Columbia MA Math Camp

Optimization

Vinayak Iyer [a]

July 22, 2020

[a] Material adapted from notes by David Thompson and Xingye Wu

Unconstrained Optimization

$\max f(x)$

Equality Constrained Optimization

$\max f(x) \quad s.t. \quad g(x) = c$

$\max x^{\frac{1}{2}} y^{\frac{1}{2}} \quad s.t. \quad x + y = 1$

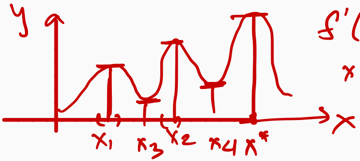
Inequality Constrained Opt

$\max f(x) \quad s.t. \quad g(x) \leq c$

May not cover. (Kuhn Tucker Theorem)

Some Definitions

x_1 & x_2 are local maxima



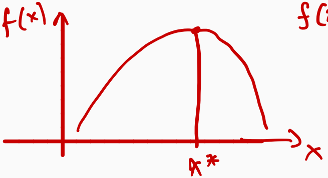
$f'(x) = 0$

$x = x_1, x_2$ or x^*, x_3 or x_4

Let $\mathcal{D} \subseteq \mathbb{R}^n$ and $f : \mathcal{D} \to \mathbb{R}$.

- A point c is a **maximum point** or **global maximum** of f if $f(c) \geq f(x)$ for all $x \in \mathcal{D}$
- A point c is a **local maximum** of f if there exists an $\epsilon > 0$ such that $f(c) \geq f(x)$ for all $x \in B_\epsilon(c)$
- If f is differentiable, a point such that $f'(c) = 0$ is a **critical point** of f

Maximum and minimum points are also called **extreme points, extremum** and **optimal points**



$f(x^*) \geq f(x) \;\forall x \in D$

& cts & differentiable

Suppose f is concave & D is a closed set & bounded

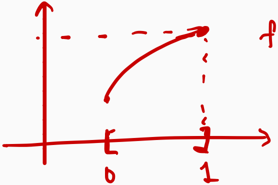
Is it necessary that at the maximizer, $x^* \in D$

$f'(x^*) = 0$?

Proposition 1.1
Let $\mathcal{D} \subseteq \mathbb{R}^n$ be an open set, and $f : \mathcal{D} \to \mathbb{R}$ a differentiable function. If f has a local extreme point *at x, then $f'(x) = 0$* $\Longleftrightarrow$. $\longrightarrow$ This condition isolates extreme points!

Notes : This could mean that x is either a maximum or minimum or neither

$f : [0,1] \to \mathbb{R}$.



$f'(1) \neq 0$

Second-Order Conditions

Twice continuously differentiable function!

Proposition 1.2
Let $\mathcal{D}$ be an open set of $\mathbb{R}^n$ and $f : \mathcal{D} \to \mathbb{R}$ a C^2 function.

- *If f has a* local maximum (minimum) *at x, the Hessian of f at x is* negative (positive) semi-definite
- *If $f'(x) = 0$ and $H(x)$ is* negative (positive) definite, *then x is a* strict local maximum (minimum)

(Proof of second result): Since $H(x)$ is negative definite, there exists $\epsilon > 0$ such that $H(\zeta)$ is negative definite for all $\zeta \in B_\epsilon(x)$. Using the exact form of Taylor's theorem, for any $y \in B_\epsilon(x)$:

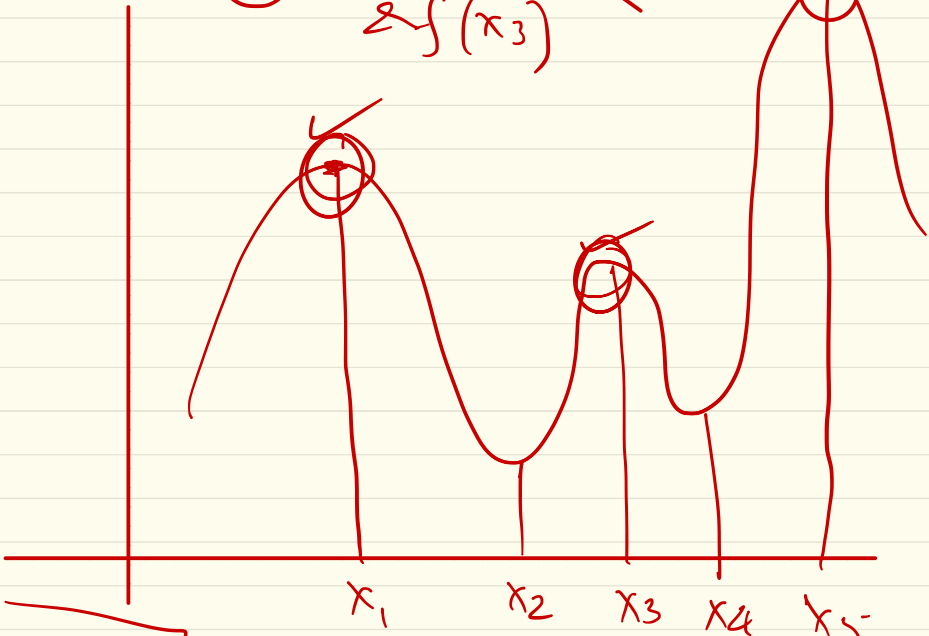
$$\begin{aligned} f(y) &= f(x) + f'(x)(y - x) + \frac{1}{2}(y - x)^T H(\zeta)(y - x) \\ &< f(x) \end{aligned}$$

Maximum:- $f'(x) = 0$ & $f''(x) \leq 0$

Hessian negative semi-definite.

Global maxima

$f(x_5) > f(x_1)$ & $f(x_3)$



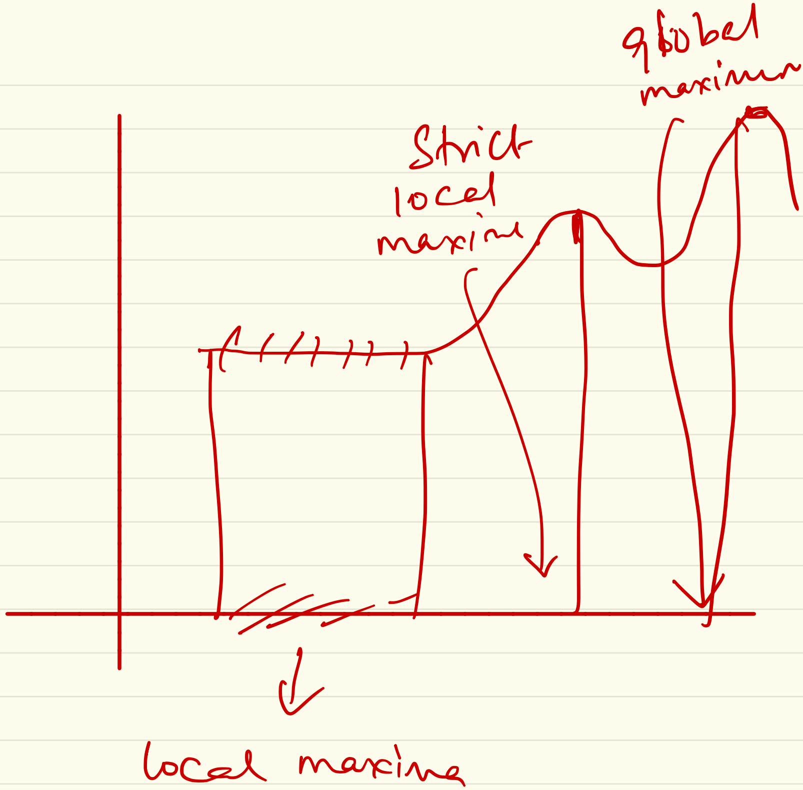
$\boxed{f'(x) = 0}$ ← $\{x_1, x_2, x_3, x_4, x_5\}$

Critical Points.

$f''(x) < 0$ ← $\{x_1, x_3, x_5\}$

$f(x_1)$, $f(x_3)$, $f(x_5)$

To find which point is the global maxima.


global maximum
Strict local maxima
local maxima

Summarizing

maxima / minima

- Critical points are a **necessary (not sufficient)** condition for extrema
- The second derivative can give us **local sufficient conditions**

To this point we've only discussed maximization over an open set

- Maxima need not exist on an open set (e.g. $f(x) = x$ on $(0, 1)$)
- If you're maximizing over a closed set S, you can decompose it as $S = int(S) \cup Boundary(S)$. Need to check the boundary

max $f(x)$
$x \in (0,1)$
$f(x) = x^{\frac{1}{2}}$
maximizer does not exist!

General recipe to maximize a function :

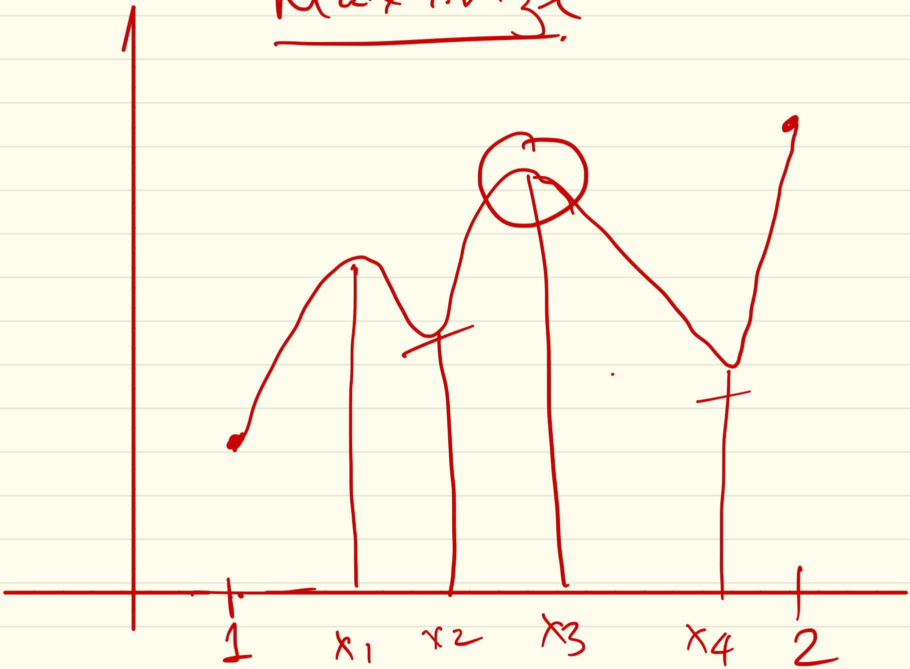
- Find all critical points. If there are many, $\boxed{\text{SOC}}$ can help filter
- Find the critical point with the largest value
- Check if the function takes on a higher value along the boundary

$f'(x) = 0$
filter out the minima!

If $f'(x)=0$ & $f''(x)<0 \Rightarrow x$ is local max

$f : [1, 2] \to \mathbb{R}$.

Maximize



(1)(a) find all critical points.

$\to \{x_1, x_2, x_3, x_4\}$

(b) Filter out based on SOC

$\to \{x_1, x_3\}$.

critical pt

(2) $f(x_1)$ & $f(x_3) \to f(x_3) > f(x_1)$

(3) Check $f(1), f(2)$ & compare to $f(x_3)$

$\longrightarrow$ 2 is the maxima

because

$f(2) \geq f(x) \quad \forall x \in [1,2]$

$$\max \; x^{\frac{1}{2}}$$

$x \in (0,1)$ — Not compact

$\Rightarrow$ max not exists necessarily.

$$f'(x) = 0 = \left[\frac{1}{2} x^{-\frac{1}{2}}\right]$$

$$x = 0 \notin (0,1)$$

Why is it not guarenteed in this simple problem that a max exists?

$x^* \in D$ $\quad$ $x^* \in (0,1)$

<u>Weierstrass</u>: If f is cts & D is compact $\Rightarrow$ max exists.

$\not\Rightarrow$ $\neg P \Rightarrow \neg Q$.

Sufficient Conditions for $\boxed{\text{Global Extrema}}$

For convex functions, optimization is dramatically simpler, as evidenced by the following proposition :

Proposition 1.3
Let $\mathcal{D}$ be a convex open set of $\mathbb{R}^n$ and $f : \mathcal{D} \to \mathbb{R}$ be a convex function. Then :

- *The set of minimizers of f is* convex
- *If f is* strictly convex, *it has* at most one minimizer

$f(x) = x^2$

$D = (0,1)$

- *Any* local minimum *of f is a* global minimum

- *If f is* differentiable, *then x is a* global minimum *of f iff $f'(x) = 0$*

The same results hold for concave functions, replacing "minimizers" with "maximizers"

Note :

- Note that we already proved the second result for quasiconvex functions.

Convex fn :- $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$

(1) Set of minimizers is a convex set

$\forall\ x_1^*, x_2^* \in S_{minimizers}$

$\lambda x_1^* + (1-\lambda) x_2^* \in S_{min}$

$\Rightarrow f(\lambda x_1^* + (1-\lambda) x_2^*) \leq f(x) \quad \forall\ x \in D$

$f(x_1^*) = f(x_2^*)$

$$\Rightarrow f(\lambda x_1^* + (1-\lambda) x_2^*) \leq f(x_1^*) = f(x_2^*) \leq f(x) \quad \forall\ x \in D.$$

$\Rightarrow \lambda x_1^* + (1-\lambda) x_2^* \in S_{minimizers}.$

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$$

f is strictly convex $\Rightarrow$ At most 1 minimizer.

<u>Pf:</u> Suppose there are 2 minimizers x_1^* & x_2^*. $f(x_1^*) = f(x_2^*)$

$$f(\underbrace{\lambda x_1^* + (1-\lambda) x_2^*}_{}) < \lambda f(x_1^*) + (1-\lambda) f(x_2^*)$$

$$= f(x_1^*)$$

which contradicts x_1^* & x_2^* being minimizers!

Hence minimizer (if exists) is unique.

Table of Contents

Optimization

Unconstrained

$$\max_{x \in D} f(x)$$

Constrained

Equality Constrained

$$\max_{x \in D} f(x) \quad \text{s.t. } g(x) = c$$

Focus

Inequality Constrained

$$\max_{x \in D} f(x) \quad \text{s.t. } g(x) \leq c$$

Briefly

Constrained Optimization Problems

- In economics, it's more common to maximize an objective function subject to some constraints
- For example, in consumer theory, you will see problems of this form:

utility　　　　　　Budget constraint

$$\max_{c_1, c_2} \log c_1 + \alpha \log c_2 \quad \text{s.t.} \quad p_1 c_1 + p_2 c_2 = M$$

function

- One approach to these problems is to put the constraints in the objective. For instance, in the above example

Substitute into objective fn

$$c_1 = \frac{M - p_2 c_2}{p_1}$$

(Rewriting budget constraint)

So we could do the unconstrained maximization problem:

$$\underbrace{\max_{c_2} \log\left(\frac{M - p_2 c_2}{p_1}\right) + \alpha \log c_2}$$

] → Do this!

[Do More Examples]

$c_2^* =$ something. How to get c_1^* after that?

$$\max_{c_2} \underbrace{\log\left(\frac{M - p_2 c_2}{p_1}\right) + \alpha \log c_2}_{}.$$

The First Order Condition wrt c_2 is :-

Setting FOC = 0

$$\Rightarrow \frac{-p_2}{M - p_2 c_2} + \frac{\alpha}{c_2} = 0$$

$$\boxed{-\frac{\cancel{p_1}}{M - p_2 c_2} \frac{p_2}{\cancel{p_1}}}$$

$$\Rightarrow \frac{M - p_2 c_2}{p_2} = \frac{c_2}{\alpha}$$

$$\Rightarrow \alpha M - p_2 c_2 \alpha = p_2 c_2$$

$$\Rightarrow \left(\frac{\alpha}{1+\alpha}\right) \frac{M}{p_2} = c_2^*$$

$$\Rightarrow c_2^* = \left(\frac{\alpha}{1+\alpha}\right) \frac{M}{p_2}$$ What is c_1^*?

$$\text{SOC} \Rightarrow \frac{\partial^2 f(x2}{\partial c^2} < 0$$

Maximizers c_1^*, c_2^* are Same !!

$$\max_{c_1, c_2} \log c_1 + \alpha \log c_2 \quad \text{s.t. } p_1 c_1 + p_2 c_2 = M$$

OR

$$\max_{c_1, c_2} c_1 c_2^{\alpha} \quad \text{s.t. } p_1 c_1 + p_2 c_2 = M$$

Will the 2 optimization problems have different solutions?

$$\log c_1 + \alpha \log c_2 = \log (c_1 c_2^{\alpha})$$

$$\max c_1^{\alpha} c_2^{1-\alpha} \quad \text{s.t. B.C.}$$

$$\max \alpha \log c_1 + (1-\alpha) \log c_2 \quad \text{s.t. B.C.}$$

If c^* maximizes $f(c)$

$\Rightarrow$ c^* will also maximize $g(f(c))$

is g is an increasing function.

Example $\log(\,)$ is an increasing fn!

$$\boxed{\max\ u(x) \iff \max\ \log(u(x)) \qquad \text{Maximizers } x^* \text{ will be the same!!}}$$

(Lagrange)

Theorem 2.1
Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ be C^1 functions, and consider the program:

$$\max_{x \in \mathbb{R}^n} f(x) \text{ s.t. } \underbrace{g(x)}_{k \times 1} = \underset{k \times 1}{0}$$

$g_1(x) = 0$
$g_2(x) = 0$
$\vdots$
$g_k(x) = 0$

equality
k constraints

If x^* is a local maximum and x^* satisfies the constraint qualification ; $\operatorname{rank}(g'(x^*)) = k$, then there exist k Lagrange multipliers $\lambda = (\lambda_1, ..., \lambda_k)^T \in \mathbb{R}^k$ *such that the first-order condition holds:*

n first order conditions!!

$f'(x^*) = \left(\frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; \ldots\right)$

$$\underbrace{f'(x^*)}_{1 \times n} + \underset{1 \times k}{\lambda^T} \underset{k \times n}{g'(x^*)} = \underset{1 \times n}{0}$$

$$g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_k(x) \end{pmatrix}$$

It is common to talk about the **Lagrangian** of a system:

$$\text{Lagrangian} \rightarrow \mathcal{L}(x, \lambda) = f(x) + \lambda^T g(x)$$

} unconstrained Optimization

The first-order conditions wrt x and λ give us the critical point of the Lagrangian.

$\mathcal{L}(x, \lambda) = f(x) + \lambda_1 g_1(x) + \lambda_2 g_2(x) + \ldots \lambda_k g_k(x)$

max wrt x, λ

Caveats!

(1) This is only necessary but not sufficient!

It doesn't say that if a critical point x^* satisfies the FOC $\Rightarrow$ it is a maxima!

(2) For (local) maxima, you still need to check SOC.

(3) OR for global maxima, you need Quasiconcavity / concavity of the Lagrangean function! $(\mathcal{L}(x, \lambda))$.

concavity of $f(x)$ is not enough in constrained optim problems!

Question :- Using Lagrangean method

(1) Solve the following problem :-

$$\max_{x,y} \; xy$$
$$\text{s.t. } x+y=6$$

(2) Solve :-

$$\max \; c_1^{\alpha} c_2^{\beta}$$
$$\text{s.t. } p_1 c_1 + p_2 c_2 = M$$

Question 1 :- $f: \mathbb{R}^2 \to \mathbb{R}$ $\quad g: \mathbb{R}^2 \to \mathbb{R}$

$$\max_{x,y} \; xy$$

$$\text{s.t. } x+y=6 \Rightarrow \underbrace{6-x-y}_{g(x)}=0$$

Step 1: Setup the Lagrangian.

$$\mathcal{L}(x,y,\lambda) = xy + \lambda(6-x-y)$$

Step 2: Find FOC's wrt x, y & λ and set them equal to 0 to find critical points !!

(1) $\frac{\partial \mathcal{L}}{\partial x}$: $y - \lambda = 0$

(2) $\partial \mathcal{L}/\partial y$: $x - \lambda = 0$

$\Rightarrow x = y$

(3) $\frac{\partial \mathcal{L}}{\partial \lambda}$: $6 - x - y = 0$

$\Rightarrow 2x = 6$

$\Rightarrow x^* = 3 = y^*$

3 equations in 3 unknowns!

Using (1) & (2) we get

$x = y$. — (4)

Plug (4) into (3) & get :-

$2x = 6$

$\Rightarrow x^* = 3$

$\Rightarrow y^* = 3$

$\lambda^* = 3$

(x^*, y^*, λ^*)

$= (3, 3, 3)$

Step3: Found a unique critical point!!

Check SOC's of $\mathcal{L}(x, y, \lambda)$ to show that this is indeed a maximum!!

(You can show Hessian is negative definite).

$$H : \begin{pmatrix} \overset{<0}{\dfrac{\partial^2 \mathcal{L}}{\partial x^2}} & \dfrac{\partial^2 \mathcal{L}}{\partial x \partial y} & \dfrac{\partial^2 \mathcal{L}}{\partial x \partial \lambda} \\ & \dfrac{\partial^2 \mathcal{L}}{\partial x^2} & \\ & & \dfrac{\partial^2 \mathcal{L}}{\partial \lambda^2} \end{pmatrix}$$

Question 2:-

Solve:-

$$\max \; c_1^{\alpha} c_2^{\beta}$$
$$\text{s.t. } p_1 c_1 + p_2 c_2 = M$$

using the Lagrangean.

$$\mathcal{L}(c_1, c_2, \lambda) = c_1^{\alpha} c_2^{\beta} + \lambda\left(M - p_1 c_1 - p_2 c_2\right)$$

Take FOC's wrt c_1, c_2, λ

$$\frac{\partial \mathcal{L}}{\partial c_1} = \alpha c_1^{\alpha-1} c_2^{\beta} - \lambda p_1 = 0 \qquad (1)$$

$$\frac{\partial \mathcal{L}}{\partial c_2} = \beta c_1^{\alpha} c_2^{\beta-1} - \lambda p_2 = 0 \qquad (2)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = M - p_1 c_1 - p_2 c_2 = 0 \qquad (3)$$

Dividing (1) & (2) we get:

$$\frac{\alpha \cancel{c_1^{\alpha-1}} c_2^{\beta}}{\beta c_1^{\alpha} \cancel{c_2^{\beta-1}}} = \frac{\cancel{\lambda} p_1}{\cancel{\lambda} p_2} \Rightarrow \frac{\alpha}{\beta} \frac{c_2}{c_1} = \frac{p_1}{p_2}$$

$$\frac{\alpha}{\beta}\frac{c_2}{c_1} = \frac{p_1}{p_2}$$

$$\Rightarrow p_1 c_1 = \frac{\alpha}{\beta} p_2 c_2 \qquad (4)$$

Substitute (4) into (3)

$$\frac{\alpha}{\beta} p_2 c_2 + p_2 c_2 = M$$

$$\Rightarrow \frac{\alpha+\beta}{\beta} p_2 c_2 = M$$

$$\Rightarrow c_2^* = \frac{\beta}{\alpha+\beta}\frac{M}{p_2}$$

Substitute c_2^* into (3)

$$p_1 c_1 + \cancel{p_2}\frac{\beta}{\alpha+\beta}\frac{M}{\cancel{p_2}} = M$$

$$\Rightarrow c_1^* = \frac{\alpha}{\alpha+\beta}\frac{M}{p_1}$$

$$(c_1^*, c_2^*) = \left(\frac{\alpha}{\alpha+\beta} \frac{M}{p_1}, \frac{\beta}{\alpha+\beta} \frac{M}{p_2} \right)$$

$$c_1^* = \frac{\alpha}{\alpha+\beta} \frac{M}{P_1}$$

Comparative Statics

(a) $$\frac{\partial c_1^*}{\partial p_1} = -\frac{\alpha}{\alpha+\beta} \frac{M}{p_1^2} < 0$$

b) $$\frac{\partial c_1^*}{\partial p_2} = 0$$

(c) $$\underline{\underline{\frac{\partial c_1^*}{\partial M}}} = \frac{\alpha}{\alpha+\beta} \frac{1}{p_1} > 0$$

A sketch of Lagrange's Theorem in two variables

- Write $x = (x_1, x_2)$. Let $x^* = (x_1^*, x_2^*)$ be a local maximum of f subject to g
- By the IFT, we can write $x_2 = h(x_1)$, with $h'(x_1) = -\frac{g_1(x_1, x_2)}{g_2(x_1, x_2)}$
- We now do unconstrained optimization of $f(x_1, h(x_1))$. The FOC is

$$f_1(x^*) + f_2(x^*)h'(x_1^*) = 0$$

- Define $\lambda = -\frac{f_2(x^*)}{g_2(x^*)}$. Then

$$\begin{aligned} f_1 + \lambda g_1(x^*) &= 0 \\ f_2 + \lambda g_2(x^*) &= 0 \end{aligned}$$

The general case is similar, just with more cumbersome matrix notation.

Comments on Lagrange's Theorem

- The Lagrange condition is a necessary condition.
- As with unconstrained optimization, there are second order conditions that let you check whether a critical point of the Lagrangian is a local maximum or minimum (see FMEA Section 3.4 for details)
- In order to get sufficient conditions for global maxima along the constraint, we need additional structure Q. Concavity of $L(x, \lambda) \Leftarrow f(x)$ Q. concave & $g(x)$ being linear
- If the constraint qualification fails, the theorem says nothing. So you need to check points where the CQ fails separately

Proposition 2.1
Let $f : \mathbb{R}^n \to \mathbb{R}$ be strictly quasiconcave and consider the program

$$\max_x f(x) \text{ s.t. } Ax = b \quad \text{Linear constraints} \tag{1}$$

where A is an $m \times n$ matrix with $m < n$. If (x^*, λ^*) is a critical point of the Lagrangian and $f'(x^*) \neq 0$, then x^* solves (1).

Proof.
The FOC of the Lagrangian implies $f'(x^*) + \lambda^T A = 0$. Suppose there were an $\hat{x}$ such that $A\hat{x} = b$ and $f(\hat{x}) > f(x^*)$. Since f is strictly quasiconcave:

$$\begin{aligned} 0 &< f'(x^*)(\hat{x} - x^*) \\ &= -\lambda^T A(\hat{x} - x^*) \\ &= 0 \end{aligned}$$

a contradiction. □

Note : $f(x)$ being strictly quasiconcave and the constraint being linear ensures that the Lagrangian is strictly quasiconcave as well (which we have shown before implies a unique maximizer) maximized utility function

max $f(x, \theta)$ s.t $g(x) = b$

Define the **"value function"** V as follows

$V(b, \theta) =$

$$V(b) = \max_x f(x) \text{ s.t. } g(x) = b$$

$\mathcal{L}(x(b,\theta), \lambda(b,\theta), b, \theta)$

Form the Lagrangian :

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

Write the solution of this problem as $x^*(b), \lambda^*(b)$. Then

$V(b) = \mathcal{L}(x^*(b), \lambda^*(b), b)$

$$V(b) = \mathcal{L}(x^*(b), \lambda^*(b), b)$$

You express x^* & λ^* as a function of the exogenous variables.

$V'(b) = \frac{\partial \mathcal{L}}{\partial b}$

$$V(b) = \mathcal{L}(x(b), \lambda(b), b)$$

$$\frac{dV}{db} = \frac{\partial \mathcal{L}}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial \mathcal{L}}{\partial \lambda}\frac{\partial \lambda}{\partial b} + \frac{\partial \mathcal{L}}{\partial b}$$

Using the chain rule, we have :

$$V'(b) = \underbrace{\frac{\partial \mathcal{L}}{\partial x}\frac{dx^*}{db} + \frac{\partial \mathcal{L}}{\partial \lambda}\frac{d\lambda^*}{db}}_{=0} + \frac{\partial \mathcal{L}}{\partial b}$$

→ Envelope Theorem !!

However, we know $\boxed{\frac{\partial \mathcal{L}}{\partial x} \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda}}$ are 0 at x^*, λ^*, so we have

$$V'(b) = \lambda^T$$

Only the direct effect of a paramter is relevant. we can ignore second order effects!

Interpretation: Lagrange multipliers measure the marginal value of loosening a constraint by 1 unit

$$\mathcal{L} = f(x) + \lambda^T(b - g(x))$$

$$V'(b) = \frac{\partial \mathcal{L}}{\partial b} = \lambda$$

$$\mathcal{L} = c_1^{\alpha} c_2^{\beta} + \lambda\left[M - p_1 c_1 - p_2 c_2\right]$$

$$\mathcal{L}(c_1^*, c_2^*, \lambda^*) =$$

$$v'(b) = \frac{\partial \mathcal{L}}{\partial M} = \lambda$$

How much utility changes if I give you 1 more dollar -

λ measures the MU of money!!