LECTURE 5

DATE : 25 OCTOBER 2021

5. CONSTRAINT OPTIMIZATION

Buy a BMW

Optimization : buy the car that maximizes your "utility"

Constraint optimization : buy the car that maximizes your utility under a (budget / time etc.) constraint

Math formulation :

$$\begin{cases} \text{maximize} & f(x_1, x_2, \dots, x_n) \\ \text{subject to} & g(x_1, x_2, \dots, x_n) = 0 \end{cases} \leftarrow \begin{cases} \text{add} \\ \text{cond} \end{cases}$$

Remark: You could have more than one constraint
$(g_1 = 0,\ g_2 = 0,\ \dots,\ g_k = 0)$

Simple but highly nontrivial example (see Exercises) :
Box with minimal surface area and fixed volume (= 1)

$$\underbrace{2x_1x_2 + 2x_2x_3 + 2x_3x_1}_{\text{surface}} \longrightarrow$$

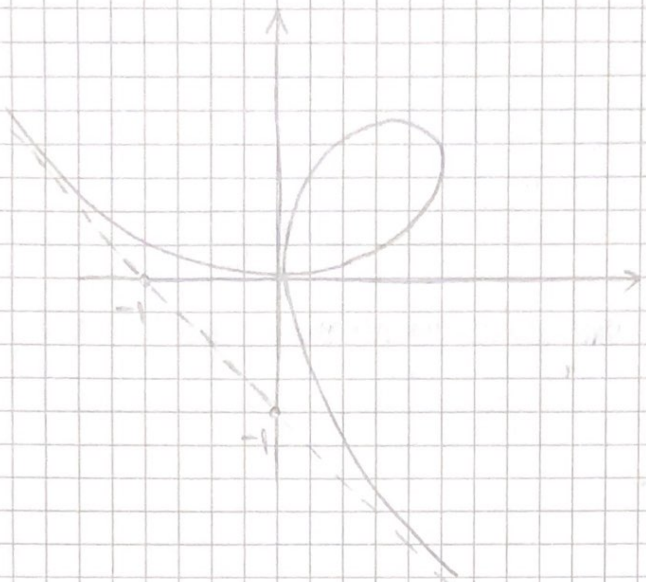
$$\underbrace{x_1x_2x_3}_{\text{volume}} = 1$$

Constraint Opt. prob. reduce to unconstr. Opt

LAGRANGE Multiplier ~~Meth~~ Meth

§. 5.1. Planar curves and the Implicit Function Theorem

"Functions" vs "Curves"



Descarte's Folium (Leaf)
= the locus of all
(x_1, x_2) with
$x_1^3 + x_2^3 - 3x_1x_2 = 0$

Descarte challenged Fermat to find the tangent to this curve (Fermat recently had discovered the "method of tangents" = find tangent using 1st derivative)

Fermat solves the tricky problem using implicit differentiation

we'll get there!

But first : How should we "represent" curves?

3 ways for describing curves

(C_1) Implicit form $\quad f(x_1, x_2) = 0$

ex : unit Circle $\quad x_1^2 + x_2^2 = 1$

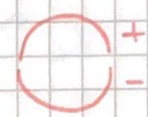
Folium $\quad x_1^3 + x_2^3 - 3x_1x_2 = 0$

(C_2) Explicit form (solve the implicit equation)

$$x_2 = \varphi_f(x_1)$$

Circle $\quad x_2 = \pm\sqrt{1 - x_1^2}$

two branches



Folium ?

(C_3) Parametric form (add a "parameter")

$$\begin{cases} x_1 = x_1(t) \\ x_2 = x_2(t) \end{cases} \qquad t \in [0, T)$$

param

Circle $\begin{cases} x_1 = \cos t \\ x_2 = \sin t \end{cases} \qquad t \in [0, 2\pi)$

Obviously $\quad x_1^2 + x_2^2 = (\cos t)^2 + (\sin t)^2 = 1$

Folium $\quad x_1 = \dfrac{3t}{1+t^3} \qquad x_2 = \dfrac{3t^2}{1+t^3}$

Remark :

explicit form = automatically parametric form $\begin{cases} x_1 = x_1 \\ x_2 = \varphi(x_1) \end{cases}$

x_1 = param

Theorem 1 (Implicit Function Theorem in $\mathbb{R}^2$)

Let $f : \mathbb{R}^2 \to \mathbb{R}$, $x^* = (x_1^*, x_2^*)$ with:

(i) $f(x^*) = f(x_1^*, x_2^*) = 0$

(ii) f continuous differentiable.

(iii) $\frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \neq 0$

Then there exist $\mathcal{Y}_\varepsilon = (x_1^* - \varepsilon, x_1^* + \varepsilon) \subset \mathbb{R}$, $\varepsilon > 0$ and a function $\varphi : \mathcal{Y}_\varepsilon \to \mathbb{R}$ such that:

(1) $\varphi(x_1^*) = x_2^*$

(2) $f(x_1, \varphi(x_1)) = 0 \quad \forall x_1 \in \mathcal{Y} \quad (\Leftrightarrow |x_1 - x_1^*| < \varepsilon)$

(3) φ is differentiable on $\mathcal{Y}_\varepsilon$ and

$$\varphi' = - \frac{\frac{\partial f}{\partial x_1}(x, \varphi(x))}{\frac{\partial f}{\partial x_2}(x, \varphi(x))}$$

"implicit differentiation"

Remark: For the general $d > 2$ setting see

D. Popa, Theorem 6.9.1

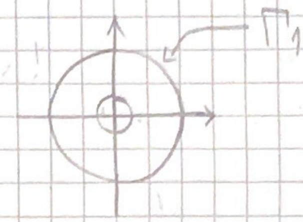
§ 5.2. Level sets (Level curves)

Def: $f : \mathbb{R}^d \to \mathbb{R}$, $c \in \mathbb{R}$

$$\Gamma_c = \{ (x_1, \dots, x_d) \in \mathbb{R}^d : f(x_1, \dots, x_d) = c \}$$

Γ_c ← c-level set of f (may be empty!)

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



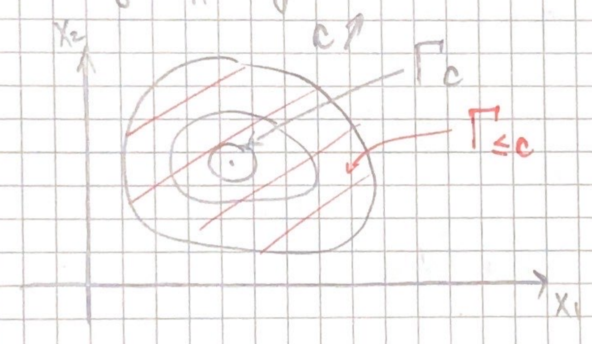
$c = 1$ $\quad x_1^2 + x_2^2 = 1$

$c = \frac{1}{4}$ $\quad x_1^2 + x_2^2 = \frac{1}{4}$

$c = -1$ $\quad x_1^2 + x_2^2 = -1$ ↯ $\quad (\Gamma_{-1} = \emptyset)$

Why are level sets important?

↳ They offer geometric intuition to Optimization



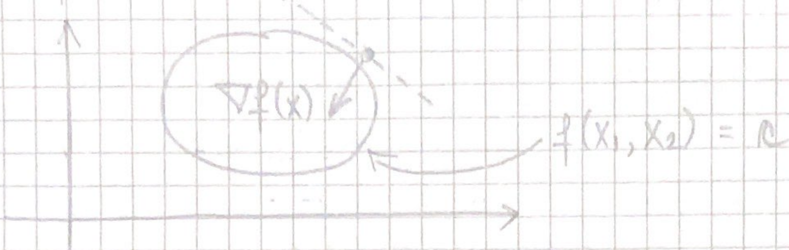
Contour plot!
Level curves = also called "Contour lines".

Remark: You can also talk about sub-level sets

$$\Gamma_c = \{ (x_1, \dots, x_d) \in \mathbb{R}^d : f(x_1, \dots, x_d) \leq c \}$$

Theorem 2: The gradient is orthogonal to level sets!

$f:\mathbb{R}^2 \to \mathbb{R}$, ∇f cont , $\Gamma_c \neq \emptyset$, $c \in \mathbb{R}$



Remark: If $\Gamma \begin{cases} x_1 = x_1(t) \\ x_2 = x_2(t) \end{cases}$ $t \in [0,T)$ is a differentiable parametric curve

Then the tangent to Γ is given by

$$\frac{d}{dt}\left(x_1(t), x_2(t)\right) = \left(x_1'(t), x_2'(t)\right)$$

($x_1, x_2 : [0,T] \to \mathbb{R}$, differentiable functions of t)

Proof of Theorem 2: If $\nabla f(x) = 0_{\mathbb{R}^2}$ ($0_{\mathbb{R}^2} \perp$ any direction)

If $\nabla f(x) \neq 0_{\mathbb{R}^2}$ (as $\frac{\partial f}{\partial x_2}(x) \neq 0$ $(*)$)

Apply Implicit Function Theorem to $f(x_1, x_2) - c$

$\exists$ (locally around x) φ : $x_2 = \varphi(x_1)$

$$f(x_1, \varphi(x_1)) - c = 0 \quad (*)$$

Now use the CHAIN rule

$$\frac{d}{dx_1} f(x_1, \varphi(x_1)) = \nabla f(x_1, \varphi(x_1)) \cdot \frac{d}{dx_1}(x_1, \varphi(x_1))$$

Implicit diff of φ

$$= \nabla f(x_1, \varphi(x_1)) \cdot \left(1, -\frac{\frac{\partial f}{\partial x_1}}{\frac{\partial f}{\partial x_2}}\right) = 0$$

§5.3. The Lagrange Multiplier Method

Constraint Optimization $\quad f(x_1, x_2) \longrightarrow \min$

$\qquad$ subj. to $\quad g(x_1, x_2) = 0$

Lagrange Multiplier Method (Idea): reduce this to unconstraint optimization (with an additional variable λ called Lagrange multiplier)

$$L(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda g(x_1, x_2) \longrightarrow \min$$

Def: $\quad f, g : \mathbb{R}^2 \to \mathbb{R}$ cont, differentiable

$x^* = (x_1^*, x_2^*)$ is called local conditional min if $g(x^*) = 0$ and $f(x^*) < f(x)$ $\forall x$ with $g(x) = 0$.

Theorem 3 (Lagrange Multiplier Method):

$f, g : \mathbb{R}^2 \to \mathbb{R}$ cont, diff-able, x^* conditional min

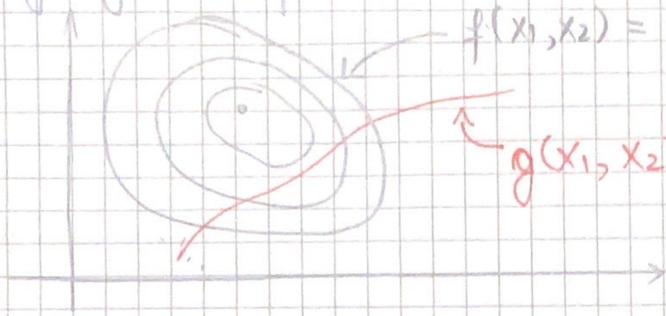
Then there exists $\lambda^* \in \mathbb{R}$ such that $(x_1^*, x_2^*, \lambda^*)$ is a local (unconditional) min for $L(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda g(x_1, x_2)$

that is $\nabla L(x_1, x_2, \lambda) = 0_{\mathbb{R}^3} \iff$

with respect to x_1, x_2, λ

$$\iff \begin{cases} \dfrac{\partial L}{\partial x_1}(x^*, \lambda^*) = 0 \\[2mm] \dfrac{\partial L}{\partial x_2}(x^*, \lambda^*) = 0 \\[2mm] \dfrac{\partial L}{\partial \lambda}(x^*, \lambda^*) = 0 \iff g = 0 \end{cases}$$

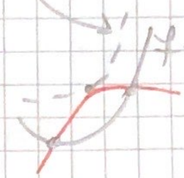
Lagrange Multiplier Method (Idea of Proof):



Geometric Insight: At the conditional min. point the $g=0$ and $f = c$ contour lines are <u>tangent to eachother</u>.

If the two wouldn't touch then x^* does <u>not</u> satisfy the condition ($g(x^*) \neq 0$)

If they intersect (nontangentially) then $\cap$ point are <u>not</u> optimal



$c^* < c$

The only "good" case is

from Theorem 3 $\nabla f \perp$ tangent but also $\nabla g \perp$ tangent (i.e. ∇f and ∇g are colinear)

$$\left.\begin{array}{l}\text{Mathematically } \exists \lambda^* \in \mathbb{R} \quad \nabla f(x^*) = \lambda^* g(x^*) \iff \overbrace{\nabla f(x^*) - \lambda^* g(x^*) = 0}^{\nabla_{x_1, x_2} L = 0} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad g = 0 \quad \left(\dfrac{\partial L}{\partial \lambda} = 0\right)\end{array}\right\}$$

$$x_1, x_2, \lambda \longrightarrow \nabla L = 0$$

Theorem 4: (D. Popa, Theorem 6.10.1)

General result

The Lagrange Multiplier Method works in dimension $d > 2$ and with multiple (compatible!) constraints

Example: The box of minimal surface and volume $= 1$

$$f(x_1, x_2, x_2) := \underbrace{2x_1x_2 + 2x_2x_3 + 2x_3x_1}_{\text{surface of box}} \longrightarrow \min$$

$$g(x_1, x_2, x_3) := \underbrace{x_1x_2x_3}_{\text{volume}} - 1 = 0$$

To find (possible) local minima solve the min problem for

$$L(x_1, x_2, x_3, \lambda) = (2x_1x_2 + 2x_2x_3 + 2x_3x_1) - \lambda(x_1x_2x_3 - 1)$$

$$\frac{\partial L}{\partial x_1} = 0 \qquad \frac{\partial L}{\partial x_2} = 0 \qquad \frac{\partial L}{\partial x_3} = 0 \qquad \frac{\partial L}{\partial \lambda} = 0$$

system of 4 equations & with 4 unknowns (x_1, x_2, x_3, λ)