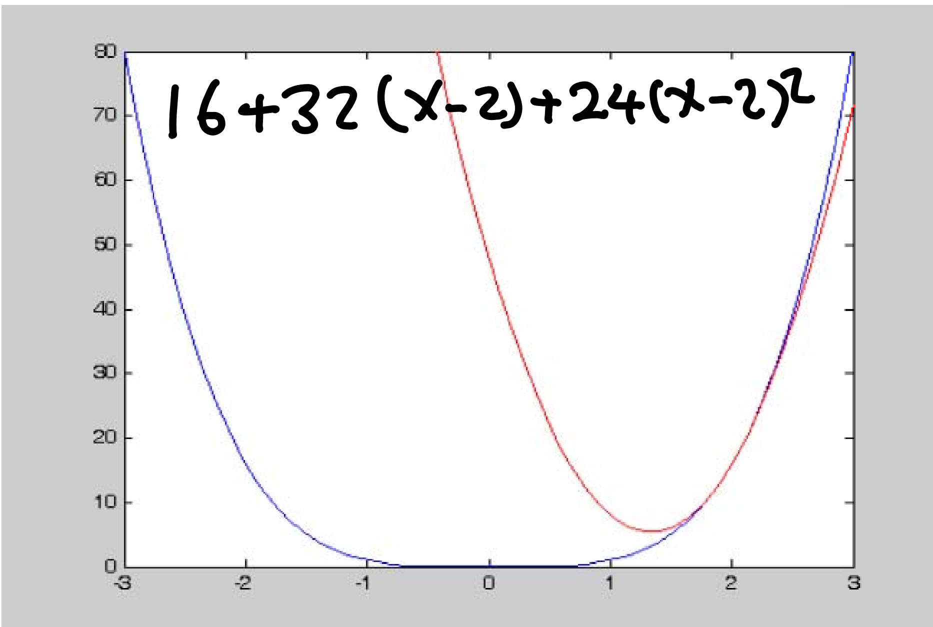
Single Variable Expansion (Cont'd)

* a first order approximation is given by $f(x+h) \simeq f(x) + h f'(x)$

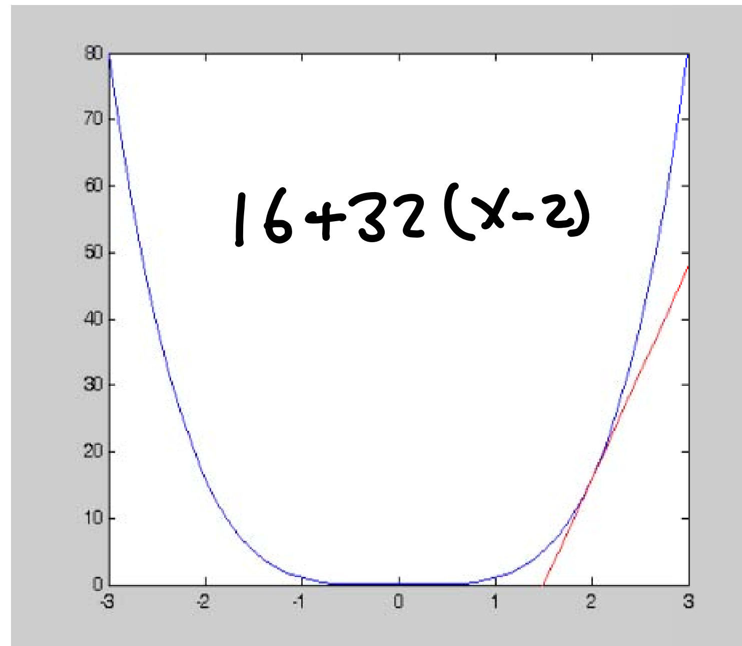
* a second order approximation is given by $f(x+h) \simeq f(x) + h f'(x) + \frac{h^2}{2!} f''(x)$

* We usually use 1st or 2nd order approximation!

Example



$f(x) = x^4$



Multi-variable Taylor Expansion

$$f(\underline{x} + \Delta\underline{x}) = f(\underline{x}) + df(\underline{x}) + \frac{1}{2!} d^2 f(\underline{x}) + \cdots$$

$$+ \cdots + \frac{1}{(n-1)!} d^{n-1} f(\underline{x}) + \frac{1}{n!} d^n f(\underline{x} + \theta \Delta\underline{x})$$

$$0 < \theta < 1$$

$d^r f$ = rth differential of the function f

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \cdots \sum_{k=1}^{n} h_1 h_2 \cdots h_k \frac{\partial^r f(\underline{x})}{\partial x_i \partial x_j \cdots \partial x_n}$$

(contains all possible derivatives of order r

($\Delta\underline{x} = [h_1 \; h_2 \cdots h_n]^T$)

Multi-variable Taylor Expansion (Cont'd)

Consider a function of 3 variables

$df(\underline{x})$ = first order differential

$$df(\underline{x}) = h_1 \frac{\partial f}{\partial x_1} + h_2 \frac{\partial f}{\partial x_2} + h_3 \frac{\partial f}{\partial x_3}$$

$$= [h_1 \quad h_2 \quad h_3] \underline{\nabla} f = \Delta \underline{x}^T \underline{\nabla} f$$

first order Taylor approximation

$$f(\underline{x} + \Delta \underline{x}) \simeq f(\underline{x}) + \Delta \underline{x}^T \underline{\nabla} f \quad \text{(a hyperplane)}$$

Multi-variable Taylor Expansion (Cont'd)

for a function of 3 variables we have

$d^2 f(\underline{x} + \Delta\underline{x})$ = second order differential

$$= h_1^2 \frac{\partial^2 f}{\partial x_1^2} + h_2^2 \frac{\partial^2 f}{\partial x_2^2} + h_3^2 \frac{\partial^2 f}{\partial x_3^2} + 2h_1 h_2 \frac{\partial^2 f}{\partial x_1 \partial x_2} + 2h_1 h_3 \frac{\partial^2 f}{\partial x_1 \partial x_3} + 2h_2 h_3 \frac{\partial^2 f}{\partial x_2 \partial x_3}$$

(all possible 2nd order derivatives)

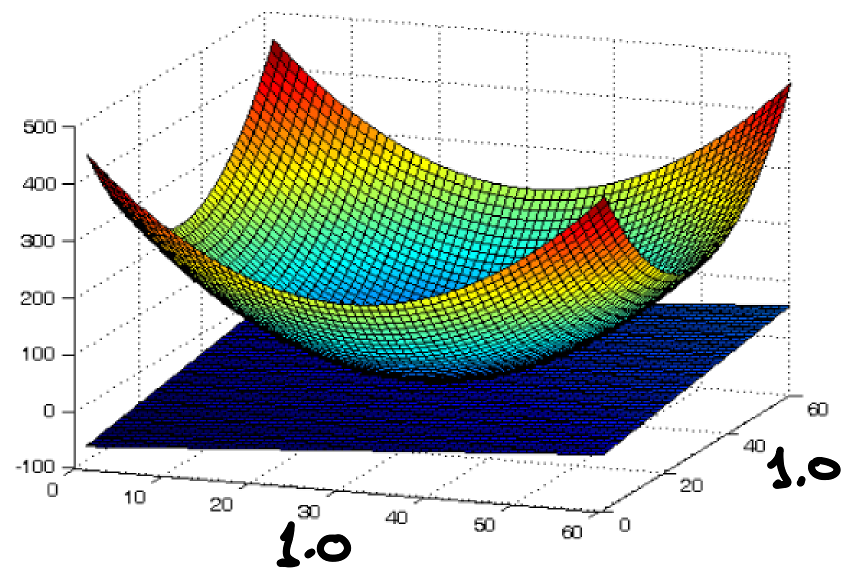
$$d^2 f = \Delta\underline{x}^T \underline{H} \, \Delta\underline{x}$$ ($\underline{H}$ is Hessian matrix)

Example

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

$$\nabla f = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

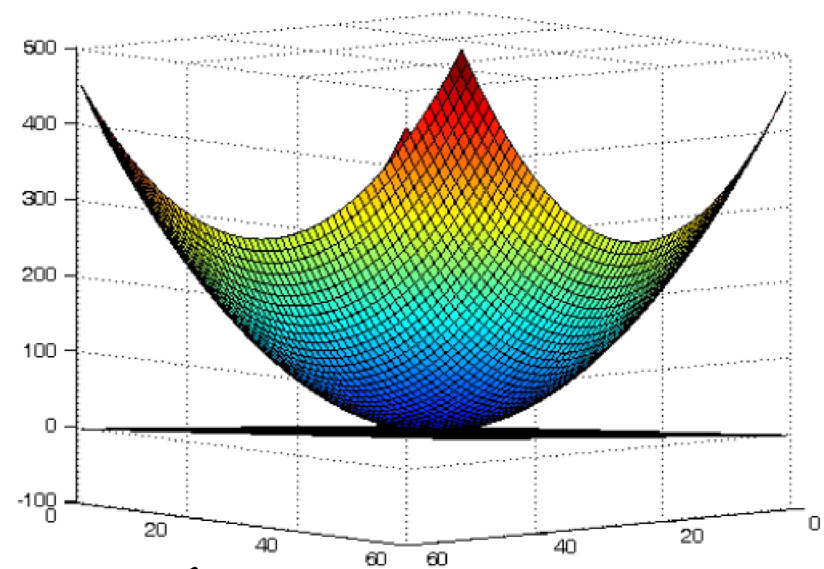
$$\nabla f(1,1) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$



$$L(x_1, x_2) = 2 + 2(x_1 - 1) + 2(x_2 - 1)$$

Example (Cont'd)

$$\underline{H} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$



$$Q(x_1, x_2) = 2 + 2(x_1 - 1) + 2(x_2 - 1) + \frac{1}{2}\begin{bmatrix}(x_1 - 1) & (x_2 - 1)\end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix}(x_1 - 1) \\ (x_2 - 1)\end{bmatrix}$$

$$Q(x_1, x_2) = 2 + 2(x_1 - 1) + (x_2 - 1) + (x_1 - 1)^2 + (x_2 - 1)^2$$

$$Q(x_1, x_2) = 2 + x_1^2 - 1 + x_2^2 - 1 = x_1^2 + x_2^2 = f(x_1, x_2)$$ Why?

Meaning of the gradient

$$f(\underline{x} + \Delta\underline{x}) = f(\underline{x}) + \Delta\underline{x}^T \underline{\nabla} f + \frac{1}{2} \Delta\underline{x}^T \underset{\sim}{H}(\underline{x} + \theta \Delta\underline{x}) \Delta\underline{x}$$

for a sufficiently small $\Delta\underline{x}$, we have

$$f(\underline{x} + \Delta\underline{x}) \approx f(\underline{x}) + \Delta\underline{x}^T \underline{\nabla} f$$

what is $\Delta\underline{x}$ that maximizes $\Delta\underline{x}^T \underline{\nabla} f$ over all $\Delta\underline{x}$ with $\|\Delta\underline{x}\| = \varepsilon$?

$$\Downarrow$$

Gradient is the direction of maximum increase!

Example

find the gradient of the Rosenbrock function

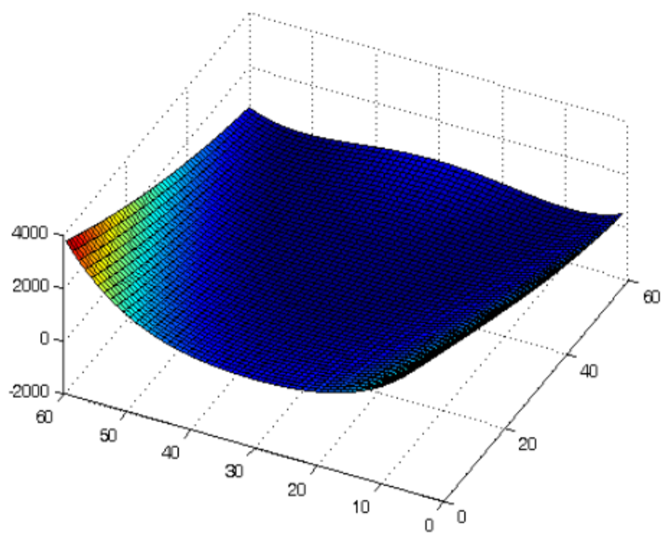
$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \quad \text{at } (1,2)$$

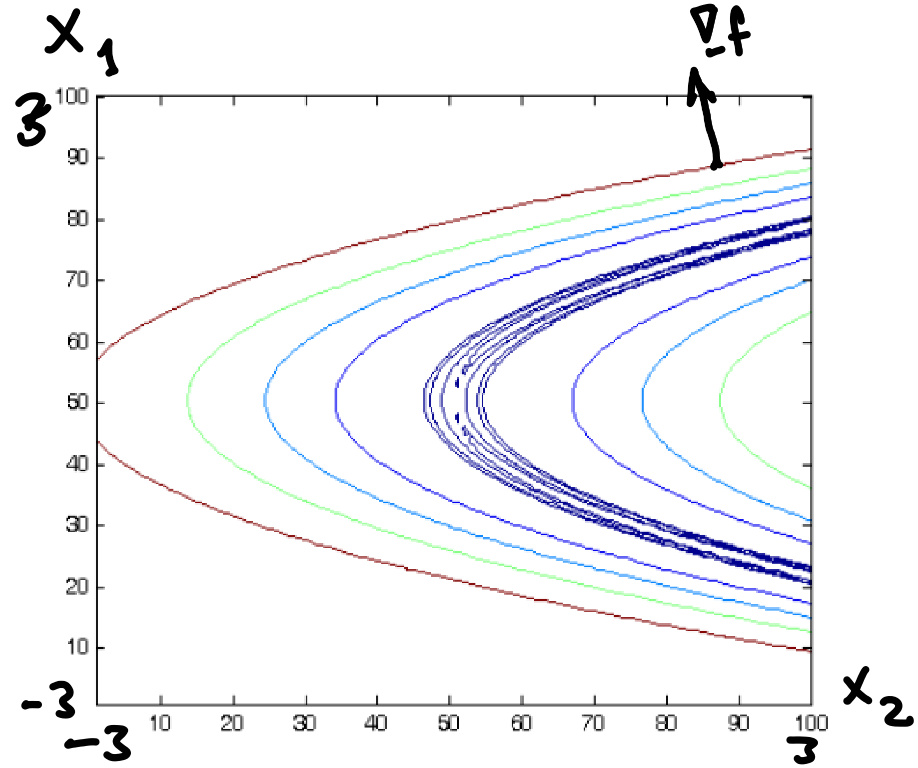
$$\frac{\partial f}{\partial x_1} = 200(x_2 - x_1^2) * -2x_1 - 2(1 - x_1) \Rightarrow \left.\frac{\partial f}{\partial x_1}\right)_{(1,2)} = -400$$

$$\frac{\partial f}{\partial x_2} = 200(x_2 - x_1^2) \Rightarrow \left.\frac{\partial f}{\partial x_2}\right)_{(1,2)} = 200$$

$$\underline{\nabla} f = \begin{bmatrix} -400 \\ 200 \end{bmatrix}$$

Example (cont'd)





$\underline{\nabla} f$ gives direction of maximum function change!

Meaning of the gradient (Cont'd)

* The gradient is normal to constant value surfaces

$$f(\underline{x} + \Delta\underline{x}) = f(\underline{x}) + \underbrace{\Delta\underline{x}^T \underline{\nabla} f}_{0}$$

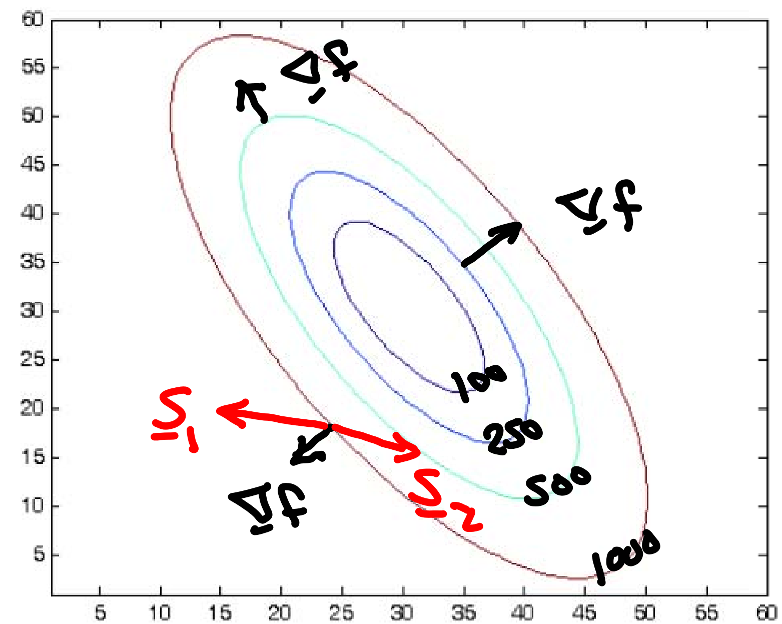
* A function increases in any direction with an angle $\theta < 90^\circ$ with the gradient

$$f(\underline{x} + \Delta\underline{x}) = f(\underline{x}) + \underbrace{\Delta\underline{x}^T \underline{\nabla} f}_{>0} \qquad (\|\Delta\underline{x}\| \text{ small enough})$$

* A function decreases along any direction with $\theta > 90^\circ$

<u>Illustration</u>

* $\underline{\nabla} f$ is a vector in the parameter space



* It changes direction from one point to another
* The function decreases the most along the direction of $-\underline{\nabla} f$ (steepest descent)

Unconstrained Classical Optimization

* Necessary Conditions of optimality must be satisfied at an optimal point $\underline{x}^*$
* If <u>necessary</u> Conditions are satisfied at a point this may not mean that it is the optimal point!
* If <u>sufficient</u> Conditions are satisfied at a point then this point MUST be the optimal point!

Unconstrained Classical Optimization (Cont'd)

* If $\underline{x}^* = \min_{\underline{x}} f(\underline{x})$ then a necessary condition is that $\underline{\nabla} f(\underline{x}^*) = \underline{0}$

Proof: $f(\underline{x}^* + \Delta\underline{x}) = f(\underline{x}^*) + \Delta\underline{x}^T \underline{\nabla} f(\underline{x}^*) + \frac{1}{2!} d^2 f(\underline{x}^* + \theta \underline{h})$

For a sufficiently small $\Delta\underline{x}$, the linear term dominates and $\underline{\nabla} f(\underline{x}^*)$ MUST be zero otherwise a direction $\Delta\underline{x}$ exists such that $\Delta\underline{x}^T \underline{\nabla} f(\underline{x}^*) < 0 \Rightarrow \underline{x}^*$ is not optimal

Sufficient Condition

If $\nabla f(\underline{x}^*) = \underline{0}$ and the Hessian matrix is positive definite then $\underline{x}^*$ is a local minimum

Proof

$$f(\underline{x}^* + \Delta\underline{x}) = f(\underline{x}^*) + \underbrace{\Delta\underline{x}^T \nabla f}_{0} + \frac{1}{2}\Delta\underline{x}^T \underline{H}(\underline{x}^* + \theta\Delta\underline{x})\Delta\underline{x}$$

$f(\underline{x}^* + \Delta x) = f(\underline{x}^*)$ + a positive term

if $\Delta\underline{x}^T \underline{H} \Delta\underline{x} \geq 0 \quad \forall \Delta\underline{x}$

Example

Find the extreme points of the function

$$f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$$

$$\frac{\partial f}{\partial x_1} = 3x_1^2 + 4x_1 = x_1(3x_1 + 4) = 0$$

$$\frac{\partial f}{\partial x_2} = 3x_2^2 + 8x_2 = x_2(3x_2 + 8) = 0$$

Solutions $(0, 0)$, $(0, -\frac{8}{3})$, $(-\frac{4}{3}, 0)$, $(-\frac{4}{3}, -\frac{8}{3})$

Example ((ont'd)

Second order derivatives are then calculated

$$\frac{\partial^2 f}{\partial x_1^2} = 6x_1 + 4, \quad \frac{\partial^2 f}{\partial x_2^2} = 6x_2 + 8, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$$

$$\underset{\sim}{H} = \begin{bmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{bmatrix}$$

at $(0, 0) \Rightarrow \underset{\sim}{H} = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix} \Rightarrow$ Local minimum

at $(0, -\frac{8}{3}) \Rightarrow \underset{\sim}{H} = \begin{bmatrix} 4 & 0 \\ 0 & -8 \end{bmatrix} \Rightarrow$ Saddle point

Example ((cont'd)

at $\left(-\frac{4}{3}, 0\right) \Rightarrow \underline{H} = \begin{bmatrix} -4 & 0 \\ 0 & 8 \end{bmatrix} \Rightarrow$ saddle point

at $\left(-\frac{4}{3}, -\frac{8}{3}\right) \Rightarrow \underline{H} = \begin{bmatrix} -4 & 0 \\ 0 & -8 \end{bmatrix} \Rightarrow$ Local maximum

Optimization with Equality Constraints

* $$\underline{x}^* = \arg\min_{\underline{x}} f(\underline{x})$$

  subject to $g_j(\underline{x}) = 0, \quad j = 1, 2, \ldots, m$

  $\underline{x} \in R^n$

* The method of direct substitution may be used to convert the problem into an unconstrained optimization problem!

Example Solve $\min_{\underline{x}} f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$

subject to $-2x_1^2 + x_2 = 4$

Solution: use $x_2 = 2x_1^2 + 4$, we get

$$f(x_1) = x_1^2 + (2x_1^2 + 4 - 1)^2 = x_1^2 + (2x_1^2 + 3)^2$$

$$\frac{\partial f}{\partial x_1} = 0 \Rightarrow 2x_1 + 2(2x_1^2 + 3) * 4x_1 = 0$$

$$x_1(1 + 4(2x_1^2 + 3)) = 0$$

$x_1 = 0 \Rightarrow x_2 = 4$ (Local minimum, Why?)

A Tricky Example

$$\min_{\underline{x}} f(\underline{x}) = x_1^2 + x_2^2 \quad \text{subject to } (x_1 - 1)^3 = x_2^2$$

by substitution, we get

$$h(x_1) = x_1^2 + (x_1 - 1)^3$$

notice that $h(-\infty) = -\infty$! (Is the problem unbounded?)

Method of Constrained Variation

$$\min_{\underline{x}} f(\underline{x}) \quad \text{subject to} \quad \underline{g}_j(\underline{x}) = 0, \quad j = 1, \ldots, m$$

m constraints in n unknowns

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n = 0 \quad \text{(why?)}$$

all constraints <u>must</u> remain active

$$dg_j = \frac{\partial g_j}{\partial x_1} dx_1 + \frac{\partial g_j}{\partial x_2} dx_2 + \cdots + \frac{\partial g_j}{\partial x_n} dx_n,$$

$$j = 1, 2, \ldots, m$$

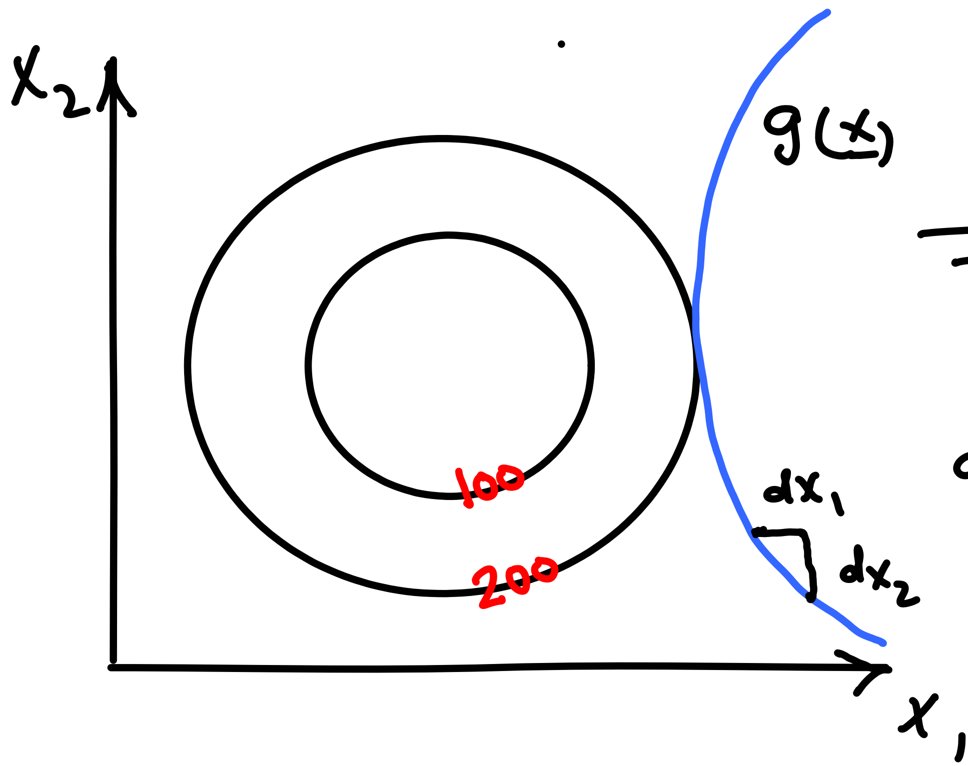
Constrained variation (Cont'd)

we express $dx_1, dx_2, \ldots, dx_m$ in terms of the remaining $(n-m)$ variations. To obtain an equation of the form

$$\alpha_1 \, dx_{m+1} + \alpha_2 \, dx_{m+2} + \cdots + \alpha_{n-m} \, dx_n = 0$$

because the variations are independent, we set the coefficients $\alpha_1 \ldots \alpha_{n-m}$ to zero and solve for the parameters.

Illustration



$$dg = 0$$

$$\frac{\partial g}{\partial x_1} dx_1 + \frac{\partial g}{\partial x_2} dx_2 = 0$$

$$\Rightarrow dx_2 = -\frac{\partial g / \partial x_1}{\partial g / \partial x_2} dx_1$$

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 = 0$$

$$\left(\frac{\partial f}{\partial x_1} + \left(-\frac{\partial g / \partial x_1}{\partial g / \partial x_2} \right) \frac{\partial f}{\partial x_2} \right) dx_1 = 0$$

Example: solve $\min_{\underline{x}} f(\underline{x}) = \dfrac{k}{x_1 x_2^2}$

s.t. $x_1^2 + x_2^2 - a^2 = 0$

solution: $\dfrac{\partial f}{\partial x_1} = -k x_1^{-2} x_2^{-2}$, $\dfrac{\partial f}{\partial x_2} = -2k x_1^{-1} x_2^{-3}$

* $\dfrac{\partial g}{\partial x_1} = 2x_1$, $\dfrac{\partial g}{\partial x_2} = 2x_2 \Longrightarrow dx_2 = -\dfrac{x_1}{x_2} dx_1$

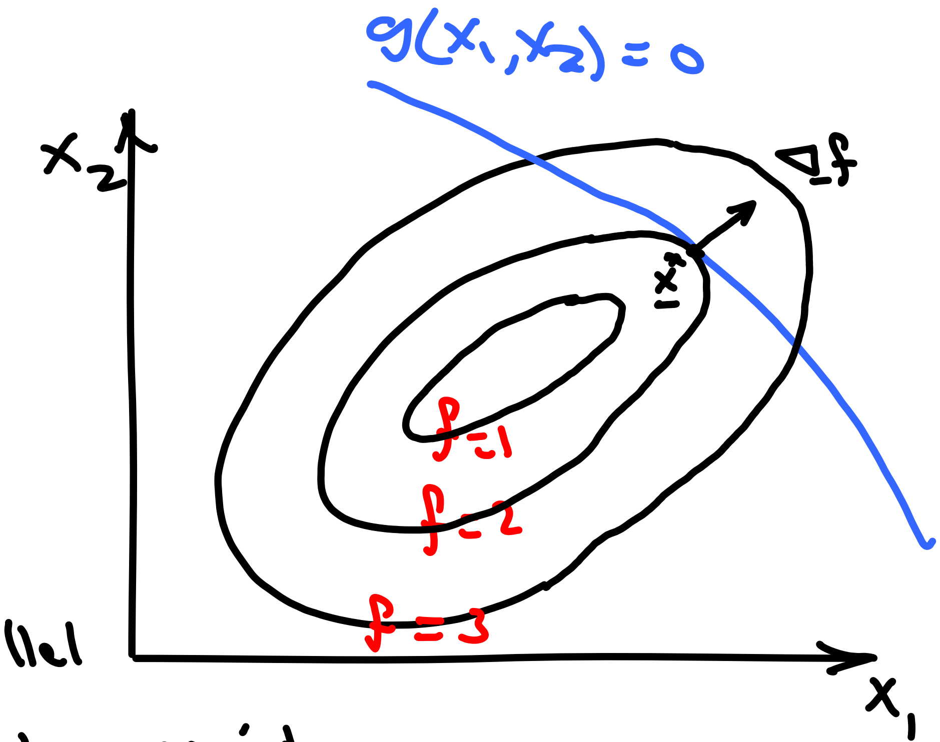
* $\dfrac{\partial f}{\partial x_1} = \dfrac{x_1}{x_2} \dfrac{\partial f}{\partial x_2} \Longrightarrow x_1^{-2} x_2^{-2} = \dfrac{2x_1}{x_2} x_1^{-1} x_2^{-3} \Longrightarrow x_2^* = \sqrt{2}\, x_1^*$

Substituting, we get $x_1^* = \dfrac{a}{\sqrt{3}}$, $x_2^* = \sqrt{2} \dfrac{a}{\sqrt{3}}$

(Physical Meaning?)

Lagrange Multipliers

* Consider the case of one constraint and two parameters
* Note that for this simple case $\underline{\nabla} f$ is parallel to $\underline{\nabla} g$ at the optimal point.



$$\Longrightarrow \quad \underline{\nabla} f + \lambda \underline{\nabla} g = \underline{0} \quad \text{at optimal point } \underline{x}^*$$

Lagrange Multipliers (Cont'd)

* For the case of more than one constraint, we have

$$\left.\begin{aligned} &\underline{\nabla} f(\underline{x}^*) + \sum_{j=1}^{m} \lambda_j \underline{\nabla} g_j(\underline{x}^*) = \underline{0} \\ \& \quad &g_j(\underline{x}) = 0, \quad j = 1, 2, \ldots, m \end{aligned}\right\} n+m \text{ eqns}$$

* We can thus define the Lagrangian to be equal to

$$L(\underline{x}, \underline{\lambda}) = f(\underline{x}) + \sum_{j=1}^{m} \lambda_j \, g_j(\underline{x})$$

$$\frac{\partial L}{\partial \underline{x}} = \underline{0} \Longrightarrow \nabla f(\underline{x}) + \sum_{j=1}^{m} \lambda_j \nabla g_j(\underline{x}) = 0$$

$$\frac{\partial L}{\partial \underline{\lambda}} = 0 \Longrightarrow g_j(\underline{x}) = 0, \quad j = 1, 2, \ldots, m$$

Example: Find the dimensions of a closed card-board box with maximum volume and given surface area A.

* solution: Volume $x_1 x_2 x_3$ and surface area is $2x_1x_2 + 2x_2x_3 + 2x_1x_3 = A$

$$L(\underline{x}, \lambda) = -x_1 x_2 x_3 + \lambda \left(x_1x_2 + x_2x_3 + x_1x_3 - A/2\right)$$

$$\frac{\partial L}{\partial x_1} = 0 \Longrightarrow -x_2 x_3 + \lambda(x_2 + x_3) = 0 \longleftarrow (1)$$

Example (cont'd)

$$\frac{\partial L}{\partial x_2} = -x_1x_3 + \lambda(x_1+x_3) = 0 \longleftarrow (2)$$

$$\frac{\partial L}{\partial x_3} = -x_1x_2 + \lambda(x_1+x_2) = 0 \longleftarrow (3)$$

$$\frac{\partial L}{\partial \lambda} = 0 \Rightarrow x_1x_2 + x_2x_3 + x_1x_3 = A/2 \longleftarrow (4)$$

Solving, we get $x_1^* = x_2^* = x_3^* = \sqrt{A/6}$, $\lambda^* = \frac{2\sqrt{A/6}}{A/6} = 2\sqrt{\frac{6}{A}}$

(Why a zero solution is ruled out?)

(What is the meaning of λ)