Second-Order Optimality Conditions & Some important results in Finite Dimensions

MTT106 Non-linear Programming

Pham Thua Tieu Thanh[b, c] Le Nhut Nam[a,b,c]

[a]Department of Computer Science, Faculty of Information Technology
[b]University of Science, Ho Chi Minh City, Vietnam
[c]Vietnam National University, Ho Chi Minh City, Vietnam

July 6, 2024

Table of contents

Optimization problem & Unconstrained optimization

Optimization problem in general

- Formally, an optimization problem in general (or abstract) form:

$$\min_{\text{s.t. } x \in \omega} f(x) \tag{1}$$

- A point that minimizes f over Ω

$$f(x) \geq f(x^*), \forall x \in \Omega \tag{2}$$

  - Maybe not exists!
  - Or maybe not unique!

Unconstrained optimization problem

- Constraint set (or feasible set): $\Omega = \mathbb{R}^n$
- Decision variables are not constrained at all. The goal is only to minimize the objective function.

Application of Unconstrained optimization problem

One popular application and its real-world applied situation:

- Application: Least square
- Real-world application: Data fitting

Problem statement:

- Input: A be $m \times n$ matrix, b be vector $m \times 1$

Solve:

$$\min_x \|Ax - b\| \tag{3}$$

Application of Unconstrained optimization problem

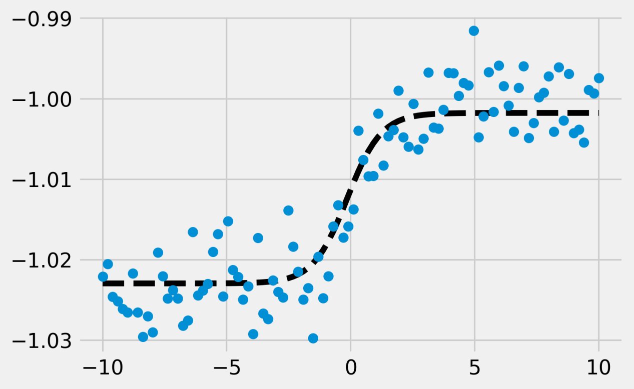

Figure 1: Example data fitting.

Second-Order Optimality Conditions

Motivation

Consider the unconstrained optimization problem of the form:

$$min\{f(x) : x \in \mathbb{R}^n\}, f : \mathbb{R}^n \longrightarrow \mathbb{R} \tag{4}$$

f non-linear.

Problem

How to find exactly minimum (or maximum) points of eq.(4)?

Definition

Positive semidefinite matrix and positive definite matrix

Let $f: A^{n\times n}, d \in \mathbb{R}^n$, if :

- $\langle Ad, d\rangle \geq 0 \implies A$ is positive semidefinite $(A \succeq 0)$.
- $\langle Ad, d\rangle > 0, d \neq 0 \implies A$ is positive definite $(A \succ 0)$.

Definition

Hessian matrix and Second-order theorem

Let $x, \overline{x} \in \mathbb{R}^n$ and $f \in C^2$. We have:

- $Hf(x) = \bigtriangledown^2 f(x) = \bigtriangledown(\bigtriangledown f(x))$ is called Hessian matrix.
- If $\overline{x}$ is a local minimizer $\implies \bigtriangledown f(\overline{x}) = 0, Hf(\overline{x}) \succeq 0$ (Second-order necessary condition).
- If $\overline{x} \in \mathbb{R}^n, \bigtriangledown f(\overline{x}) = 0, Hf(\overline{x}) \succ 0 \implies \overline{x}$ is strict local minimizer (Second-order sufficient condition).
- If $U \subseteq \mathbb{R}^n$ is a open convex set, $\overline{x} \in U, \bigtriangledown f(\overline{x}) = 0, Hf(\overline{x}) \succeq 0 \implies x$ is a global minimizer (Second-order sufficient condition for a global minimizer).
- If $\bigtriangledown f(\overline{x}) = 0, Hf(x)$ is indefinite $\implies \overline{x}$ is a **saddle point** (Second-order sufficient condition for a saddle point)

Discussion

Proof: Second-order necessary condition for a local minimizer

- Let $\bar{x}$ is a local minimizer, $|t|$ is small enough, $\forall d \in \mathbb{R}^n$, Then:

$$f(\bar{x}+td) - f(\bar{x}) \geq 0 \tag{5}$$

- Because $f \in C^2$, then:

$$f(\bar{x}+td) = f(x) + t\langle \triangledown f(\bar{x}), d\rangle + \frac{t^2}{2}\langle \triangledown^2 f(\bar{x})\,d, d\rangle + o(t^2) \tag{6}$$

- We have $\langle \triangledown f(\bar{x}), d\rangle = 0$. Then:

$$0 \leq f(\bar{x}+td) - f(\bar{x}) = \frac{t^2}{2}\langle \triangledown^2 f(\bar{x})\,d, d\rangle + o(t^2) \implies \triangledown^2 f(\bar{x})\,d, d\rangle \geq 0 \implies \triangledown^2 f(\bar{x}) \succ 0 \tag{7}$$

Discussion

Proof: Second-order sufficient condition for a local minimizer

- If $\bigtriangledown^2 f(\bar{x}) \succ 0$, let λ is a smallest eigenvalue of $\bigtriangledown^2 f(\bar{x})$
  $\Longrightarrow \langle \bigtriangledown^2 f(\bar{x}) d, d\rangle \geq \lambda ||d||^2, \forall d \in \mathbb{R}^n$.
- We have $\langle \bigtriangledown f(\bar{x}), d\rangle = 0$. Then

$$f(\bar{x} + td) = f(\bar{x}) + \frac{t^2}{2}\langle \bigtriangledown^2 f(\bar{x}) d, d\rangle + o(t^2) \Longrightarrow \frac{f(\bar{x} + td) - f(\bar{x})}{t^2} \geq \frac{\lambda ||d||^2}{2} + \frac{o(t^2)}{t^2} \quad (8)$$

- $|t|$ is small enough $\Longrightarrow f(\bar{x} + td) - f(\bar{x}) > 0, \forall 0 \neq d \in \mathbb{R}^n \Longrightarrow \bar{x}$ is strict local minimizer of f in $\mathbb{R}^n$.

Discussion

Proof: Second-order sufficient condition for a global minimizer

Let $y \in U \implies \exists z \in (x, y)$:

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}(y - x)^T Hf(z)(y - x) \tag{9}$$

Because $\nabla f(x) > 0, Hf(x) \succeq 0 \implies f(y) \geq f(x)$. Thus, x is a global minimizer of f on U.

Discussion

Consequence

Since: $max\{f(x)|x \in \mathbb{R}^n\} = -min\{-f(x)|x \in \mathbb{R}^n\}$, we have:

- If $\bar{x}$ is a local maximizer of $f(x) \implies \triangledown f(\bar{x}) = 0$ and $\triangledown^2 f(\bar{x}) \preceq 0$
- If $\triangledown f(\bar{x}) = 0$ and $\triangledown^2 f(\bar{x}) \prec 0 \implies \bar{x}$ is a strict local maximizer of $f(x)$

Meaning of Second-Order Optimality Conditions

- The second order condition is a filter that helps identify the nature of stationary points is a local minimum, local maximum, or saddle point. The result of the second derivative at a point x tells us the slope of the tangent line.
- Useful in practice: Optimize for Machine Learing, Deep learing algorithmns.

How to find ?

Let $x \in A^n, f: A^n \to \mathbb{R}$:

- We differentiate once to find $\bigtriangledown f(x)$.
- Let $\bigtriangledown f(x) = 0$, we find all critical points.
- We differentiate twice to find $\bigtriangledown^2 f(x)$.
- For each point x_0 in step (2), calculate $\bigtriangledown^2 f(x_0)$. If:
  - $\bigtriangledown^2 f(x_0) \succ 0 \implies x_0$ is a local minimizer (if A convex x_0 is a global minimizer).
  - $\bigtriangledown^2 f(x_0) \prec 0 \implies x_0$ is a local maximizer.
  - $\bigtriangledown^2 f(x_0)$ is indefinite $\implies x_0$ is a saddle points.

Computational example

Problem 1: Find all critical points of $f(x)$

$$f(x) = x^5 - 5x \tag{10}$$

Proof:
- Let $x \to +\infty \implies f(x) \to +\infty$; $x \to -\infty \implies f(x) \to -\infty$. Thus, $f(x)$ hasn't global critical points.
- $f'(x) = 5x^4 - 5 = 0 \Leftrightarrow x = \pm 1$.
- $f''(x) = 20x^3, f''(1) = 20 > 0, f''(-1) = -20 < 0 \implies x_1 = 1$ is a local minimizer, $x_2 = -1$ is a local maximizer.

Computational example

Problem 2: Find all minimizers and maximizers of $f(x, y)$

$$f(x, y) = \frac{1}{4}(x^4 - 4xy + y^4) \tag{11}$$

Proof:

- We have:

$$\nabla f(x, y) = \begin{pmatrix} x^3 & -y \\ -x & +y^3 \end{pmatrix}, \nabla^2 f(x, y) = \begin{pmatrix} 3x^2 & -1 \\ -1 & +3y^2 \end{pmatrix} \tag{12}$$

- $\nabla f(x, y) = 0 \Leftrightarrow (x, y) \in \{(0, 0)^T, (1, 1)^T, (-1, -1)^T\}$. We have:

$$\nabla^2 f(0, 0) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \nabla^2 f(1, 1) = \nabla^2 f(-1, -1) = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} \tag{13}$$

- Because $\nabla^2 f(1, 1) \succ 0$ and $\nabla^2 f(-1, -1) \succ 0 \implies (\pm 1, \pm 1)^T$ is strict local minimizers. $\mathbb{R}^2$ is convex set $\implies (\pm 1, \pm 1)^T$ also is global minimizers.
- Because $\nabla^2 f(0, 0) \notin \{\succeq 0, \preceq 0\} \implies (0, 0)^T$ not is local critical points.

Computational example

Problem 3: Consider the family of problems

$$min f(x, y) = x^2 + y^2 + \beta xy + x + 2y \tag{14}$$

Proof:

- We have:

$$\bigtriangledown f(x, y) = \begin{pmatrix} 2x + \beta y + 1 \\ 2y + \beta x + 2 \end{pmatrix}, \bigtriangledown^2 f(x, y) = \begin{pmatrix} 2 & \beta \\ \beta & 2 \end{pmatrix} \tag{15}$$

- If $\beta \neq \pm 2 \implies \bigtriangledown f(x, y) = 0 \Leftrightarrow (x^*, y^*) = (2\beta - 2, \beta - 4)/(4 - \beta^2)$.
- If $\beta = \pm 2$, we have an inconsistent system of equations. Therefore, no critical points exist for $\beta = \pm 2$
- Let $A = Hf(x, y), \lambda$ is a eigenvalue of A. We have:

$$det(A - \lambda I) = (2 - \lambda)^2 - \beta^2 = 0 \Leftrightarrow \lambda = 2 \pm \beta \tag{16}$$

- If $-2 < \beta < 2 \implies A \succ 0, \forall \lambda > 0 \implies (x^*, y^*)$ is global minimizers.
- If $|\beta| > 2 \implies \lambda_1 \lambda_2 < 0 \implies (x^*, y^*)$ is saddle points.

Important results in finite dimensions

What is implicit function?

If a function is written in the form of

$$y = f(x),\ \text{e.g.,}\ y = 2x^3 \tag{17}$$

is called an **explicit function**.
And sometimes functions are given in the form

$$y - f(x) = 0\ \text{e.g.,}\ y - 2x^3 = 0 \tag{18}$$

is called an **implicit function**.
In general, **implicit function** are written in a general form as

$$F(y, x) = 0 \tag{19}$$

[1]While we can always change an explicit function into an implicit function (by taking $f(x)$ to the other side of the equality) the reverse is not always true

Motivation

Question

1. Given a solution to a system of equations, are there other solutions nearby?
   => The analytic version of the implicit function theorem.
2. What does the set of all solutions look like near a given solution?
   => The geometric version of the implicit function theorem.

Implicit function theorem

Theorem: Implicit function theorem

Suppose that $U \subseteq \mathbb{R}^n$, and $V \subseteq \mathbb{R}^m$ are open sets, and $\mathbf{F} : U \times V \to \mathbb{R}^m$ is a function of class C^1. Let $(x_0, y_0) \in U \times V$ is a point such that $\mathbf{F}(x_0, y_0) = 0$ and $D_y\mathbf{F}(x_0, y_0) : \mathbb{R}^m \to \mathbb{R}^m$, the derivative of $\mathbf{F}$ w.r.t y, is nonsingular, i.e $D_y\mathbf{F}(x_0, y_0) \neq 0$. Then

- There exists neighborhoods $U_1 \ni x_0$ and $V_1 \ni y_0$ and a C^1 mapping $y : U_1 \to V_1$ such that a point $(x, y) \in U_1 \times V_1$ satisfies $\mathbf{F}(x, y) = 0$ if and only if $y = \mathbf{f}(x)$. The derivative of y at x_0 is given by

$$D_y(x_0) = -D_y\mathbf{F}(x_0, y_0)^{-1} D_x\mathbf{F}(x_0, y_0) \tag{20}$$

- Moreover, if $\mathbf{F}$ is k-times continously differentiable, i.e., $\mathbf{F} \in C^k$, then $\mathbf{f}(x) \in C^k$.

Proof of Implicit function theorem

At first, we need to setup something...

If necessary, considering the function

$$(x, y) \mapsto f(x + x_0, y + y_0) - f(x_0, y_0) \tag{21}$$

And, let

$$f(x) = (f_1(x, y), \ldots, f_m(x, y)) \tag{22}$$

Proof of Implicit function theorem

Df is continous $\Rightarrow$ exist neighborhoods $U_0 \in \mathbb{R}^n$, $V_0 \in \mathbb{R}^m$ that

$$\begin{bmatrix} \nabla_y f_1(x, y_1)^\top \\ \nabla_y f_2(x, y_2)^\top \\ \vdots \\ \nabla_y f_m(x, y_m)^\top \end{bmatrix} \tag{23}$$

is invertible for all $(x, y_i) \in U_0 \times V_0$.

Proof of Implicit function theorem

Question: Is every $x \in U_0$, then there exists at most one $y \in V_0$ such that $f(x, y) = 0$. By

contradiction, suppose that $\exists y, z \in V_0, y \neq z, f(x, y) = f(x, z) = 0$. Due the mean value theorem, $\exists y_i \in (y, z)$ that

$$f_i(x, z) - f_i(x, y) = \langle \nabla_y f_i(x, y_i), z - y \rangle \tag{24}$$

And the previous matrix is non-singular, so $y = z$.

Proof of Implicit function theorem

Let $\overline{B}_r(0) \subseteq V_0$

- $f(0, y) \neq 0, \forall y \in S_r(0) := \{y \in \mathbb{R}^l : \|y\| = r\}$ due to $f(0, 0) = 0$.
- $\exists \alpha > 0, \|f(0, y) \geq \alpha\|, \forall y \in S_r(0)$ due to f is continous.

Proof of Implicit function theorem

Consider the function

$$F(x, y) := \|f(x, y)\|^2 = \sum_{i=1}^{m} f_i(x, y)^2 \tag{25}$$

that satisfies the properties

- $F(0, y) \geq \alpha > 0, \forall y \in S_r(0)$
- $F(0, 0) = 0$

Proof of Implicit function theorem

Because F is continous function, $\exists U_1 \subseteq U_0$ of $0 \in \mathbb{R}^n$ such that

$$F(x, y) \geq \frac{\alpha}{2}, F(x, 0) \leq \frac{\alpha}{2} \forall x \in U_1, y \in S_r(0) \tag{26}$$

fixed $x \in U_1$, function $y \mapsto F(x, y)$ achieves its minimum on $\overline{B}_r(0)$ at $y(x)$ in the interior of $\overline{B}_r(0)$

$$D_y F(x, y(x)) = 2 D_y f(x, y(x)) f(x, y(x)) = 0 \tag{27}$$

And the matrix $D_y f(x, y(x))$ is non-singular, so that

$$f(x, y(x)) = 0 \tag{28}$$

Proof of Implicit function theorem

Writting $\Delta y := y(x + \Delta x) - y(y)$, by the mean value theorem

$$0 = D_x f(\widetilde{x}, \widetilde{y})\Delta x + D_y f(\widetilde{x}, \widetilde{y})\Delta y \tag{29}$$

for some point $(\widetilde{x}, \widetilde{y})$ one the line segment between $(x, y(x))$ and $(x + \Delta x, y(x + \Delta x))$. And as $\|\Delta x\| \to 0$, $\|\Delta y\| \to 0$, thus lead $y(x)$ is continous. And by Taylors formula and continuity of

$y(x)$, we prove that $y(x)$ is Fréchet differentiable. If $f \in C^2$, then $y(x) \in C^2$. And in general, if

C^k, we prove by induction on k that $y(x)$ is C^k.

Discussion about the analytic meaning

Problem

Solving the equation below:

$$f(x, y) = 0 \tag{30}$$

for y as a function of x, and say $y = f(x)$

If we have a solution $b = f(a)$, then in principle it is possible to solve for x near a, if the crucial hypothesis $D_y f(a, b) \neq 0$ holds.
In this meaning, Implicit function theorem is a theorem about the **possibility of solving a system of nonlinear equations**.

Discussion about the geometric meaning

There are 3 natural ways to represent a curve $S \subseteq \mathbb{R}^n$

- As a graph
- As a level set
- Parametrically

Discussion about the geometric meaning

Considering the case $n = 2$

- Graph:

$$\left.\begin{array}{c} S = \{(x,y) \in \mathbb{R}^2 : \ y = f(x) \text{ for } x \in I\} \\ \text{OR} \\ S = \{(x,y) \in \mathbb{R}^2 : \ x = f(y) \text{ for } y \in I\} \end{array}\right\} \tag{31}$$

  for some $f : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval.

- Level set, is, a set of the form

$$S = \{(x,y) \in U : F(x,y) = c\} \tag{32}$$

  for some open $U \subseteq \mathbb{R}^2$, some $F : U \to \mathbb{R}$, some $c \in \mathbb{R}$.

- Parametrically, in the form

$$S = \{\mathbf{f}(t) : \ t \in I\} \tag{33}$$

  for some interval $I \subseteq \mathbb{R}$, and some $\mathbf{f} : I \to \mathbb{R}^2$

Discussion about the geometric meaning

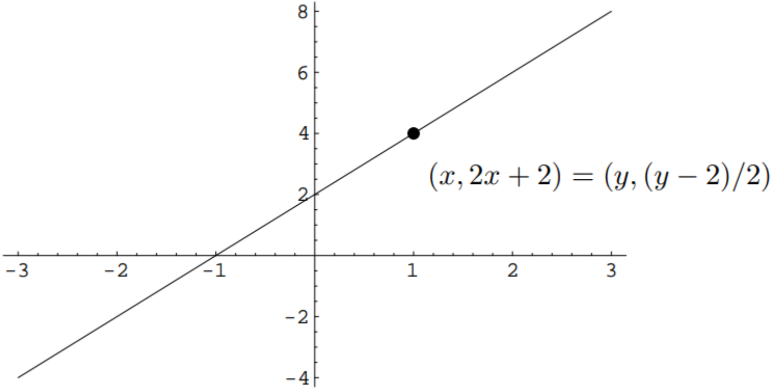

Figure 2: The line $2xy + 2 = 0$ as the graph of $f(x) = 2x + 2$ or of $g(y) = (y2)/2$

Discussion about the geometric meaning

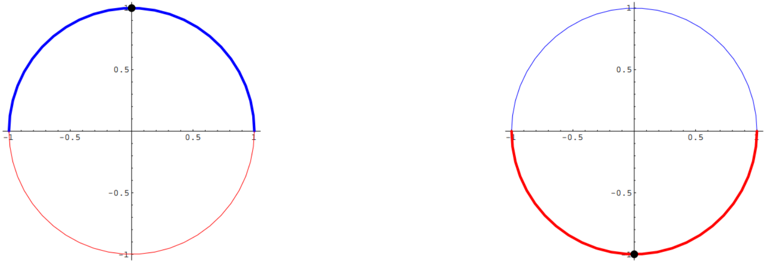

Figure 3: The thick arcs are the graphs of $x \mapsto \pm\sqrt{1-x^2}$

Discussion about the geometric meaning

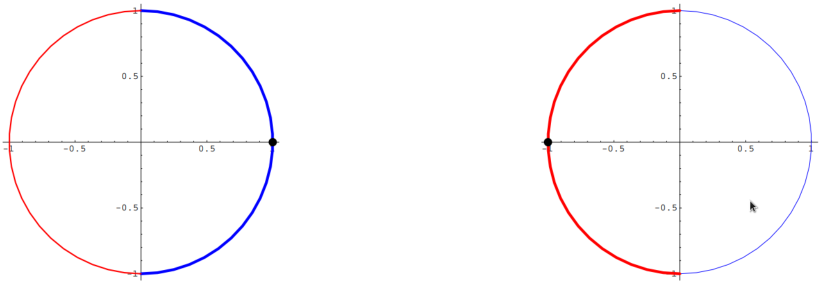

Figure 4: The thick arcs are the graphs of $y \mapsto \pm\sqrt{1-y^2}$

Discussion about the geometric meaning

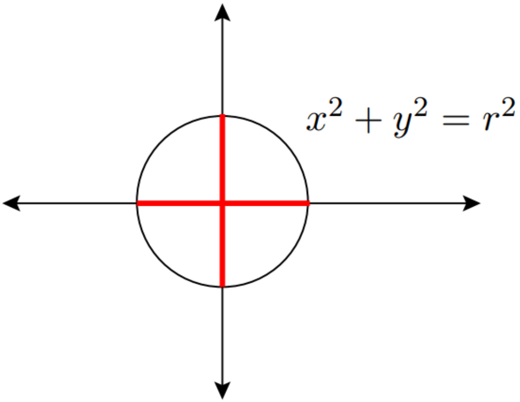

Figure 5: Can the thick line segments be a graph?

Why the Implicit Function Theorem is a great theorem?

Fact

With given k nonlinear equations in k unknowns:

- It is **impossible to solve**.
- It is often **impossible to determine whether it has any solutions**.

The Implicit Function Theorem allows us to (partly) reduce impossible questions about systems of nonlinear equations to straightforward questions about systems of linear equations.

Some special cases of the implicit function theorem

Theorem: Implicit function theorem, $m = n = 1$

Suppose that F is real-valued C^1 function defined for all (x, y) in open set $U \times V \subseteq \mathbb{R}^2$.

- If $f(a, b) = 0$ and $\partial_y f(a, b) \neq 0$, then the equation
$$F(x, y) = 0 \tag{34}$$
implicitly determines y as a C^1 function of x, i.e., $y = f(x)$, for x near a. Moreover $f(a) = b$.
- If $f(a, b) = 0$ and $\partial_x f(a, b) \neq 0$, then the equation
$$F(x, y) = 0 \tag{35}$$
implicitly determines x as a C^1 function of x, i.e., $x = f(y)$, for y near b. Moreover $f(b) = a$.

Some special cases of the implicit function theorem

Theorem: Implicit function theorem, $n = 2, m = 1$

Suppose that F is a scalar function of class C^1 function defined for all (x, y, z) in open set $U \times V \subseteq \mathbb{R}^3$.

- If $f(a, b, c) = 0$ and $\partial_z f(a, b, c) \neq 0$, then the equation

$$F(x, y, x) = 0 \tag{36}$$

  implicitly determines z as a C^1 function of (x, z), i.e., $z = f(x, y)$ for (x, y) near (a, b). Moreover, $f(a, b) = c$.
- Similarly, we also have results in the cases: $f(a, b, c) = 0$ and $\partial_y f(a, b, c) \neq 0$; and $f(a, b, c) = 0$ and $\partial_x f(a, b, c) \neq 0$

Some special cases of the implicit function theorem

Theorem: Implicit function theorem, $n = 1, m = 2$

Suppose that $\mathbf{F} = (F_1, F_2)$ is function $U \times V \to \mathbb{R}^2$ of class C^1 defined for all (x, y, z) in open set $U \times V \subseteq \mathbb{R}^3$.

- If $\mathbf{F}(a, b, c) = \mathbf{0}$ and

$$\begin{vmatrix} \partial_y F_1 & \partial_z F_1 \\ \partial_y F_2 & \partial_z F_2 \end{vmatrix} \neq 0 \tag{37}$$

  then

$$\mathbf{F}(x, y, z) = \mathbf{0} \Leftrightarrow \begin{cases} F_1(x, y, z) = 0 \\ F_2(x, y, z) = 0 \end{cases} \tag{38}$$

  implicitly determines (y, z) as a C^1 function of x, i.e., $(y, z) = \mathbf{f}(x)$, for x near a. Moreover, $\mathbf{f}(b, c) = f(a)$.
- Similarly, we also have results in the remain cases.

Computational example 1

Problem 01

Consider the equation

$$F(x, y, z) = xy + xz\ln(yz) = 1 \tag{39}$$

We know that $(1, 1, 1)$ is a solution. Does the equation implicitly determine z as a function $f(x, y)$ for (x, y) near $(1, 1)$ with $f(1, 1) = 1$? If so, find a formula for $\partial_x f(1, 1)$, and evaluate it at $(1, 1) = (1, 1)$.

Computational example 1

We have

$$\partial_z F = xln(yz) + x \tag{40}$$

And at $(x, y, z) = (1, 1, 1)$, $F(1, 1, 1) = 1$. So the Implicit Function Theorem guarantees that there is a function $f(x, y)$, defined for (x, y) near $(1, 1)$ such that

$$F(x, y, z) = 1 \text{ when } z = f(x, y) \tag{41}$$

To find $\partial_x f$, by using the original equation that defines z as a function of (x, y) to differentiate both sides with respect to x.

$$y + zln(yz) + x\frac{\partial z}{\partial x}ln(yz) + \frac{xz}{yz}y\frac{\partial z}{\partial x} = 0 \Leftrightarrow \frac{\partial z}{\partial x} = -\frac{y + zln(yz)}{x + xln(yz)} \tag{42}$$

Evaluating at $(x, y, z) = (1, 1, 1)$, and solving for $\frac{\partial z}{\partial x}$. We have:

$$1 + \frac{\partial z}{\partial x} = 0 \Leftrightarrow \frac{\partial z}{\partial x} = -1 \tag{43}$$

Computational example 2

Problem 02

Consider the system of equations

$$F_1(x, y, u, v) = xye^u + \sin(v - u) = 0 \quad (44)$$

$$F_2(x, y, u, v) = (x + 1)(y + 2)(u + 3)(v + 4) - 24 = 0 \quad (45)$$

We know that $(0, 0, 0, 0)$ is a solution. Does the system of equations implicitly determine $(u, v$ as a function of (x, y), i.e., $(u, v) = \mathbf{f}(x, y)$ for (x, y) near $(0, 0)$? If so, find a formula for $\partial_x \mathbf{f}(x, y)$ at $(x, y) = (0, 0)$

Computational example 2

Let $\mathbf{F} = \binom{F_1}{F_2}$. Then

$$\begin{pmatrix} \partial_u F_1 & \partial_v F_1 \\ \partial_u F_2 & \partial_v F_2 \end{pmatrix} = \begin{pmatrix} xye^u - \cos(v-u) & \cos(v-u) \\ (x+1)(y+2)(v+4) & (x+1)(y+2)(u+3) \end{pmatrix} \tag{46}$$

At $(x, y, u, v) = (0, 0, 0, 0)$,

$$\begin{pmatrix} \partial_u F_1(0,0,0,0) & \partial_v F_1(0,0,0,0) \\ \partial_u F_2(0,0,0,0) & \partial_v F_2(0,0,0,0) \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 8 & 6 \end{pmatrix} \tag{47}$$

This matrix is invertible, so the theorem guarantees that the equations implicitly determine (u, v) as a function of (x, y)

Computational example 2

Find $\partial_x \mathbf{f} = \binom{\partial_x f_1}{\partial_x f_2}$, where $\binom{u}{v} = \mathbf{f}(x, y) = \binom{f_1(x,y)}{f_2(x,y)}$. Considering equations below:

$$\begin{aligned} xye^u + \sin(v - u) &= 0 \\ (x+1)(y+2)(u+3)(v+4) - 24 &= 0. \end{aligned} \tag{48}$$

and differentiate everything with respect to x:

$$\begin{aligned} ye^u + (xye^u - \cos(v-u))\frac{\partial u}{\partial x} + \cos(v-u)\frac{\partial v}{\partial x} &= 0 \\ (y+2)(u+3)(v+4) + (x+1)(y+2)(v+4)\frac{\partial u}{\partial x} + (x+1)(y+2)(u+3)\frac{\partial v}{\partial x} &= 0. \end{aligned} \tag{49}$$

Computational example 2

At $(x, y, u, v) = (0, 0, 0, 0)$,

$$\begin{pmatrix} -1 & 1 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \end{pmatrix} = \begin{pmatrix} 0 \\ -24 \end{pmatrix}. \tag{50}$$

And

$$\begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \end{pmatrix} = \frac{1}{-14} \begin{pmatrix} 6 & -1 \\ -8 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ -24 \end{pmatrix} = -\begin{pmatrix} 12/7 \\ 12/7 \end{pmatrix}. \tag{51}$$

What is Transformations?

Example

Fact Let U, and V be two open subsets of $\mathbb{R}^n$. Considering functions as

$$\mathbf{f} : U \to V \tag{52}$$

is called transformation.
If $\mathbf{f}$ is a bijection (that is, both one-to-one and onto). Then implies that $\mathbf{f}^{-1} : V \to U$ exists.
And both $\mathbf{f}$, $\mathbf{f}^{-1}$ are class of C^1.

Example

Considering the transformation of Cartesian grid by using the linear mapping $\mathbf{f}(x, y) = (2y - x, x + y)$.

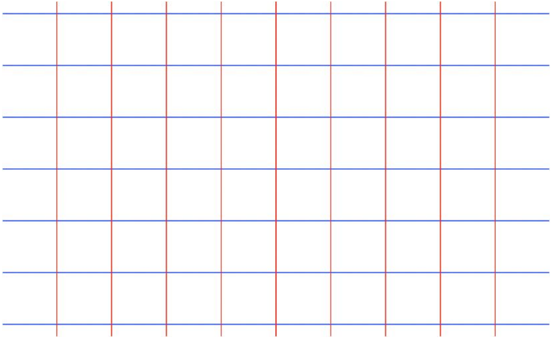

Figure 6: Before transformation.

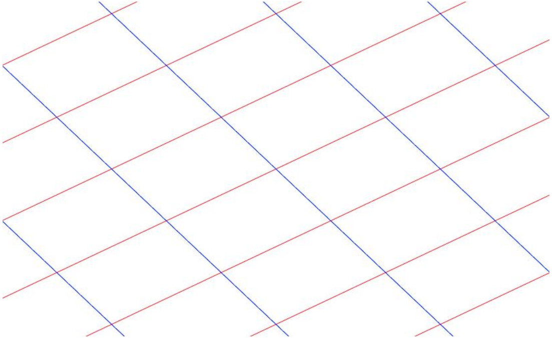

Figure 7: After transformation.

The Inverse Function Theorem

Theorem: The Inverse Function Theorem

Let $\mathbf{f}$ be a C^1 map from a neighborhood of $x_0 \in \mathbb{R}^n$ into $\mathbb{R}^n$. If $D\mathbf{f}(x_0)$ is non-singular, then there exist neighborhoods $U \ni x_0$ and $V \ni y_0 = \mathbf{f}(x_0)$ such that $\mathbf{f} : U \to V$ is a C^1 diffeomorphism[1], and

$$D\mathbf{f}^{-1}(y) = D\mathbf{f}(x)^{-1} \text{ for all} (x, y) \in U \times V, y = \mathbf{f}(x) \tag{53}$$

Moreover, if $\mathbf{f}$ is C^k, then $\mathbf{f}$ is a C^k diffeomorphism on U.

[1] A diffeomorphism is an isomorphism of smooth manifolds. It is an invertible function that maps one differentiable manifold to another such that both the function and its inverse are continuously differentiable.

Proof of The Inverse Function Theorem

We can define the function

$$F(x, y) = f(x) - y \tag{54}$$

and, find that

$$D_x F(x_0, y) = Df(x_0) \tag{55}$$

is non-singular. And apply the Implicit function theorem to F.

The usage of Implicit Function Theorem

Example

Problem Suppose that given $\mathbf{f}: U \to V$, and $\mathbf{y} \in V$, find $\mathbf{x}$ by solving

$$\mathbf{f}(\mathbf{x}) = \mathbf{y} \tag{56}$$

Example

Fact This problem is often an impossible problem to solve handle.

The usage of Implicit Function Theorem

The Inverse Function Theorem says

If we know that $\mathbf{f}(\mathbf{a}) = \mathbf{b}$, then for $\mathbf{y}$ near $\mathbf{b}$, the solvability of the **nonlinear system** can be established by considering a **much easier** question about linear algebra, whether the matrix $D\mathbf{f}(a)$ is invertible.

Computational example

Problem 03

Determine whether the system

$$\begin{cases} u(x, y, z) = x + xyz \\ v(x, y, z) = y + xy \\ w(x, y, z) = z + 2x + 3z^2 \end{cases} \tag{57}$$

can be solved for x, y, z in terms of u, v, w near $p = (0, 0, 0)$.

Computational example

Let $\mathbf{F}(x, y, z) = (u, v, w)$, then

$$D\mathbf{F}(p) = \begin{pmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{pmatrix}(p) = \begin{pmatrix} 1+yz & xz & xy \\ y & 1+x & 0 \\ 2 & 0 & 1+6z \end{pmatrix}(p) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \tag{58}$$

Due to,

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{vmatrix} = 1 \neq 0 \tag{59}$$

By the Inverse Function Theorem, the inverse function $F^{-1}(u, v, w)$ exists near $p = (0, 0, 0)$, i.e., we can solved for x, y, z in terms of u, v, w near $p = (0, 0, 0)$.

The Lyusternik theorem

Definition of tangent direction

Let M be a non-empty subset of $\mathbb{R}^n$ and $x \in M$. A vector $d \in R^n$ is called a tangent direction of M at x if there exist a sequence $x_n \in M$ converging to x and a non-negative seqence α_n such that

$$\lim_{n \to \infty} \alpha_n (x_n - x) = d \tag{60}$$

The tangent cone of M at x, donated by $T_M(x)$, is the set of all tangent direction of M at x.

The Lyusternik theorem

Theorem: The Lyusternik theorem

Let $\mathbf{F} : U \to \mathbb{R}^m$ be C^1 map, where $U \subset \mathbb{R}^n$ be an open set. Let $M = \mathbf{F}^{-1}(\mathbf{F}(x_0))$ be the level set of a point $x_0 \in U$. If the derivative $D\mathbf{F}(x_0)$ is a linear map onto $\mathbb{R}^m$, then the tangent cone of M is the null space of the linear map $D\mathbf{F}(x_0)$, that is,

$$T_M(x_0) = \{d \in \mathbb{R}^n : D\mathbf{F}(x_0)d = 0\} \tag{61}$$

The proof of Lyusternik theorem

If nescessary, we will consider the function

$$x \mapsto f(x + x_0) - f(x_0) \tag{62}$$

Assume that $x_0 = 0$, and $f(x_0) = 0$. And, define $A := Df(0)$

The proof of Lyusternik theorem

Now, we will prove that

$$T_M(0) \subseteq \text{Ker} A \tag{63}$$

If $d \in T_M(0)$, then there exists points $x(t) = td + o(t) \in M$. We have:

$$\begin{aligned} f(x) &= 0 \\ f(x(t)) &= 0 \\ f(td + o(t) + 0) &= 0 \\ tDf(0)(d) + o(t) + f(0) &= 0 \\ Df(0)(d) + \frac{o(t)}{t} &= 0 \\ \lim_{t \to \infty} \left(Df(0)(d) + \frac{o(t)}{t} \right) &= 0 \\ Df(0)(d) &= 0 \end{aligned} \tag{64}$$

The proof of Lyusternik theorem

Idea to prove the reverse inclusion: the equation $f(x) = 0$ can be written as $f(y, z) = 0$in a form that is suitable for applying the implicit function theorem.

The proof of Lyusternik theorem

Define $K := \text{Ker}A$, and $L := K^{\perp}$. Since $A := Df(0)$ is onto $\mathbb{R}^m$, we can identify K and L with $\mathbb{R}6n - m$ and $\mathbb{R}^m$ respectively, by introducing a suitable basis in $\mathbb{R}^n$.
For a point $x \in \mathbb{R}^n$ in form that $x = (y, z) \in K \times L$, we have $A = [D_y f(0), D_z f(0)]$, and

$$0 = A(K) = \{A(d_1, 0) : d_1 \in R^{n-m} = D_y f(0)(\mathbb{R}^{n-m}\} \tag{65}$$

so that $D_y f(0) = 0$. And due to A has rank m, so that $D_z f(0)$ is non-singular.

The proof of Lyusternik theorem

Following the Implicit function theorem, there exists neighborhoods $U_1 \subseteq \mathbb{R}^m$, and $U_2 \subseteq \mathbb{R}^{n-m}$ around the origin and a C^1 map $\alpha : U_1 \to U_2, \alpha(0) = 0$, such that $x = (y, z) \in U_1 \times U_2$ satisfies $f(x) = 0$ if and only if $z = \alpha(y)$. The equation $f(x) = 0$ can be written as $f(y, \alpha(y))$. Differentiating this equation

$$D_y f(y, \alpha(y)) + D_z f(y, \alpha(y) D_\alpha(y) = 0 \tag{66}$$

At $x = 0$, $D_y f(0) = 0$, and $D_z f(0)$ is non-singular, so that $D_\alpha(0) = 0$

The proof of Lyusternik theorem

If $|y|$ is small:

$$\alpha(y) = \alpha(0) + D\alpha(0)y + o(y) = o(y) \tag{67}$$

Let $d = (d_1, 0) \in K$, as $t \to 0$, the point $x(t) := (td_1, \alpha(td_1)) = (td_1, o(t))$ lies in M. And that is $f(x(t)) = 0$, and satisfies

$$\frac{x(t) - td}{t} = \frac{(0, o(t))}{t} \to 0 \tag{68}$$

This implies that $K \subseteq T_M(0)$.

The usage of the Lyusternik theorem

The usage of the Lyusternik theorem (finite version)

- Application in multi-objective optimization [2]

Conclusion

Conclusion & Future working

Conclusion

- Fully presentation about second conditional optimality
- Fully presentation about implicit function theorem, inverse function theorem, and Lyusternik theorem.

Future work

- Give more detail about the proof of these theories and theorems.
- Give more examples and applications of these theories and theorems.

References I

O. Güler.
Foundations of optimization, volume 258.
Springer Science & Business Media, 2010.

B. Jiménez and V. Novo.
A finite dimensional extension of lyusternik theorem with applications to multiobjective optimization.
Journal of mathematical analysis and applications, 270(2):340–356, 2002.