

Recitation Class for Extrema

Sometimes it is tiring to prepare RC (as reading the slides is very tiring), but it is an interesting process for me to "re-learn" the knowledge with you. Mathematics is not something far away from life, linear algebra and multi-variable are just around us. I hope what you get is not only some ways to calculate but a way to see the world. You know, the world is changing quickly, humans are developing at a rapid speed in language, image, and sound processing. You can find that many things can be seen as kind of linear, such as language. Maybe your life has no relation to physics problems, or even you don't care about engineering. But if you see the world as numbers or vectors, there are high-dimension spaces everywhere. Hope you have fun in VV285!



Stokes' theorem

Newton-Leibniz formula

$$\int_b^a f(x)dx = F(b) - F(a) \quad (1)$$

Green's theorem

Let $R \subset \mathbb{R}^2$ be a bounded, simple region and $\Omega \supset R$ an open set containing R . Let $F : \Omega \rightarrow \mathbb{R}^2$ be a continuously differentiable vector field. Then

$$\int_{\partial R^*} F d\vec{s} = \int_R \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) dx \quad (2)$$

where ∂R^* denotes the boundary curve of R with positive (counter-clockwise) orientation.

Stokes theorem

Let $\Omega \subset \mathbb{R}^3$ be an open set, $\mathcal{S} \subset \Omega$ a parametrized, admissible surface in \mathbb{R}^3 with boundary $\partial \mathcal{S}$ and let $F : \Omega \rightarrow \mathbb{R}^3$ be a continuously differentiable vector field. Then

$$\int_{\partial \mathcal{S}^*} F d\vec{s} = \int_{\mathcal{S}^*} \text{rot } F d\vec{A} \quad (3)$$

where the orientations of the boundary curve $\partial \mathcal{S}^*$ and the surface \mathcal{S}^* are chosen according to right hand law.

Guass's theorem

Let $R \subset \mathbb{R}^n$ be an admissible region and $F : \overline{R} \rightarrow \mathbb{R}^n$ a continuously differentiable vector field. Then

$$\int_R \text{div } F dx = \int_{\partial R^*} F d\vec{A}. \quad (4)$$

Exercise

Find the integral: $\int \int x^2 dydz + y^2 dx dz + z^2 dx dy$ on the surface $S : x^2 + y^2 + z^2 = a^2$ with the help of Guass's theorem.

General Case (Only for interest and not talked about)

For $v = P\vec{i} + Q\vec{j} + R\vec{k}$, we denote

$$\int_L \vec{v} \cdot d\vec{r} = \int_L Pdx + Qdy + Rdz = \int_L \omega_v^1 \quad (5)$$

$$\int \int_s \vec{v} \cdot d\vec{S} = \int \int_S Pdy \wedge dz + Qdx \wedge dz + Rdx \wedge dy = \int \omega_v^2 \quad (6)$$

So we can know

$$\begin{aligned} \oint_{\partial S} \vec{v} \cdot d\vec{S} &= \int \int_S \nabla \times \vec{v} \cdot d\vec{S} \\ \oint_{\partial S} \omega_v^1 &= \int_S \omega_{\nabla \times \vec{v}}^2 \end{aligned} \quad (7)$$

Notice that

$$d\omega_v^1 = \omega_{\nabla \times \vec{v}}^2 \quad (8)$$

We know

$$\int_{\partial S} \omega_v^1 = \int_S d\omega_v^1 \quad (9)$$

Similarly, we can deduce from Guass's theorem that

$$\int_{\partial S} \omega_v^2 = \int_S d\omega_v^2 \quad (10)$$

Generally, to n-dimension, we have

$$\int_{\partial \Omega} \omega = \int_{\Omega} d\omega \quad (11)$$

where Ω is a r dimension surface in \mathbb{R}^n

Green Identity

Let $R \subset \mathbb{R}^n$ be an admissible region and $u, v : \overline{R} \rightarrow \mathbb{R}$ be twice continuously differentiable potential functions. Then

$$\int_R \langle \nabla u, \nabla v \rangle dx = - \int_R u \cdot \Delta v dx + \int_{\partial R^*} u \frac{\partial v}{\partial n} dA \quad (12)$$

and

$$\int_R (u \cdot \Delta v - v \cdot \Delta u) dx = \int_{\partial R^*} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) dA. \quad (13)$$

(12) is commonly called *Green's first identity* and (13) *Green's second identity*.

Some Applications of VV285

- Machine Learning. An MLP(Multilayer Perceptron), or even a complex DNN(Deep Neural Network) is essentially a function with a set of parameters. To train the network, we essentially optimize these parameters to obtain a low score of the Loss function so that our model can best fit reality. The most famous training method is called Gradient descent. To avoid over-fitting, we also use regularization to constrain the network.
- Optimal control is a branch of control theory that deals with finding the best control inputs for a given system to optimize a certain objective or criterion. It involves determining the control strategy that minimizes or maximizes a performance measure, subject to system dynamics and constraints.
- Power System
- Operational Research (IOE)

Second Derivative

We can write

$$Df : x \mapsto (\nabla f(x))^T = Df|_x. \quad (14)$$

Hence, $Df = (\cdot)^T \circ \nabla f$ and we can differentiate Df by the chain rule. The derivative of the map

$$\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \Big|_x \\ \vdots \\ \frac{\partial f}{\partial x_n} \Big|_x \end{pmatrix} \quad (15)$$

can be easily calculated: assuming sufficient smoothness of f , it is just the Jacobian of ∇f .

Remark: We represent $Df|_x$ using $\nabla f(x)$ because we are familiar with how to differentiate a column vector.

We hence have

$$D(\nabla f)|_x = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} \Big|_x & \frac{\partial^2 f}{\partial x_2 \partial x_1} \Big|_x & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \Big|_x \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} \Big|_x & \frac{\partial^2 f}{\partial x_2 \partial x_n} \Big|_x & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \Big|_x \end{pmatrix} \in \text{Mat}(n \times n; \mathbb{R}). \quad (16)$$

Where

$$\frac{\partial^2 f}{\partial x_i \partial x_j} := \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j} \quad (17)$$

is the second partial derivative of f concerning x_j (first) and x_i (second). The matrix in equation is important enough to warrant a special name: It is called the *Hessian* of f and denoted by $\text{Hess } f(x)$.

Schwarz's Theorem

Let (X, V) be normed vector spaces and $\Omega \subset X$ an open set. Let $f \in C^2(\Omega, V)$. Then $D^2f|_x \in \mathcal{L}^{(2)}(X \times X, V)$ is symmetric for all $x \in \Omega$,

$$D^2f(u, v) = D^2f(v, u), \quad \text{for all } u, v \in X. \quad (18)$$

which means $\text{Hess } f(x) = (\text{Hess } f(x))^T$, implying that the Hessian of f at x is a symmetric matrix. Writing out the components of $\text{Hess } f(x)$, this means that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}. \quad (19)$$

In other words, if f is twice continuously differentiable, the order of differentiation in the second-order partial derivatives does not matter.

Free Extrema

Do you find it is sounds easy for you to find an extrema? Yes! In VV186, you have already learned the meaning of derivatives and extreme points. You can say, all the saddle points and extreme points you find is the result of linear approximation. Furthermore, to check whether the point is an extrema, second derivative is a necessity, which I believe you have already known in your VV186. Second derivative, is actually an quadratic approximation, which means you knows more properties locally about the function.

For *Quadratic Approximation*. Let $\Omega \subset \mathbb{R}^n$ be an open set and f in $C^2(\Omega, \mathbb{R})$. Then, as $h \rightarrow 0$,

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \text{Hess } f(x)h, h \rangle + o(h^2). \quad (20)$$

You can think about what the case will be like if all the things is one dimention. It is Taylor Expansion.

Quadratic Form

To better deal with the third item $\langle \text{Hess } f(x)h, h \rangle$, we introduce *quadratic form*.

Let $A \in \text{Mat}(n \times n, \mathbb{R})$. Then the *quadratic form induced by (A)* is defined as the map

$$Q_A := \langle \cdot, A(\cdot) \rangle, \quad x \mapsto \langle x, Ax \rangle = \sum_{j,k=1}^n a_{jk} x_j x_k, \quad x \in \mathbb{R}^n \quad (21)$$

Clearly, $Q_A(\lambda x) = \lambda^2 Q_A(x)$ for any $\lambda \in \mathbb{R}$. Note also that Q_A is continuous, because it is a polynomial in x_1, \dots, x_n .

Definition. A quadratic form Q_A induced by a matrix $A \in \text{Mat}(n \times n, \mathbb{R})$ is called

- *positive definite* if $Q_A(x) > 0$ for all $x \neq 0$,
- *negative definite* if $Q_A(x) < 0$ for all $x \neq 0$,
- *indefinite* if $Q_A(x_0) > 0$ for some $x_0 \in \mathbb{R}^n$ and $Q_A(y_0) < 0$ for some $y_0 \in \mathbb{R}^n$.

A matrix (A) is said to be negative definite / positive definite / indefinite if the induced quadratic form Q_A has the corresponding property.

Whether a matrix is a positive definite/negative definite/indefinite can be judged in some other ways, which is somewhat related to the eigenvector which will be introduced later. You can find it at the end of this RC and it is only some extension that is not related to the exams.

For matrix A which is defined on $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, $A \in \text{Mat}(2 \times 2, \mathbb{R})$, suppose A is symmetric, there is a way to judge whether Q_A is positive definite:

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Let $\Delta = \det A$. Then:

- (i) A positive definite $\Rightarrow a > 0$ and $\Delta > 0$
- (ii) A negative definite $\Rightarrow a < 0$ and $\Delta > 0$
- (iii) A indefinite $\Rightarrow \Delta < 0$

Exercise

Find the Quadratic form of $f(x, y) = 4x^2 + 4xy + y^2$ and judge whether it is positive definite.

Find the Quadratic form of $f(x, y) = 4x^2 + 12xy + 10y^2 + 6yz + 2z^2 + 4xz$ and judge whether it is positive definite.

Finding Free Extrema

Let $\Omega \subset \mathbb{R}^n$ and $f : \Omega \rightarrow \mathbb{R}$.

1. f is said to have a *(global) maximum* at $\xi \in \Omega$ if
$$x \in \Omega \quad \Rightarrow \quad f(x) \leq f(\xi).$$
2. f is said to have a *strict (global) maximum* at $\xi \in \Omega$ if
$$x \in \Omega \setminus \{\xi\} \quad \Rightarrow \quad f(x) < f(\xi).$$

3. f is said to have a *local maximum* at $\xi \in \Omega$ if there exists a $\delta > 0$ such that
$$x \in \Omega \cap B_\delta(\xi) \quad \Rightarrow \quad f(x) \leq f(\xi).$$
4. f is said to have a *strict local maximum* at $\xi \in \Omega$ if there exists a $\delta > 0$ such that
$$x \in \Omega \cap B_\delta(\xi) \setminus \{\xi\} \quad \Rightarrow \quad f(x) < f(\xi).$$

The function f is said to have a (strict) *global/local minimum* at ξ if $-f$ has a (strict) global/local maximum at ξ .

Essential Condition. Let $\Omega \subset \mathbb{R}^n$, $f : \Omega \rightarrow \mathbb{R}$, and $\xi \in \text{int } \Omega$. Assume that all partial derivatives of f exist at ξ and that f has a local extremum (maximum or minimum) in ξ . Then

$\nabla f(\xi) = 0$. If f is differentiable at ξ , this implies $Df|_\xi = 0$.

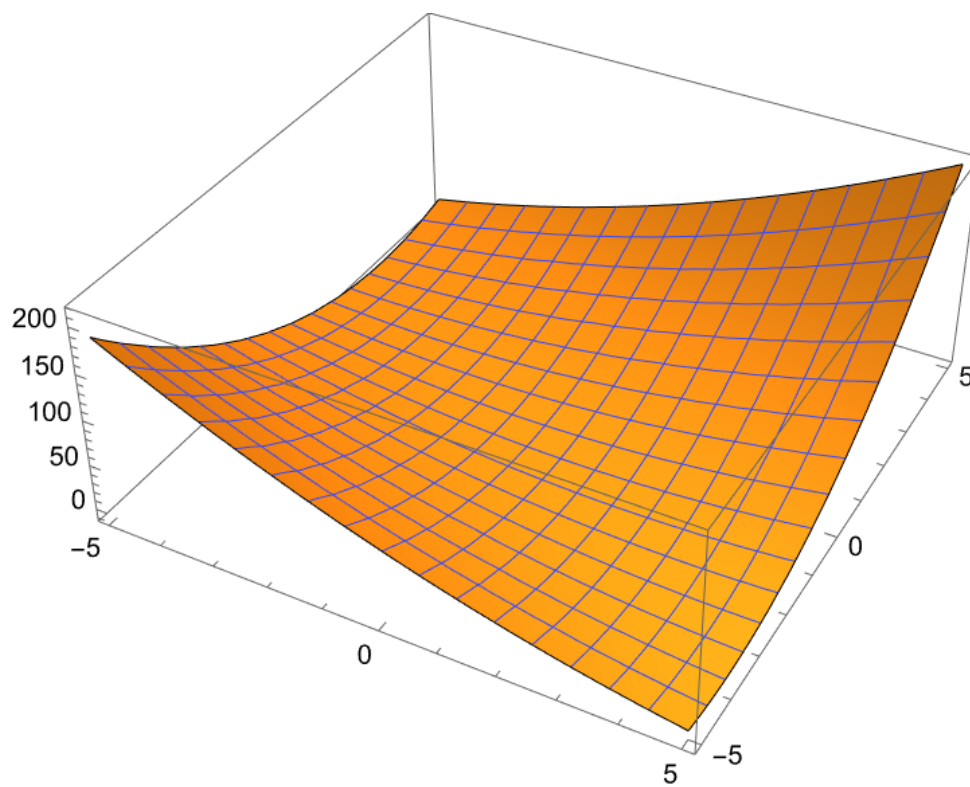
Hessian & Extrema. Let $\Omega \subset \mathbb{R}^n$ be open, $f \in C^2(\Omega)$, and $\xi \in \Omega$. Let $\nabla f(\xi) = 0$ (i.e., $Df|_\xi = 0$).

1. If $\text{Hess } f|_\xi$ is positive definite, f has a strict local minimum at ξ .
2. If $\text{Hess } f|_\xi$ is negative definite, f has a strict local maximum at ξ .
3. If $\text{Hess } f|_\xi$ is indefinite, f has no extremum at ξ .

Note: A critical point is not necessarily an extremum. $(0, 0)$ is a critical point for $y = x^3$. However, it is merely a *saddle point*.

Exercise

Find the extrema of $f(x, y) = x^2 + 3y^2 + 4xy + x + 2y + 1$.



Finding Constrained Extrema

The Normal Form of Question

In general, constrained optimization problems can be classified into equality constraints and inequality constraints. Here, we will focus on equality constraints. The general form of such problems is to find the extremum of $f(x)$ subject to the constraints $g_i(x) = 0$, where $i = 1, 2, 3, \dots, n$. We can write the problem as:

$$\begin{aligned} & \min(\max) f(x) \\ & s. t. g_i(x) = 0, i = 1, 2, 3, \dots, n \end{aligned} \quad (22)$$

Lagrange Multiplier Rule

Let $\Omega \subset \mathbb{R}^n$ be open and $f \in C^1(\Omega, \mathbb{R})$, $g \in C^1(\Omega, \mathbb{R}^m)$, $m < n$. Assume that f has a local extremum on the set $E = \{x \in \mathbb{R}^n : g(x) = 0\}$ at $\xi \in E$. Assume further that in

$$Dg|_{\xi} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix}, \quad (23)$$

there exists a submatrix consisting of m columns whose determinant does not vanish. Then there exist m numbers (called Lagrange multipliers)

$$\begin{aligned} & \lambda_1, \dots, \lambda_m \in \mathbb{R} \\ & Df|_{\xi} + \sum_{i=1}^m \lambda_i Dg_i|_{\xi} = 0. \end{aligned} \quad (24)$$

In order to find the constrained extrema, we need to solve the $m + n$ equations

These equations are equivalent to the following:

$$\begin{aligned} \frac{\partial f}{\partial x_i} + \lambda_1 \frac{\partial g_1}{\partial x_1} + \cdots + \lambda_m \frac{\partial g_m}{\partial x_i} &= 0, & i = 1, \dots, n, \\ g_j(x) &= 0, & j = 1, \dots, m. \end{aligned} \quad (25)$$

$$F(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) := f(x) + \lambda_1 g_1(x) + \cdots + \lambda_m g_m(x). \quad (26)$$

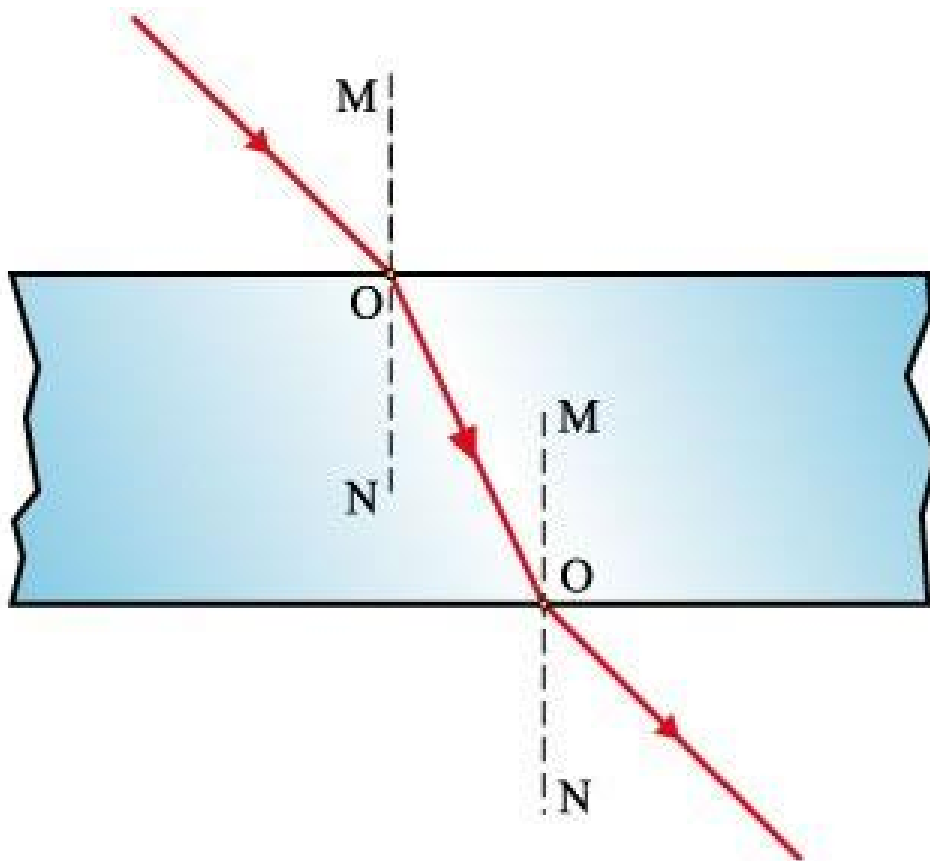
Then at an extremal point, all partial derivatives of F will vanish.

Exercise

Find the extrema of $f(x, y) = x^2 + 2y^2$, *s. t.* $x + y = 1$.

Use the Lagrange multiplier rule to show Snile's Law:

$$n_1 \cdot \sin(\theta_1) = n_2 \cdot \sin(\theta_2) \quad (27)$$



Do you remember in your physics course, Lagrange seems to be a big name? Yes! There is a subject called Lagrange mechanics and Lagrange quantity. **Joseph Louis Lagrange** is a great man in math, physics, astronomy, and the foundation of science. Born in Italy, he worked in France and Germany and was elected as a member of Fellow of the Royal

Society of London. He is a legend, and you can find his stories yourself. An interesting claim by him is, Mechanics is only a branch of Analysis.

Extra Content on Quadratic Form

Some of you may not know the eigenvector. Doesn't matter. You can have some ideas after solving this problem. (You can see this also a review of physics, and you may meet this exercise if you take VV286)

Exercise 2.7 Rigid-Body Dynamics

When a rigid body $B \subset \mathbb{R}^3$ rotates about the origin of a coordinate system the angular momentum \vec{L} is obtained to the angular velocity $\vec{\omega}$ through the relation

$$\vec{L} = I\vec{\omega}.$$

Here $I \in \text{Mat}(3 \times 3; \mathbb{R})$ is a symmetric matrix known as the *tensor of inertia*. Hence, the angular momentum depends on the axis of rotation defined by $\vec{\omega}$. The *principal axes of rotation* are precisely those axes for which the direction \vec{L} coincides with the direction of $\vec{\omega}$. Thus, they can be found by solving the eigenvalue problem $I\vec{\omega} = \lambda\vec{\omega}$.

The tensor of inertia $I = (I_{ij})$ for a solid body B with mass density $\varrho: B \rightarrow \mathbb{R}$ is given by

$$I_{ii} = \iiint_B (x_j^2 + x_k^2) \varrho(x) dx \quad \text{with } (i, j, k) \text{ pairwise distinct,} \quad I_{ij} = \iiint_B x_i x_j \cdot \varrho(x) dx$$

for $i, j, k = 1, 2, 3$.

- i) Consider a cube of side length $a > 0$, $B = [0, a]^3 \subset \mathbb{R}^3$, with mass m and constant density. Show that

$$I = \frac{ma^2}{12} \begin{pmatrix} 8 & -3 & -3 \\ -3 & 8 & -3 \\ -3 & -3 & 8 \end{pmatrix}.$$

(4 Marks)

- ii) Find the principal axes of rotation about the origin.

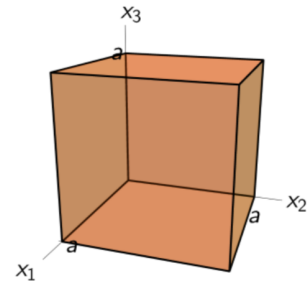
(3 Marks)

- iii) The *rotational energy* of B is given by the quadratic form

$$T = \frac{1}{2} \langle \vec{\omega}, I\vec{\omega} \rangle.$$

Which axis yields the highest value for the rotational energy? Which one the lowest? (See Exercise 2.6.)

(2 Marks)



Exercise 2. f.

i) $\rho = \frac{m}{a^3}$

$$I_{11} = \iiint_B (x_1^2 + x_3^2) \rho \, dX = \rho \iiint_B (x_1^2 + x_3^2) \, dX = \rho \left(\iiint_B x_1^2 \, dX + \iiint_B x_3^2 \, dX \right)$$

$$= \frac{m}{a^3} \left(a^2 \int_0^a x_1^2 \, dx_1 + a^2 \int_0^a x_3^2 \, dx_3 \right) = \frac{m}{a^3} \cdot a^4 \cdot \frac{2}{3} a^3 = \frac{2ma^2}{3}$$

$$I_{11} = I_{12} = - \iiint_B x_1 x_2 \rho \, dX = -\rho \iint_{\substack{x_1 \in [0,a] \\ x_2 \in [0,a]}} x_1 x_2 \, dx = -\rho \cdot \int_0^a \left(\int_0^a x_1 x_2 \, dx_1 \right) dx_2$$

$$= -\rho \int_0^a \frac{1}{2} a^2 x_2 \, dx_2 = -\frac{1}{4} a \cdot a^4 \cdot \rho = -\frac{1}{4} ma^2.$$

Similarly. $I_{11} = I_{13} = I_{23} = I_{31} = -\frac{1}{4} ma^2.$

$$\Rightarrow I = \frac{1}{12} ma^2 \begin{bmatrix} 8 & -3 & -3 \\ -3 & 8 & -3 \\ -3 & -3 & 8 \end{bmatrix}.$$

ii). For the matrix $A = \begin{bmatrix} 8 & -3 & -3 \\ -3 & 8 & -3 \\ -3 & -3 & 8 \end{bmatrix}.$

Suppose λ is its eigenvalue.

$$\det \begin{bmatrix} 8-\lambda & -3 & -3 \\ -3 & 8-\lambda & -3 \\ -3 & -3 & 8-\lambda \end{bmatrix} = 0.$$

$$(8-\lambda)^3 - 27 - 27 - 9(8-\lambda) \times 3 = 0. \Rightarrow (8-\lambda)^3 - 54 - 27(8-\lambda) = 0.$$

$$\Rightarrow \lambda_1 = \lambda_2 = 11, \quad \lambda_3 = 2.$$

① For λ_1, λ_2
For $\begin{bmatrix} 8 & -3 & -3 \\ -3 & 8 & -3 \\ -3 & -3 & 8 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = 11 \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$

$$\Rightarrow \begin{cases} 8u_1 - 3u_2 - 3u_3 = 11u_1 \\ -3u_1 + 8u_2 - 3u_3 = 11u_2 \\ -3u_1 - 3u_2 + 8u_3 = 11u_3 \end{cases} \Rightarrow u_1 + u_2 + u_3 = 0.$$

let two eigen vectors be $\left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$ and $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right).$

$$\textcircled{2} \text{ For } \lambda_3 = 2, \begin{bmatrix} 8 & -3 & -3 \\ -3 & 8 & -3 \\ -3 & -3 & 8 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = 2 \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$\text{let } (u_1, u_2, u_3) = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)$$

ζ_0 : principle axes are

$$\left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{2}}, 0 \right), \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{6}} \right) \text{ and } \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right).$$

$$\text{iii)} \quad T = \frac{1}{2} \langle \vec{\omega}, I \vec{\omega} \rangle.$$

From Exercise 2.6 we know, this occurs when x is eigenvectors

After testing we found for $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right) \Rightarrow$ lowest

$\left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{6}} \right)$ and $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{6}}, 0 \right) \Rightarrow$ highest.

The eigenvector and eigenvalue of a matrix A can be described as, if

$$\lambda v = Av \quad (28)$$

Then λ is called an eigenvalue and v is called a eigenvector. You can find in the exercise above, The eigenvectors can give the highest or lowest rotation energy. And also, you may discover the rotation energy is a quadratic form.

For quadratic forms, if all the eigenvalues are positive, then it is positive definite.

References

- VV285, slide. Horst Hohberger
- RC, slides. TA-Pingbang Hu
- RC, slides. TA-Chen Yuxiang

For further questions, you can contact me through WeChat.



Fragments

芬兰



扫一扫上面的二维码图案，加我为朋友。