

珞珈数学

总二十卷

武汉大学珞珈数学研习会

目录

基础数学论文

1

施图姆定理的证明和改进 3

从复环面到复 Abel 簇 (节选) 7

关于不完备内积空间的一些讨论 25

椭圆方程的极大值原理与 Harnack 不等式 (节选) 27

应用数学论文

51

从极值到变分 53

小波分析原理及应用研究 62

优秀数模论文

69

2014 Mathematical Contest in Modeling The Keep-Right-Except-To-Pass Rule ... 71

编委

分管组织

珞珈数学研习会

负责人

刘智彬

主编

张如心

副主编

鲍恒涛 王颖

技术指导

钱立龙

征稿及初审

李静 周鑫 郝子墨

校对

鲍恒涛 王颖

文字编辑

曹鑫林 王文博 梁琪 陈喆 刘政云 朱果 罗永石

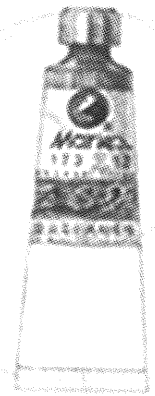
美术编辑

潘悦然

特别鸣谢

李彬老师 胡元明老师

基
础
数
学
论
文



施图姆定理的证明和改进

基础数学系 李彬

¹ **摘要** 谈到施图姆定理, 一些同学或许未必记得定理的内容, 甚至对这个定理没什么深刻的印象。这并不奇怪, 因为它在我们的学习过程中只作为工具短暂地出现过, 无论是教材上还是教师的授课过程中通常都不会提到它的证明 (甚至可能根本不会提到这个定理)。

在回忆施图姆定理的内容之前, 我们先来追根溯源。我们知道代数学研究的核心问题是代数方程的求解, 例如线性代数关心的是一次方程 (组) 的求解, 对一元多项式方程的求解也最终催生了抽象代数这门学科。如果只关心方程根的存在性或者个数多少, 那么关于复系数一元多项式, 我们有代数学基本定理保证了其一定有根; 而对于实系数一元多项式, 分析理论里的介值定理可以帮助我们定位实根所在的区间; 如果运气足够好, 恰好选到两个数 $a < b$, 使得 $f(a)f(b) < 0$, 那么区间 (a, b) 中必有实根。施图姆定理可谓对介值定理的一种强大推广, 它的内容如下:

若有非常值多项式 $f(x) \in \mathbb{R}[x]$, 令 $f_0(x) = f(x), f_1(x) = f'(x)$ 。利用欧几里得带余除法可得到施图姆序列 $f_0(x), f_1(x), \dots, f_m(x)$:

$$\begin{aligned} f_0(x) &= d_1(x)f_1(x) - f_2(x), & \deg(f_2) < \deg(f_1), \\ f_1(x) &= d_2(x)f_2(x) - f_3(x), & \deg(f_3) < \deg(f_2), \\ & \dots, \\ f_{m-2}(x) &= d_{m-1}(x)f_{m-1}(x) - f_m(x), & \deg(f_m) < \deg(f_{m-1}), \\ f_{m-1}(x) &= d_m(x)f_m(x). \end{aligned}$$

对于任意实数 c , 记 $V_f(c)$ 为 $f_0(c), f_1(c), \dots, f_m(c)$ 去掉其中的 0 后剩余子序列的变号数 (相邻两项异号即为一个变号)。

Theorem(施图姆定理). 对于任意实数 $a < b$, 若 a, b 均不是常值多项式 $f(x)$ 的重根, 则在区间 $(a, b]$ 中共有 $V_f(a) - V_f(b)$ 个不同的实根。

(比对一下陈志杰主编的《高等代数与解析几何 (下册)》[1] 中的该定理, 不难发现定理条件有区别, 区间也不同, 事实上教材是有问题的, 孟子言: 尽信书不如无书, 诚不我欺!)

Sturm 关于该定理的原始证明我没有找到, 但是有些资料, 比如 Jacobson 的著作 [2] 或者维基百科中都可以查到该定理的证明。本文中我想给出一个代数味道浓厚些, 不同与之前版本的证明。(篇幅所限, 只能证明梗概, 细节待有心读者完善。)

第 1 节

准备工作

证明施图姆定理之前, 我们先来谈另一个看似无关实则紧密联系的问题: 如何通过主子式计算实对称矩阵的正负惯性指数?

对于满秩的 n 阶实对称阵 A , 若能找到一个主子式序列 D_1, D_2, \dots, D_n 满足 D_i 为 D_{i+1} 的 i 阶子式, 且 D_i 均非 0, 则不难证明 A 的负惯性指数 $q(A)$ 就等于 D_1, D_2, \dots, D_n 的变号个数 $+\delta_{-1, \text{sgn}(D_1)}$ 。

¹ 作者简介: 李彬, 博士, 2010 年毕业于清华大学数学科学系, 现为武汉大学数学与统计学院讲师, 研究方向为李代数、量子群及表示理论。

对于秩为 r 的 n 阶实对称阵 A , 一个主子式序列 D_1, D_2, \dots, D_n 满足 D_i 为 D_{i+1} -的 i 阶子式, $D_r \neq 0$ 且 D_1, \dots, D_r 中任意相邻两项不同时为 0, 则我们称之为 A 的标准主子式序列。这样的序列一定是存在的! (只需证明如下结论: 秩为 r 的实对称阵必有 r 阶非 0 主子式, 另外, n 阶满秩实对称阵必有 $n-1$ 或 $n-2$ 阶非 0 主子式。)

如下引理可以帮助我们由标准主子式序列出发, 计算实对称阵的负惯性指数:

$$q(A) = \delta_{-1, \text{sgn}(D_1)} + \sum_{i=1}^{r-1} \delta_{-1, \text{sgn}(D_i D_{i+1})} + \sum_{i=1}^{r-1} \delta_{0, D_i}.$$

Lemma 1.1. 若 A 为 n 阶实对称阵, $r(A) \geq n-1$, $B = \begin{pmatrix} A & \alpha \\ \alpha^T & b \end{pmatrix}$ 且 $|A|$ 与 $|B|$ 不全为 0, 则

$$q(B) = \begin{cases} q(A) & \text{若 } |A||B| > 0 \text{ 或 } |B| = 0 \\ q(A) + 1 & \text{若 } |A||B| < 0 \text{ 或 } |A| = 0 \end{cases}$$

下面, 我们来定义任意两个实值多项式 $f(x), g(x)$ 所对应的施图姆序列, 即令 $f_0(x) = f(x), f_1(x) = g(x)$, 通过欧几里得除法得到的满足如下条件的多项式序列 f_0, \dots, f_m :

$$\begin{aligned} f_0(x) &= d_1(x)f_1(x) - f_2(x), & \deg(f_2) < \deg(f_1), \\ f_1(x) &= d_2(x)f_2(x) - f_3(x), & \deg(f_3) < \deg(f_2), \\ & \dots, \\ f_{m-2}(x) &= d_{m-1}(x)f_{m-1}(x) - f_m(x), & \deg(f_m) < \deg(f_{m-1}), \\ f_{m-1}(x) &= d_m(x)f_m(x). \end{aligned}$$

关于此序列我们不难发现以下若干简单事实。

- (a) 若 a 为 $f(x)$ 和 $g(x)$ 的公共根, 则 $f_i(a)$ 均为 0; 否则, 任意相邻两项 $f_i(a)$ 和 $f_{i+1}(a)$ 不同时为 0。
 - (b) $f_m(x) = \gcd(f(x), g(x))$, 因此 a 是 $f(x), g(x)$ 的公共根当且仅当 $f_m(a) = 0$ 。
 - (c) 若有某 $1 \leq i \leq m-1$ 满足 $f_i(a) = 0$ 且 $f_m(a) \neq 0$, 则 $f_{i-1}(a)f_{i+1}(a) < 0$ 。
- 对于任意实数 c , 记 $V_f(c)$ 为 $f_0(c), f_1(c), \dots, f_m(c)$ 去掉其中的 0 后剩余子序列的变号数。

对于 $f(x), g(x) \in \mathbb{R}[x]$, 我们定义如下矩阵

$$\begin{pmatrix} d_1(x) & 1 & & & \\ & 1 & d_2(x) & & \\ & & & \ddots & \\ & & & & 1 \\ & & & & & 1 & d_m(x) \end{pmatrix}$$

不妨称之为 Sturm 矩阵, 记为 $S_{f,g}(x)$ 。若令 $\tilde{f}_i = \frac{f_i}{f_m}$, 则 $\tilde{f}_0, \tilde{f}_1, \dots, \tilde{f}_m = 1$ 恰为 \tilde{f}_0, \tilde{f}_1 所对应的施图姆序列, 且有 $S_{f_0, f_1}(x) = S_{\tilde{f}_0, \tilde{f}_1}(x)$ 。若 a 不是 $f(x), g(x)$ 的公共根, 则有

$$V_{f,g}(a) = V_{\tilde{f}, \tilde{g}}(a)$$

其中 $\tilde{f} = \tilde{f}_0, \tilde{g} = \tilde{f}_1$ 。

给定 $x \in \mathbb{R}, S_{f,g}(x)$ 为实对称阵, 取其最后 i 行和最后 i 列交叉处的主子式记为 $D_i(x)$, 如下引理易证明

Proposition 1.2. 对任意 $1 \leq i \leq m$, $D_i(x) = \tilde{f}_{m-i}$ 。

由上述关于施图姆序列的事实可知, 若 a 非 $f(x), g(x)$ 的公共根, 则 $D_1(a), \dots, D_m(a)$ 为 $S_{f,g}(a)$ 的一个标准主子式序列, 进而利用公式 (1) 我们可以得到下列结论。

Theorem 1.3. 若 a 非 $f(x), g(x)$ 的公共根, 则变号数 $V_{f,g}(a)$ 恰为矩阵 $S_{f,g}(a)$ 的负惯性指数。

第2节

施图姆定理的证明

设 $a_1, \dots, a_k \in \mathbb{R}$ 为 $\tilde{f}_0(x)$ 的全部不同实根且有

$$a_1 < a_2 < \dots < a_k.$$

令 $a_0 = -\infty, a_{k+1} = +\infty$, 对于任意 $0 \leq i \leq k$, 记 I_i 为开区间 (a_i, a_{i+1}) 。

Proposition 2.1. $q(S_{f,g}(x))$ 在区间 I_i 为常值 (记为 q_i)。

proof. 对于任意 $x \in \mathbb{R}$, 由于 $S_{f,g}(x)$ 为实对称阵, 故 $S_{f,g}(x)$ 有 m 个实特征值

$$\lambda_1(x) \leq \lambda_2(x) \leq \dots \leq \lambda_m(x).$$

又知 $|S_{f,g}(x)| = \tilde{f}_0(x) = \lambda_1(x) \cdots \lambda_m(x) \neq 0$, 且由 $\tilde{f}_0(x)$ 的连续性知 $|S_{f,g}(x)|$ 在区间 I_i 中不变号。另外每个 $\lambda_i(x)$ 都连续地依赖于 x , 因此 $\lambda_i(x)$ 在 I_i 中不变号, 从而 $S_{f,g}(x)$ 在区间 I_i 中负特征值的个数为定值, 即负惯性指数为定值。

Proposition 2.2. 若 $c \in \mathbb{R}$ 为 \tilde{f}_0 的根, 则存在 $\varepsilon > 0$ 使得 $f(x), g(x)$ 在 $(c - \varepsilon, c) \cup (c, c + \varepsilon)$ 上均非 0; 进一步有, 对于 $x \in (c - \varepsilon, c) \cup (c, c + \varepsilon)$,

$$q(S_{f,g}(x)) = \begin{cases} q(S_{f,g}(c)) & \text{若 } f(x)g(x) > 0 \\ q(S_{f,g}(c)) + 1 & \text{若 } f(x)g(x) < 0 \end{cases}$$

(提示: 注意到 $D_{m-1}(x) = \tilde{f}_1(x)$ 在 c 点附近非 0, 故其对应的 $m-1$ 阶主子矩阵在 c 点附近正负惯性指数为定值, 再利用 Lemma 1.1.)

对于 $1 \leq j \leq k$, 令 $q(S_{f,g}(a_j)) = t_j$, 我们有以下推论。

Corollary 2.3. 对于 $1 \leq i \leq k$,

$$t_i = \begin{cases} q_{i-1} & \text{若在某区间 } (a_i - \varepsilon, a_i) \text{ 上 } f(x)g(x) > 0 \\ q_{i-1} + 1 & \text{若在某区间 } (a_i - \varepsilon, a_i) \text{ 上 } f(x)g(x) < 0 \end{cases}$$

$$= \begin{cases} q_i & \text{若在某区间 } (a_i, a_i + \varepsilon) \text{ 上 } f(x)g(x) > 0 \\ q_i + 1 & \text{若在某区间 } (a_i, a_i + \varepsilon) \text{ 上 } f(x)g(x) < 0 \end{cases}$$

若 $g(x) = f'(x)$, 则 $\tilde{f}_0(x)$ 与 $f(x)$ 恰有完全相同的实根且重数均为 1。易知实值函数 $f^2(x)$ 在 a_i 处取到最小值, 故存在某 $\varepsilon > 0$, 使得在区间 $(a_i, a_i + \varepsilon)$ 上 $(f^2(x))' = 2f(x)f'(x) > 0$, 在区间 $(a_i - \varepsilon, a_i)$ 上 $(f^2(x))' = 2f(x)f'(x) < 0$ 。在先前讨论的基础上, 我们给出如下定理。

Theorem 2.4(施图姆定理). 若 $a_1 < a_2 < \dots < a_k$ 为 $f(x) \in \mathbb{R}[x]$ 的全部不同实根,

(i) 对于 $1 \leq i \leq k$, 在区间 $[a_i, a_{i+1})$ 上 $q(S_{f,f'}(x)) \equiv q_i$, 在区间 $(-\infty, a_1)$ 上 $q(S_{f,f'}(x)) \equiv q_0$;

(ii) 对于任意 $0 \leq i \leq k-1$, $q_i - q_{i-1} = 1$;

(iii) 对于任意 $a < b$, $f(x)$ 在区间 $(a, b]$ 中不同实根的个数为 $q(S_{f,f'}(a)) - q(S_{f,f'}(b))$ 。

细心的话大家会发现此定理与本文最初给的定理形式不同, 这里并不要求 a, b 不是 $f(x)$ 的重根, 这说明我们所定义的 Sturm 矩阵的负惯性指数比施图姆序列的变号数要更一般, 在计算实根个数时也更强大。

第 3 节

References

- 1 陈志杰,《高等代数与解析几何(下册)》,第 2 版,高等教育出版社,2008.
- 2 Nathan Jacobson, Basis Algebra I, 2nd Edition, W.H Freeman and Com- pany, New York, 1985.

从复环面到复 Abel 簇 (节选)

2010 级弘毅班 余红杰

第 1 节

引言

一个 Abel 簇是指一个具有群结构的射影代数簇, 且群法则是正则映射紧复。而由于一个紧复李群必是复环, 再结合周炜良定理, 可知可嵌入射影空间的复环就是此类对象。因此复线性空间成为它的万有覆盖, 复环与 Abel 簇的理论将很大程度划归为对复线性空间与其内格的讨论。

本文中最重要的定理是 Appell-Humbert 定理, 它将完全决定了复环上线丛的结构。并且从 Appell-Humbert 定理我们可以看到复环与线性代数及 Abel 簇的紧密联系。因此我们这篇文章可以看做对 Appell-Humbert 定理的讨论, 第一章可视为为了证明 Appell-Humbert 定理所作的铺垫, 而后我们将在第二章一开始证明 Appell-Humbert 定理, 并把剩下的内容专注于对 Appell-Humbert 定理作一些应用。读者将会发现, 在理解了 Appell-Humbert 定理之后, 其余的定理都是显而易见的了。

为了更好的理解本文, 读者需要具备基本的复几何的知识, 比如 [13] 的前三章以及 [7] 的前四章中所涉及到的知识; 以及基本的层论与 Čech 上同调的知识, 这些可在 [5] 中找到, [13] 的第二章也可作为快速的导论。最后, 读者需要具有基本的范畴与同调代数知识, 可参考文献 [3] 与 [6]。

由于篇幅限制我们不可能面面俱到, 能力所限不足之处更不可避免。但还是希望通过本文的编排, 能够使读者感觉到除了个别需要稍微思考的定理之外, 每个结论都是显而易见的, 但是我们最后达到的目标却是很不平凡的!

第 2 节

复环的几何

2.1

复环上的同调

我们分别用 $H_r(X, \mathbb{Z}), H^r(X, \mathbb{Z})$ 指代 r 阶整系数奇异同调与奇异上同调, $H_{DR}^r(X, \mathbb{C})$ 为 r 阶复系数 de Rham 上同调. 记 $\varepsilon, \varepsilon^n$ 分别为光滑函数层与 n 阶 (复值) 光滑微分形式层, \mathbb{Z}, \mathbb{C} 为通常意义下的局部常值层.

我们下面来求出复环的一些上同调, 以供使用:

首先给出一些我们需要的定理, 这些定理的我们不详细证明, 因为任何一本标准教科书都可以找到证明, 只要读者承认本节前几个结论, 就可以不妨碍理解地继续阅读.

定理 1 de Rham-Weil 定理

[13] 设 $0 \rightarrow \mathcal{M} \rightarrow \mathcal{A}^*$ 为 X 上的层 \mathcal{M} 的消解, 且所有的 \mathcal{A}^q 都是零调的 (即 $H^{>0}(X, \mathcal{A}^q) = 0$), 则

$$H^n(X, \mathcal{M}) \cong H^n(\Gamma(\mathcal{A}^*))$$

且若 $0 \rightarrow \mathcal{N} \rightarrow \mathcal{B}^*$ 是另一组 \mathcal{N} 的零调消解, 使得

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathcal{M} & \longrightarrow & \mathcal{A}^* & & \\ & & \downarrow f & & \downarrow & & \\ 0 & \longrightarrow & \mathcal{N} & \longrightarrow & \mathcal{B}^* & & \end{array}$$

交换, 则上面定义的同构给出如下的交换图:

$$\begin{array}{ccc} H^n(X, \mathcal{M}) & \xrightarrow{\sim} & H^n(\Gamma(\mathcal{A}^*)) \\ \downarrow f_* & & \downarrow \\ H^n(X, \mathcal{N}) & \xrightarrow{\sim} & H^n(\Gamma(\mathcal{B}^*)) \end{array}$$

且当 f 是同构时, 该图所有的映射都是同构.

推论 1

$$\begin{aligned} H^r(X, \mathbb{Z}) &\cong H^r(X, \mathbb{Z}) \\ H^r(X, \mathbb{C}) &\cong H^r(H^0(X, \varepsilon^*)) = H_{DR}^r(X, \mathbb{C}) \end{aligned}$$

定理 2 de Rham 定理

设 X 为一流形. 则如下定义的映射是有意义的

$$\begin{aligned} f : H_r(X, \mathbb{Z}) \times H_{DR}^r(X, \mathbb{C}) &\longrightarrow \mathbb{C} \\ (c, [\omega]) &\longmapsto \int_c \omega \end{aligned}$$

且该映射可使我们得到

$$H_{DR}^r(X, \mathbb{C}) \cong \text{Hom}_{\mathbb{Z}}(H_r(X, \mathbb{Z}), \mathbb{C}) = H^r(X, \mathbb{C})$$

其中最后的一个等号由万有系数定理得到, 因此我们有一个交换图:

$$\begin{array}{ccc} H^r(X, \mathbb{Z}) & \longrightarrow & H^r(X, \mathbb{C}) \\ \downarrow \wr & & \downarrow \wr \\ H^r(X, \mathbb{Z}) & \xrightarrow{i} & H_{DR}^r(X, \mathbb{C}) \end{array}$$

我们把 $H_{DR}^r(X, \mathbb{C})$ 中包含在 i 中的形式称为整的微分形式.

定理 3 *DD* 引理

设 X 为一个 Kähler 流形, (也就是说其上存在一个闭的 (1,1)-形式) 对于一个微分形式 ω 下面几点等价:

- a) $\omega \in \text{Im}d$
- b) $\omega \in \text{Im}\partial$
- c) $\omega \in \text{Im}\bar{\partial}$
- d) $\omega \in \text{Im}\partial\bar{\partial}$

这个定理依赖于 Hodge 分解, 请参见 [13]. 注意到复空间 V 是 Kähler 流形, 因此复环 $X = V/\Gamma$ 也是一个 Kähler 流形.

设 Γ 是 V 中的格, 让 Γ 以平移的方式作用在 V 上, 则此作用是无不动点的不连续作用. 商空间 V/Γ 同胚于 $(S^1)^n$, 其中 $n = \dim V$ 且自然投射 $\pi: V \rightarrow V/\Gamma$ 是覆盖映射. 由于 V 单连通, 这是万有覆盖. 因此

$$\Gamma \cong \pi_1(V/\Gamma)$$

其中 Γ 中的一点与连接 0 到这点的线段投射在 V/Γ 上的圈所对应的等价类相对应.

定理 4

由万有系数定理,

$$H^1(X, \mathbb{Z}) \cong \text{Hom}_{\mathbb{Z}}(\pi_1(X), \mathbb{Z}) = \Gamma^{\vee}$$

因此我们利用 Künneth 公式

$$H^r(X, (\mathbb{Z})) \cong \bigwedge^r \Gamma^{\vee}$$

证: 见 [10]. □

我们将通过 $\pi: V \rightarrow X$ 将 V 与 $T_e X$ 等同起来, 然后按照惯例, 用 V 内的平移 r_x 将 V 与 $T_x X$ 等同起来.

定理 5

令 $X = V/\Gamma$ 为一复环, 其上的全纯函数层记为 \mathcal{O}_X , 则:

$$\begin{array}{ccc} H^r(X, \mathbb{C}) & \longrightarrow & H^r(X, \mathcal{O}_X) \\ \downarrow \wr & & \downarrow \wr \\ \bigwedge^r V^* & \longrightarrow & \bigwedge^r \bar{V}^{\vee} \end{array}$$

交换. 具体的构造将在定理证明中指出.

证: 为了证明该定理, 我们考虑 Dolbeault 消解, 以及 de Rham 消解:

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathcal{O} & \longrightarrow & \varepsilon^{0,0} & \xrightarrow{\bar{\partial}} & \varepsilon^{0,1} & \xrightarrow{\bar{\partial}} & \varepsilon^{0,2} & \xrightarrow{\bar{\partial}} & \dots \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathcal{O} & \longrightarrow & \varepsilon^{0,0} & \xrightarrow{\bar{\partial}} & \varepsilon^{0,1} & \xrightarrow{\bar{\partial}} & \varepsilon^{0,2} & \xrightarrow{\bar{\partial}} & \dots \end{array}$$

其中 $\varepsilon^{p,q}$ 指 C^∞ 复值 (p,q) 型微分形式层 ($\varepsilon^{0,0}$ 就是为光滑函数层 ε)，这是一个好层，故零调. 因此由 De Rahm-Weil 定理，我们有一个映射

$$\begin{array}{ccc} H^r(X, \mathcal{O}_X) & \xrightarrow{\sim} & H^r(H^0(X, \varepsilon^{0,*})) \\ \downarrow & & \downarrow \\ H^r(X, \mathbb{C}) & \xrightarrow{\sim} & H^r(H^0(X, \varepsilon^{0,*})) \end{array} \quad (1)$$

利用如下同构:

$$\phi_p: \varepsilon \otimes_{\mathbb{C}} \bigwedge^p V^* \longrightarrow \varepsilon^p$$

它把 $\sum f_i \otimes \alpha_i$ 变成 $\sum f_i \omega_{\alpha_i}$ 这里 ω_{α} 指平移不变形式 $x \mapsto \tau_x^* \alpha$ ，于是我们可以把 \int_X 扩张成 $\varepsilon^p \rightarrow \bigwedge^p V^*$ ，它将 $\varepsilon_{p,q}$ 映入 $\bigwedge^{p,q} V^*$.

取 V 的一组基，利用覆盖对 X 参数化，不妨设 $\int_X dx = 1$ ，定义映射如下:

$$\begin{array}{ccc} \int_X: \varepsilon & \longrightarrow & \mathbb{C} \\ & & g \longmapsto \int_X g \end{array}$$

我们通过平移不变形式将 $\bigwedge^p V^*$ 等同于 ε^p 的一个子空间. 当 ω 为恰当的时候, $\int_X \omega = 0$ 因此 \int_X 产生出一个同态 $H^p(X, \mathbb{C}) \rightarrow \bigwedge^p V^*$. 现在由于每个平移不变形式是闭的. 我们知道这个同态是满的. 为了证明这个同态是单的, 只需作如下考虑: □

引理 1

设 ω 是 X 上一个闭的微分形式, $a \in X$, τ_a 为 X 上的平移, 则 $\tau_a^* \omega - \omega$ 是恰当的.

证: 我们设 $\tilde{a} \in V$, 于是 $\tau_{\tilde{a}}$ 乃是由 V 上的平移 $\tau_{\tilde{a}}$ 诱导, 置

$$\eta_{\tilde{a}}(x)(x_1, \dots, x_{r-1}) = \int_0^1 \pi^* \omega(x + t\tilde{a})(\tilde{a}, x_1, \dots, x_{r-1}) dt$$

于是 $\eta_{\tilde{a}}$ 是一个 Γ -周期的 $(r-1)$ -形式, 且满足:

$$\begin{aligned} d\eta_{\tilde{a}}(x)(x_1, \dots, x_r) &= \sum_{j=1}^r (-1)^{j-1} (\eta'_{\tilde{a}}(x) \cdot x_j)(x_1, \dots, \hat{x}_j, \dots, x_r) \\ &= \sum_{j=1}^r (-1)^{j-1} \int_0^1 (\pi^* \omega'(x + t\tilde{a}) \cdot x_j)(\tilde{a}, x_1, \dots, \hat{x}_j, \dots, x_r) dt \\ &= \int_0^1 (\pi^* \omega'(x + t\tilde{a}) \cdot a)(x_1, \dots, x_r) dt \\ &= \pi^* \omega(x + \tilde{a})(x_1, \dots, x_r) - \pi^* \omega(x)(x_1, \dots, x_r) \end{aligned}$$

其中, 第三个等号处用到:

$$\begin{aligned} 0 &= d\pi^* \omega(x + t\tilde{a})(\tilde{a}, x_1, \dots, x_r) \\ &= (\pi^* \omega'(x + t\tilde{a}) \cdot \tilde{a})(x_1, \dots, x_r) - \sum_{j=1}^r (-1)^{j-1} (\pi^* \omega'(x + t\tilde{a}) \cdot x_j)(\tilde{a}, x_1, \dots, \hat{x}_j, \dots, x_r) \end{aligned}$$

因此我们有:

$$d\eta_{\tilde{a}}(x) = \pi^* \omega(x + \tilde{a}) - \pi^* \omega(x)$$



即 $\eta_{\bar{a}}$ 在 X 上得到的微分形式为 $\tilde{\eta}_a$, 则上式意味着:

$$d\tilde{\eta}_a(x) = \omega(x+a) - \omega(x)$$

这表明 $\tau_a^*\omega - \omega$ 是恰当的. □

则

$$\begin{aligned} \int_X \omega - \omega &= \int_X (\tau_{-x}^*\omega - \omega) dx \\ &= \int_X (d\tilde{\eta}_x) dx \\ &= d \left(\int_x \tilde{\eta}_x dx \right) \end{aligned}$$

是恰当的.

因而若 ω 在 \int_X 下的像为 0, 则 ω 本身是恰当的, 因此上面定义的 \int_X 是一个同构: $H_{DR}^p(X, \mathbb{C}) \xrightarrow{\sim} \Lambda^p V^*$. 由 $\partial\bar{\partial}$ -引理, d 恰当等价于 $\bar{\partial}$ 恰当, 因此上面的证明可直接得到: \int_X 将定义 $H^r(H^0(X, \epsilon^{0,*}))$ 到 $\Lambda^r \bar{V}^V$ 的同构, 因而定理中指出的交换图的确是交换的.

2.2

复环上的除子、线丛、完备线性系

我们回忆一些定义以确定记号.

定义 1

复流形上的秩为 1 的全纯向量丛称为线丛. 记 X 上的所有线丛在张量作乘法下组成的群为 $Pic(X)$.

定义 2

$H^0(X, \mathcal{K}_X^*/\mathcal{O}_X^*)$ 中的元素称为 X 上的 Cartier 除子.

注意到我们有一个正合列

$$0 \rightarrow \mathcal{O}_X^* \mathcal{K}_X^* \rightarrow \mathcal{K}_X^*/\mathcal{O}_X^* \rightarrow 0$$

于是得到长正合列, 下面为我们关注的那一部分

$$H^0(X, \mathcal{K}_X^*) \rightarrow Div(X) \rightarrow Pic(X)$$

把上面映射中第一个记为 $() : f \mapsto (f)$, 第二个记为 $\mathcal{O} : D \mapsto \mathcal{O}(D)$ 称两个除子 D_1 与 D_2 线性等价, 若

$$\exists f \in \mathcal{K}_X^*, s.t. D_1 - D_2 = (f)$$

或者等价的

$$\mathcal{O}(D_1) = \mathcal{O}(D_2)$$

记为

$$D_1 \sim D_2$$

现在给定复环 X 上的除子 D , 我们用 $|D|$ 来指代所有的与 D 线性等价的有效除子, 注意到 $|D|$ 只依赖于 $\mathcal{O}(D)$:

$$|D| = \{Z(s) | s \in H^0(X, \mathcal{O}(D))\}$$

我们把 $|D|$ 称为 D 或 $\mathcal{O}(D)$ 的完全线性系 (也记作 $|\mathcal{O}(D)|$).

性质 1

令:

$$L(D) = \{f \in H^0(X, \mathcal{K}_X) | D + (f) \geq 0\}$$

则我们有自然的对应:

$$|D| \cong \mathbb{P}(L(D)) \cong \mathbb{P}H^0(X, \mathcal{O}(D))$$

证: 见 [4]. □

定义 3

现在设 D 是一有效除子, 注意到存在 $s \in H^0(X, \mathcal{O}(D))$, 使得 $Z(s) = D$, 我们把 s 的零点称为 D 中的点; 在不至混淆的情况下, 我们就用 D 表示 D 中所有点组成的集合 (注意到这是 X 的一个子簇). 并定义完全线性系 $|D|$ 的基点如下:

$$Base(|D|) := \bigcap_{D' \in |D|} D'$$

定义 4

设 G 是一个群, 令 M 为一个 G -模 (即一个 $\mathbb{Z}[G]$ -模). 定义

$$C^p(G, M) = Hom_G(G^p, M)$$

则 $C^p(G, M)$ 自然地成为一个 G -模, 我们再定义边缘同态 $\delta: C^p \rightarrow C^{p+1}$, 使之构成一个上链复形:

$$\begin{aligned} \delta f(\sigma_0, \dots, \sigma_p) &= \sigma_0 f(\sigma_1, \dots, \sigma_p) + \sum_{i=0}^{p-1} (-1)^{i+1} f(\sigma_0, \dots, \sigma_i \cdot \sigma_{i+1}, \dots, \sigma_p) \\ &\quad + (-1)^{p+1} f(\sigma_0, \dots, \sigma_{p-1}) \end{aligned}$$

直接验证可得 $\delta^2 = 0$, 于是我们有了一个 G -模上链复形:

$$0 \longrightarrow C^0(G, M) \cong M \xrightarrow{\delta} C^1(G, M) \xrightarrow{\delta} C^2(G, M) \xrightarrow{\delta} \dots$$

定义 5

定义上述链复形的上同调为 G -模 M 上的同调:

$$H^p(G, M) := H^p(C^*(G, M))$$



特别的, 注意到

$$H^0(G, M) = \{m \in M \mid \delta m = m, \forall \delta \in G\} := M^G$$

此外, 设 $f: M \rightarrow N$ 为 G -模同态, 自然的 f 诱导出 $C^p(G, M) \rightarrow C^p(G, N)$ 且 f 与边缘算子显然可交换, 因此诱导出上同调之间的映射:

$$H^*(G, M) \rightarrow H^*(G, N)$$

我们将只关心群的上同调中的 H^1 , 但是下面的定理基本上对于高阶上同调也是对的, 读者可以自己证明, 或者参见群的上同调有关的书籍, 比如 [9].

定理 6

若 M 是平凡 G -模 (i.e. Γ 以相同的方式作用在 M 上), 则 $H^1(G, M) = \text{Hom}_G(G, M)$

证: 直接验证. □

回到复环上来, 设 $X = V/\Gamma, \mathcal{F}$ 为 X 上的层. 注意到 $\pi: V \rightarrow X$ 是一个覆盖, 因此 X 上的线丛可以拉回到 V 上成为 $\pi^*\mathcal{F}, \forall \gamma \in \Gamma$, 平移映射 $\tau_\gamma: V \rightarrow V$, 产生了一个层的拉回映射 $\tau_\gamma^*\pi^*\mathcal{F} \rightarrow \pi^*\mathcal{F}$ 最后注意到 $\tau_\gamma^*\pi^*\mathcal{F} = \pi^*\mathcal{F}$ 过渡到整体截面即是

$$\tau_\gamma^*: H^0(V, \pi^*\mathcal{F}) \rightarrow H^0(V, \pi^*\mathcal{F})$$

而显然

$$\tau_{\gamma+\gamma'}^* = \tau_\gamma^* \tau_{\gamma'}^*$$

于是我们定义了一个群同态

$$\Gamma \rightarrow \text{End}(H^0(V, \pi^*\mathcal{F}))$$

从而 $H^0(V, \pi^*\mathcal{F})$ 通过这种方式成为了一个 Γ -模.

引理 2

若 M 是一个 Abel 群, 记 \underline{M} 为取之于 M 的 X 上的局部常值层, 则 $\pi^*\underline{M}$ 是 V 上的局部常值层, 还是记为 \underline{M} , 且 $H^0(V, \underline{M})$ 是一个平凡 Γ -模.

证: 直接验证. □

注意到由 Poincaré-Grothendieck $\bar{\partial}$ 引理 (参见 [7]), 我们有 $H^{>0}(\mathbb{C}^n, \mathcal{O}_{\mathbb{C}^n}) = 0$, 利用层的正合列

$$0 \rightarrow \underline{\mathbb{Z}} \xrightarrow{i} \mathcal{O}_X \xrightarrow{e^{2i\pi(\cdot)}} \mathcal{O}_X^* \rightarrow 0$$

以及 $H^{>0}(\mathbb{C}^n, \underline{\mathbb{Z}}) = 0$ 得到 $H^{>0}(\mathbb{C}^n, \mathcal{O}_{\mathbb{C}^n}^*) = 0$. 这意味着 \mathbb{C}^n 上的所有线丛都是平凡的. 而由于 $\pi: V \rightarrow X$ 是覆盖映射, 于是 X 上的线丛到 V 的拉回都同构于线丛 $V \times \mathbb{C} \rightarrow V$

性质 2

我们有同构: $H^1(\Gamma, H^0(V, \mathcal{O}_V^*)) \cong \text{Pic}(X)$ 其中 $H^1(\Gamma, H^0(V, \mathcal{O}_V^*))$ 是群 Γ 的上同调.

证: 设 L 是 X 上的线丛, 取同构

$$\chi: \pi^*L \rightarrow V \times \mathbb{C}$$

那么 Γ 在 π^*L 上的典范作用将通过 χ 变为在 $V \times \mathbb{C}$ 上的作用, 注意到这个作用有如下形式:

$$\gamma \cdot (z, v) = (z + \gamma, e_\gamma(z) \cdot v), \forall \gamma \in \tau, \forall (z, v) \in V \times \mathbb{C}$$

这里 e_γ 是 V 上的无处为零的全纯函数, 即在 $H^0(X, \mathcal{O}_V^*)$, 由于 $(\gamma_1 + \gamma_2) \cdot (z, v) = \gamma_1 \cdot (\gamma_2 \cdot (z, v))$ 我们得到:

$$\begin{cases} e_{\gamma_1 + \gamma_2}(z) = e_{\gamma_1}(z + \gamma_2)e_{\gamma_2}(z) \\ e_0(z) \equiv 1 \end{cases} \quad (2)$$

注意到假如我们选取另一个从 π^*L 到 $V \times \mathbb{C}$ 的同构 χ' , 则将会有 V 上的无处为零的全纯函数 f , 使得

$$\begin{array}{ccc} & & V \times \mathbb{C} \\ & \nearrow \chi & \downarrow (z, v) \mapsto (z, f(z) \cdot v) \\ \pi^*L & & V \times \mathbb{C} \\ & \searrow \chi' & \end{array}$$

是交换的. 则

$$e'_\gamma(z) = f(z + \gamma)e_\gamma(z)f^{-1}(z) \quad (3)$$

注意到满足 (2.9) 的如下的群同态

$$\begin{aligned} e : \tau &\longrightarrow H^0(V, \mathcal{O}_V^*) \\ \gamma &\longmapsto e_\gamma \end{aligned}$$

□

在乘法下组成的群记恰好就是 $Z^1(\Gamma, H^0(V, \mathcal{O}_V^*))$. 而满足 (2.10) 的 e 与 e' 的商, 即 $e' \cdot e^{-1}$ 生成的子群恰好 $B^1(\Gamma, H^0(V, \mathcal{O}_X^*))$. 因此 $H^1(\Gamma, H^0(V, \mathcal{O}_V)) = \frac{Z^1(\Gamma, H^0(V, \mathcal{O}_X^*))}{B^1(\Gamma, H^0(V, \mathcal{O}_X^*))}$.

下面来证明定理:

给定一组满足 (2.8) 的 (e_γ) , 我们可以如下构造一个线丛:

我们利用 (e_γ) 构造 $V \times \mathbb{C}$ 上的一个作用, 于是诱导出了一个商空间 L , 它实际上是 X 上的线丛:

选择 (U_i) 为 X 的在 V 覆盖下的基本开邻域组成的覆盖, 于是

$$\pi^{-1}(U_i) = \bigcup_{\gamma \in \Gamma} O_i^\gamma,$$

其中 $O_i^\gamma = O_i^0 + \gamma, \pi$ 限制在 $O_i^\gamma = O_i^0 + \gamma$ 是同胚. 注意到我们可以取得 U_i 足够小, 使得对每个 $U_i \cap U_j \neq \emptyset$, 存在唯一的一个 γ_{ij} 使得 $(O_i^0 + \gamma_{ij}) \cap O_j^0 \neq \emptyset$. 对每个 $U_{ij} = U_i \cap U_j \neq \emptyset$, 定义: $\phi : f_{ij} = e_{\gamma_{ij}}(\pi|_{O_i^0}^{-1}(z))$. 我们来验证这样定义的 f_{ij} 定义了一个线丛, 为此假设 $U_i \cap U_j \cap U_k \neq \emptyset$, 则 $\gamma_{ij} + \gamma_{jk} + \gamma_{ki} = 0$. 利用 (2.8) 我们有

$$\begin{aligned} f_{ij}(z)f_{jk}(z)f_{ki}(z) &= e_{\gamma_{ij}}(\pi|_{O_i^0}^{-1}(z))e_{\gamma_{jk}}(\pi|_{O_j^0}^{-1}(z))e_{\gamma_{ki}}(\pi|_{O_k^0}^{-1}(z)) \\ &= e_{\gamma_{ij}}(\pi|_{O_i^0}^{-1}(z))e_{\gamma_{jk}}(\pi|_{O_i^0}^{-1}(z) + \gamma_{ij})e_{\gamma_{ki}}(\pi|_{O_i^0}^{-1}(z) + \gamma_{ik}) \\ &= e_{\gamma_{ij} + \gamma_{jk}}(\pi|_{O_i^0}^{-1}(z))e_{\gamma_{ki}}(\pi|_{O_i^0}^{-1}(z) + \gamma_{ik}) \\ &= e_{-\gamma_{ki}}(\pi|_{O_i^0}^{-1}(z))e_{\gamma_{ki}}(\pi|_{O_i^0}^{-1}(z) - \gamma_{ki}) \\ &= 1 \end{aligned}$$

注意到这样我们定义了一个映射 $\tilde{\phi} : Z^1(\Gamma, H^0(V, \mathcal{O}_V^*)) \rightarrow H^1(X, \mathcal{O}_X^*)$, 显然由我们的构造方式, 这是一个群同态. 根据我们前面的讨论, 这是一个满射, 且核恰好为 $B^1(\Gamma, H^0(V, \mathcal{O}_V^*))$. 因此过渡到商得到同构 $\phi : H^1(\Gamma, H^0(V, \mathcal{O}_V^*)) \xrightarrow{\sim} H^1(X, \mathcal{O}_X^*)$.

由于上面的证明只用到了 $H^1(\mathbb{C}^n, \mathbb{Z}) = 0$, 上述定理的证明可逐字翻译成如下定理:

定理 7

设 \mathcal{F} 是复环 X 上的层, 若 $H^1(V, \pi^* \mathcal{F}) = 0$, 则有一相对于层变量的自然同构:

$$H^1(\Gamma, H^0(V, \pi^* \mathcal{F})) \cong H^1(X, \mathcal{F})$$

证: 只有自然性没有涉及到, 而这通过上述证明也是显然的. \square

由于我们还不知道 $H^1(\Gamma, H^0(V, \mathcal{O}_V))$ 是什么样的, 所以还不能算彻底弄清楚了环面上的线丛. 我们现在构造一些具体而特别的线丛, 后面将证明这构造出了所有线丛:

注意到设 H 是一个 Hermitian 形式, α 为相应的半特征. 令

$$e_\gamma(z) = \alpha(\gamma) e^{\frac{\pi}{2} H(z, z) + \pi H(z, \gamma)}, \forall z \in V.$$

则 (e_γ) 满足 (2.8). 于是利用这组 (e_γ) 我们可以定义一个线丛, 我们把这样定义的线丛记为 $L(H, \alpha)$, 注意到

$$L(H_1, \alpha_1) \otimes L(H_2, \alpha_2) \cong L(H_1 + H_2, \alpha_1 \alpha_2)$$

把这样的线丛生成的 $\text{Pic}(X)$ 的子群记为 $P(\Gamma)$.

2.3

Appell-Humbert

考虑复环 X 上的局部常数值层 \mathbb{Z} , 我们有一个层的短正合列:

$$0 \longrightarrow \mathbb{Z} \xrightarrow{i} \mathcal{O}_X \xrightarrow{e} \mathcal{O}_X^* \longrightarrow 0$$

其中 i 是嵌入, e 是指数映射 $f_U \mapsto e^{2i\pi f_U}$

于是得到上同调群的长正合列, 下面是我们关注的那一部分:

$$H^0(X, \mathcal{O}_X^*) \xrightarrow{\delta} H^1(X, \mathbb{Z}) \xrightarrow{H^1(i)} H^1(X, \mathcal{O}_X) \xrightarrow{H^1(e)} H^1(X, \mathcal{O}_X^*) \xrightarrow{j} H^2(X, \mathbb{Z})$$

由于紧复流形上全纯函数为常数, $H^0(X, \mathcal{O}_X) \rightarrow H^0(X, \mathcal{O}_X^*)$ 是满的, 因此 $\delta = 0$, 而第二个连接同态

$$j: H^1(X, \mathcal{O}_X^*) \rightarrow H^2(X, \mathbb{Z})$$

称之为第一陈类. 由下面定理, 这个与复几何的第一陈类是一致的:

定理 8

下面的图是交换的

$$\begin{array}{ccc} & & H^2(X, \mathbb{Z}) \\ & \nearrow j & \downarrow \\ H^1(X, \mathcal{O}_X^*) & & \\ & \searrow c_1 & \\ & & H^2_{DR}(X, \mathbb{C}) \end{array}$$

证: 参见 [13] 或者 [4]

□

因此我们在不致混淆的情况下也把 j 记为 c_1 , 注意到前一章所做出的上同调之间的典范同构, 在复环上我们将用不变形式来表示 c_1 的像.

我们把 c_1 的核空间记为 $Pic^0(X)$, c_1 的像空间记为 $NS(X)$.

定理 9

$$c_1(L(H, \alpha)) = \Im H$$

证: 我们指出证明的关键: 注意到 $L(H, \alpha)$ 上有一个自然的 Hermitian 度量, 实际上 V 上的平凡线丛 $V \times \mathbb{C}$ 上的度量 $\|(x, t)\| = e^{\frac{\alpha}{2} \Im H(x, x)} |t|$ 在由 (H, α) 定义的 Γ 作用下保持不变, 因而诱导出 $L(H, \alpha)$ 上的 Hermitian 度量. 于是由第一陈类的性质, 我们直接计算可得. □

定理 10

通过前面的内容, 我们已经得到如下交换图:

$$\begin{array}{ccccccc}
 0 & \longrightarrow & H^1(X, \mathbb{Z}) & \longrightarrow & H^1(X, \mathcal{O}_X) & \longrightarrow & H^1(X, \mathcal{O}_X^*) & \longrightarrow & H^2(X, \mathbb{Z}) \\
 & & \downarrow \wr & & \downarrow \wr & & \downarrow \wr & & \downarrow \wr \\
 0 & \longrightarrow & \Gamma^\vee & \longrightarrow & \bar{V}^\vee & \longrightarrow & Pic(X) & \xrightarrow{c_1} & \Lambda^2 \Gamma^\vee
 \end{array}$$

在此交换图下诱导如下交换图

$$\begin{array}{ccccccc}
 0 & \longrightarrow & Pic^0(X) & \longrightarrow & Pic(X) & \longrightarrow & NS(X) & \longrightarrow & 0 \\
 & & \downarrow \wr & & \downarrow \wr & & \downarrow \wr & & \\
 0 & \longrightarrow & Hom(\Gamma, U(1)) & \longrightarrow & P(\Gamma) & \longrightarrow & \Lambda^{1,1} V^* \cap \Lambda^2 \Gamma^\vee & \longrightarrow & 0
 \end{array}$$

证: 考虑层的长正合列

$$\begin{array}{ccccccc}
 0 & \longrightarrow & \mathbb{Z} & \longrightarrow & \mathbb{R} & \longrightarrow & U(1) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 0 & \longrightarrow & \underline{\mathbb{Z}} & \longrightarrow & \mathcal{O}_X & \longrightarrow & \mathcal{O}_X^* & \longrightarrow & 0
 \end{array}$$

产生的长正合列, 取其中一部分:

$$\begin{array}{ccccccc}
 H^1(X, \mathbb{R}) & \xrightarrow{e} & H^1(X, U(1)) & \xrightarrow{\eta} & H^2(X, \mathbb{Z}) & \longrightarrow & H^2(X, \mathbb{R}) \\
 \downarrow & & \downarrow & & & & \\
 H^1(X, \mathcal{O}_X) & \longrightarrow & H^1(X, \mathcal{O}_X^*) & & & &
 \end{array}$$

由 De Rham 定理我们可以得到第一排最后一个映射是单射, 因此 $\eta = 0$, 此即说明 e 是满射. 于是中间



的列映射是下面交换图的一部分:

$$\begin{array}{ccc} H^1(X, \mathbb{R}) & \xrightarrow{\sim} & \text{Hom}_{\mathbb{R}}(V, \mathbb{R}) \\ \downarrow & & \downarrow \wr \\ H^1(X, \mathcal{O}_X) & \xrightarrow{\sim} & \bar{V}^{\vee} \end{array}$$

这也表明它是同构.

注意到对于局部常值层 $\underline{U}(1)$, $H^0(V, \underline{U}(1)) = U(1)$, 由我们前面的构造它们是一个平凡的 Γ -模. 因此直接计算得:

$$H^1(\Gamma, H^0(V, \underline{U}(1))) = H^1(\Gamma, \underline{U}(1)) = \text{Hom}(\Gamma, U(1))$$

而根据定理 (2.5.2) 我们有下列的交换图:

$$\begin{array}{ccc} H^1(X, \underline{U}(1)) & \xrightarrow{\sim} & \text{Hom}(\Gamma, U(1)) \\ \downarrow & & \downarrow \varepsilon \\ H^1(X, \mathcal{O}_X^*) & \xrightarrow{\sim} & H^1(\Gamma, H^0(V, \mathcal{O}_X^*)) \end{array}$$

根据定理 (2.5.1) ε 定义如下,

$$\begin{aligned} \varepsilon : \text{Hom}(\Gamma, U(1)) &\longrightarrow H^1(\Gamma, H^0(V, \mathcal{O}_X^*)) \\ \alpha &\longmapsto L(0, \alpha) \end{aligned}$$

我们注意到这是一个单射, 结合 ε 是满射, 我们知道了

$$\text{Pic}^0(X) \cong \text{Hom}(\Gamma, U(1))$$

然后, $NS(X) \subseteq \Lambda^{1,1}V^* \cap \Lambda^2\Gamma^{\vee}$ 由陈类的定义直接成立. 而由于 $\forall \omega \in \Lambda^{1,1}V^* \cap \Lambda^2\Gamma^{\vee}$, 由第一章知道它是某个 Hermitian 形式 $-H$ 的虚部, 由前面引理我们知道与之相应的半特征 α , 于是 $L(H, \alpha)$ 是一个第一类陈类为 ω 的线丛. 因此

$$NS(X) = \bigwedge^{1,1} V^* \cap \bigwedge^2 \Gamma^{\vee}$$

最后, 注意到 $L \otimes L(-c_1(L), \alpha^{-1})$ 的第一陈类为 0, 因此存在唯一一个 $K \in \text{Pic}^0(X)$, 使得 $L \otimes L(-c_1(L), \alpha^{-1}) = K$, 由 $\text{Pic}^0(X)$ 的描述, 可设 $K = L(0, \beta)$, 于是 $L = L(c_1(L), \alpha \cdot \beta)$, 这即是说:

$$\text{Pic}(X) \cong P(\Gamma)$$

□

推论 2

(Appell-Humbert) 复环 X 上每个线丛都同构于某个 $L(H, \alpha)$.

2.4

Theta 函数与复环的若干对应

定理 11

在 $\tau^{H, \alpha}$ 与 $H^0(X; L(H, \alpha))$ 之间有一个自然的同构.

证: 设 $L = L(H, \alpha)$ 是复环 X 上的线性丛, L 上的整体截面与 V 上的全纯某些函数通过如下交换图对应:

$$\begin{array}{ccc} V & \xrightarrow{x \mapsto (x, \theta(x))} & V \times \mathbb{C} \\ \downarrow & & \downarrow \\ X & \xrightarrow{x \mapsto s(x)} & L \end{array}$$

而根据我们的构造, 如下交换图应当交换,

$$\begin{array}{ccc} V \times \mathbb{C} & \xrightarrow{\gamma} & V \times \mathbb{C} \\ \pi \searrow & & \swarrow \pi \\ & L & \end{array}$$

因此, 我们有:

$$\begin{array}{ccc} (x, \theta(x)) & \xrightarrow{\gamma} & V \times \mathbb{C} \\ \pi \searrow & & \swarrow \pi \\ & L & \end{array}$$

于是

$$\theta(x + \gamma) = \alpha(\gamma) e^{\frac{\alpha}{2} H(x, x) + \pi H(x, \gamma)} \theta(x)$$

这也就是说 θ 应当是一个 Theta 函数 □

推论 3

在一个复环 X 上, 每个有效除子都是 Theta 除子.

证: 设 D 是一个有效除子, 于是 $\mathcal{O}(D)$ 是带有全局截面 s 的线丛, 且 $Z(s)=D$, 由 Appell-Humbert 定理可设 $\mathcal{O}(D) = L(H, \alpha)$; 由上面的对应, 存在 $\theta \in \tau^{H, \alpha}$, 使得 θ 与 s 通过上面的方式对应, 于是我们有 $Z(\theta) = Z(s) = D$ □

注意到若 $\theta_1, \theta_2 \in \tau^{H, \alpha}$, 则 $\frac{\theta_1}{\theta_2}$ 是一个 Γ -周期的亚纯函数, 因此将定义一个 X 上的亚纯函数. 事实上, 至此我们已然有了该结论的反面也对:

推论 4

设 X 是一个复环, 则 X 上的每一个亚纯函数都是两个 Theta 函数的商.

证: 设 f 是 X 上的一个亚纯函数, 设 $(f) = D_1 - D_2$, 这里 D_1, D_2 是两个有效除子. 设对于 Theta 函数 θ_1, θ_2 , 使得 $D_i = Z(\theta_i), i = 1, 2$, 则 $(f \cdot \theta_2 / \theta_1) = 0$, 因此 $f \cdot \theta_1 / \theta_2$ 是一个常数 C . 这表明

$$f = \frac{C\theta_1}{\theta_2}$$

□

2.5

线丛上的截面

定理 12

(Riemann-Roch 定理) 设 L 为环 X 上的一线丛, 且其第一陈类 $c_1(L)$ 正定, 则我们有

$$\dim H^0(X, L) = pf(c_1(L)) > 0$$

证: 这已经是显然的了! □

定理 13

令 $f: X \rightarrow Y$ 是一个同源, L 是 Y 上的线丛:

$$\dim H^0(X, f^*L) = \deg f \cdot \dim H^0(Y, L)$$

证: 我们可设 $X = V/\Gamma, Y = V/\Gamma'$, 且 $\Gamma \subseteq \Gamma'$, 设 $L = L(H, \alpha), \omega = \Im H$, 注意到有

$$Pf_{(\omega)} = [\Gamma' : \Gamma] Pf_{(\Gamma')}(\omega)$$

而 $[\Gamma', \Gamma] = \deg(f)$, 因此定理得证 □

第 3 节

作为 Abel 簇的复环

3.1

嵌入

性质 3

令 D 为复环上的除子, $\forall x, y \in X$ 记 $\tau_{-x}^* D$ 为 $(D+x)$, 则:

$$(D+x+y) + D \sim (D+x) + (D+y)$$

证: 只需要验证, $\mathcal{O}(D+x) \otimes \mathcal{O}(D+y) \cong \mathcal{O} \otimes \mathcal{O}(D+x+y)$, 这等价于 $\tau_{-x}^* \mathcal{O}(D) \otimes \tau_{-y}^* \mathcal{O}(D) \cong \tau_{-x-y}^* \mathcal{O}(D) \otimes \mathcal{O}(D)$ 这不过是平方定理的内容罢了. □

引理 5

线性系 $|D|$ 中一定包含约化除子.

证: 令 $D = \sum n_j D_j$, 若某个 $n_j > 1$, 则用

$$E_j = (D_j + x_1) + \cdots + (D_j + x_{n_j})$$

来代替 D_j , 其中 $x_1 + \cdots + x_{n_j} = 0$, 且它们两两不同. 我们一定可以选取适当的 x_j 使得上述所有除子两两不同;

事实上若 $F_1 + x = F_2$, 则

$$x \in \bigcap_{f_1 \in F_1} (F_2 - f_1)$$

这是无处稠密的集合, 因此我们可以做到. □

性质 4

若 $n \geq 2$ 为一整数, 则 $|nD|$ 无基点 (也就是说基点集合为空集).

证: 任取一点 $y \in X$, 我们想找一个除子 $E \in |nD|$ 使得它不通过 y . 注意到 $y \in (D - x)$ 当且仅当 $x \in (D - y)$, 因此与上引理的证明方法相同, 我们总可以取 $x_1 \cdots x_{n-1}$, 使得 y 不包含于 $E = (D + x_1) + \cdots + (D + x_{n-1}) + (D - \sum x_j)$, 由于 E 在 $|nD|$ 中, 证毕. □

性质 5

设 D 是 X 上的有效除子, 设 $\mathcal{O}(D)$ 的第一陈类正定, 则 $|D|$ 的子集:

$$\{E \in |D| \mid \exists x \in X, E + x = E\}$$

是有限线性 $|D|$ 中真线性子空间的并集, 因此是零测的.

设 L 是 X 上的线丛, 取 $H^0(X, L)$ 的一组基 $s_0, \dots, s_r, \forall x \in X \setminus \text{Base}(|L|), [s_0(x), \dots, s_r(x)]$ 作为齐次坐标定义了 \mathbb{P}^r 中的一点, 于是我们可以定义如下映射: :

$$\begin{aligned} \psi_L : X \setminus \text{Base}(|L|) &\longrightarrow \mathbb{P}^r \\ x &\longmapsto [s_0(x), \dots, s_r(x)] \end{aligned}$$

事实上对于不同的基的选取, 我们的定义只相差 \mathbb{P}^r 中的一个线性变换, 因此在忽略掉一个线性变换的意义下 ψ_L 只依赖于 L . ψ_L 已然是全纯的, 现在我们想考虑复环的嵌入问题, 有如下重要结论:

定理 14 Lefschets 定理

令 L 是 X 上第一陈类正定的线丛, 则当 $n \geq 3, \psi_L$ 定义了一个 X 到映射空间的嵌入 $\varphi : X \hookrightarrow \mathbb{P}^n$.

证: 注意到这不过是上述定理的直接推论, 参考 [4] 或 [8]. □

推论 5 Riemann-条件

$X = V/\Gamma$ 可以映射到复射影空间的充分必要条件是: 存在 V 一组 \mathbb{C} -基 $\mathfrak{B}, \tau \in \mathfrak{H}_g$ 以及正整数 d_1, d_2, \dots, d_g 满足 $d_1|d_2|\dots|d_g$, 使得在基底 \mathfrak{B} 下:

$$\Gamma = \tau\mathbb{Z}^g \oplus \Delta\mathbb{Z}^g$$

其中 Δ 是以 d_1, d_2, \dots, d_g 的对角矩阵.

证: 由 Lefschets 定理, 当且仅当其上存在第一陈类正定的复线丛, 而由定理 Appell-Humbert 定理与定理 (2.1.5) 知道这是正确的. \square

3.2

Abel 簇

定义 6

我们把 X 上的整的 Kähler 形式称为 X 上的极化. 一个 Abel 簇是指其上存在极化的复环, 极化 Abel 簇是指赋予了极化的 Abel 簇, 即一个双元组 (X, ω) 其中 X 是一个复环, ω 是一个极化.

根据我们前面已经证明的, 一个复环是 Abel 簇当且仅当它可以嵌入射影空间.

定义 7

我们把一个幺模的极化称为主极化, 带有主极化的复环称为主极化 Abel 簇.

这样是由第一章第一节的知识, 我们知道对于一个主极化的 Abel 簇, $\varphi_\omega: X \rightarrow \hat{X}$ 是一个同构.

性质 6

设 X 是一个极化 Abel 簇, L 是 X 上的一个线丛, 那么存在一个主极化 Abel 簇 Y , 其上的一个线丛 M , 一个同源 $u: X \rightarrow Y$, 使得 $L \cong u^*M$.

证: 设 $X = V/\Gamma$ 是一个 Abel 簇, ω 为其上的一个极化, 按照定理 (2.1.5) 取 Γ 的一组基 $(\gamma_1, \dots, \gamma_{2g})$. 设 Γ' 为 V 中由 $\gamma_1/d_1, \dots, \gamma_g/d_g, \gamma_{g+1}, \dots, \gamma_{2g}$ 生成的格. 于是 Kähler 形式 ω 在 Γ' 是整的, 且是幺模的, 因此定义了一个 $Y = V/\Gamma'$ 上的主极化. 于是典范的满射 $u: X \rightarrow Y$ 是一个同源. 设 L 是 X 上第一陈类为 ω 的线丛, 而 M 是 Y 上第一陈类为 ω 的线丛. 于是 L 与 u^*M 有相同的第一陈类, 存在 $x \in X$ 使得 $L \cong \tau_x^*(u^*M)$. 因此

$$L \cong \tau_x^*(u^*M) \cong u^*(\tau_{u(x)}^*M)$$

\square

下面讨论 Abel 簇的结构:

定理 15 Poincaré 约化定理

设 X 是一个 Abel 簇, Y 是 X 的子 Abel 簇, 则存在 X 的子簇 Z , 使得 $Y \cap Z$ 是有限的, 且 $Y + Z = X$. 因此 $Y \times Z$ 与 X 同源, 我们把 Z 称作 Y 在 X 中的补.

证: 令 $X = V/\Gamma$ 以及 $Y = W/M$, 其中 W 是 V 的一个子空间, $M = W \cap L$ 为 W 中的格, 令 ω 为 X 的一个极化, H 为相应的 Hermitian 形式, 则 H 是正定的. 有第一章第一节, 我们知道 $V = W^\perp \oplus W$, 且 $W^\perp \cap \Gamma$ 是 W^\perp 中的格. 因此 $Z = \frac{W^\perp}{W^\perp \cap \Gamma}$ 满足条件: 由 $\frac{W}{\Gamma} \cong \frac{W^\perp + \Gamma}{\Gamma}$ 以及 $\frac{W^\perp}{W^\perp \cap \Gamma} \cong \frac{W^\perp + \Gamma}{\Gamma}$, 我们知道 $Y \cap Z = \frac{(W + \Gamma) \cap (W^\perp + \Gamma)}{\Gamma}$ 而 $(W + L) \cap (W^\perp + \Gamma)$ 是 V 中包含 Γ 的离散群, 因而是格, 故它们的交有限, 而显然其和为 X , 故 Z 满足条件. \square

定义 8

我们称一个不包含任何一个子复环的复环是单的.

注意到 Poincaré 约化定理促使所有的 Abel 簇都可以在同源意义下完全分解成单 Abel 簇的直积. 下面我们进一步考察极化 Abel 簇的分解:

定理 16

设 (X, ω) 是一个极化 Abel 簇, 包含子 Abel 簇 X_1, X_2, Y_1 和 Y_2 使得

$$(X, \omega) = (X_1, \omega|_{X_1}) \times (X_2, \omega|_{X_2}) = (Y_1, \omega|_{Y_1}) \times (Y_2, \omega|_{Y_2})$$

对于 $i = 1, 2$, 我们有

$$(X_i, \omega|_{X_i}) \cong (X_i \cap Y_1, \omega|_{X_i \cap Y_1}) \times (X_i \cap Y_2, \omega|_{X_i \cap Y_2})$$

证: 参见 [2]. \square

3.3**Abel 化**

设 $X = V/\Gamma$ 为一个复环, 我们如下定义它的 Abel 化: 令 N 为 Γ 上所有 Theta 函数的 Riemann 形式的核的交集, 于是存在 Theta 函数 $\theta_1, \dots, \theta_r$ 使得它们的 Riemann 形式的交集就是 N . 注意到 $\theta_1, \dots, \theta_r$ 也是个 Theta 函数, 而且其核就是 N . 由定理 (2.1.5) 得在 V/N 中的像 Γ_{ab} 将是一个格, 我们定义 X 的 Abel 化为 $X_{ab} = V_{ab}/\Gamma_{ab}$

定理 17

X_{ab} 是一个 Abel 簇.

证: 由于 ω 是 Γ_{ab} 的正定的整的 Kähler 形式, 因此由 Lefschets 嵌入 X_{ab} 是 Abel 簇. \square

性质 7

a) 通过 $\Xi: X \rightarrow X_{ab}$, 我们有一个同构的

$$\text{Div}(X) \cong \text{Div}(X_{ab})$$

b) Ξ 诱导了亚纯函数环之间的同构:

$$\mathcal{M}(X_{ab}) \cong \mathcal{M}(X)$$

推论 6

Abel 簇范畴与复环范畴等价.

- [1] 佟文廷, 同调代数引论, 高等教育出版社。
- [2] O. Debarre *Ttores et Variétés Abéliennes Complexes*, Société Mathématiques de France
- [3] Danilov & Shokurrov, *Algebraic Curves, Algebraic Manifolds and Schemes*, EMS (Algebraic Geometry 1), Springer.
- [4] Griffith & Harris, *Principles of Algebraic Geometry*, Wiley.
- [5] R. Godement, *Topologie algébrique et Théorie des Faisceaux*, Hermann, Paris.
- [6] A. Grothendieck, *Sur Quelques Points d'Algèbre Homologique*. Tohoku, 1957
- [7] D. Huybechts, *Complex Geometry. Undergraduate Text*, Springer.
- [8] G.R. Kempf, *Abelian Varieties and Theta Functions*, Cambridge University Press
- [9] S. Mac Lane, *Homology*, Classics in Mathematics, Springer.
- [10] D. Mumford, *Abelian Varieties*, Oxford University Press
- [11] A. Polishchuk, *Abelian Varieties, Theta Functions, and the Fourier Transform*, Cambridge University Press.
- [12] Birkenhake & Lange, *Complex Abelian Varieties*, Springer
- [13] R.O. Wells, *Differential Analysis on Complex Manifolds*, Graduate Text in Mathematics 65, Springer.

关于不完备内积空间的一些讨论

2011 级弘毅班 张怡烽

设 $(E, \langle \cdot, \cdot \rangle)$ 为内积空间, E^* 为其对偶空间. 若 E 完备, 由 Riesz 表示定理, 我们有 $E^* = E$. 这里等号理解为:

$$\forall f \in E^*, \exists y_f \in E, \text{ st. } f(\cdot) = \langle \cdot, y_f \rangle.$$

本文讨论 E 不完备的情形.

应当指出, 不完备的内积空间是存在的, 例如 Hilbert 空间 $(l_2, \langle \cdot, \cdot \rangle)$ 的子空间 $(l_1, \langle \cdot, \cdot \rangle)$ 就是不完备的. 但是类似度量空间的完备化定理, 我们有

定理 1

对任意内积空间 E , 必存在 Hilbert 空间 \tilde{E} , 使得 E 与 \tilde{E} 的某一稠密子空间等距同构, 并且在等距同构意义下这样的 \tilde{E} 是唯一的.

Proof. 定义 $\{\{x_n\} : \{x_n\} \text{ 为 } E \text{ 中 Cauchy 列}\}$. 按通常方式将 \hat{E} 作成线性空间. 在 \hat{E} 上定义二元函数:

$$\forall \{x_n\}, \{y_n\} \in \hat{E} : \langle \{x_n\}, \{y_n\} \rangle = \lim_n \langle x_n, y_n \rangle.$$

容易验证定义的合理性. 令

$$\hat{E}_0 = \{\{x_n\} \in \hat{E} : \langle \{x_n\}, \{x_n\} \rangle = 0\}.$$

作 $\tilde{E} = \hat{E}/\hat{E}_0$. 则 \tilde{E} 成为内积空间, 并且 E 与 \tilde{E} 的稠密子集 $\tilde{E}_0 = \{\widetilde{\{x_n\}} : x_n = x, \exists x \in E\}$ 等距同构. 最后利用定义直接证明 \tilde{E} 的完备性和唯一性. 定理得证. \square

称 Hilbert 空间 \tilde{E} 为 E 的完备化.

设 $f(\cdot) = \langle \cdot, y \rangle \in \tilde{E}^*$, 将其限制在 E 上成为 E 上的连续线性泛函. 反之, 对任意 $f_0 \in E^*$, 由 Hahn-Banach 定理, 存在 $f \in \tilde{E}^* : f|_E = f_0$, 且 $\|f\| = \|f_0\|$. 由 E 在 \tilde{E} 中的稠密性, 这样的 f 是唯一的. 设 $f(\cdot) = \langle \cdot, y_f \rangle$, 定义

$$\begin{aligned} \Phi : E^* &\longrightarrow \tilde{E} \\ f_0(\cdot) &\longmapsto y_f. \end{aligned}$$

以上分析表明 Φ 是合理定义的保范同构映射, 进而有

定理 2

设 E 为内积空间, \tilde{E} 为 E 的完备化, 则 $E^* = \tilde{E}$.

定理 2 表明, 在不完备的内积空间中, Riesz 表示定理不再成立. 而通常在证明表示定理时, 完备性条件体现在投影定理的运用中, 因此在不完备的内积空间中, 投影定理也将不再成立, 即

定理 3

设 E 为不完备的内积空间, 则存在闭子空间 $F \subset E$ 以及 $x \in E$, 使得 x 在 F 上的投影 $P_F(x) = 0$.

Proof. 任取 $y \in \tilde{E} \setminus E, f(\cdot) = \langle \cdot, y \rangle \in \tilde{E}^*$. 考虑 $F = \ker f$, 取 $x \in E \setminus F$. 设投影 $P_F(x)$ 存在, 则 $y_0 := x - P_F(x) \in F^\perp$, 且 $f(y_0) = f(x) \neq 0$. 容易证明 $f(\cdot) = \langle \cdot, \frac{\overline{f(y_0)}}{\|y_0\|^2} y_0 \rangle$. 但 $\frac{\overline{f(y_0)}}{\|y_0\|^2} y_0 \in E$, 此与 f 的选取矛盾. 定理得证. \square

可见, 在内积空间中, Riesz 定理或投影定理的正确性均等价于 E 的完备性.

椭圆方程的极大值原理与 Harnack 不等式 (节选)

2010 级弘毅班 段晓谕

摘要 本文主要介绍和讨论了二阶椭圆型偏微分方程中的极大值原理和 Harnack 不等式. 第一节是二阶椭圆型方程的极大值原理, 包括弱极大值原理、Hopf 引理、强极大值原理和 Alexandrov 极大值原理等重要和经典的结论. 第二节是关于 Harnack 不等式的介绍, 主要是针对非散度型二阶椭圆方程的, 由 Krylov 和 Safanov 给出的证明.

关键词: 二阶椭圆偏微分方程; 极大值原理; Harnack 不等式; 弱解; Moser 迭代

第 1 节

二阶线性椭圆方程的极大值原理

本章主要介绍二阶线性椭圆方程古典解的极大值原理, 包括弱极值原理, Hopf 引理, 强极值原理以及 Alexandrov 极大值原理, 它们将成为之后几章的重要工具. 本章内容主要参考自 [4] 的前两章以及 [9] 的第二章.

1.1

Laplace 方程的极大值原理

我们先从最简单的一类二阶线性椭圆方程——Laplace 方程说起.

定理 1 弱极值原理

设 U 为 \mathbb{R}^n 中的有界开集, $u \in C^2(U) \cap C(\bar{U})$.

(1) 若 u 在 U 中下调和, 即 $\Delta u \geq 0$, 则

$$\max_{\bar{U}} u \leq \max_{\partial U} u.$$

(2) 若 u 在 U 中上调和, 即 $\Delta u \leq 0$, 则

$$\min_{\bar{U}} u \geq \min_{\partial U} u.$$

证明. (1) 先考虑特殊情况: 假设 u 在 U 中满足 $\Delta u > 0$.

如果 u 在点 $x_0 \in U$ 达到极大值, 即

$$u(x_0) = \max_{x \in \bar{U}} u(x),$$

则由多元微积分的定理知, u 在 x_0 的梯度向量 $\nabla u(x_0) = 0$, 且 Hessian 矩阵 $D^2u(x_0)$ 半负定. 因此 $\Delta u(x_0) = \text{Tr}(D^2u(x_0)) \leq 0$, 与假设矛盾. 因此 u 只能在 ∂U 上取得在 \bar{U} 中的最大值, 此时当然也有

$$\max_{\bar{U}} u \leq \max_{\partial U} u.$$

对于一般情况 $\Delta u \geq 0$, 设 $U \subset B(0, R)$.

$\forall \varepsilon > 0$, 考虑辅助函数

$$u^\varepsilon(x) = u(x) - \varepsilon(R^2 - |x|^2),$$

显然 $u^\varepsilon \leq u$. 计算 u^ε 的一阶和二阶导数得

$$u_i^\varepsilon(x) = u_i(x) + 2\varepsilon x_i, \quad u_{ii}^\varepsilon = u_{ii} + 2\varepsilon.$$

因此

$$\Delta u^\varepsilon = \Delta u + 2n\varepsilon > 0.$$

从而由特殊情况的结论知

$$\max_{\bar{U}} u^\varepsilon \leq \max_{\partial U} u^\varepsilon.$$

因此

$$\begin{aligned} \max_{\bar{U}} u &= \max_{\bar{U}} (u^\varepsilon + \varepsilon(R^2 - |x|^2)) \leq \max_{\bar{U}} u^\varepsilon + \varepsilon R^2 \\ &\leq \max_{\partial U} u^\varepsilon + \varepsilon R^2 \leq \max_{\partial U} u + \varepsilon R^2. \end{aligned}$$

令 $\varepsilon \rightarrow 0$ 即可得结论.

(2) 将结论 (1) 应用到 $-u$ 即可. □

注 1

弱极值原理的意思是说, 如果 u 在 U 中下调和, 那么 u 一定能在 U 的边界上取到 u 在 \bar{U} 中的最大值.

证:

设 U 为 \mathbb{R}^n 中的有界开集, $u, v \in C^2(U) \cap C(\bar{U})$,

(1) 若在 U 中有 $\Delta u \leq 0$, 在 ∂U 上有 $u \leq 0$, 则在 U 中也有 $u \leq 0$;

(2)(比较原理) 若在 U 中有 $\Delta u \leq \Delta v$, 在 ∂U 上有 $u \leq v$, 则在 U 中也有 $u \leq v$. □

从以上的弱极值原理我们可以证明 Poisson 方程的 Dirichlet 问题的解的惟一性.

证:

设 U 为 \mathbb{R}^n 中的有界开集, 且有 $f \in C(U)$ 及 $g \in C(\partial U)$, 则 Dirichlet 问题

$$\begin{cases} -\Delta u = f, & x \in U, \\ u = g, & x \in \partial U \end{cases}$$

最多存在一个解 $u \in C^2(U) \cap C(\bar{U})$. □

证明. 假如存在两个解 u_1 和 u_2 , 则在 U 中有 $\Delta u_1 = \Delta u_2$, 在 ∂U 上有 $u_1 = u_2$, 故由比较原理, 在 U 中有 $u_1 \leq u_2$ 和 $u_2 \leq u_1$, 故 $u_1 \equiv u_2$. \square

引理 1 Hopf 引理

B_R 为 \mathbb{R}^n 中以原点为中心, R 为半径的开球, $u \in C^2(B_R) \cap C^1(\overline{B_R})$, 在 B_R 中有 $\Delta u \geq 0$, 且存在 $x_0 \in \partial B_R$, 使得 $\forall x \in B_R, u(x_0) > u(x)$, 则有 $\frac{\partial u}{\partial \nu}(x_0) > 0$. (ν 为 B_R 在 x_0 处的单位外法向.)

证明. 构造辅助函数 $w(x) = e^{-\mu|x|^2} - e^{-\mu R^2}$, 其中 $\mu > 0$ 待定. 可知 $w(x) \leq 1$. 计算其一阶和二阶导数得

$$w_i(x) = -2\mu x_i e^{-\mu|x|^2}, w_{ii}(x) = -2\mu e^{-\mu|x|^2} + 4\mu^2 x_i^2 e^{-\mu|x|^2},$$

故有

$$\Delta w = \mu(4\mu|x|^2 - 2n)e^{-\mu|x|^2}.$$

记 $D := B_R \setminus \overline{B_{R/2}}$, 则 $\partial D = \{|x| = R\} \cup \{|x| = R/2\}$. 因此取 $\mu > \frac{2n}{R^2}$ 时, 有

$$\Delta w > \mu(\mu R^2 - 2n)e^{-\mu|x|^2} > 0.$$

令 $v(x) := u(x) - u(x_0) + \varepsilon w(x)$, $\varepsilon > 0$ 待定. 因此在 D 中有 $\Delta v = \Delta u + \varepsilon \Delta w > 0$.

在 $\{|x| = R\}$ 上, $w = 0$, 故由 u 的连续性, 有 $v(x) \leq 0$;

在 $\{|x| = R/2\}$ 上, 由 u 的连续性, 取 $\varepsilon > 0$ 足够小, 使得在 $\{|x| = R/2\}$ 上均有 $u(x) \leq u(x_0) - \varepsilon$, 又因为 $w(x) \leq 1$, 从而有

$$v(x) = u(x) - u(x_0) + \varepsilon w(x) \leq -\varepsilon + \varepsilon w(x) \leq -\varepsilon + \varepsilon \leq 0.$$

因此在 ∂D 上 $v \leq 0$. 故由推论 1.1 可知, 在 D 中也有 $v \leq 0$. 而 $v(x_0) = 0$, 故 v 在 x_0 处达到极大值, 因此有

$$\frac{\partial v}{\partial \nu}(x_0) \geq 0.$$

因此

$$\begin{aligned} \frac{\partial u}{\partial \nu}(x_0) &\geq -\varepsilon \frac{\partial w}{\partial \nu}(x_0) = -\varepsilon \langle Dw(x_0), \frac{1}{R}x_0 \rangle \\ &= -\varepsilon(-2\mu e^{-\mu|x_0|^2} \cdot \frac{|x_0|^2}{R}) = 2\varepsilon\mu R e^{-\mu R^2} > 0. \end{aligned}$$

\square

注 2

由 Hopf 引理中 x_0 点处的条件已经可以得出 $\frac{\partial u}{\partial \nu}(x_0) \geq 0$, 因此 Hopf 引理实际上是说, 在加上 u 是下调和的条件后, 可以得出严格的不等号.

注 3

上面 Hopf 引理中条件其实只需要 $u \in C^2(B_R) \cap C^1(B_R \cup \{x_0\})$ 即可, 这里只是为了叙述方便. 我们还可以继续减弱条件为 $u \in C^2(B_R) \cap C(B_R \cup \{x_0\})$, 则结论也相应地减弱为, 对 x_0 点处任意的满足 $\nu \cdot \mathbf{n}(x_0) > 0$ 的外法向 ν , 都有

$$\liminf_{t \rightarrow 0^+} \frac{u(x_0) - u(x_0 - t\nu)}{t} > 0.$$

证: [Hopf 引理 (一般区域上)]

设 U 为 \mathbb{R}^n 中的有界开集, $u \in C^2(U) \cap C^1(\bar{U})$, 在 U 中有 $\Delta u \geq 0$, 且存在 $x_0 \in \partial U$, 使得 $\forall x \in U, u(x_0) > u(x)$, 如果 U 在 x_0 处还有内切球条件, 即存在开球 $B \subset U$, 使得 $x_0 \in \partial B$, 那么就有 $\frac{\partial u}{\partial \nu}(x_0) > 0$. (ν 为 U 在 x_0 处的单位外法向.) \square

注 4

如果 ∂U 在 x_0 附近是 C^2 的, 那么 U 在 x_0 处自动地有内切球条件.

由 Hopf 引理可以证明强极值原理.

定理 2 强极值原理

设 U 为 \mathbb{R}^n 中的连通有界开集, $u \in C^2(U) \cap C(\bar{U})$ 在 U 上满足 $\Delta u \geq 0$. 若 u 在 $x_0 \in U$ 取得在 \bar{U} 上的最大值, 则 u 在 \bar{U} 上是常数.

证明. 记

$$M = \max_{\bar{U}} u.$$

考虑集合 $D = \{x \in U \mid u(x) = M\}$, 由条件知 D 非空. 由函数 $u(x)$ 的连续性, D 相对于 U 是闭的. 下面证 D 相对于 U 是开的, 从而由 U 连通且 D 非空, 可知 $D = U$, 从而得到结论.

设 $x_* \in D$, 则存在球 $B(x_*, 2r) \subset U$. 如果 x_* 不是 D 的内点, 那么存在 $\tilde{x} \in (U \setminus D) \cap B(x_*, r)$. 记 $d = \text{dist}\{\tilde{x}, \bar{D}\} = \inf_{y \in \bar{D}} |y - \tilde{x}|$, 则 $d \leq \text{dist}\{\tilde{x}, x_*\} \leq r$, 因此 $B(\tilde{x}, d) \subset B(x_*, 2r) \subset U$, 从而 $B(\tilde{x}, d) \subset U \setminus D$, 即在 $B(\tilde{x}, d)$ 中均有 $u(x) < M$.

记 $x_0 \in \partial B(\tilde{x}, d) \cap D$, 则 $u(x_0) = M$. 因此 x_0 为 u 的极值点, $\nabla u(x_0) = 0$. 但由 Hopf 引理 (引理 1.1), 存在一个方向 ν , 使得 $\frac{\partial u}{\partial \nu}(x_0) > 0$, 于是导出矛盾, 从而 x_0 是 D 的内点, D 相对于 U 是开的. 证毕. \square

1.2

方程 $-\Delta u + c(x)u = 0$ 的极大值原理

在讨论一般二阶线性椭圆方程的极大值原理前, 我们先讨论一个相对简单的情形, 即方程

$$-\Delta u + c(x)u = 0, x \in U,$$

其中 U 为 \mathbb{R}^n 中的有界开集, $c(x)$ 为 U 中有界非负函数.

定理 3 弱极值原理

设 U 为 \mathbb{R}^n 中的有界开集, $u \in C^2(U) \cap C(\bar{U})$, c 为 U 中有界非负函数. 定义椭圆算子 L 为 $Lu := -\Delta u + cu, \forall u \in C^2(U)$. 若在 U 中有 $Lu \leq 0$, 则 u 必在 ∂U 上达到它在 \bar{U} 中的非负最大值, 即

$$\max_{\bar{U}} u \leq \max_{\partial U} u^+,$$

其中 $u^+(x) = \max\{u(x), 0\}$.

证明. 证明与定理1.1的证明类似.

先考虑特殊情况: 假设 u 在 U 中满足 $Lu = -\Delta u + cu < 0$.

如果 u 在点 $x_0 \in U$ 达到其在 \bar{U} 中的非负最大值, 即

$$u(x_0) = \max_{x \in \bar{U}} u(x) \geq 0,$$

则 $D^2u(x_0)$ 半负定. 因此 $\Delta u(x_0) \leq 0$, 又因为 $c(x) \geq 0, \forall x \in U$, 故有

$$Lu(x_0) = -\Delta u(x_0) + c(x_0)u(x_0) \geq 0,$$

与假设矛盾. 因此在此情况下 u 只能在 ∂U 上取得在 \bar{U} 中的非负最大值, 此时当然也有

$$\max_{\bar{U}} u \leq \max_{\partial U} u^+.$$

对于一般情况 $-\Delta u + cu \leq 0$, 设 $U \subset B(0, R)$.

$\forall \varepsilon > 0$, 仍然考虑辅助函数

$$u_\varepsilon(x) = u(x) - \varepsilon(R^2 - |x|^2),$$

显然 $u_\varepsilon \leq u$. 且

$$\Delta u_\varepsilon = \Delta u + 2n\varepsilon,$$

又由于 $c \geq 0$, 故

$$Lu_\varepsilon = -\Delta u_\varepsilon + cu_\varepsilon = Lu - 2n\varepsilon - c\varepsilon(R^2 - |x|^2) \leq Lu - 2n\varepsilon < 0.$$

从而由特殊情况的结论知

$$\max_{\bar{U}} u_\varepsilon \leq \max_{\partial U} u_\varepsilon^+.$$

因此

$$\begin{aligned} \max_{\bar{U}} u &= \max_{\bar{U}} (u_\varepsilon + \varepsilon(R^2 - |x|^2)) \leq \max_{\bar{U}} u_\varepsilon + \varepsilon R^2 \\ &\leq \max_{\partial U} u_\varepsilon^+ + \varepsilon R^2 \leq \max_{\partial U} u^+ + \varepsilon R^2. \end{aligned}$$

令 $\varepsilon \rightarrow 0$ 即可得结论. 证毕. □

引理 2 Hopf 引理

B_R 为 \mathbb{R}^n 中以原点为中心, R 为半径的开球, $u \in C^2(B_R) \cap C^1(\bar{B}_R)$, c 为 B_R 中有界非负函数. L 定义同上. 在 B_R 中有 $Lu \leq 0$, 且存在 $x_0 \in \partial B_R$, 使得 $u(x_0) \geq 0$, 且 $\forall x \in B_R, u(x_0) > u(x)$, 则有 $\frac{\partial u}{\partial \nu}(x_0) > 0$. (ν 为 B_R 在 x_0 处的单位外法向.)

证明. 证明与引理1.1的证明类似.

仍构造辅助函数 $w(x) = e^{-\mu|x|^2} - e^{-\mu R^2}$, 其中 $\mu > 0$ 待定. 可知 $w(x) \leq 1$. 且

$$\Delta w = (4\mu^2|x|^2 - 2\mu n)e^{-\mu|x|^2}.$$

故由 $c(x) \geq 0$ 可得

$$\begin{aligned} Lw &= -\Delta w + cw = -(4\mu^2|x|^2 - 2\mu n - c)e^{-\mu|x|^2} - ce^{-\mu R^2} \\ &\leq -(4\mu^2|x|^2 - 2\mu n - c)e^{-\mu|x|^2}. \end{aligned}$$

记 $D := B_R \setminus \overline{B_{R/2}}$, 则 $\partial D = \{|x| = R\} \cup \{|x| = R/2\}$, 且在 D 中, 有 $|x| > \frac{R}{2}$, 进而有

$$Lw \leq -(\mu^2 R^2 - 2\mu n - c)e^{-\mu|x|^2}.$$

由 c 有界, 可取 μ 足够大, 使得在 D 中有 $Lw < 0$.

令 $v(x) := u(x) - u(x_0) + \varepsilon w(x)$, $\varepsilon > 0$ 待定. 因此在 D 中有

$$\begin{aligned} Lv &= -\Delta v + cv = -\Delta u - \varepsilon \Delta w + cu - cu(x_0) + c\varepsilon w \\ &= Lu + \varepsilon Lw - cu(x_0) < 0. \end{aligned}$$

之后用与引理1.1相同的方式证明, 可得在 ∂D 上 $v \leq 0$, 由上面的弱极值定理 1.2 可知, 在 D 中也有 $v \leq 0$. 而 $v(x_0) = 0$, 故 $v(x_0) = \max_D v$, 因此有 $\frac{\partial v}{\partial \nu}(x_0) \geq 0$. 进而有

$$\frac{\partial u}{\partial \nu}(x_0) \geq -\varepsilon \frac{\partial w}{\partial \nu}(x_0) = 2\varepsilon \mu R e^{-\mu R^2} > 0.$$

证明完毕. □

定理 4 强极值原理

设 U 为 \mathbb{R}^n 中的连通有界开集, $u \in C^2(U) \cap C(\overline{U})$, c 为 U 中有界非负函数. 定义椭圆算子 L 为 $Lu = -\Delta u + cu, \forall u \in C^2(U)$. 若在 U 中有 $Lu \leq 0$, 且 u 在 U 内取得在 \overline{U} 上的非负最大值, 则 u 在 \overline{U} 上是常数.

证明. 利用 Hopf 引理 1.2, 证明的其他部分与定理 1.1 的证明完全相同. □

1.3

一般二阶线性椭圆方程的极大值原理

我们考虑讨论一般二阶线性椭圆方程的极大值原理如下.

定义 1

本节我们定义算子 L 为

$$Lu := \sum_{i,j=1}^n a_{ij}(x) \partial_{ij} u + \sum_{i=1}^n b_i(x) \partial_i u + c(x)u, \forall u \in C^2(U),$$

有时也简记为

$$Lu := \sum_{i,j} a_{ij} u_{ij} + \sum_i b_i u_i + cu, \forall u \in C^2(U),$$

其中 $\{a_{ij}\}_{i,j=1}^n, \{b_i\}_{i=1}^n, c$ 均为 U 中逐点定义的有界函数, U 为 \mathbb{R}^n 中的一个有界区域.

定义 2 椭圆算子

定义算子 L 如定义 1.3 那样, 则

- (1) 设 $x \in U$, 称 L 在 x 是椭圆的, 若矩阵 $(a_{ij}(x))$ 正定.
 (2) 称 L 在 U 中是严格椭圆的, 若 $\exists 0 < \lambda < \infty$, 使得 $\forall x \in U, \forall \xi \in \mathbb{R}^n$, 有

$$\sum_{i,j=1}^n a_{ij}(x)\xi_i\xi_j \geq \lambda|\xi|^2.$$

- (3) 称 L 在 U 中是一致椭圆的, 若 $\exists 0 < \lambda \leq \Lambda < \infty$, 使得 $\forall x \in U, \forall \xi \in \mathbb{R}^n$, 有

$$\lambda|\xi|^2 \leq \sum_{i,j=1}^n a_{ij}(x)\xi_i\xi_j \leq \Lambda|\xi|^2.$$

注 5

上面定义中的 (2) 表明 $\forall x \in U$, 矩阵 $(a_{ij}(x))$ 的特征值均大于等于 λ ; (3) 表明 $\forall x \in U$, 矩阵 $(a_{ij}(x))$ 的特征值均属于 $[\lambda, \Lambda]$.

定理 5 弱极值原理

设 U 为 \mathbb{R}^n 中的有界区域, $u \in C^2(U) \cap C(\bar{U})$. 定义算子 L 如定义 1.3 那样, 且 L 在 U 中严格椭圆, 其中 $c(x) \leq 0, \forall x \in U$. 若在 U 中有 $Lu \geq 0$, 则

$$\max_{\bar{U}} u \leq \max_{\partial U} u^+,$$

其中 $u^+(x) = \max\{u(x), 0\}$.

证明. 证明与定理 1.2 的证明基本类似, 只是修改了辅助函数.

先考虑特殊情况: 假设 u 在 U 中满足 $Lu > 0$.

如果 u 在点 $x_0 \in U$ 达到其在 \bar{U} 中的非负最大值, 即

$$u(x_0) = \max_{x \in \bar{U}} u(x) \geq 0,$$

则 $\nabla u(x_0) = 0$, 矩阵 $B := D^2u(x_0) = (u_{ij}(x_0))$ 半负定. 又因为 $c(x) \leq 0, \forall x \in U$, 故有

$$\begin{aligned} Lu(x_0) &= \sum_{i,j} a_{ij}(x_0)u_{ij}(x_0) + \sum_i b_i(x_0)u_i(x_0) + c(x_0)u(x_0) \\ &= \sum_{i,j} a_{ij}(x_0)u_{ij}(x_0) + c(x_0)u(x_0) \\ &\leq \sum_{i,j} a_{ij}(x_0)u_{ij}(x_0). \end{aligned}$$

而由 L 的椭圆性条件, $A := (a_{ij}(x_0))$ 正定, 故 AB 半正定. 又由 B 的对称性可知,

$$Lu(x_0) \leq \sum_{i,j} a_{ij}(x_0)u_{ij}(x_0) = \text{Tr}(AB) \leq 0.$$

与假设矛盾. 因此在此情况下 u 只能在 ∂U 上取得在 \bar{U} 中的非负最大值, 此时当然也有

$$\max_{\bar{U}} u \leq \max_{\partial U} u^+.$$

对于一般情况 $Lu \geq 0$, 考虑辅助函数

$$w(x) = u(x) + \varepsilon e^{\mu x_1},$$

其中 $\varepsilon, \mu > 0$ 待定. 则有

$$Lw = Lu + \varepsilon(a_{11}\mu^2 + b_1\mu + c)e^{\mu x_1}.$$

由椭圆性条件知, $a_{11} \geq \lambda$. 故取 μ 足够大, 使得在 U 中有 $a_{11}\mu^2 + b_1\mu + c > 0$, 从而在 U 中有 $Lw > 0$. 从而由特殊情况的结论知

$$\max_{\bar{U}} w \leq \max_{\partial U} w^+.$$

因此

$$\max_{\bar{U}} u \leq \max_{\bar{U}} w \leq \max_{\partial U} w^+ \leq \max_{\partial U} u^+ + \max_{\partial U} \varepsilon e^{\mu x_1}.$$

由 U 的有界性, 令 $\varepsilon \rightarrow 0$ 即可得

$$\max_{\bar{U}} u \leq \max_{\partial U} u^+,$$

于是就完成了证明. □

证:

设 U 为 \mathbb{R}^n 中的有界区域, $u, v \in C^2(U) \cap C(\bar{U})$. 定义算子 L 如定义 1.3 那样, 且 L 在 U 中严格椭圆, 其中 $c(x) \leq 0, \forall x \in U$.

- (1) 若在 U 中有 $Lu \leq 0$, 在 ∂U 上有 $u \leq 0$, 则在 U 中也有 $u \leq 0$;
- (2)(比较原理) 若在 U 中有 $Lu \leq Lv$, 在 ∂U 上有 $u \leq v$, 则在 U 中也有 $u \leq v$.
- (3) $\forall f \in C(U), \forall g \in C(\partial U)$, Dirichlet 问题

$$\begin{cases} Lu = f, & x \in U, \\ u = g, & x \in \partial U \end{cases}$$

最多存在一个解 $u \in C^2(U) \cap C(\bar{U})$. □

引理 3 Hopf 引理

B_R 为 \mathbb{R}^n 中以原点为中心, R 为半径的开球, $u \in C^2(B_R) \cap C^1(\bar{B}_R)$. 定义算子 L 如定义 1.3 那样, 且 L 在 U 中严格椭圆, 其中 $c(x) \leq 0, \forall x \in U$. 在 B_R 中有 $Lu \geq 0$, 且存在 $x_0 \in \partial B_R$, 使得 $u(x_0) \geq 0$, 且 $\forall x \in B_R, u(x_0) > u(x)$, 则有 $\frac{\partial u}{\partial \nu}(x_0) > 0$. (ν 为 B_R 在 x_0 处的单位外法向.)

证明. 与引理 1.2 的证明类似.

仍构造辅助函数 $w(x) = e^{-\mu|x|^2} - e^{-\mu R^2}$, 其中 $\mu > 0$ 待定. 可知 $w(x) \leq 1$. 且

$$\partial_{ij} w(x) = (4\mu^2 x_i x_j - 2\mu \delta_{ij}) e^{-\mu|x|^2}.$$

故由 $c(x) \leq 0$ 及椭圆性条件可得

$$\begin{aligned} Lw &= (4\mu^2 \sum_{i,j} a_{ij} x_i x_j - 2\mu \sum_{i,j} a_{ij} \delta_{ij} - 2\mu \sum_i b_i x_i + c) e^{-\mu|x|^2} - c e^{-\mu R^2} \\ &\geq [4\mu^2 \lambda |x|^2 - 2\mu \sum_i (a_{ii} + b_i x_i) + c] e^{-\mu|x|^2}. \end{aligned}$$

记 $D := B_R \setminus \overline{B_{R/2}}$, 则在 D 中有

$$Lw \geq [\lambda R^2 \mu^2 - 2\mu \sum_i (a_{ii} + b_i x_i) + c] e^{-\mu|x|^2}.$$

由 c 有界, $\lambda > 0$, 可取 μ 足够大, 使得在 D 中有 $Lw > 0$. 之后与引理 1.2 的证明完全相同. \square

定理 6 强极值原理

设 U 为 \mathbb{R}^n 中的连通有界区域, $u \in C^2(U) \cap C(\overline{U})$. 定义算子 L 如定义 1.3 那样, 且 L 在 U 中严格椭圆, 其中 $c(x) \leq 0, \forall x \in U$. 若在 U 中有 $Lu \geq 0$, 且 u 在 U 内取得在 \overline{U} 上的非负最大值, 则 u 在 \overline{U} 上是常数.

证明. 利用 Hopf 引理 1.3, 证明的其他部分与定理 1.1 的证明完全相同. \square

1.4

Aleksandrov 极大值原理

本节均设 U 为 \mathbb{R}^n 中一有界区域.

定义 3 上接触集、下接触集

设 $u \in C(U)$, 定义 u 的上接触集为

$$\Gamma_u^+ = \{y \in U \mid \exists p = p(y) \in \mathbb{R}^n, \text{使得 } \forall x \in U, \text{有 } u(x) \leq u(y) + p \cdot (x - y)\}.$$

同样地, 可以定义 u 的下接触集为

$$\Gamma_u^- = \{y \in U \mid \exists p = p(y) \in \mathbb{R}^n, \text{使得 } \forall x \in U, \text{有 } u(x) \geq u(y) + p \cdot (x - y)\}.$$

注 6

- (1) 若 $u \in C^1(U)$, 则 $\Gamma_u^+ = \{y \in U \mid \forall x \in U, u(x) \leq u(y) + \nabla u(y) \cdot (x - y)\}$.
- (2) 若 $u \in C^2(U)$, 则 Hessian 矩阵 D^2u 在 Γ_u^+ 中半负定.

引理 4

设 $g \in L^1_{loc}(\mathbb{R}^n)$ 为非负函数, 则 $\forall u \in C^2(U) \cap C(\bar{U})$, 有

$$\int_{B(0, \frac{M}{d})} g(y) dy \leq \int_{\Gamma_u^+} g(\nabla u) |det(D^2 u)| dx,$$

其中 Γ_u^+ 为 u 的上接触集, $M = \sup_U u - \sup_{\partial U} u^+$, $d = \text{diam}(U) = \sup_{x, y \in U} |x - y|$.

证明. 注意要证的不等式与函数 u 加减任意常数无关, 故不妨设 u 在 ∂U 上恒非正, 从而 $M = \sup_U u$.

考虑映射 $\phi_\varepsilon := \nabla u - \varepsilon Id : U \rightarrow \mathbb{R}^n$, 其中 Id 为恒同映射. 则有 $\nabla \phi_\varepsilon = D^2 u - \varepsilon I_n$, 其为 Γ_u^+ 上的负定矩阵. 故由变量替换可知

$$\int_{\phi_\varepsilon(\Gamma_u^+ \cap U^+)} g(y) dy = \int_{\Gamma_u^+ \cap U^+} g(\phi_\varepsilon) |det(D^2 u - \varepsilon Id)| dx.$$

令 $\varepsilon \rightarrow 0$, 得

$$\int_{\nabla u(\Gamma_u^+ \cap U^+)} g(y) dy \leq \int_{\Gamma_u^+ \cap U^+} g(\nabla u) |det(D^2 u)| dx.$$

下面我们证明: $B(0, \frac{M}{d}) \subset \nabla u(\Gamma_u^+ \cap U^+)$, 即 $\forall a \in \mathbb{R}^n$, 且 $|a| < \frac{M}{d}$, $\exists x \in \Gamma_u^+ \cap U^+$, 使得 $a = \nabla u(x)$.

假设 u 在 $x_0 \in U$ 处达到 $\sup_U u = M$, 则 $\forall a \in \mathbb{R}^n$, 且 $|a| < \frac{M}{d}$, 考虑 $L(x) := u(x) - a \cdot (x - x_0)$, 则 $L(x_0) = M$, 且 $\forall x \in \partial U, L(x) \leq -a \cdot (x - x_0) \leq |a| |x_0 - x| < M$. 因此 L 的极大值在某 $x_a \in U$ 达到, 且 $L(x_a) \geq M$. 故

$$u(x_a) = L(x_a) + a \cdot (x_a - x_0) \geq M + a \cdot (x_a - x_0) > 0,$$

而且对于任意的 $x \in U$, 由 $L(x) \leq L(x_a)$, 知

$$u(x) = L(x) + a \cdot (x - x_0) \leq L(x_a) + a \cdot (x - x_0) = u(x_a) + a \cdot (x - x_a).$$

因此 $x_a \in \Gamma_u^+ \cap U^+$ 且 $a = \nabla u(x_a)$. 从而有

$$\int_{B(0, \frac{M}{d})} g(y) dy \leq \int_{\nabla u(\Gamma_u^+ \cap U^+)} g(y) dy \leq \int_{\Gamma_u^+ \cap U^+} g(\nabla u) |det(D^2 u)| dx.$$

最后再由 u 加减常数不改变 ∇u , 得到所求证的结果. \square

定理 7

(1) $\forall u \in C^2(U) \cap C(\bar{U})$, 有

$$\sup_U u \leq \sup_{\partial U} u^+ + \frac{d}{\omega_n^{\frac{1}{n}}} \left(\int_{\Gamma_u^+} |det(D^2 u)| dx \right)^{\frac{1}{n}},$$

其中 ω_n 为 \mathbb{R}^n 中半径为 1 的球的体积.

(2) 设算子 L 定义为 $Lu := \sum_{i, j=1}^n a_{ij} u_{ij}$, 其中 $a_{ij} \in C(U)$, $A = (a_{ij})$ 对称, 且 $\forall x \in U, (a_{ij}(x))$ 正定. 记 $D = \det A, D^* = \sqrt[n]{D}$. 设 $u \in C^2(U) \cap C(\bar{U})$ 在 U 中满足 $Lu \geq f$, 其中 $f \in C(U), \frac{f^-}{D^*} \in L^n(U)$, 则有

$$\sup_U u \leq \sup_{\partial U} u^+ + \frac{d}{n\omega_n^{\frac{1}{n}}} \left\| \frac{f^-}{D^*} \right\|_{L^n(\Gamma_u^+)}.$$

证明. (1) 在引理1.4中令 $g \equiv 1$, 则有

$$\omega_n \left(\frac{\sup_U u - \sup_{\partial U} u^+}{d} \right)^n \leq \int_{\Gamma_u^+} |\det(D^2u)| dx,$$

因此证得所求结论.

(2) 由于 $A = (a_{ij})$ 正定, 在 Γ_u^+ 上有

$$\det(A) \cdot \det(-D^2u) \leq \left(-\frac{1}{n} \sum_{i,j=1}^n a_{ij}u_{ij} \right)^n \leq \left(\frac{-f}{n} \right)^n \leq \left(\frac{f^-}{n} \right)^n,$$

即

$$\det(-D^2u) \leq \left(\frac{f^-}{nD^*} \right)^n.$$

代入 (1) 即得. □

定理 8 Aleksandrov 极大值原理

设算子 L 定义为 $Lu := \sum_{i,j=1}^n a_{ij}u_{ij} + \sum_{i=1}^n b_iu_i + cu$, 其中 $a_{ij} \in C(U)$, $A = (a_{ij})$ 对称, 且 $\forall x \in U, (a_{ij}(x))$ 正定. 记 $D = \det A, D^* = \sqrt[n]{D}$. 设 $u \in C^2(U) \cap C(\bar{U})$ 在 U 中满足 $Lu \geq f$, 其中 $f \in C(U), \frac{|b|}{D^*}, \frac{f}{D^*} \in L^n(U)$, 且在 U 中有 $c \leq 0$, 则

$$\sup_U u \leq \sup_{\partial U} u^+ + Cd \left\| \frac{f^-}{D^*} \right\|_{L^n(\Gamma_u^+)},$$

其中

$$C = \left\{ \exp \left[\frac{2^{n-2}}{\omega_n n^{n-1}} \left(\left\| \frac{b}{D^*} \right\|_{L^n(\Gamma_u^+)} + 1 \right) \right] - 1 \right\}^{\frac{1}{n}}.$$

证明. 由 $Lu \geq f$ 的条件可知, 在 $U^+ := \{x \in U | u(x) > 0\}$ 中有

$$\begin{aligned} -\sum_{i,j=1}^n a_{ij}u_{ij} &\leq \sum_{i=1}^n b_iu_i + cu - f \\ &\leq \sum_{i=1}^n b_iu_i - f \quad (\text{因 } c \leq 0, \text{ 且在 } U^+ \text{ 上有 } u > 0) \\ &\leq |b| \cdot |\nabla u| + \frac{f^-}{\varepsilon} \cdot \varepsilon \quad (\text{Cauchy 不等式, 其中 } \varepsilon > 0 \text{ 为待定常数}) \\ &\leq \left(|b|^n + \left(\frac{f^-}{\varepsilon} \right)^n \right)^{\frac{1}{n}} \cdot \left(|\nabla u|^n + \varepsilon^n \right)^{\frac{1}{n}} \cdot (1+1)^{\frac{n-2}{n}} \quad (\text{H\"older 不等式}), \end{aligned}$$

故有

$$\left(-\sum_{i,j=1}^n a_{ij}u_{ij} \right)^n \leq \left(|b|^n + \left(\frac{f^-}{\varepsilon} \right)^n \right) \cdot \left(|\nabla u|^n + \varepsilon^n \right) \cdot 2^{n-2}. \quad (1)$$

现在取辅助函数

$$g(x) = \frac{1}{|x|^n + \varepsilon^n},$$

并应用引理1.4以及 (1) 式可得,

$$\begin{aligned} \int_{B(0, \frac{M}{d})} g &\leq \int_{\Gamma_u^+ \cap U^+} \frac{|\det(D^2u)|}{|\nabla u|^n + \varepsilon^n} \\ &\leq \int_{\Gamma_u^+ \cap U^+} \frac{1}{D(|\nabla u|^n + \varepsilon^n)} \cdot \left(-\frac{1}{n} \sum_{i,j=1}^n a_{ij} u_{ij}\right)^n \\ &\leq \int_{\Gamma_u^+ \cap U^+} \frac{2^{n-2}}{D \cdot n^n} \cdot \left(|b|^n + \left(\frac{f^-}{\varepsilon}\right)^n\right) \\ &= \frac{2^{n-2}}{n^n} \left(\left\| \frac{b}{D^*} \right\|_{L^n(\Gamma_u^+ \cap U^+)}^n + \varepsilon^{-n} \left\| \frac{f^-}{D^*} \right\|_{L^n(\Gamma_u^+ \cap U^+)}^n \right). \end{aligned}$$

而

$$\begin{aligned} \int_{B(0, \frac{M}{d})} g(y) dy &= \int_{B(0, \frac{M}{d})} \frac{dx}{|x|^n + \varepsilon^n} = \omega_n \int_0^{\frac{M}{d}} \frac{r^{n-1} dr}{r^n + \varepsilon^n} \\ &= \frac{\omega_n}{n} \log(r^n + \varepsilon^n) \Big|_{r=0}^{\frac{M}{d}} = \frac{\omega_n}{n} \log \left[\left(\frac{M}{d\varepsilon}\right)^n + 1 \right], \end{aligned}$$

因此

$$\left(\frac{M}{d\varepsilon}\right)^n \leq \exp \left[\frac{2^{n-2}}{\omega_n n^{n-1}} \left(\left\| \frac{b}{D^*} \right\|_{L^n(\Gamma_u^+ \cap U^+)}^n + \varepsilon^{-n} \left\| \frac{f^-}{D^*} \right\|_{L^n(\Gamma_u^+ \cap U^+)}^n \right) \right] - 1,$$

故有

$$\sup_U u \leq \sup_{\partial U} u^+ + d\varepsilon \left\{ \exp \left[\frac{2^{n-2}}{\omega_n n^{n-1}} \left(\left\| \frac{b}{D^*} \right\|_{L^n(\Gamma_u^+ \cap U^+)}^n + \varepsilon^{-n} \left\| \frac{f^-}{D^*} \right\|_{L^n(\Gamma_u^+ \cap U^+)}^n \right) \right] - 1 \right\}^{\frac{1}{n}}.$$

若 $f^- \equiv 0$, 则令 $\varepsilon \rightarrow 0^+$ 即可得到结论;

若 $f^- \not\equiv 0$, 则取

$$\varepsilon = \left\| \frac{f^-}{D^*} \right\|_{L^n(\Gamma_u^+ \cap U^+)},$$

则得所求结论

$$\sup_U u \leq \sup_{\partial U} u^+ + d \left\{ \exp \left[\frac{2^{n-2}}{\omega_n n^{n-1}} \left(\left\| \frac{b}{D^*} \right\|_{L^n(\Gamma_u^+)}^n + 1 \right) \right] - 1 \right\}^{\frac{1}{n}} \left\| \frac{f^-}{D^*} \right\|_{L^n(\Gamma_u^+)}.$$

证明完毕. □

第 2 节

椭圆方程的 Harnack 不等式

2.1

Laplace 方程的 Harnack 不等式

以下均设 B_r 为 R^n 中以原点为中心, 以 r 为半径的开球.

引理 5 微分 Harnack 不等式/Li-Yau 不等式

设在 B_1 中有 $\Delta u = 0$ 及 $u > 0$, 则存在常数 $C = C(n) > 0$, 使得

$$\sup_{x \in \overline{B_{\frac{1}{2}}}} |\nabla \log u(x)| \leq C.$$

证明. 为计算方便, 我们将研究 $|\nabla \log u(x)|^2$ 的上确界.

记 $v = \log u$, $w = |\nabla v|^2 = |\nabla \log u|^2$, 则有

$$v_i = \frac{u_i}{u}, \quad v_{ii} = \frac{u_{ii}}{u} - \left(\frac{u_i}{u}\right)^2 = \frac{u_{ii}}{u} - (v_i)^2,$$

$$w_i = 2 \sum_{k=1}^n v_k v_{ki}, \quad w_{ii} = 2 \sum_{k=1}^n v_{ik}^2 + 2 \sum_{k=1}^n v_k v_{iik},$$

故由 $\Delta u = 0$ 的条件可得

$$\Delta v = \frac{\Delta u}{u} - |\nabla u|^2 = -|\nabla u|^2 = -w,$$

$$\Delta w = 2 \sum_{i,k=1}^n v_{ik}^2 + 2 \sum_{k=1}^n v_k \partial_k (\Delta v) = 2 \sum_{i,k=1}^n v_{ik}^2 - 2 \sum_{k=1}^n v_k w_k,$$

引入截断函数 $\varphi \in C_0^2(B_1)$, 且 φ 在 B_1 上非负, 在 $\overline{B_{\frac{1}{2}}}$ 上恒为 1, 则由 Cauchy 不等式得,

$$\begin{aligned} \Delta(\varphi w) &= w \Delta \varphi + 2 \nabla \varphi \cdot \nabla w + \varphi \Delta w \\ &= w \Delta \varphi + 4 \sum_{i,k=1}^n \varphi_i v_k v_{ki} + \varphi \left(2 \sum_{i,k=1}^n v_{ik}^2 - 2 \sum_{k=1}^n v_k w_k \right) \\ &= w \Delta \varphi + 4 \sum_{i,k=1}^n \varphi_i v_k v_{ki} + 2 \varphi \sum_{i,k=1}^n v_{ik}^2 - 2 \sum_{k=1}^n v_k (\varphi w)_k + 2 \sum_{k=1}^n v_k \varphi_k w \\ &= |\nabla v|^2 \Delta \varphi + 2 \cdot \sum_{i,k=1}^n \frac{2 \varphi_i v_k}{\sqrt{\varphi}} \cdot \sqrt{\varphi} v_{ki} + 2 \varphi \sum_{i,k=1}^n v_{ik}^2 - 2 \nabla v \cdot \nabla(\varphi w) - 2 |\nabla v|^2 \nabla v \cdot \nabla \varphi \\ &\geq |\nabla v|^2 \Delta \varphi - \frac{4 |\nabla \varphi|^2 |\nabla v|^2}{\varphi} - \sum_{i,k=1}^n \varphi v_{ik}^2 + 2 \varphi \sum_{i,k=1}^n v_{ik}^2 - 2 \nabla v \cdot \nabla(\varphi w) - 2 |\nabla v|^3 |\nabla \varphi| \\ &= \varphi \sum_{i,k=1}^n v_{ik}^2 + \left(\Delta \varphi - \frac{4 |\nabla \varphi|^2}{\varphi} \right) |\nabla v|^2 - 2 \nabla v \cdot \nabla(\varphi w) - 2 |\nabla \varphi| |\nabla v|^3. \end{aligned}$$

而

$$\sum_{i,k=1}^n v_{ik}^2 \geq \sum_{i=1}^n v_{ii}^2 \geq \frac{1}{n} \left(\sum_{i=1}^n v_{ii} \right)^2 = \frac{1}{n} (\Delta v)^2 = \frac{1}{n} |\nabla v|^4,$$

故

$$\Delta(\varphi w) + 2 \nabla v \cdot \nabla(\varphi w) \geq \frac{\varphi}{n} |\nabla v|^4 + \left(\Delta \varphi - \frac{4 |\nabla \varphi|^2}{\varphi} \right) |\nabla v|^2 - 2 |\nabla \varphi| |\nabla v|^3.$$

为使 $\frac{|\nabla \varphi|^2}{\varphi}$ 在 B_1 中有意义, 令 $\eta = \sqrt[4]{\varphi} \in C_0^2(B_1)$, 则有

$$|\nabla \varphi| = 4 \eta^3 |\nabla \eta|, \quad \Delta \varphi = 4 \eta^3 \Delta \eta + 12 \eta^2 |\nabla \eta|^2,$$

故

$$\begin{aligned}\Delta(\eta^4 w) + 2\nabla v \cdot \nabla(\eta^4 w) &\geq \frac{1}{n}\eta^4 |\nabla v|^4 - 8\eta^3 |\nabla \eta| |\nabla v|^3 + (4\eta^3 \Delta \eta - 52\eta^2 |\nabla \eta|^2) |\nabla v|^2 \\ &= \frac{1}{n}(\eta |\nabla v|)^4 - 8|\nabla \eta|(\eta |\nabla v|)^3 + (4\eta \Delta \eta - 52|\nabla \eta|^2)(\eta |\nabla v|)^2.\end{aligned}$$

将右端视为关于 $\eta |\nabla v|$ 的一个 4 次多项式, 其首项系数为正; 其余系数依赖于 η , 因此为 x 的有界函数. 故存在 $C = C(n, \eta) > 0$, 使得

$$\frac{1}{2n}t^4 - 8|\nabla \eta|t^3 + (4\eta \Delta \eta - 52|\nabla \eta|^2)t^2 \geq -C, \forall t \in \mathbb{R}.$$

故取 $t = \eta |\nabla v|$, 可得

$$\Delta(\eta^4 w) + 2\nabla v \cdot \nabla(\eta^4 w) \geq \frac{1}{2n}(\eta |\nabla v|)^4 - C = \frac{1}{2n}\eta^4 w^2 - C.$$

假设 $\eta |\nabla v|$ 在 $x_0 \in \overline{B_1}$ 达到最大值, 由 η 的定义可知, $x_0 \notin \partial B_1$, 故

$$\nabla(\eta^4 w)(x_0) = 0, \Delta(\eta^4 w)(x_0) \leq 0,$$

因此

$$\eta^4 w^2(x_0) \leq 2nC.$$

如果 $w(x_0) \geq 1$, 那么 $\eta^4 w(x_0) \leq 2nC$; 如果 $w(x_0) < 1$, 则有 $\eta^4 w(x_0) \leq \eta^4(x_0)$. 总之存在 $\tilde{C} = \tilde{C}(n, \eta) > 0$, 使得在 B_1 中恒有 $\eta^4 w \leq \tilde{C}$. 因此由函数 η 的构造, 可得

$$\sup_{x \in B_{\frac{1}{2}}} |\nabla \log u(x)| = \sup_{x \in B_{\frac{1}{2}}} |w|^{\frac{1}{2}} = \sup_{x \in B_{\frac{1}{2}}} |\eta^4 w|^{\frac{1}{2}} \leq C,$$

其中 $C = C(n) > 0$. □

定理 9 Harnack 不等式

设在 B_1 中有 $\Delta u = 0$ 及 $u > 0$, 则存在常数 $C = C(n) > 0$, 使得

$$\sup_{B_{\frac{1}{2}}} u \leq C \inf_{B_{\frac{1}{2}}} u.$$

证明. $\forall x_1, x_2 \in B_{\frac{1}{2}}$, 有 $|x_1 - x_2| \leq 1$. 故由引理 2.1 可得,

$$\begin{aligned}\log \left(\frac{u(x_1)}{u(x_2)} \right) &= \log u(x_1) - \log u(x_2) \\ &= \log u(tx_1 + (1-t)x_2) \Big|_{t=0}^1 \\ &= \int_0^1 \frac{d}{dt} \log u(tx_1 + (1-t)x_2) dt \\ &= \int_0^1 \nabla \log u(tx_1 + (1-t)x_2) \cdot (x_1 - x_2) dt \leq \tilde{C}.\end{aligned}$$

故

$$u(x_1) \leq C u(x_2).$$

因此

$$\sup_{B_{\frac{1}{2}}} u \leq C \inf_{B_{\frac{1}{2}}} u.$$

证毕. □

证: [Harnack 不等式]

设 U 为 \mathbb{R}^n 中的连通开集, 在 U 中有 $\Delta u = 0$ 及 $u > 0$. 则 $\forall V \subset\subset U, \exists C = C(n, V, U) > 0$, 使得

$$\sup_V u \leq C \inf_V u.$$

□

证明. $\forall V \subset\subset U$, 取 $R = \frac{1}{4} \text{dist}(V, \partial U)$, 由于 V 紧, 存在有限个球

$$B(x_1, R), \dots, B(x_N, R)$$

覆盖 V , 分别记为 B_1, \dots, B_N . 由 R 的取法知这些球均包含于 U 中.

现在任取 $y_1, y_2 \in V$, 不妨设 $y_1 \in B_k, y_2 \in B_{k+m}, m \geq 1$. 并找出球 $B_{k+1}, \dots, B_{k+m-1}$, 使得 $B_j \cap B_{j+1} \neq \emptyset, j = k, k+1, \dots, k+m-1$. 对 B_k, \dots, B_m 运用定理2.1 可得

$$\begin{aligned} u(y_1) &\leq \sup_{B_k} u \leq C \inf_{B_k} u \quad (\text{由定理2.1}) \\ &\leq C \sup_{B_{k+1}} u \quad (\text{因 } B_k \cap B_{k+1} \neq \emptyset) \\ &\leq C^2 \inf_{B_{k+1}} u \quad (\text{再由定理2.1}) \\ &\leq \dots \leq C^{m+1} \inf_{B_{k+m}} u \leq C^{m+1} u(y_2). \end{aligned}$$

由于 y_1 与 y_2 的选取是任意的, 以及 $m \leq N$, 故得

$$\sup_V u \leq C^{N+1} \inf_V u.$$

证毕.

□

2.2

Krylov 和 Safanov 的 Harnack 不等式

若无额外说明, 本节均有以下定义, 之后不再重复:

(1) $B_R(x) := \{y \in \mathbb{R}^n \mid |y - x| < R\}, Q_R(x) := \{y \in \mathbb{R}^n \mid |y_i - x_i| < \frac{R}{2}, i = 1, \dots, n\}, \forall x \in \mathbb{R}^n, \forall R > 0$. 当中心 $x = 0$ 时, 分别简记它们为 B_R 和 Q_R .

(2) 算子 L 定义为 $Lu := \sum_{i,j=1}^n a_{ij} u_{ij}$, 其中 $A = (a_{ij})$ 对称. 称此 L 为 \mathbb{R}^n 中某有界开区域 U 上的一致椭圆算子, 若 $a_{ij} \in C(U)$, 且 $\exists 0 < \lambda \leq \Lambda < \infty$, 使得 $\forall x \in U, \forall \xi \in \mathbb{R}^n$, 有

$$\lambda |\xi|^2 \leq \sum_{i,j=1}^n a_{ij}(x) \xi_i \xi_j \leq \Lambda |\xi|^2.$$

Harnack 不等式的证明一般分为两部分: 下解的估计和上解的估计, 分别对应于下面的定理2.2和定理2.2.

将二者结合起来, 就可以得到 Harnack 不等式. 下面我们先来看其中相对容易的下解部分.

定理 10 局部极值原理/下解的上界估计

设 L 为 B_R 上的一致椭圆算子. 如果 $u \in C^2(B_R) \cap C(\overline{B_R})$ 在 B_R 中满足 $Lu \geq f$, 其中 $f \in C(B_R) \cap L^\infty(B_R)$, 那么 $\forall p > 0$, 存在常数 $C = C(n, p, \lambda, \Lambda) > 0$, 使得

$$\sup_{B_{\frac{R}{2}}} u \leq C \left\{ R^{-\frac{n}{p}} \|u^+\|_{L^p(B_R)} + R \|f^-\|_{L^\infty(B_R)} \right\}.$$

证明. 只需证明 $R = 1$ 时的结论, 即 $\forall p > 0, \exists C = C(n, p, \lambda, \Lambda) > 0$, 使得

$$\sup_{B_{\frac{1}{2}}} u \leq C \{ \|u^+\|_{L^p(B_1)} + \|f^-\|_{L^n(B_1)} \}.$$

(因为如果已经证明了 $R = 1$ 时的结论, 那么对于一般的 $R > 0$, 只需令

$$\tilde{u}(x) = u(Rx), \tilde{f}(x) = R^2 f(Rx), x \in B_1,$$

则由 $Lu \leq f$ 于 B_R 的条件可得 $\tilde{L}\tilde{u} \leq \tilde{f}$ 于 B_1 , 其中 \tilde{L} 定义为

$$\tilde{L}w(x) := \sum_{i,j} \tilde{a}_{ij}(x)w_{ij}(x), x \in B_1,$$

其中 $\tilde{a}_{ij}(x) = a_{ij}(Rx)$, 这种变换不改变一致椭圆性, 即 \tilde{L} 为 B_1 上的一致椭圆算子, 从而利用 $R = 1$ 时的结论可得对于一般的 R 的结论. 这种技巧可称为 Rescaling 技巧, 在这里可以用来将常数 C 与球的半径 R 分离开. 在之后还会用到这个方法, 到时就不再像这样详加说明.)

引入截断函数 $\eta(x) := (1 - |x|^2)^\beta, x \in B_1$, 其中 $\beta \geq 1$ 待定. 从而在 B_1 上有 $0 < \eta < 1$.

计算可得

$$\begin{aligned} \eta_i(x) &= -2\beta x_i(1 - |x|^2)^{\beta-1}, \\ \eta_{ij}(x) &= 4\beta(\beta - 1)(1 - |x|^2)^{\beta-2}x_i x_j - 2\delta_{ij}\beta(1 - |x|^2)^{\beta-1}, \end{aligned}$$

故

$$\begin{aligned} |\nabla\eta| &= 2\beta(1 - |x|^2)^{\beta-1}|x| \leq 2\beta\eta^{1-\frac{1}{\beta}}, \\ L\eta &= 4\beta(\beta - 1)(1 - |x|^2)^{\beta-2} \sum_{i,j=1}^n a_{ij}x_i x_j - 2\beta(1 - |x|^2)^{\beta-1} \sum_{i=1}^n a_{ii} \\ &\geq 4\beta(\beta - 1)\lambda|x|^2(1 - |x|^2)^{\beta-2} - 2\beta n\Lambda(1 - |x|^2)^{\beta-1} \\ &\geq -2\beta n\Lambda\eta^{1-\frac{1}{\beta}}. \end{aligned}$$

令 $w = \eta u$, 则 $Lw = \eta Lu + uL\eta + 2 \sum_{i,j} a_{ij}\eta_i u_j$.

考虑 w 在 B_1 中的上接触集

$$\Gamma_w^+ := \{y \in B_1 \mid \forall x \in B_1, w(x) \leq w(y) + \nabla w(y) \cdot (x - y)\},$$

则在 Γ_w^+ 上有 $w > 0$ (自然也有 $u > 0$). $\forall y \in \Gamma_w^+$, 从 y 出发沿 $-\nabla w(y)$ 方向作射线, 交 ∂B_1 于 x_0 , 即

$$x_0 = y - \frac{1}{|\nabla w(y)|} |\nabla w(y)|.$$

将其代入 $w(x_0) \leq w(y) + \nabla w(y) \cdot (x_0 - y)$ 可得

$$0 \leq w(y) - |x_0 - y| |\nabla w(y)|.$$

从而 $\forall y \in \Gamma_w^+$, 有

$$\begin{aligned} |\nabla w(y)| &\leq \frac{w(y)}{|x_0 - y|} \leq \frac{w(y)}{1 - |y|} = \frac{(1 - |y|)w(y)}{1 - |y|^2} \\ &\leq \frac{2w(y)}{1 - |y|^2} = (2w\eta^{-\frac{1}{\beta}})(y). \end{aligned}$$

因此在 Γ_w^+ 上有

$$\begin{aligned} |\nabla u| &= \frac{1}{\eta} |\nabla w - u \nabla \eta| \leq \frac{1}{\eta} (|\nabla w| + u |\nabla \eta|) \\ &\leq \frac{1}{\eta} (2w\eta^{-\frac{1}{\beta}} + w\eta^{-1} \cdot 2\beta\eta^{1-\frac{1}{\beta}}) = 2(\beta+1)\eta^{-1-\frac{1}{\beta}}w. \end{aligned}$$

故由椭圆性条件,

$$\left| 2 \sum_{i,j} a_{ij} \eta_i u_j \right| \leq \Lambda |\nabla \eta| |\nabla u| \leq \Lambda \cdot 2\beta\eta^{1-\frac{1}{\beta}} \cdot 2(\beta+1)\eta^{-1-\frac{1}{\beta}}w = 4\beta(\beta+1)\Lambda\eta^{-\frac{2}{\beta}}w.$$

又因为在 Γ_w^+ 中有 $u > 0$, 故

$$uL\eta \geq -2u\beta n\Lambda\eta^{1-\frac{1}{\beta}} = -2\beta n\Lambda\eta^{-\frac{1}{\beta}}w \geq -2\beta n\Lambda\eta^{-\frac{2}{\beta}}w.$$

从而在 Γ_w^+ 中有

$$\begin{aligned} L(w^+) &= Lw = \eta Lu + uL\eta + 2 \sum_{i,j} a_{ij} \eta_i u_j \\ &\geq \eta f - 2\beta n\Lambda\eta^{-\frac{2}{\beta}}w - 4\beta(\beta+1)\Lambda\eta^{-\frac{2}{\beta}}w \\ &\geq -\eta f^- - C_1\eta^{-\frac{2}{\beta}}w \geq -(C_1\eta^{-\frac{2}{\beta}}w^+ + f^-), \end{aligned}$$

其中 $C_1 = C_1(n, \Lambda, \beta) > 0$. 故由 Aleksandrov 极大值原理 (定理1.4), 存在常数 $C_2 > 0$, 使得

$$\begin{aligned} \sup_{B_1} w^+ &\leq \sup_{\partial B_1} w^+ + C_2 \|C_1\eta^{-\frac{2}{\beta}}w^+ + f^-\|_{L^n(\Gamma_w^+)} \\ &\leq C_3 \|\eta^{-\frac{2}{\beta}}w^+\|_{L^n(B_1)} + C_4 \|f^-\|_{L^n(B_1)}. \end{aligned}$$

因此 $\forall p > 0$, 若 $0 < p < n$, 则令 $\beta = \frac{2n}{p} > 2$, 于是有

$$\begin{aligned} \sup_{B_1} w^+ &\leq C_3 \|\eta^{-\frac{2}{\beta}}(w^+)^{1-\frac{2}{\beta}+\frac{2}{\beta}}\|_{L^n(B_1)} + C_4 \|f^-\|_{L^n(B_1)} \\ &\leq C_3 (\sup_{B_1} w^+)^{1-\frac{2}{\beta}} \|\eta^{-\frac{2}{\beta}}(w^+)^{\frac{2}{\beta}}\|_{L^n(B_1)} + C_4 \|f^-\|_{L^n(B_1)} \\ &= C_3 (\sup_{B_1} w^+)^{1-\frac{2}{\beta}} \|(w^+)^{\frac{2}{\beta}}\|_{L^n(B_1)} + C_4 \|f^-\|_{L^n(B_1)} \\ &= C_3 (\sup_{B_1} w^+)^{1-\frac{2}{\beta}} \|u^+\|_{L^{\frac{2n}{\beta}}(B_1)}^{\frac{2}{\beta}} + C_4 \|f^-\|_{L^n(B_1)}. \end{aligned}$$

利用 Young 不等式

$$ab \leq \frac{1}{r}a^r + \frac{1}{s}b^s \quad (a, b > 0, r, s > 1, \frac{1}{r} + \frac{1}{s} = 1),$$

取

$$a = \varepsilon (\sup_{B_1} w^+)^{1-\frac{2}{\beta}}, b = \frac{1}{\varepsilon} \|u^+\|_{L^{\frac{2n}{\beta}}(B_1)}^{\frac{2}{\beta}}, r = \frac{\beta}{\beta-2}, s = \frac{\beta}{2} > 1,$$

其中 $\varepsilon > 0$ 待定. 可得

$$\sup_{B_1} w^+ \leq C_3 \left\{ (1 - \frac{2}{\beta}) \varepsilon^r (\sup_{B_1} w^+)^r + \frac{2}{\beta} \varepsilon^{-s} \|u^+\|_{L^{\frac{2n}{\beta}}(B_1)}^{\frac{2}{\beta}} \right\} + C_4 \|f^-\|_{L^n(B_1)}.$$

取 $\varepsilon > 0$ 使得 $C_3(1 - \frac{2}{\beta})\varepsilon^r \leq \frac{1}{2}$, 从而有

$$\begin{aligned} \sup_{B_1} w^+ &\leq C_5 \left\{ \|u^+\|_{L^{\frac{2n}{\beta}}(B_1)}^{\frac{2}{\beta}} + \|f^-\|_{L^n(B_1)} \right\} \\ &= C_5 \left\{ \|u^+\|_{L^p(B_1)} + \|f^-\|_{L^n(B_1)} \right\}. \end{aligned}$$

因此有

$$\begin{aligned} \sup_{B_{\frac{1}{2}}} u &\leq \sup_{B_{\frac{1}{2}}} u^+ = \sup_{B_{\frac{1}{2}}} \left(\frac{w^+}{(1-|x|^2)^\beta} \right) \leq \left(\frac{4}{3} \right)^\beta \sup_{B_{\frac{1}{2}}} w^+ \leq \left(\frac{4}{3} \right)^{\frac{2n}{p}} \sup_{B_1} w^+ \\ &\leq C \left\{ \|u^+\|_{L^p(B_1)} + \|f^-\|_{L^n(B_1)} \right\}, \end{aligned}$$

其中 $C = C(n, p, \lambda, \Lambda) > 0$.

而若 $p \geq n$, 则由 Hölder 不等式和 $p < n$ 时的结论可直接得出. □

而对于上解部分, 我们需要以下几个引理, 称为 Calderon-Zygmund 分解, 这是证明 Harnack 不等式的一个重要途径.

我们将 Q_1 平分为 2^n 个立方体, 分得的每个小立方体称为第一代立方体, 之后将每个第一代立方体 2^n 等分, 得到第二代立方体, \dots , 以此类推, 将所有这些立方体统称为二进方体. 我们注意到, 任意一个第 $(k+1)$ 代立方体 Q 一定是从某个第 k 代立方体 \tilde{Q} 得来, 我们将 \tilde{Q} 称为 Q 的前身. 我们有以下的引理:

引理 6

设 $f \in L^1(Q_1)$ 为非负函数, $\delta > 0$ 为常数, 使得 $\frac{1}{|Q_1|} \int_{Q_1} f < \delta$, 则在 Q_1 中存在一系列互不重叠的二进方体 $\{Q^j\}_{j \geq 1}$, 满足

- (1) 在 $Q_1 \setminus \bigcup_j Q^j$ 中几乎处处有 $f \leq \delta$;
- (2) $\delta \leq \frac{1}{|Q^j|} \int_{Q^j} f < 2^n \delta$.

证明. 将 Q_1 分割为 2^n 个二进方体, 并保留那些满足

$$\frac{1}{|Q|} \int_Q f \geq \delta \tag{2}$$

的立方体 Q , 对其余的立方体继续分割, 并继续保留满足 (2) 式的立方体. 记 $\{Q^j\}_{j \geq 1}$ 为此无穷过程中保留下来的立方体, 则 $\forall x \in Q_1 \setminus \bigcup_j Q^j, \exists \{Q^{k'}\}_{k \geq 1}$ 包含 x , 且满足

$$\frac{1}{|Q^{k'}|} \int_{Q^{k'}} f < \delta, \quad \lim_{k' \rightarrow \infty} \text{diam}(Q^{k'}) = 0.$$

由 Lebesgue 密度定理, 在 $Q_1 \setminus \bigcup_j Q^j$ 中几乎处处有 $f \leq \delta$, 即 (1). 又注意到每个 Q^j 的前身 \tilde{Q}^j 一定不满足 (2) 式, 即

$$\frac{1}{|\tilde{Q}^j|} \int_{\tilde{Q}^j} f < \delta,$$

故

$$\frac{1}{|Q^j|} \int_{Q^j} f = \frac{2^n}{|\tilde{Q}^j|} \int_{\tilde{Q}^j} f \leq \frac{2^n}{|\tilde{Q}^j|} \int_{\tilde{Q}^j} f < 2^n \delta,$$

即得 (2). □

引理 7

设 A 与 B 为 2 个可测集, 满足 $A \subset B \subset Q_1$, 如果它们还满足下面两点:

- (1) $\exists \delta \in (0, 1)$, 使得 $|A| < \delta$;
- (2) 对任意二进方体 Q , 只要 $|A \cap Q| \geq \delta|Q|$, Q 的前身 \tilde{Q} 就满足 $\tilde{Q} \subset B$, 那么一定有 $|A| \leq \delta|B|$.

证明. 由(1)可知, $\exists \delta \in (0, 1)$, 使得 $\int_{Q_1} \chi_A < \delta$, 其中 χ_A 为集合 A 的特征函数. 故由引理 2.2, 存在一列互不重叠的二进方体 $\{Q^j\}_{j \geq 1}$, 使得除了一个零测集之外, 有

$$A \subset \bigcup_j Q^j, \delta \leq \frac{|A \cap Q^j|}{|Q^j|} < 2^n \delta,$$

而且对每个 Q^j 的前身 \tilde{Q}^j , 均有

$$\frac{|A \cap \tilde{Q}^j|}{|\tilde{Q}^j|} < \delta.$$

由(2)可知, $A \subset \bigcup_j \tilde{Q}^j \subset B$, 将这些 \tilde{Q}^j 重排, 使它们互不重叠, 则有

$$|A| \leq \sum_j |A \cap \tilde{Q}^j| < \sum_j \delta |\tilde{Q}^j| \leq \delta |B|.$$

这就完成了证明. □

引理 8

设 L 为 $B_{2\sqrt{n}}$ 上的一致椭圆算子. 如果 $u \in C^2(B_{2\sqrt{n}}) \cap L^\infty(B_{2\sqrt{n}})$ 在 $B_{2\sqrt{n}}$ 中满足 $Lu \leq f$ 以及 $u \geq 0$, 其中 $f \in C(B_{2\sqrt{n}}) \cap L^n(B_{2\sqrt{n}})$, 那么就有以下结果:

(1) 存在均只依赖于 n, λ, Λ 的常数 $\varepsilon_0 > 0, \mu \in (0, 1)$ 和 $M > 1$, 使得只要有

$$\inf_{Q_3} u \leq 1, \quad \|f\|_{L^n(B_{2\sqrt{n}})} \leq \varepsilon_0, \quad (3)$$

就有

$$|\{u > M\} \cap Q_1| \leq 1 - \mu.$$

(2) 在(1)中选取的 ε_0, μ, M 进而还满足: 只要有(3)式成立, 就有

$$|\{u > M^k\} \cap Q_1| \leq (1 - \mu)^k, \quad \forall k \in \mathbb{N}^+.$$

(3) 存在均只依赖于 n, λ, Λ 的常数 $\varepsilon_0, \gamma, C > 0$ (其中 ε_0 取得与(1)和(2)中相同), 使得只要有(3)式成立, 就有

$$|\{u > t\} \cap Q_1| \leq Ct^{-\gamma}, \quad \forall t > 0.$$

证明. (1) 注意到 $B_{\frac{1}{4}} \subset B_{\frac{1}{2}} \subset Q_1 \subset Q_3 \subset B_{\frac{3}{2}\sqrt{n}} \subset B_{2\sqrt{n}}$, 这里选取 $B_{2\sqrt{n}}$ 来讨论只是为了避免分数 $\frac{3}{2}$ 给过程带来的繁琐.

定义 $\varphi(x) = -M(1 - \frac{|x|^2}{4n})^\beta, x \in B_{2\sqrt{n}}$, 其中 $\beta > 2, M > 1$ 待定. 计算可得

$$\begin{aligned} \varphi_i(x) &= \frac{M\beta}{2n} x_i \left(1 - \frac{|x|^2}{4n}\right)^{\beta-1} \\ \varphi_{ij}(x) &= \frac{M\beta}{2n} \left(1 - \frac{|x|^2}{4n}\right)^{\beta-1} \delta_{ij} - \frac{M\beta(\beta-1)}{4n^2} \left(1 - \frac{|x|^2}{4n}\right)^{\beta-2} x_i x_j. \end{aligned}$$

因此

$$\begin{aligned} L\varphi &= \frac{M\beta}{2n} \left(1 - \frac{|x|^2}{4n}\right)^{\beta-2} \left[\left(1 - \frac{|x|^2}{4n}\right) \sum_{i=1}^n a_{ii} - \frac{\beta-1}{2n} \sum_{i,j=1}^n a_{ij} x_i x_j \right] \\ &\leq \frac{M\beta}{2n} \left(1 - \frac{|x|^2}{4n}\right)^{\beta-2} \left[\left(1 - \frac{|x|^2}{4n}\right) n\Lambda - \frac{\beta-1}{2n} \lambda |x|^2 \right]. \end{aligned}$$

在 $B_{2\sqrt{n}}$ 中定义 $w = u + \varphi$, 则当 $|x| \geq \frac{1}{4}$ 时, 取 $\beta = \beta(n, \lambda, \Lambda)$ 足够大, 使得在 $B_{2\sqrt{n}} \setminus Q_1$ 中有 $L\varphi \leq 0$. 从而 $\exists \eta \in C_0^\infty(Q_1), 0 \leq \eta \leq C_1$, 其中 $C_1 > 0$ 为常数, 使得 $Lw \leq f + \eta$, 即 $L(-w) \geq -(f + \eta)$.

由 Aleksandrov 极大值原理 (定理1.4), 有

$$\begin{aligned} \sup_{B_{2\sqrt{n}}}(-w) &\leq \sup_{\partial B_{2\sqrt{n}}}(-w)^+ + C_2 \|f + \eta\|_{L^n(\Gamma_w^+ \cap B_{2\sqrt{n}} \cap U^-)} \\ &= C_2 \|f + \eta\|_{L^n(\Gamma_w^- \cap B_{2\sqrt{n}} \cap U^-)}. \end{aligned}$$

其中 $U^- = B_{2\sqrt{n}} \cap \{w \leq 0\}$, Γ_w^+ 为 $-w$ 的上接触集, Γ_w^- 为 w 的下接触集. 选取 M 足够大, 使得在 Q_3 中 $\varphi \leq -2$, 从而

$$\inf_{Q_3} w = \inf_{Q_3} (u + \varphi) \leq \inf_{Q_3} u - 2,$$

进而有

$$\begin{aligned} 2 - \inf_{Q_3} u &\leq -\inf_{Q_3} w \leq -\inf_{B_{2\sqrt{n}}} w = \sup_{B_{2\sqrt{n}}}(-w) \leq C_2 \|f + \eta\|_{L^n(\Gamma_w^- \cap B_{2\sqrt{n}} \cap U^-)} \\ &\leq C_3 \left\{ \|f\|_{L^n(B_{2\sqrt{n}})} + |\Gamma_w^- \cap Q_1 \cap U^-|^{\frac{1}{n}} \right\}. (\text{因 } \eta \in C_0^\infty(Q_1).) \end{aligned}$$

故取 $\varepsilon_0 > 0$ 满足 $C_3 \varepsilon_0 \leq \frac{1}{2}$, 取 $\mu \in (0, 1)$ 满足 $\mu < (\frac{1}{2C_3})^n$, 则当

$$\inf_{Q_3} u \leq 1, \quad \|f\|_{L^n(B_{2\sqrt{n}})} \leq \varepsilon_0$$

时, 有

$$1 \leq C_3 \varepsilon_0 + C_3 |\Gamma_w^- \cap Q_1 \cap U^-|^{\frac{1}{n}} \leq \frac{1}{2} + C_3 |\Gamma_w^- \cap Q_1 \cap U^-|^{\frac{1}{n}}.$$

而在 U^- 上有 $w \leq 0$, 从而 $u \leq -\varphi \leq M$, 进而有

$$\frac{1}{2} \leq C_3 |\{u \leq M\} \cap Q_1|^{\frac{1}{n}},$$

即

$$|\{u > M\} \cap Q_1| \leq 1 - \mu.$$

(2) 我们仍取 (1) 中取定的 ε_0, μ, M , 并设

$$\inf_{Q_3} u \leq 1, \quad \|f\|_{L^n(B_{2\sqrt{n}})} \leq \varepsilon_0,$$

将利用归纳法来证明本命题.

当 $k = 1$ 时, 此即 (1) 已证得的结论. 现假设

$$|\{u > M^{k-1}\} \cap Q_1| \leq (1 - \mu)^{k-1}, k \in \mathbb{N}^+,$$

并记 $A = \{u > M^k\} \cap Q_1, B = \{u > M^{k-1}\} \cap Q_1$, 我们将利用引理2.2来证明 $|A| \leq (1 - \mu)|B|$.

首先显然有 $A \subset B$, 其次由 (1), 有

$$|A| = |\{u > M^k\} \cap Q_1| \leq |\{u > M\} \cap Q_1| \leq 1 - \mu.$$

对于引理2.2的条件 (2), 我们声明: 对于任意立方体 $Q_r(x_0) \subset Q_1$ (其中 $r \leq \frac{1}{2}$), 如果 $|A \cap Q_r(x_0)| > (1 - \mu)|Q_r(x_0)|$, 那么 $(Q_{3r}(x_0) \cap Q_1) \subset B$.

事实上, 定义

$$\begin{aligned} \tilde{u}(x) &:= \frac{1}{M^{k-1}} u(x_0 + rx), x \in B_{2\sqrt{n}}, \\ \tilde{f}(x) &:= \frac{r^2}{M^{k-1}} f(x_0 + rx), x \in B_{2\sqrt{n}}, \\ \tilde{L}g &:= \sum_{i,j=1}^n a_{ij}(x_0 + rx) g_{ij}(x), \forall g \in C^2(B_{2\sqrt{n}}), \end{aligned}$$

由于

$$|x_0| + 2\sqrt{nr} < 1 + 2\sqrt{nr} \leq 1 + \sqrt{n} < 2\sqrt{n},$$

以上定义均有意义, 且有 $\tilde{u} \geq 0$ 以及 $\tilde{L}\tilde{u} \leq \tilde{f}$, 其中 \tilde{L} 仍为一致椭圆算子. 由

$$\begin{aligned} \{\tilde{u} > M\} \cap Q_1 &= \{x \in Q_1 \mid \tilde{u}(x) > M\} \\ &= \{x \in Q_1 \mid u(x_0 + rx) > M^k\} \\ &= \{u > M^k\} \cap Q_r(x_0) \end{aligned}$$

可知,

$$|\{\tilde{u} > M\} \cap Q_1| = \frac{1}{r^n} |\{u > M^k\} \cap Q_r(x_0)| = \frac{|A \cap Q_r(x_0)|}{|Q_r(x_0)|} > 1 - \mu.$$

又由变量替换可得

$$\|\tilde{f}\|_{L^n(B_{2\sqrt{n}})} = \frac{r}{M^{k-1}} \|f\|_{L^n(B_{2\sqrt{nr}}(x_0))} \leq \|f\|_{L^n(B_{2\sqrt{n}})} \leq \varepsilon_0,$$

因此由 (1) 可知,

$$\inf_{Q_3} \tilde{u} > 1,$$

从而在 $Q_{3r}(x_0)$ 中有 $u > M^{k-1}$, 即 $(Q_{3r}(x_0)) \cap Q_1 \subset B$. 于是声明得证. 最后由引理2.2, $|A| \leq (1 - \mu)|B|$, 再加上假设就可得

$$|\{u > M^k\} \cap Q_1| \leq (1 - \mu)^k.$$

由归纳假设, 可知

$$|\{u > M^k\} \cap Q_1| \leq (1 - \mu)^k, \forall k \in \mathbb{N}^+.$$

(3) 对于 (1) 和 (2) 中所取的 ε_0, μ 和 M , 可知 $\forall t > 0, \exists k_0 \in \mathbb{Z}$, 使得 $M^{k_0-1} \leq t < M^{k_0}$, 从而 $\frac{\log t}{\log M} < k_0$. 因此

$$\begin{aligned} |\{u > t\} \cap Q_1| &\leq |\{u > M^{k_0-1}\} \cap Q_1| \leq (1 - \mu)^{k_0-1} \leq (1 - \mu)^{\frac{\log t}{\log M} - 1} \\ &= \frac{1}{1 - \mu} e^{\frac{\log t}{\log M} \cdot \log(1 - \mu)} = \frac{1}{1 - \mu} t^{\frac{\log(1 - \mu)}{\log M}} = Ct^{-\gamma}, \end{aligned}$$

其中 $C, \gamma > 0$. □

有了以上的准备作铺垫, 我们就可以很快地完成上解部分的证明.

定理 11

设 L 为 $B_{2\sqrt{n}}$ 上的一致椭圆算子. 若 $u \in C^2(B_{2\sqrt{n}}) \cap L^\infty(B_{2\sqrt{n}})$ 在 $B_{2\sqrt{n}}$ 中满足 $Lu \leq f$ 以及 $u \geq 0$, 其中 $f \in C(B_{2\sqrt{n}}) \cap L^n(B_{2\sqrt{n}})$, 则存在常数 $p = p(n, \lambda, \Lambda) > 0$ 和 $C = C(n, \lambda, \Lambda) > 0$, 使得

$$\|u\|_{L^p(B_{\frac{1}{2}})} \leq C \left\{ \inf_{B_{\frac{1}{2}}} u + \|f\|_{L^n(B_{2\sqrt{n}})} \right\}.$$

证明. 令

$$c_0 = \inf_{Q_3} u + \delta + \frac{1}{\varepsilon_0} \|f\|_{L^n(B_{2\sqrt{n}})},$$

其中 ε_0 与引理2.2中取得相同. 又设

$$u_\delta = \frac{1}{c_0} u, f_\delta = \frac{1}{c_0} f,$$

则有 $Lu_\delta \leq f_\delta$, $\inf_{Q_3} u_\delta \leq 1$, 以及

$$\|f_\delta\|_{L^n(B_{2\sqrt{n}})} = \frac{\|f\|_{L^n(B_{2\sqrt{n}})}}{\inf_{Q_3} u + \delta + \frac{1}{\varepsilon_0} \|f\|_{L^n(B_{2\sqrt{n}})}} \leq \varepsilon_0.$$

因此由引理2.2的 (3) 可知, 存在常数 $\gamma, C_1 > 0$, 使得

$$|\{u_\delta > t\} \cap Q_1| \leq C_1 t^{-\gamma}, \forall t > 0,$$

即

$$|\{u > c_0 t\} \cap Q_1| \leq C_1 (c_0 t)^{-\gamma} c_0^\gamma, \forall t > 0.$$

故 $\forall t > 0$,

$$\begin{aligned} |\{u > t\} \cap Q_1| &\leq C_1 t^{-\gamma} c_0^\gamma \\ &= C_1 t^{-\gamma} (\inf_{Q_3} u + \delta + \frac{1}{\varepsilon_0} \|f\|_{L^n(B_{2\sqrt{n}})})^\gamma \\ &\leq C_2 t^{-\gamma} (\inf_{Q_3} u + \delta + \|f\|_{L^n(B_{2\sqrt{n}})})^\gamma. \end{aligned}$$

令 $\delta \rightarrow 0$, 可得

$$|\{u > t\} \cap Q_1| \leq C_2 t^{-\gamma} (\inf_{Q_3} u + \|f\|_{L^n(B_{2\sqrt{n}})})^\gamma, \forall t > 0.$$

$\forall t > 0$, 定义 $A(t) := \{u > t\} \cap Q_1$. 令 $p > 0$ 待定, 则 $\forall \tau > 0$, 由 Fubini 定理交换积分次序可得

$$\begin{aligned} \int_{Q_1} u^p dx &= \iint_{\{x \in Q_1, 0 < t < u(x)\}} p t^{p-1} dt dx \\ &= p \int_0^\infty t^{p-1} |A(t)| dt \\ &= p \int_0^\tau t^{p-1} |A(t)| dt + p \int_\tau^\infty t^{p-1} |A(t)| dt \\ &\leq p \int_0^\tau t^{p-1} dt + p \int_\tau^\infty t^{p-1} C_2 t^{-\gamma} (\inf_{Q_3} u + \|f\|_{L^n(B_{2\sqrt{n}})})^\gamma dt \\ &= \tau^p + C_2 (\inf_{Q_3} u + \|f\|_{L^n(B_{2\sqrt{n}})})^\gamma p \int_\tau^\infty t^{p-\gamma-1} dt. \end{aligned}$$

取 $p < \gamma$, 则

$$\int_{Q_1} u^p dx \leq C_3 \left\{ \tau^p + \tau^{p-\gamma} (\inf_{Q_3} u + \|f\|_{L^n(B_{2\sqrt{n}})})^\gamma \right\}.$$

最后令

$$\tau = \inf_{Q_3} u + \|f\|_{L^n(B_{2\sqrt{n}})},$$

即可得

$$\|u\|_{L^p(Q_1)} \leq C \left\{ \inf_{Q_3} u + \|f\|_{L^n(B_{2\sqrt{n}})} \right\},$$

进而有

$$\|u\|_{L^p(B_{\frac{1}{2}})} \leq C \left\{ \inf_{B_{\frac{1}{4}}} u + \|f\|_{L^n(B_{2\sqrt{n}})} \right\}.$$

证毕. □

定理 12 Harnack 不等式

设 L 为 B_R 上的一致椭圆算子. 如果 $u \in C^2(B_R) \cap C(\overline{B_R})$ 为方程 $Lu = f, x \in B_R$ 的一个非负解, 其中 $f \in C(B_R) \cap L^\infty(B_R)$, 那么存在常数 $C = C(n, \lambda, \Lambda) > 0$, 使得

$$\sup_{B_{\frac{R}{2}}} u \leq C \left\{ \inf_{B_{\frac{R}{2}}} u + R \|f\|_{L^n(B_R)} \right\}.$$

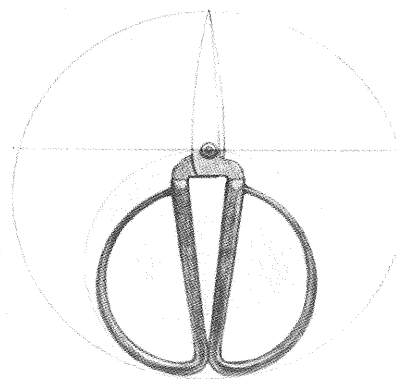
证明. 将定理2.2和定理2.2的结论进行放缩组合, 再利用紧集的覆盖即可得. \square

第3节

参考文献

- [1] Robert A Adams and John JF Fournier. Sobolev spaces, volume 140. Academic press, 2003.
- [2] Lawrence C Evans. Partial differential equations (graduate studies in mathematics, vol. 19). Instructor, 2009.
- [3] David Gilbarg and Neil S Trudinger. Elliptic partial differential equations of second order, volume 224. springer, 2001.
- [4] Qing Han and Fanghua Lin. Elliptic partial differential equations, volume 1. American Mathematical Soc., 2011.
- [5] J Jost. Partial differential equations (vol. 214 of graduate texts in mathematics. springer-verlag, new-york, 2002). Translated and revised from the, 1998.
- [6] Jürgen Moser. On harnack' s thm for elliptic differential equations. Communications on Pure and Applied Mathematics, 14(3):577-591, 1961.
- [7] Richard Schoen, Leon Simon, and Shing-Tung Yau. Curvature estimates for minimal hypersurfaces. Acta Mathematica, 134(1):275-288, 1975.
- [8] Zhuoqun Wu, Jingxue Yin, and Chunpeng Wang. Elliptic & parabolic equations. World Scientific, 2006.
- [9] 周蜀林, 偏微分方程, 北京大学出版社, 2005.
- [10] 陈亚浙, 吴兰成, 二阶椭圆型方程与椭圆型方程组, 科学出版社, 2006.

应用数学论文



从极值到变分

2012 级弘毅班 袁航 陈明洁 基地班 刘梦晨

摘要：本文从极值问题入手，推广数学分析中的函数极值等问题，通过变分的观点探讨更广义的极值问题。

关键词：极值 变分 E-L 方程 最速曲线 导数 测地线 极小曲面

第 1 节

引言

变分法的关键定理是欧拉-拉格朗日方程。它对应于泛函的临界点。在寻找函数的极大和极小值时，在一个解附近的微小变化的分析给出一阶的一个近似。它不能分辨是找到了最大值或者最小值（或者都不是）。变分法在理论物理中非常重要：在拉格朗日力学中，以及在最小作用量原理在量子力学的应用中。变分法提供了有限元方法的数学基础，它是求解边界值问题的强力工具。它们也在材料学中研究材料平衡中大量使用。而在纯数学中的例子有，黎曼在调和函数中使用狄力克雷原理。最优控制的理论是变分法的一个推广。同样的材料可以出现在不同的标题中，例如希尔伯特空间技术，摩尔斯理论，或者辛几何。变分一词用于所有极值泛函问题。微分几何中的测地线的研究是很显然的变分性质的领域。极小曲面（肥皂泡）上也有很多研究工作，称为 Plateau 问题。

本文是一篇介绍变分法的专题论文，选择了几个具有代表性的问题说明变分法的应用。

首先，我们介绍经典的最速降线问题，历史上变分法从这个问题开始。接着我们讨论一些理论和实际问题，它们往往我们需要知道的并不仅仅是函数的极值性质，而是最终寻求的是极值函数：它们使得某个泛函取得极大或极小值。

其次，我们介绍一个实际中抽象出来的一个模型问题，以阐述变分法的一个关键定理——欧拉-拉格朗日方程 (Euler-Lagrange equation)，它主要针对理论物理中的问题。

最后，我们讨论微分几何中一些理论问题，长度是曲面的泛函，我们知道测地线满足一种极值性质，即“曲面上联结两点的最短线是测地线”。幸运的是，为了挖掘出这些性质，并没有逃出数学分析中最基本的求导数求极值的思想！只是我们对函数的选取和导数的处理有了更多的要求和认识。换句话说，测地线是长度泛函的临界点；对应地，面积（体积）是曲面（超曲面）的泛函，我们还将论证面积泛函的临界点是极小曲面。

第 2 节

最速降线

历史上，变分法可能是最速曲线 (brachistochrone curve) 问题开始出现的。

问题 意大利科学家伽利略在 1630 年提出一个分析学的基本问题——“铅直平面内给定不在一条垂直线上的两个点 A, B ，如图，求连接它们的光滑曲线，使质点在重力作用下沿该曲线以最短时间从 A 点滑到 B 点（摩擦力不计）”。他说这曲线是圆，可是这是一个错误的答案。

瑞士数学家约翰·伯努利在 1696 年再提出这个最速降线的问题 (problem of brachistochrone), 征求解答。次年已有多位数学家得到正确答案, 其中包括牛顿、莱布尼兹、洛必达和伯努利兄弟。牛顿用非凡的微积分技巧解出了最速降线方程, 约翰·伯努利用光学的办法巧妙的也解出最速降线方程, 雅各布·伯努利用比较麻烦的办法解决了这个问题。这问题的正确答案是连接两个点上凹的唯一一段旋轮线或圆滚线。旋轮线与 1673 年荷兰科学家惠更斯讨论的摆线相同。因为钟表摆锤作一次完全摆动所用的时间相等, 所以摆线 (旋轮线) 又称等时曲线。

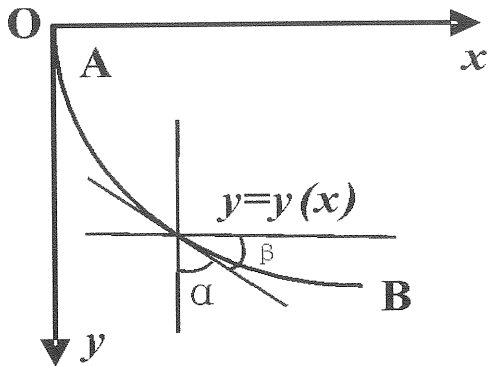


图 1

数学家十分关注最速降线问题, 大数学家欧拉也在 1726 年开始发表有关的论著, 在雅各布·伯努利方法的基础上, 1744 年最先给了这类问题的普遍解法, 并产生了变分法这一新数学分支。现在来看, 雅各布的方法是更有意义和价值的。

首先我们假设质点在滑动过程中不考虑空气阻力。

尽管 A, B 两点间的最短距离是连接它们的直线, 但是沿直线运动时速度增长较慢, 如果沿一条陡峭的曲线下滑, 虽然路径加长, 但运动速度增长很快。

为了求这条运动时间最短的曲线, 在图 3 中将 A 点取为坐标原点 $(0,0)$, B 点坐标为 (x_1, y_1) , 连接 A, B 的曲线记为 $y(x)$, 于是曲线上的弧长为 $ds = \sqrt{1+y'^2}dx$, 根据能量守恒定律, 质点在曲线 $y(x)$ 上任一点的速度 $\frac{ds}{dt}$ 满足 $\frac{1}{2}m(\frac{ds}{dt})^2 = mgy$, 其中 m 是质点的质量, g 是重力加速度。将上面 ds 的关系代入, 得到 $dt = \sqrt{\frac{1+y'^2}{2gy}}dx$, 于是质点沿曲线 $y(x)$ 从 A 点滑到 B 点的时间可表示为

$$J(y(x)) = \int_0^{x_1} \sqrt{\frac{1+y'^2}{2gy}} dx$$

$y(x)$ 在 A, B 两个端点应有

$$y(0) = 0, y(x_1) = y$$

最速降线问题归结为 $y(x)$, 在满足 (2) 的条件下, 使 (1) 的 $J(y(x))$ 达到最小。

约翰·伯努利设想质点也像光线那样按从 A 到 B 耗时最少的路径滑行, 根据光学原理 (史奈尔折射定律) 得

$$\frac{\sin \alpha}{v} = c$$

c 为常数, 由能量守恒定律得

$$v = \sqrt{2gy}$$

由几何关系得

$$\sin \alpha = \cos \beta = \frac{1}{\sec \beta} = \frac{1}{\sqrt{1 + \tan^2 \beta}} = \frac{1}{\sqrt{1 + y'^2}}$$

由 (3)、(4)、(5) 得

$$y(1 + y'^2) = b$$



其中 $b = \frac{1}{2gc^2}$

令

$\tan \varphi = \left(\frac{y}{b-y}\right)^{\frac{1}{2}}$, 则 $y = b \sin^2 \varphi$, $dy = b \sin 2\varphi d\varphi$

$dx = \left(\frac{y}{b-y}\right)^{\frac{1}{2}} dy = 2b \sin^2 \varphi d\varphi$, 积分后得

$$x = \frac{b}{2} (2\varphi - \sin 2\varphi) + c_1$$

由曲线过原点知, $\varphi = 0$ 时, $x = y = 0$, 于是 $c_1 = 0$, 故

$$\begin{cases} x = \frac{b}{2} (2\varphi - \sin 2\varphi) \\ y = \frac{b}{2} (1 - \cos 2\varphi) \end{cases}$$

令 $a = \frac{b}{2}$, $\theta = 2\varphi$ 则

$$\begin{cases} x = a(\theta - \sin \theta) \\ y = a(1 - \cos \theta) \end{cases}$$

也可令 $y' = \cot t$, 则 $y = \frac{b}{1+\cot^2 \varphi} = b \sin^2 \varphi$, $dy = b \sin 2\varphi d\varphi$,

$$dx = \frac{1}{y'} dy = 2b \sin^2 \varphi d\varphi$$

第 3 节

一个最短路线问题

引言 这是一个实际问题抽象出来的数学模型问题, 我们介绍它说明欧拉-拉格朗日方程 (Euler-Lagrange equation)。它对应于泛函的临界点。下面的推导某种角度上反应了变分法的应用: 对应一个泛函, 我们找到一个“最佳状态”, 对此的任意一个小的扰动或变动, 都会偏离这个最佳状态。在本问题中, 就是图中的 ABC 弧, 手法是对任意的其他曲线用不等式控制的方法说明它不能比所选的 ABC 弧更佳。

我们将发现这里的手法与后面黎曼几何中测地线变分的问题是类似的。

问题 设一辆汽车停止于 A 处并垂直于 AB 方向, 此汽车可转弯的最小圆半径为 R , 求不倒车时由 A 移到 B 的最短路线。将汽车视为一个点, 汽车行走的路线视为一条曲线。

解 (1) 首先讨论 $AB > 2R$ 的情形。以 AB 为 Y 轴正向, 作一半径为 R 的圆 Γ 与 X 轴切于 A 点, 问题就是要找一条最短曲线连结 AB , 在 A 点切于 X 轴正向, 且任一点的曲率半径不小于 R 。

直观上不难猜测出最短路径。从 B 点向圆 Γ 做切线 BC , 那么由 A 点沿圆弧 AC 移到 C 点, 再沿直线移到 B 点, 这就是最短路径 (如图 2 所示)。

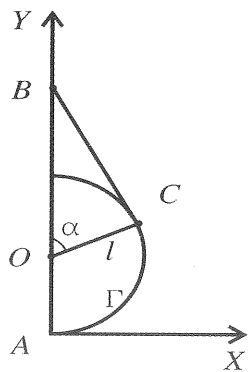


图 2

为了证明这一事实, 作一条直线 l 通过圆 Γ 的中心 O 和 C 点。假设汽车沿某一条曲线 Γ_1 由 A 点移到

B 点, 因 A, B 分别在直线 l 两侧, Γ_1 与 l 必有一交点 C_1 , Γ_1 被分成弧 AC_1 和弧 BC_1 两段. 因 BC 与 l 垂直, 弧 BC_1 的长度必不小于线段 BC 的长度 (当且仅当弧 BC_1 与线段 BC 重合时才可能相等). 设弧 AC_1 的参数方程为 $x = x(s), y = y(s), x(0) = 0, y(0) = 0$, 其中 s 为弧长. 在点 $(x(s), y(s))$ 处, 曲线的切线与 x 轴的夹角记为 θ , 依条件有

$$\left| \frac{d\theta}{ds} \right| \leq \frac{1}{R}$$

当 $s = 0$ 时, $\theta = 0$, 故

$$-\int_0^s \frac{1}{R} ds \leq \int_0^\theta d\theta \leq \int_0^s \frac{1}{R} ds$$

从而

$$|\theta| \leq \frac{s}{R}$$

研究曲线上点与直线的距离 (在 l 的右边为正)

$$J(s) = x(s) \cos \alpha - (y(s) - R) \sin \alpha, \alpha = \angle BOC$$

因为

$$\frac{dx}{ds} = \cos \theta, \frac{dy}{ds} = \sin \theta$$

故

$$x(s) = \int_0^s \cos \theta(t) dt, y(s) = \int_0^s \sin \theta(t) dt$$

因此

$$J(s) = \cos \alpha \int_0^s \cos \theta(t) dt - \sin \alpha \left(\int_0^s \sin \theta(t) dt - R \right) = \int_0^s \cos(\theta(t) + \alpha) dt + R \sin \alpha$$

当 $t \geq 0$ 时, 有 $|\theta(t)| \leq \frac{t}{R}$. 当 $0 \leq t \leq R(\pi - \alpha)$ 时, $-\frac{t}{R} + \alpha \leq \theta(t) + \alpha \leq \frac{t}{R} + \alpha \leq \pi$.

故

$$\cos(\theta(t) + \alpha) \geq \cos\left(\frac{t}{R} + \alpha\right)$$

故当 $0 \leq s \leq R(\pi - \alpha)$ 时,

$$J(s) \geq \int_0^s \cos\left(\frac{t}{R} + \alpha\right) dt + R \sin \alpha = R \sin\left(\frac{s}{R} + \alpha\right) \geq 0$$

这就是说, 当汽车移动距离不超过 $R(\pi - \alpha)$ 即弧 AC 的长度时, 它不可能超过直线 l . 因此弧 AC_1 的长度至少为 $R(\pi - \alpha)$, 并且只有当弧 AC_1 与 AC 完全重合时, 它的长度才能等于 $R(\pi - \alpha)$.

总结上述讨论, 知曲线 Γ_1 的长度必不小于 $R(\pi - \alpha) + R \tan \alpha$, 并且只有当 Γ_1 与 ACB 重合时才可能相等. 因此 ACB 是唯一的最短路径.

(2) 若 B 点在圆 Γ 内, 即 $AB < 2R$, 则应过 A 点作一半径 R 的圆, 其圆心在 BA 延长线上, 再过 B 点作一圆, 半径为 R , 且与前圆切于点 C , 则最短路径是弧 AC 和弧 CDB 所组成的曲线 (如图 3 所示).

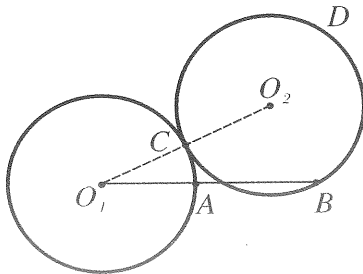


图 3

第4节

曲线长度的变分

引言 数学分析中的求一个函数的极值问题是往往是通过求一阶和二阶导数来确定的,一阶导数找出可疑点,二阶导数判断极大还是极小。(详见《数学分析中的典型问题与方法》裴礼文著)。然而这一种看似简单的方法却可以推广到很多更深刻的极值问题中去。

下面我们从黎曼几何中看看变分的应用我们假定读者懂得基本的微分几何与微分流形知识。这一节我们讨论关于测地线的变分问题,先给出一些基本定义和概念:

定义1

设 (M, g) 为一黎曼流形(或伪黎曼流形),则仿射联络 ∇ 在满足以下条件时是列维——奇维塔联络:

1. 无挠率: 也就是, 对任何向量场 X, Y 我们有 $\nabla_X Y - \nabla_Y X = [X, Y]$, 其中 $[X, Y]$ 是向量场 X 和 Y 的李括号。
2. 与度量相容: 也就是, 对任何向量场 X, Y, Z 我们有 $X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$ 。

定义2

设 (M, g) 为一黎曼流形(或伪黎曼流形),则仿射联络 ∇ 在满足以下条件时是列维——奇维塔联络。

1. 无挠率: 也就是, 对任何向量场 X, Y 我们有 $\nabla_X Y - \nabla_Y X = [X, Y]$, 其中 $[X, Y]$ 是向量场 X 和 Y 的李括号。
2. 与度量相容: 也就是, 对任何向量场 X, Y, Z 我们有 $X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$ 。

定义3

在一个黎曼流形 M 上, 一条曲线 $\gamma: [a, b] \rightarrow M$ 若符合常微分方程 $\nabla_{\gamma'} \gamma' = 0$, 就称之为测地线。其中 ∇ 是 M 上的列维——奇维塔联络(或称黎曼联络)。方程左边为曲线在流形上的加速度向量, 所以方程是说测地线是在流形上加速度为零的曲线, 也因此测地线必定是等速曲线。以上方程用局部坐标表示为

$$\frac{d^2 \gamma^\lambda}{dt^2} + \Gamma_{\mu\nu}^\lambda \frac{d\gamma^\mu}{dt} \frac{d\gamma^\nu}{dt} = 0$$

其中 $\Gamma_{\mu\nu}^\lambda$ 是 M 的黎曼度量的克里斯托费尔符号。

定义4

设是黎曼流形中一条光滑曲线段。若有 $\varepsilon > 0$ 以及光滑映射 $\Gamma: [a, b] \times (-\varepsilon, \varepsilon) \rightarrow M$, 满足 $\Gamma(t, 0) = \gamma(t)$, 则称 Γ 是曲线 γ 的一个变分。记 $\gamma_s: [a, b] \rightarrow M$, 使得 $\gamma_s(t) = \Gamma(t, s)$, 称之为变分曲线。令 $U(t) = d\Gamma_{(t, s)}(\partial_s)|_{s=0}$, 称为 Γ 的变分向量场。

启发的问题：曲面上联结两点的最短线是测地线？这里极值性体现在“最短”这两个字。这样，我们第一步应该描述出“最短”这两个字。

定义 5

设 $\gamma: [a, b] \rightarrow M$ 是黎曼流形 (M, g) 上一条分段光滑曲线。如果对于 M 中联结 $\gamma(a)$ 与 $\gamma(b)$ 的任意一条分段光滑曲线 σ ，都有 $L(\sigma) \geq L(\gamma)$ ，则称 γ 联结 $\gamma(a)$ 与 $\gamma(b)$ 的一条最短线。其中 $L(\cdot)$ 是长度泛函，定义为 $L(\gamma) := \int_a^b |\dot{\gamma}'(t)| dt$ 。（注： $\dot{\gamma}'(t)$ 为切向量。）

定理 1

曲面上联结两点的最短的单位速度光滑曲线是测地线。

注 1

单位速度指切向量始终模长为 1；更一般的可以考虑分段光滑，详见 P96, John M. Lee, GTM176-Riemannian Manifold.

这里的任意性很大，我们的方法是变分法，考虑这条最短线的任意一个“扰动”。为方便起见，我们使扰动固定端点且变化光滑。不妨我们放在三维空间中的二维曲面上试着想象，就好像有无数条曲线组合成了一个梭形的小曲面（即一个变分），并且每一条变分曲线都对应了一个参数 s ，记作 γ_s 。观察！我们有一个关于 s 的一元函数 $l(s) := L(\gamma_s)$ 。这样，我们完全可以预见到数学分析里那种基本的求导数的方法可以照搬到对 $l(s)$ 的处理上。如果假定 $s = 0$ 对应原最短曲线，那么简单地由数学分析的知识，“最短”的已知条件就有一个简单的必要条件： $l'(0) = 0$ ！

事实上，只有如何演算才是难点，黎曼几何中的第一变分公式是长度泛函一阶导数的对应物。

命题 1 第一变分公式

设 $\gamma: [a, b] \rightarrow M$ 是一条单位速度光滑曲线。如果 $\Gamma: [a, b] \times (-\varepsilon, \varepsilon) \rightarrow M$ 是一个变分，设 $l(s) := L(\gamma_s)$ ，则有：

$$l'(0) = \langle \dot{\gamma}', U \rangle \Big|_a^b - \int_a^b \langle U, \nabla_{\dot{\gamma}'} \dot{\gamma}' \rangle dt$$

其中 U 是变分向量场， ∇ 是黎曼联络。

证：

简单的记， $T(t, s) = \partial_t \Gamma(t, s)$, $S(t, s) = \partial_s \Gamma(t, s)$ 。利用一个简单关系 $D_s \partial_t \Gamma = D_t \partial_s \Gamma$ （类似于偏导数交

换顺序, 写出局部坐标系不难证明, 在此略去), 有:

$$\begin{aligned} l'(s) &= \frac{d}{ds} L(\gamma_s) = \int_a^b \frac{\partial}{\partial s} \langle T, T \rangle^{1/2} dt \\ &= \int_a^b \frac{1}{2} \langle T, T \rangle^{-1/2} 2 \langle D_s T, T \rangle dt \\ &= \int_a^b \frac{1}{|T|} \langle D_t S, T \rangle dt \end{aligned}$$

注意 $U(t) = d\Gamma_{(t,s)}(\partial_s)|_{s=0} = S(t, 0)$, $\gamma'(t) = d\Gamma_{(t,s)}(\partial_t)|_{s=0} = T(t, 0)$, 我们有:

$$\begin{aligned} l'(0) &= \frac{d}{ds} \Big|_{s=0} L(\gamma_s) = \int_a^b \langle D_t U, \gamma' \rangle dt \\ &= \int_a^b \left(\frac{d}{dt} \langle U, \gamma' \rangle \right) - \langle u, D_t \gamma' \rangle dt \\ &= \langle \gamma', U \rangle \Big|_a^b - \int_a^b \langle U, \nabla_{\gamma'} \gamma' \rangle dt \end{aligned}$$

命题获证。 □

回到定理的证明, 注意一个事实: 若 U 是一个变分向量场, 则 φU 是另一个变分的变分向量场, 其中 φ 是一个光滑函数。取 $U = \phi \nabla_{\gamma'} \gamma'$, 其中使得在 (a, b) 上 $\phi > 0$, 在 $t = a, b$ 处 $\phi = 0$ 。这样我们有 $0 = -\int_a^b \phi |\nabla_{\gamma'} \gamma'|^2 dt$, 但右边非正故处处为零, 即有 $\nabla_{\gamma'} \gamma' = 0$ ——测地线的定义式。至此, 定理 1.2 得证。

仔细推敲上面的证明不难发现, 我们并没有完整的用到“最短”这个条件, 而仅仅只用到一阶导数等于零这样一个性质。例如“最长”曲线似乎也可以同理证明是一条测地线。

因此, 我们得到的准确的结果如下:

推论 1

一条单位速度光滑曲线 $\gamma: [a, b] \rightarrow M$ 是长度泛函 $L(\cdot)$ 的临界点 (即导数为零的“点”) 当且仅当它是一条测地线。

换句话说, 我们只用到了“最”, 没用到“短”。为克服这一点, 类比数学分析中的问题, 我们势必要涉及到二阶导数才行。事实上, 黎曼几何中的第二变分公式是长度泛函二阶导数的对应物! 它涉及曲率张量、共轭点、雅可比场等概念, 并且有许多深刻的结果, 如 Morse 指标定理、Bonnet-Myers 定理, 在此略去, 详见 p185, John M. Lee, GTM176-Riemannian Manifold。注意, 在单位速度的假定下, 可以用能量泛函 $E(\gamma) := \int_a^b |\gamma'(t)|^2 dt$ 代替长度泛函, 详见 p191, do Carmo, Riemannian Geometry。

第 5 节

曲面面积的变分

这一节我们讨论与面积变分相关的一些问题, 为简单起见, 我们的主要对象是三维空间中的曲面, 希望类似上面的讨论得到“最小”的曲面所应有的性质。

设 Σ 是 E^3 的曲面, D 是 Σ 上的一个紧致区域, \vec{e}_3 是 Σ 的单位法向, 设 f 是 Σ 上的函数, 支撑集包含在 D 是中. 曲面族

$$\vec{r}_t = \vec{r} + tf \cdot \vec{e}_3, -\varepsilon < t < \varepsilon$$

称为 Σ 的法向变分, 记作 Σ_t (这里 \vec{r}_t 为位置向量). 同样可以定义变分向量场: $\frac{d\vec{r}_t}{dt} |_{t=0}$ 称为变分 Σ_t 的变分向量场. 对上面的法向变分而言, 变分向量场就是 $f \cdot \vec{e}_3$.

设 $\{\vec{r}, \vec{e}_1, \vec{e}_2, \vec{e}_3\}$ 是 Σ 的活动标架, \vec{e}_3 是 Σ 的单位法向; $\{w_1, w_2, w_{12}, w_{23}, w_{13}\}$ 是相应的诸微分形式, 且 $w_{13} = h_{11}w_1 + h_{12}w_2$, $w_{23} = h_{21}w_1 + h_{22}w_2$, 其中 $(h_{\alpha\beta})_{\alpha,\beta=1,2}$ 是曲面 Σ 的第二基本形式. 对上面的变分式求微分并利用曲面的结构方程, 我们有:

$$\begin{aligned} d\vec{r}_t &= d\vec{r} + tdf \cdot \vec{e}_3 + tf \cdot d\vec{e}_3 \\ &= (\vec{e}_1 + tf_1\vec{e}_3 + tfh_{11}\vec{e}_1 + tfh_{21}\vec{e}_2)w_1 + (\vec{e}_2 + tf_2\vec{e}_3 + tfh_{12}\vec{e}_1 + tfh_{22}\vec{e}_2)w_2 \end{aligned}$$

其中 f_1, f_2 由 $df = f_1w_1 + f_2w_2$ 定义. 曲面 Σ_t 的面积元为

$$dA_t = J_t w_1 \wedge w_2$$

其中

$$\begin{aligned} J_t &= |[(1 + tfh_{11})(1 + tfh_{22}) - t^2 f^2 h_{12}^2] \cdot \vec{e}_3 \\ &\quad + [t^2 f f_1 h_{12} - (1 + tfh_{11}) \cdot t f_2] \cdot \vec{e}_2 \\ &\quad + [t^2 f f_2 h_{21} - (1 + tfh_{22}) \cdot t f_1] \cdot \vec{e}_1| \end{aligned}$$

将上式按 t 展开到二项式, 有

$$J_t \sim 1 + 2tHf + \frac{t^2}{2}(2Kf^2 + f_1^2 + f_2^2)$$

其中 H, K 分别是平均曲率和高斯曲率.

结合 dA_t 与 J_t 的表达式, 得到

$$\begin{aligned} \frac{dA_t}{dt} |_{t=0} &= J_t |_{t=0} \frac{w_1 \wedge w_2}{dt} \\ &= \int_{\Sigma} \frac{d}{dt} J_t |_{t=0} w_1 \wedge w_2 \\ &= 2 \int_{\Sigma} H f w_1 \wedge w_2 \end{aligned}$$

由 f 的任意性不难发现如下定理:

定理 2

Σ 是面积泛函的临界点当且仅当 $H = 0$ 即 Σ 是极小曲面.

从这里, 我们也看出为什么将平均曲率为零的曲面称为极小曲面.

第6节

后记

历史上, 变分法可能是从约翰·伯努利 (1696) 提出最速曲线 (brachistochrone curve) 问题开始出现的. 它立即引起了约翰·伯努利和洛必达的注意, 但欧拉首先详尽的阐述了这个问题. 他的贡献始于 1733 年, 他的《变分原理》(Elementa Calculi Variationum) 寄予了这门科学这个名字. 拉格朗日对这个理论的贡献非常大. 拉格朗日 (1786) 确定了一种方法, 但在对极大和极小的区别不完全令人满意. 牛顿和莱布尼茨也是在早期关注这一学科. 对于这两者的区别 Vincenzo Brunacci(1810), Carl Friedrich Gauss(1829), Simeon Poisson(1831), Mikhail Ostrogradsky(1884), 和 Carl Jacobi(1837) 都曾做出过贡献. Sarrus(1842) 的由 Cauchy(1844) 浓缩和修改的是一个重要的具有一般性的成就. Strauch(1849), Jellett(1850), Otto Hesse(1857), Alfred Clebsch(1858), 和 Carll(1885) 写了一些其他有价值的论文和研究报告, 但可能那个世纪最重要的成果是 Weierstrass 所取得的. 他关于这个理论的著名教材是划时代的, 并且他可能是第一个将变分法置于一个稳固而不容置疑的基础上的. 1900 发表的第 20 和 23 个希尔伯特 (Hilbert) 促进了更深远的发展. 在 20 世纪 David Hilbert, Emmy Noether, Leonida Tonelli, Henri Lebesgue 和 Jacques Hadamard 等人做出重要贡献. Marston Morse 将变分法应用在 Morse 理论中. Lev Pontryagin, Ralph Rockafellar 和 Clarke 广义变分法理想控制论发展了新的数学工具.

小波分析原理及应用研究

2011 级信计班 陶发艳

摘要 小波分析是数学与应用数学领域三十年来迅速发展的一新的分支,在理论和应用方面都取得了令人瞩目的成就。本文从连续小波变换、离散小波变换及多分辨率分析三方面阐述了小波分析的原理;从信号的奇异性检测、信号的消噪处理、数据图像压缩、图像处理等方面研究了小波分析的部分应用;并对小波分析未来的发展及应用进行了展望。本文对小波分析的原理及应用研究具有一定的现实意义。

第 1 节

前言

小波分析是近年迅速发展起来的新兴学科,具有深刻的理论意义和广泛的应用范围。小波分析是一种信号的时间-尺度(时间-频率)分析方法,它具有多分辨率分析的特点,而且在时频两域都具有表征信号局部特征的能力,是一种窗口大小固定不变,但其形状可以改变的时频局部化分析方法。即在低频部分具有较高的频率分辨率和较低的时间分辨率,在高频部分具有较高的时间分辨率和较低的频率分辨率,很适合于探测正常信号中夹带的瞬变反常信号并分析其成分,所以被誉为分析信号的显微镜。由于小波具有多分辨率分析的能力,可以对信号和图像在不同尺度上进行分解,在小波域进行去噪、压缩处理后,作反变换得到去噪和压缩后的信号和图像。小波分析用于非平稳信号和图像的处理优于传统的傅立叶变换,已被许多应用领域的事实所证实。因此,自小波分析诞生到现在不过 10 年的时间,就在诸如地球物理勘探、信号信息处理、图像处理、语音分割与合成、故障诊断、雷达信号分析等领域内取得了很佳的应用效果。

第 2 节

小波分析的基本原理

小波函数源于多分辨率分析,其基本思想是将 L^2 中的函数 $f(x)$ 表示为一系列逐次逼近表达式,其中每一个都是 $f(x)$ 经过平滑后的形式,它们分别对应不同的分辨率。多分辨率分析,又称多尺度分析,是建立在函数空间概念基础上的理论,其思想的形成来源于工程。创建者 Mallat.S 是在研究图像处理问题时建立这套理论的。当时人们研究图像的一种很普遍的方法是将图像在不同尺度下分解,并将结果进行比较,以取得有用的信息。Meyer 正交小波基的提出,使得 Mallat 想到是否用正交小波基的多尺度特性将图像展开,以得到图像不同尺度间的“信息增量”。这种思想导致了多分辨率分析理论的建立。Mar 不仅为正交小波基的构造提供了一种简单的方法,而且为正交小波变换的快速算法提供了理论依据。其思想又同多采样率滤波器组不谋而合,使我们又可将小波变换同数学滤波器的理论结合起来。因此,多分辨率分析在正交小波变换理论中具有非常重要的地位。小波,通俗地讲即小区域的波,是一种特殊的、长度有限、平均值为 0 的波形。它有两个特点:一是“小”,即在时域具有紧支集;二是正负交替的“波动性”,也即支流分量为 0。

2.1

连续小波变换

设 $\Psi(t) \in L^2(\mathbb{R})$, 且其 Fourier 变换 $\Psi(\omega)$ 满足可允许性条件即

$$\int_{-\infty}^{+\infty} \frac{\|\Psi(\omega)\|^2}{\|\omega\|} d\omega < +\infty$$

则称 $\Psi(t)$ 为一基本小波或者小波母函数。

根据小波函数的定义, 小波函数一般在时域具有紧支集或近似紧支集, 即函数的非零值定义域具有有限的范围, 这即所谓“小”的特点; 另一方面, 根据可容性条件可知 $\Psi(\omega)|_{\omega \rightarrow +\infty} = 0$, 即直流分量为 0, 因此小波又具有正负交替的波动性。常见的小波母函数有 Haar 小波、Marr 小波、线性样条小波。下面简单介绍这三种小波。

1. Haar 小波

设

$$f(x) = \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & o.w \end{cases} \quad (1)$$

显然 $h(t) \in L^2(\mathbb{R})$, 又因为

$$\hat{h}(\omega) = \int_{-\infty}^{+\infty} h(t)e^{-i\omega t} dt = 4e^{-i\frac{\omega}{2}\pi} \frac{\sin^2 \frac{\omega}{4}}{\omega}$$

$$\int_{-\infty}^{+\infty} \frac{\|h(\omega)\|^2}{\|\omega\|} d\omega = 16 \int_{-\infty}^{+\infty} \frac{\sin^4 \frac{\omega}{4}}{\|\omega\|^3} d\omega < +\infty$$

可见, 函数满足可允许性条件, 所以是一个基本小波。这是 A.Harr 于 1910 年最先给出的, 是小波分析理论中一个非常简单且具有代表性的小波例子。

2. Marr 小波 (也称为墨西哥帽子小波)

设 $g(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$, 令 $\Psi(t) = -\frac{d^2}{dt^2} g(t) = \frac{1}{\sqrt{2\pi}} (1-t^2)e^{-\frac{t^2}{2}}$, 易知 $\Psi(t) \in L^2(\mathbb{R})$ 。又 $g(\omega) = e^{-\frac{\omega^2}{2}}$, 根据 Fourier 变换的微分形式, 有 $\hat{\Psi}(\omega) = -(i\omega)^2 \hat{g}(\omega) = \omega^2 e^{-\frac{\omega^2}{2}}$, 因为

$$\int_{-\infty}^{+\infty} \frac{\|\Psi(\omega)\|^2}{\|\omega\|} d\omega = 2 \int_{-\infty}^{+\infty} \omega^3 e^{-\omega^2} d\omega = 1 < +\infty$$

满足可允许条件, 所以 Marr 小波是一个基本小波。

3. 线性样条小波

设 $\Psi(t) = h^* \zeta(t)$, 其中 $\zeta(t) = \begin{cases} 1, 0 \leq t < 1 \\ 0, o.w \end{cases}$, 而 $h(t)$ 是 Harr 小波, 易知

$$\Psi(x) = \begin{cases} t & 0 \leq t < \frac{1}{2} \\ 1-t & \frac{1}{2} \leq t < \frac{3}{2} \\ t-2 & \frac{3}{2} \leq t < 2 \\ 0 & o.w \end{cases} \quad (2)$$

显然中 $\zeta(t)$ 满足可允许条件, 故 $\Psi(t)$ 是小波母函数。因为 $\Psi(t)$ 是逐段线性的连续函数, 故称之为线性样条函数。

连续小波变换

设 $f(t) \in L^2(\mathbb{R})$, $\Psi(t)$ 为一基本小波, 令 $\Psi_{ab}(t) = \frac{1}{\sqrt{|a|}} \Psi(\frac{t-b}{a})$ 其中 $a, b \in \mathbb{R}, a \neq 0$, 则称函数 $\Psi_{ab}(t)$ 为小波母函数 $\Psi(t)$ 生成的依赖于参数的连续小波, 而称 $\omega_f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \overline{\Psi(\frac{t-b}{a})} dt$ 为连续小波变换。

2.2

离散小波变换

离散小波

$$\Psi_{j,k}(t) = a_0^{\frac{j}{2}} \Psi(a_0^j t - kb_0), j, k \in \mathbb{Z}$$

其中 $a_0 > 0, b_0 > 0$, \mathbb{Z} 表示整数全体所构成的集合。则称 $\Psi_{j,k}(t)$ 为离散小波。为了便于计算机处理, 总是取 $a_0 = 2, b_0 = 1$ 。

离散小波变换

设 $f(t) \in L^2(\mathbb{R})$, $\Psi(t)$ 为一基本小波, 另 $C_f(j, k) = \int_{-\infty}^{+\infty} f(t) \overline{\Psi_{j,k}(t)} dt, j, k \in \mathbb{Z}$, 则称 $C_f(j, k)$ 为 $f(t)$ 的离散小波变换。

正交小波

设 $\Psi(t) \in L^2(\mathbb{R})$, $\Psi(t)$ 为一基本小波, 且 $\Psi_{j,k}(t) = 2^{\frac{j}{2}} \Psi(2^j t - k), j, k \in \mathbb{Z}$, 构成的一个标准正交基, 即 $\{\Psi_{j,k}(t)\}_{j,k \in \mathbb{Z}}$ 是 $L^2(\mathbb{R})$ 中的一个完备的标准正交系:

$$\int_{-\infty}^{+\infty} \Psi_{j,k}(t) \overline{\Psi_{j',k'}(t)} dt = \begin{cases} 1 & (j, k) = (j', k') \\ 0 & (j, k) \neq (j', k') \end{cases} \quad (3)$$

则称 $\Psi(t)$ 是一个正交小波, 而称 $\{\Psi_{j,k}(t)\}_{j,k \in \mathbb{Z}}$ 是 $L^2(\mathbb{R})$ 的正交小波基。进一步, 若 $\{\Psi_{j,k}(t)\}_{j,k \in \mathbb{Z}}$ 构成 $L^2(\mathbb{R})$ 的正交小波基, 则对于 $L^2(\mathbb{R})$ 中的任意一个函数 $f(t)$ 都有展开式 $f(t) = \sum_{j,k \in \mathbb{Z}} C_f(j, k) \Psi_{j,k}(t)$ 。

2.3

多分辨率分析

Meyer 于 1986 年创造性地构造了具有一定衰减性的光滑函数, 其二进制伸缩与平移构成 $L^2(\mathbb{R})$ 的规范正交基, 才使小波得到真正的发展。1988 年 Mallat 在构造正交小波基时提出了多分辨率分析的概念, 从空间的概念上形象地说明了小波的多分辨率特性, 将之前的所有正交小波基的构造法统一起来, 给出了正交小波的构造方法以及正交小波变换的快速算法, 即 Mallat 算法。Mallat 算法在小波分析中地位相当于快速傅立叶变换算法在经典傅立叶分析中的地位。

正交多分辨率分析

Hilbert 空间 $L^2(\mathbb{R})$ 中的一列闭子空间 $\{V_j\}_{j \in \mathbb{Z}}$ 称为一个正交多分辨率分析 (记为 OMRA), 如果满足

1. 嵌套性: $V_j \subseteq V_{j+1} (j \in \mathbb{Z})$
2. 伸缩性: $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}$
3. 隔离性: $\bigcap_{j \in \mathbb{Z}} V(j) = \{0\}$
4. 稠密性: $\overline{\bigcup_{j \in \mathbb{Z}} V(j)} = L^2(\mathbb{R})$
5. 正交性: 存在 $\varphi(t) \in V(0)$, 使得 $\{\varphi(t-k), k \in \mathbb{Z}\}$ 是 V_0 的特殊标准正交基, 其中的 $\varphi(t)$ 称为该 OMRA 的尺度函数, $V(j)$ 称为 $L^2(\mathbb{R})$ 的逼近子空间或尺度空间。

广义多分辨率分析

Hilbert 空间 $L^2(\mathbb{R})$ 中的一列闭子空间 $\{V_j\}_{j \in \mathbb{Z}}$ 称为一个广义多分辨率分析 (记为 GMRA), 如果满足

1. 嵌套性: $V_j \subseteq V_{j+1} (j \in \mathbb{Z})$
2. 伸缩性: $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}$
3. 隔离性: $\bigcap_{j \in \mathbb{Z}} V(j) = \{0\}$
4. 稠密性: $\overline{\bigcup_{j \in \mathbb{Z}} V(j)} = L^2(\mathbb{R})$
5. Rize 基: 存在 $g(t) \in V(0)$ 使得 $\{g(t-k), k \in \mathbb{Z}\}$ 是 V_0 的一个 Rize 基, 其中的 $g(t)$ 称为该 GMRA 的尺度函数。

第 3 节

小波分析的应用

随着小波理论的日益成熟, 人们对小波分析的实际应用越来越重视。小波分析的应用范围极广, 遍及自然科学、应用科学乃至社会经济等许多领域。它包括: 数学领域的许多学科; 信号分析、影像处理; 量子力学、理论物理; 军事电子对抗与武器的智能化; 电脑分类与识别; 音乐与语言的人工合成; 医学成像与诊断; 地震勘探数据处理; 大型机械的故障诊断等方面; 例如, 在数学方面, 它已用于数值分析、构造快速数值方法、曲线曲面构造、微分方程求解、控制论等。在信号分析方面的滤波、去噪声、压缩、传递等。在影像处理方面的影像压缩、分类、识别与诊断, 去污等。在医学成像方面的减少 B 超、CT、核磁共振成像的时间, 提高解析度等。下面仅就目前小波分析的部分应用领域作一些探讨。

3.1

信号奇异性检测

信号中的奇异点及不规则的突变部分经常带有比较重要的信息,它是信号的重要特征之一。比如,在故障诊断中,故障通常表现为输出信号发生突变,因而对突变点的检测在故障诊断中有着重要的意义。由于傅里叶变换将信号变换成纯频域中的信号,而使它不具有时间分辨的能力,故对信号在时域中的突变点根本无法检测出来。而小波变换具有良好的空间局部化性质,利用小波变换对信号的奇异性及奇异性位置和奇异度的大小都有比较准确的判断。Grossman 利用 Hardy 函数给出了一种测定信号奇异性的方法,他把小波变换分为:小波变换相位函数空 $U(s, x)$ 和小波变换模函数 $Q(s, x)$,但由于相位函数不是测定奇异性的充要条件,1992 年由 Mallat 等人基于计算机视觉研究中的多尺度变换提出了一种更加有效的方法,用小波变换的模极大值检测信号的奇异点,在分析典型心电图 (ECG) 时得到比较理想的效果。小波分析检测信号突变点的一般方法是:对信号进行多尺度分析,在信号出现突变时,其小波变换后的系数具有模量极大值,因而可以通过对模量极大值点的检测来确定故障发生的时间点。

3.2

信号的消噪处理

去噪是信号处理领域经常遇到的问题,最简单的也比较通用的去噪方法是对信号直接进行低通或带通滤波。虽然这种方法简单,易于实现,但它不能滤去有效频带内的噪声,并且滤波器带宽的选择与高分辨率是相矛盾的。小波分析比经典的滤波方法更具有灵活性。以小波变换为基础的时变信号去噪方法是把含噪声信号放在二维平面上,利用信号和噪声表现出的截然不同的特性进行分时分频处理。多尺度小波下信号与噪声的不同特征的理为提高 MR 图像质量提供了新的理论根据。1994 年,Xue 等人根据这个理论提出了一种 MR 图像的去噪算法,它的基本思想是:通过比较相邻尺度小波系数的相关与较小尺度上的小波系数的绝对值大小来区分信号和噪声,能在较好地保留边缘的同时滤除噪声。Coifman 和 Donoho 利用非线性阈值小波处理方法可以有效的提取出被强大背景噪声所淹没了的有用信号。通过小波的信号消噪处理不但能获得较高的信噪比而且能够保持良好的时间分辨率。

3.3

数据图像压缩

对于图像来说,如果需要快速或实时传输以及大量存储,就需要对图像数据进行压缩。所谓压缩就是去掉图像信息中的各种冗余而保留有用信息。小波分析进行数据压缩是通过只保存模极大值的信息来实现的。虽然模极大值包含的原信息并不完备,但在一定程度上由极大值重建的信号可以很好的逼近原信号。基于小波分析的图像压缩方法很多,目前比较成功的有小波包最好基方法、小波域纹理模型方法、小波变换零树压缩、小波变换向量量化压缩等。用小波分析方法进行数据图像压缩时,压缩比高,压缩速度快,压缩后能保持信号与图像的基本特征不变且在传递过程中可以抗干扰。美国耶鲁大学以 R.Coifman 教授等人为代表的小波研究小组用小波分析对美国联邦调查局存贮的三亿指纹进行数据压缩,取得了二十倍有益的成果,因为存储光盘的减少而节省的费用也有三千万美元,由指纹传输时间缩短为原来的二十分之一所创造的价值更是无法估计。

3.4

图像处理

在图像处理中,小波分析的应用是很成功的。1988 年 Mallat 将计算机视觉领域内的多分辨分析引入小波分析中提出了快速小波算法 Mallat 算法,并将这一理论用于图像分析和完全重构。小波变换可以将一幅图像分解为大小位置和方向都不同的分量。在做逆变换之前可以改变小波变换域中某些系数的大小,从而能够有选择的放大所感兴趣的分量而减小不需要的分量(这就是我们通常所说的图像增强)。小波变

换可以有效的进行图像融, 图像融合是多传感器信息融合领域的一个重要分支, 它是指将来自同一目标的不同传感器的信息通过一定的算法融合到一幅图上, 从而获得比在单幅图上更完整、更精确的信息。图像融合在军事(如军事侦察、识别伪装)和非军事(如医疗诊断、遥感、计算机技术等)领域得到广泛的应用。小波变换方法还可以进行图像的边缘检测、图像匹配、图像目标识别及图像细化等, 其中二进小波变换对图像的拼接和镶嵌具有很好的效果。图像处理是小波变换理论非常有前景的应用领域。

3.5

其它应用

小波分析在地球物理勘探中对寻找地壳物质物性参数的奇异性具有非常的意义。小波分析是数值分析强有力的工具, 能简洁有效的求解偏微分方程、积分方程以及线性非线性问题。它还在医学领域中有着广泛的应用, 利用它可以有效的识别淋巴细胞微核, 有效的检验药品以及其他各种化合物的毒性。人们利用小波方法对平面叶栅叶型进行优化设计, 从而有效地解决了流体力学中传统方法难以解决的难度较大的问题。

第4节

小波分析的展望

小波分析已对许多学科产生了多方面的影响, 并已激起了众多科学家和科技工作者的极大热情。从诞生到现在不过短短十几年的时间, 已经取得了巨大的成就和发展, 它正在世界上带来一场局部化革命。但由于以下几点原因可以说小波分析应用的真正高潮尚未到来。首先, 小波理论尚不完善, 除一维小波理论比较成熟外其它各类小波如高维小波、向量小波中的正交小波、双正交小波、二进小波等的构造和基本性质理论有待进一步研究。其次, 现在国内外虽然已有一些较好小波基的选取方法, 但缺乏系统规范的最佳小波基的选取原则, 因此小波基的优化选择仍是小波理论进一步研究的重要内容。再者目前小波分析软件尚不成熟和完善, 还没有大型、系统、权威的小波分析软件。为了使小波应用的深度和广度得到进一步拓展, 除了小波理论的研究外还要再挖掘有前景的应用领域。小波分析在图像数据压缩方面人们期待用小波能实现更高压缩比, 更高重现度图像的压缩。分析非平稳非线性问题的关键是对非线性小波分析的进一步研究, 需要小波分析与神经网络、模糊数学、分形分析、遗传优化等方法相结合, 形成小波神经网络、小波模糊神经网络、小波分形等方法。另外, 人们也期待着小波与其它理论的进一步的广泛综合应用, 来解决更多的生产实际问题。

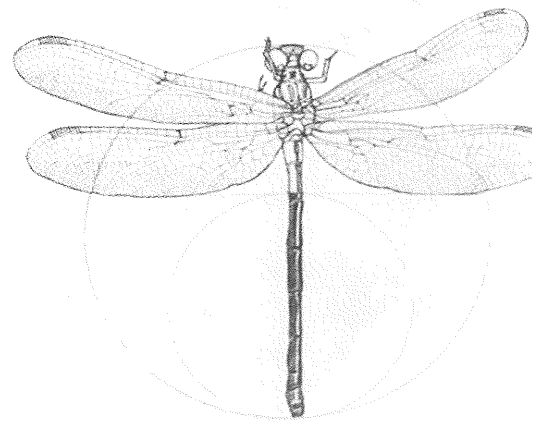
第5节

参考文献

- [1] 彭玉华. 小波变换与工程应用 [M]. 北京: 科学出版社, 2000.
- [2] 刘贵忠, 邸双亮. 小波分析及其应用 [M]. 西安: 西安电子科技大学出版社, 1992.
- [3] 李建平主编. 小波分析与信号处理——理论、应用及软件实现 [M]. 重庆: 重庆出版社, 2001.
- [4] 秦前清, 杨宗凯. 实用小波分析 [M]. 西安: 西安电子科技大学出版社, 1994.
- [5] 程正兴. 小波分析算法与应用 [M]. 西安: 西安交通大学出版社, 1998.
- [6] 杨福生. 小波变换的工程分析与应用 [M]. 北京: 科学出版社, 1999.
- [7] Daubechies, I. 小波十讲 [M]. 北京: 国防工业出版社, 2004.5.
- [8] 樊启斌. 小波分析 [M]. 武汉: 武汉大学出版社, 2008.10.
- [9] 孙延奎. 小波分析及其应用 [M]. 北京: 机械工业出版社, 2005.3.

- [10] 成礼智, 王红霞, 罗勇. 小波的理论与应用 [M]. 北京: 科学出版社, 2004.
- [11] 衡彤. 小波分析及其应用研究. 国家自然科学基金.
- [12] 刘美素, 李书光. 小波分析的理论发展及应用.[J]. 河北理工学院学报.2005.5 (2) .


数
模
优
秀
论
文





2014 Mathematical Contest in Modeling

The Keep-Right-Except-To-Pass Rule

 Wang Chang Qu Hezuo ¹

1 abstract

² We construct four models to study the performance of Keep-Right-Except-To-Pass Rule (**KREP**). The safe driving patterns in light and heavy traffic, the safety evaluation model, and the traffic flow model. Model 1 and 2 are prepared for Model 3 and 4.

Model 1 deals with the safe driving modes in the light traffic. We calculate the safe distance between front and rear cars during following, overtaking, and returning process. We analyze the different positional and motion relationships among the overtaking car, following car, and leading car. Then we get relationships of the safe distances, velocity and relative velocity. To illustrate the model, we calculate the safe distance of three velocities as a test case.

Model 2 discusses the safe driving patterns in the heavy traffic. In heavy traffic, two vehicles are within a relatively fixed distance, and the velocity is smaller. Under this circumstance, we revise the Model 1, adding one more side conditions between overtaking car and following car in the initial lane. Besides we change the values of some parameters in the former model. Then we get new results and make a comparison.

Model 3 estimates the safety evaluation system. We set a danger function to reflect the possibility of accidents in following, overtaking, and returning processes relating to velocity, relative velocity and distance. Different processes have different danger coefficient. Considering all the factors, we obtain composite danger functions in light/heavy traffic with/without the rule. Model application uses the former models to solve other questions. When in light traffic and with high speed limits, it is necessary to carry out the KREP rule, while in other situations the rule has little effect. The KREP rule can also help improve the emergency response and driving comfort. The application for left norm country can be seen as symmetrical to it is in right norm country, so the rule can be carried out to be keep-Left-Except-To-Pass. In the end, the application of intelligent system would significantly reduce the danger coefficient, and increase both the safety and traffic flow and other factors.

Model 4 studies the traffic flow. First we discuss the relationship of traffic flow, velocity, and traffic density. Then comparing to the no-overtaking condition we know that the KREP rule can promote the traffic flow but not effectively. Later using the danger function we acquire the correlation of safety, velocity and traffic flow, and we conclude that the rule has a better performance in light traffic than in heavy traffic, and can help increase the safety but not very effective either.

Model application uses the former models to solve other questions. When in light traffic and with high speed limits, it is necessary to carry out the KREP rule, while in other situations the rule has little effect. The KREP rule can also help improve the emergency response and driving comfort. The application for left norm country can be seen as symmetrical to it is in right norm country, so the rule can be carried

¹Wang Chang, 2011 级统计班王畅; Qu Hezuo, 2011 级电气院瞿合作。

²本文除章节标题样式改动外, 其他样式均符合 MCM 要求。

out to be keep-Left-Except-To-Pass. In the end, the application of intelligent system would significantly reduce the danger coefficient, and increase both the safety and traffic flow and other factors.

2 Introduction

In order to examine the rules, we should deal with four main problems:

- The relationship of safety and traffic flow
- The quantization of driving safety or the rate to cause accident
- The influence of different traffic conditions and different speed limits
- The improvement made by intelligent system

To tackle the first problem, we classify the different accidents during following, overtaking, and returning. And calculate the safety distances, which mean the minimum distance the driver must keep to avoid accident. Within the distance, there is a high possibility to cause an accident. And the farther away from the two automobile, the higher the safety is. However, the additional safety is at the cost of fewer cars on the road and lower traffic flow. Otherwise, when traffic flow is pretty high, the distance between two cars is relatively small, which would make it more dangerous to overtaking.

For the second question, we can set a danger function related to absolute velocity, relative velocity, and the distance. The faster the car is, the more it is likely to cause an accident. The greater difference of speed between two cars the more dangerous they are. And the greater the distance the car keep with cars around it, the safer it is. After precise analysis of safe distances we would know danger coefficient and get the final danger function.

For the third question in light traffic, the cars can have an arbitrarily speed (under speed limit of cause), when the limit is high, overtaking may be hard to see, however, when the limit is low, drivers may all at the limit and cannot overtaking even they want to speed up. When the traffic is heavy, the traffic flow moves slowly, in this situation if the speed limit is high, it has little affection at all, but if the speed limit is low, it may have more influence.

Since human judgment lack precise, and the possibility to accident is relatively high especially when high speed and overtaking. An intelligent system can improve the precise to drive in the straight line or overtaking. Then we can reduce the danger coefficient, and revise our former models.

3 Assumption

- The multi-lane freeways have 8 lanes in total and four lanes in same direction, the outermost lane is emergency standby lane and used only in emergency situations, such as stopping or traffic accident disposition.
- The multi-lane freeways have good conditions and would not affect the velocity for physical problems such as broken stones or pavement irregularity. The freeways are infinitely lone straight ways without traffic lights.
- Under normal circumstances, the cars have a constant speed, and when they overtaking, they would have a constant acceleration.
- The drivers all obey the traffic rules and would not exceed the speed limit.
- The drivers would not drink-driving or fatigue driving.



Table 1: Variables

Symbol	Description
Q	The traffic flow
v	The traffic density
$ $	The traffic vehicle
C	The maximum of the traffic flow on one lane
θ	The car steering angle
l	The length of the vehicle
ω	The width of the vehicle
ω_0	The width of the vehicle
a^0	The braking accelerate
a	The acceleration of overtaking
t	Time
S	The displacement in the x-axis direction
D	The danger coefficient

4 Variable and Definition

4.1 Variables

4.2 Definitions

Initial lane The lane the car is driving in the beginning.

Target lane The left lane of the initial lane where drivers are required to move on when passing another vehicle.

Traffic flow The traffic formed by the continuously road traveling.

Traffic density The intensity of the vehicles on the road, calculated as the number of vehicles on a unit length at a time.

5 Model 1 :Safe driving models in The Light Traffic

5.1 Statement

Multi-lanes freeways can be considered as the accumulation of double-lanes- model, so we can simply discuss the following and overtaking behavior in two-lane situation. We first discuss following safe distance in the same lane considering the following car has the higher, same, and slower speed than the leading car. Then we consider the speed-up overtaking and speed-down overtaking in order to get the safe distance.

Next, considering the absolute velocity, relative velocity, and the running distance, we establish a danger coefficient to reflect the safety of following and overtaking in different situations. Later, we can get the correlations between danger coefficient and ideal traffic flows. Then we can discuss the effect of speed limits and other factor as emergency solution. We can examine the model utilizing the data in the traffic office.

5.2 Safe Following Distance in Same Lane

(1)The following car is faster than the leading car

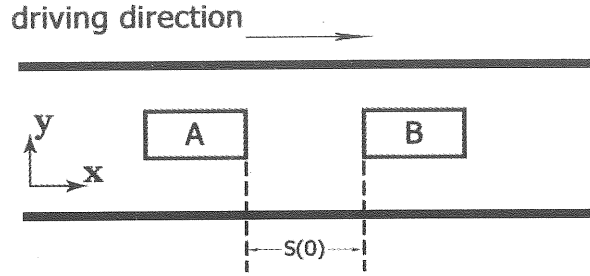


Figure 1: Safe following model

As is shown in **Figure 1**, assuming A and B are two cars driving on the freeway in the same direction at the speed of V_A and V_B , and the initial distance between the cars A and B is $S(0)$, the travel distances of A and B are L_A and L_B . The braking accelerations of A and B are a_A^0 and a_B^0 , L_s is the safe distance between two stopping cars. The time for the leading car to stop is t_l and for the following car to reflect to the leading car's behavior is t_r .

The leading car starts to stop at a constant braking acceleration, the braking distance is

$$L_B = \frac{V_B^2}{2a_B}$$

The following car sees the signal lighting of the leading car shining and begins to brake, the reflect time is t_r , during this time, the car is still running at the initial speed. And after t_r , the following car begins to brake, the distance is

$$L_A = V_A t_r + \frac{V_A^2}{2a_A}$$

So, in this situation, the safe distance between the following car and leading car is

$$L_1 = V_A t_r + \frac{V_A^2}{2a_A} - \frac{V_B^2}{2a_B} + L_s \quad (1)$$

(2) The following car has the same speed of the leading car

Let $V_A = V_B$, the distance is

$$L_2 = V_A t_r + V_A^2 \left(\frac{1}{2a_A} - \frac{1}{2a_B} \right) + L_s \quad (2)$$

5.3 Safe overtaking distance

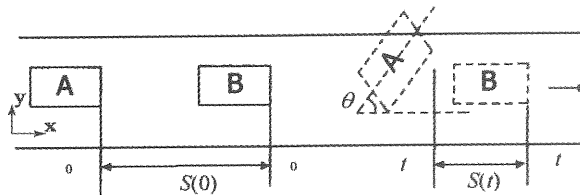


Figure 2: The vehicle changing environment

When the following car wants to overtake the leading car, the vehicle changing environment is shown as **Figure 2**, it may cause an accident in three situations. [2]

(1) Safe distance to the leading car in the same lane

The most likely hitting point is when the right corner hits the left corner of the leading car, we assume

that the following car use t_h to change its direction, and a is the acceleration when change lanes. The length of the following vehicle A is l and the width of leading vehicle B is ω . When the car steer, the length in the x-axis changes as the **Figure 5** shows.

The change in the horizontal length is

$$\Delta S = \frac{1}{2}\omega \sin\theta$$

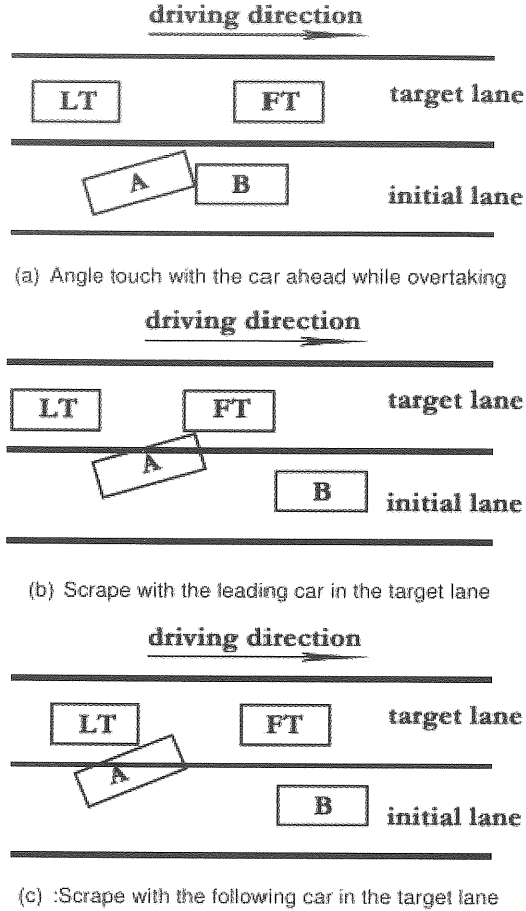


Figure 3: Three accidents might happen while overtaking

To avoid angle touch, the distance relationship is

$$S(0) + S_B - l \geq S_A + \frac{1}{2}\omega \sin\theta$$

Where S_A refers to the longitudinal journey of vehicle A within t_h , and S_B refers to the journey of B. The following car A reach the hitting point at t_h , and in the x-axis direction, the car is in the constant acceleration. So the longitudinal journey is

$$S_A = V_A t_h + \frac{1}{2} a t_h^2$$

The leading car would not change its movement pattern and still run in the initial speed, So the displacements of car B is

$$S_B = V_B t_h$$

putting the formula above into the inequality, we get the minimize initial distance to avoid the angle hitting is

$$S_{sl} = (V_A - V_B)t_h + \frac{1}{2} a t_h^2 + l + \frac{1}{2}\omega \sin\theta. \quad (3)$$

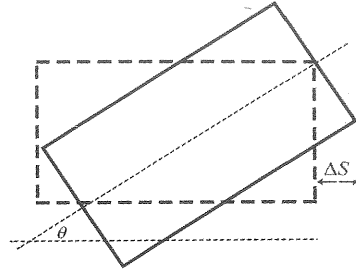
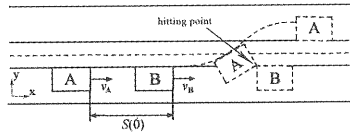


Figure 4: Safe model with leading car in the same lane Figure 5: Schematic horizontal length change

(2) Safe distance to the leading car in the target lane

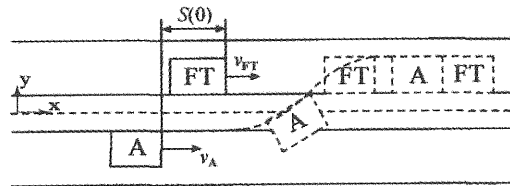


Figure 6: Safe model with the leading car in the target lane

Let the leading car in the target lane to be car **LT**, it vehicle is V_{LT} , corresponding distance is S_{LT} . We assume that when the car reaches the target lane, to decrease the danger, the speed should equal to the lane' s traffic speed.

To avoid scrape with the leading car in the target lane, the distance relationship is

$$S(0) + S_{LT} - l \geq S_A + \frac{1}{2}\omega \sin\theta$$

The overtaking car A speeding up at accelerate a , and the terminal vehicle is V_{LT} .So the time is

$$t_{LT} = \frac{V_{LT} - V_A}{a}$$

and the displacement of A in the x-axis direction is

$$S_A = \frac{V_{LT}^2 - V_A^2}{2a}$$

The leading car in the target lane is also in the uniform rectilinear motion, and the displacement is $S_{LT} = V_{LT}t_{LT}$.

Then we get the minimize safety distance between the overtaking car and the leading car in the target lane

$$S_{it} = -\frac{(V_{LT} - V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta \tag{4}$$

(3) Safe distance to the following car in the target lane

Just like the situation in (2), we take the following car in the target lane to be car **FT**, it is running at a speed of V_{FT} , the displacement of it is S_{FT} . If the distance between car **LF** and A is too short, when A turns it might incur rear-end accident too.

To avoid the rear-end accident, the distance relationship is

$$S(0) + S_A - l \geq S_{FT} + \frac{1}{2}\omega \sin\theta$$

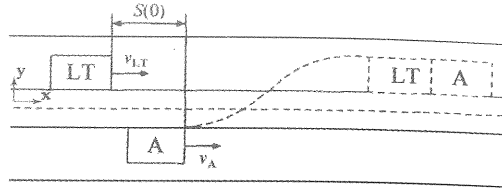


Figure 7: Safe model with the leading car in the target lane

The overtaking car is in the Uniformly accelerated rectilinear motion, and the terminal vehicle is now V_{FT} . So the time to reach the position now is $t_{FT} = \frac{V_{FT}-V_A}{2a}$, and the displacement of A in the x-axis direction is $S_A = \frac{V_{FT}^2 - V_A^2}{2a}$.

The following car FT in the target lane is still in the uniform rectilinear motion, and the displacement is

$$S_{FT} = V_{FT}t_{FT}$$

Then we get the minimize safety distance between the overtaking car and the leading car in the target lane

$$S_{ft} = \frac{(V_{FT} - V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta. \quad (5)$$

We can conclude that if the overtaking car driver wants to pass the vehicle ahead safely, avoiding the accidents with cars around it, the distance must have the demand as follows:

$$\begin{cases} S_{st} = (V_A - V_B)t_h + \frac{1}{2}at_h^2 + l + \frac{1}{2}\omega \sin\theta \\ S_{lt} = -\frac{(V_{LT}-V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta \\ S_{ft} = \frac{(V_{FT}-V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta \end{cases} \quad (6)$$

5.4 Safe returning distance model

As the rule implies,

“They move one lane to the left, pass, and return to their former travel lane.”

After overtaking, the car should return to its initial lane to continue the journey. When the traffic condition is ideal, cars can move at a high speed and a larger space, after passing, the vehicle of the car is higher than it used to be that is to say, the car is now faster than cars in its initial lane.

Considering that the car cannot keep driving in the passing lane and light traffic precondition, a small number of the cars would occupy the passing lane at the same time, so there is almost impossible for the car to hit the fore-and-aft cars in the passing lane, the accident would only occur among the cars we study (we name it the overtaking car) and the cars in the initial lanes.

When the traffic flow is small, there is no need for the overtaking car reduces its higher speed back to the initial speed, so it can keep the vehicle and only to change the direction. And it may cause accident with fore-and-aft cars in the initial lane, which is shown in **Figure5**.

(1) Safe distance to the following car in the initial lane

We name the following car in the initial lane to be car FI, it is running at a speed of V_{FI} , the displacement of it is S_{FI} . If the distance between car LI and A is too short, the two car might hit.

To avoid the rear-end accident, the distance relationship is

$$S(0) + S_A - l \geq S_{FI} + \frac{1}{2}\omega \sin\theta$$

We simplify the situation that car A is uniform rectilinear motion, and the direction has θ angle from the horizon, the speed is V_A . Let the time to be t_1 .

In the x-axis direction $S_A = V_A t_1 \cos\theta$

In the y-axis direction $\omega_0 = V_A t_1 \sin\theta$

Where the ω_0 is the width of one lane of freeway. The following car FI in the initial lane is still in the uniform rectilinear motion, and the displacement is

$$S_{FI} = V_{FI} t_1$$

Then we get the minimize safety distance between the overtaking car and the following car in the initial lane

$$S_{fi} = \frac{V_{FI}}{V_A} \omega_0 \sin\theta - \omega_0 \cot\theta + l + \frac{1}{2} \omega \sin\theta. \quad (7)$$

(2) Safe distance to the leading car in the initial lane

If the overtaking car faces a car LI ahead of it in horizon, and the space is small, it of great possible to cause an accident. If the leading car has a speed V_{LI} higher than V_A , then the overtaking car A may choose to follow this car. Otherwise, the car would exceed the car.

a. $V_{LI} > V_A$

To avoid the rear-end accident, the distance relationship is

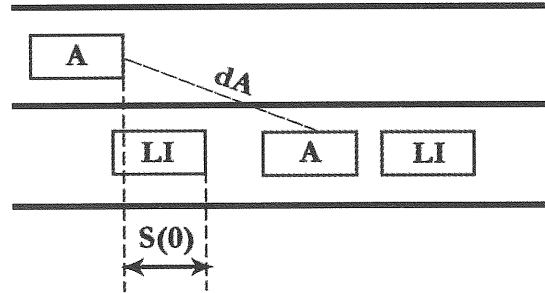


Figure 8: Safe model of the leading car in the initial lane faster than overtaking car

$$S(0) + S_{LI} - l \geq S_A + \frac{1}{2} \omega \sin\theta$$

The displacement in the x-axis direction

$$S_A = \omega_0 \cot\theta$$

The leading car LI in the initial lane 's displacement is

$$S_{LI} = \frac{V_{LI}}{V_A} \omega_0 \sin\theta$$

Then we get the minimize safety distance between the overtaking car and the following car in the initial lane

$$S_{ii} = \left(\cot\theta - \frac{V_{LI}}{V_A} \sin\theta \right) \omega_0 + l + \frac{1}{2} \omega \sin\theta \quad (8)$$

b. $V_{LI} < V_A$

The overtaking car would exceed the car. And to avoid the rear-end accident, the distance relationship is

$$S(0) + S_{LI} + \frac{1}{2} \omega \sin\theta \geq S_A - l$$

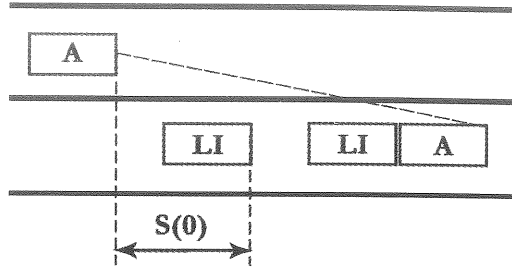
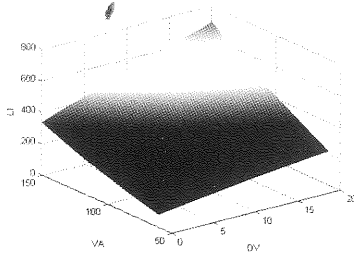
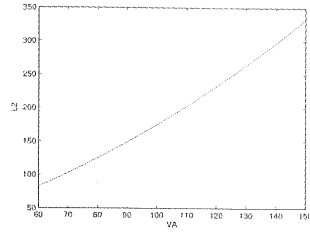


Figure 9: Safe model of the leading car in the initial lane slower than overtaking car

Figure 10: The value of L_1 Figure 11: The value of L_2

The displacement in the x-axis direction and the leading car is the same as situation a. Then we get the minimize safety distance between the overtaking car and the leading car in the initial lane

$$S'_{li} = (\cot\theta - \frac{V_{LI}}{V_A} \sin\theta)\omega_0 - l - \frac{1}{2}\omega \sin\theta \quad (9)$$

In conclusion, we get our safe following, overtaking, returning model.

$$\begin{cases} L_1 = V_A t_r + \frac{V_A^2}{2a_A} - \frac{V_B^2}{2a_B} + L_s \\ L_2 = V_A t_r + V_A^2 \left(\frac{1}{2a_A} - \frac{1}{2a_B} \right) + L_s \end{cases}$$

$$\begin{cases} S_{sl} = (V_A - V_B)t_h + \frac{1}{2}at_h^2 + l + \frac{1}{2}\omega \sin\theta \\ S_{lt} = -\frac{(V_{LT} - V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta \\ S_{ft} = \frac{(V_{FT} - V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta \end{cases}$$

$$\begin{cases} S_{fi} = \frac{V_{FI}}{V_A} \sin\theta - \omega_0 \cot\theta + l + \frac{1}{2}\omega \sin\theta \\ S_{li} = (\cot\theta - \frac{V_{LI}}{V_A} \sin\theta)\omega_0 + l + \frac{1}{2}\omega \sin\theta & (V_{LI} > V_A) \\ S_{li'} = (\cot\theta - \frac{V_{LI}}{V_A} \sin\theta)\omega_0 - l - \frac{1}{2}\omega \sin\theta & (V_{LI} < V_A) \end{cases}$$

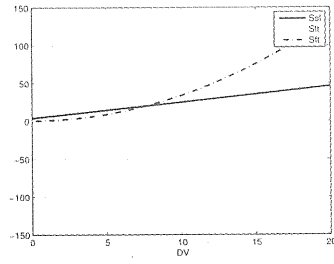
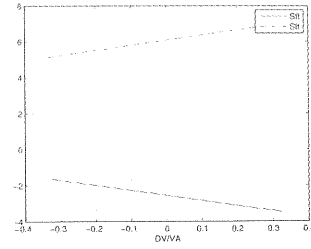
5.5 Model Simulate calculation analysis

We obtain the values of some parameters in the model, and through simulate calculation, we get the safe distance related to relative velocities.

$a = 1.5m/s^2$, $a_A = 7.8m/s^2$, $a_B = 6.8m/s^2$, $t_r = 0.8s$, $L_s = 0.5m$, $t_h = 2.1s$, $l = 4.2m$, $\omega = 1.8m$, $\theta = 4^\circ$, $\omega_0 = 3.75m$ and the models are as follows.

(1) Following model

The safety following distance is in the quadratic relationship of velocity and relative velocity. The higher velocity and relative velocity is the longer initial distance is required to ensure emergency stop of the leading car.

Figure 12: The value of S_{ft}, S_{sl}, S_{ft} Figure 13: The value of S_{ft} and S_{st}

(2) Overtaking model

The safe distance with following car in the same lane is in linear relationship with relative speed. The greater the difference, the longer safe distance is.

The safe distance with cars in the target lane is in the quadratic correlation with relative speed.

(3) Returning model

The safe distance with cars in the target lane is related to the proportion of relative velocity and absolute velocity. The bigger the proportion is, the larger distance is required for the following car in the distance.

5.6 Model testing

We assign the velocity as 80, 100, 120 m/s^2 .

Table 2: Get following distance. ($\Delta V = 0$)

v	L_2
80	124.788
100	174.699
120	232.148

Table 3: Overtaking distance ($\Delta V = 20m/s^2$)

S_{sl}	S_{lt}	S_{ft}
49.57	-129.07	137.596

$S_{lt} < 0$ means that it is safe to overtake with the leading car in the target lane.

The result has small changes when the velocity varies, and the returning distance model might has some weakness for we make the returning process idealistic, when in the reality, the returning time is 3 ~ 5s.

However, in the model it takes the car $t = \frac{\omega_0}{V_A \sin \theta} = 0.67s (V_A = 80)$.

6 Model 2 : Safe Driving Models in The Heavy Traffic

6.1 Statement

When there are many cars in the freeways, and the vehicles and the space is limited, there are more drivers want to change their lanes but would more likely to cause accidents. Based on the Model1 analysis,

Table 4: Returning distance ($\Delta V = 20m/s^2$)

v	S_{fl}	S_{lt}
80	-280.274	284.54
120	-280.296	284.56

the reasons to cause accident when following or overtaking is the same, so we can simply change the parameter values to measure the rule's performance in the heavy traffic. However, when the car returns to its initial lane in the heavy traffic, it might be difficult, since there may be little space for it to insert, and the car slows down to watch for a time to return, which may cause accidents with the car all in the passing lane, and we should discuss this possibility in the heavy traffic model.

6.2 Safe returning model

In order to insert to its initial lane, the car should reduce the vehicle to the initial speed. And the car following may hit.

To avoid hitting, the leading car and the following car should reach the same horizontal point at different lane as is shown in figure.

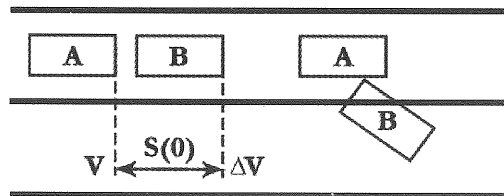


Figure 14: Returning model in high traffic

$$S(0) + S_B - \frac{1}{2}\omega \sin\theta \geq S_A$$

The following car A is in the uniform rectilinear motion, so the displacement is

$$S_A = V_A t_b$$

where t_b is the time of this movement

The leading car B is in the Deceleration uniform linear motion, the final velocity is the initial lane vehicle V_h , reducing acceleration is a'

$$t_b = \frac{V_B - V_h}{a'}$$

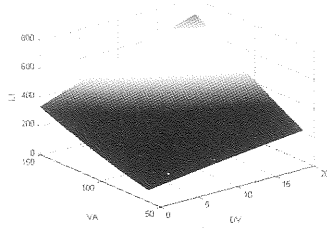
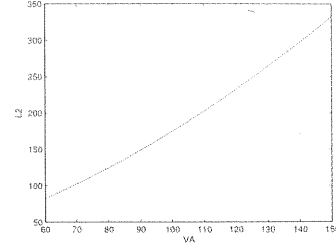
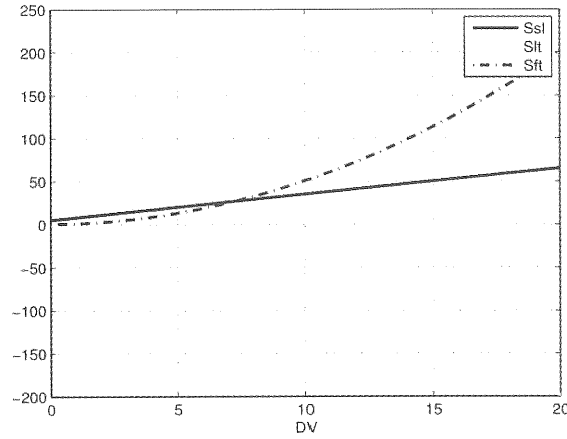
$$S_B = \frac{(V_B - V_h)^2}{2a'}$$

So we get the minimize safe distance

$$S_h = V_A \frac{V_B - V_h}{a'} - \frac{(V_B - V_h)^2}{2a'} + \frac{1}{2}\omega \sin\theta \quad (10)$$

If we assume that $V_A = V_B$

Then the safe distance is $S_h = \frac{V_B^2 - V_h^2}{2a'} + \frac{1}{2}\omega \sin\theta$

Figure 15: The value of L_1 Figure 16: The value of L_2 Figure 17: The value of S_{ft}, S_{sl}, S_{lt}

6.3 Model Simulate calculation analysis

Given the premise that the traffic is heavy, we re-assign the parameter in the model 1 as follows: $a = 1.0m/s^2, a_A = 7.8m/s^2, a_B = 6.8m/s^2, t_r = 0.8s, L_s = 0.5m, t_h = 3.0s, l = 4.2m, \omega = 1.8m, \theta = 4^\circ, \omega_0 = 3.75m, a = 2m/s^2$ and get the relationships as follows.

(1)Following model

$$\begin{cases} L_1 = V_A t_r + \frac{V_A^2}{2a_A} - \frac{V_B^2}{2a_B} + L_s \\ L_2 = V_A t_r + V_A^2 \left(\frac{1}{2a_A} - \frac{1}{2a_B} \right) + L_s \end{cases} \quad (11)$$

The following safe distance function is the same as the light traffic, but in heavy traffic, the velocity is lower, so the distance is shorter.

(2)Overtaking model

$$\begin{cases} S_{sl} = (V_A - V_B)t_h + \frac{1}{2}at_h^2 + l + \frac{1}{2}\omega \sin\theta \\ S_{lt} = -\frac{(V_{LT} - V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta \\ S_{ft} = \frac{(V_{FT} - V_A)^2}{2a} + l + \frac{1}{2}\omega \sin\theta \end{cases} \quad (12)$$

As the overtaking acceleration decreases and the time the car reach the hitting point increases, besides the relative velocity is smaller, so the distance would declines compared to the light situation.

(3)Returning model

$$\begin{cases} S_{fi} = \frac{V_{FL}}{V_A} \omega_0 \sin\theta - \omega_0 \cot\theta + l + \frac{1}{2} \omega \sin\theta \\ S_{li} = (\cot\theta - \frac{V_{Ll}}{V_A} \sin\theta) \omega_0 \\ S_h = \frac{V_h^2 - V^2}{2a'} + \frac{1}{2} \omega \sin\theta \end{cases} \quad (13)$$

As the velocity decreases much more than the relative velocity in the heavy traffic, $\frac{\Delta V}{V}$ increases, so the safe distance longer that is in heavy traffic, drivers must keep longer distance with front car.

7 Model 3 :Safety Evaluation system

After the preparing work of establishing the basis model 1 and 2, we now build the safety evaluation system and then study the tradeoffs between safety and traffic flow.

7.1 Safety Evaluation System

The possibility to cause the accident is closely related to the absolute speed, the relative speed to the cars around, and the distance between the aft-and-fore cars. We make D to represent the danger coefficient, which reflects the accident possibility. When the speed is low, the additional one unit of speed would not incur too many accident possibility, however, when the speed is relatively high, the one additional speed would incur much difference. Besides, the higher relative speed, the more likely to cause an accident, and danger happens when the cars are within the safety must be obviously high than the cars' distance is greater than safety distance. Additionally, the longer distance the less likely to incur an accident.

We can simply take our safety evaluation system as follows

$$D = c_1 V^2 + c_2 \Delta V + \frac{c_3}{S}$$

Where V represents the absolute speed of the car, and ΔV means the relative speed, and S means the distance between aft-and-fore cars.

$\Delta V > 0$ means the car is faster than the car ahead or slower than the car behind.

$\Delta V < 0$ means the car is slower than the car ahead or faster than the car behind.

The total danger include following danger, overtaking danger, and returning danger. And each danger is also related to the speed and distance.

(1) Following danger coefficient

$$D_f = c_{f1} V^2 + c_{f2} \Delta V + \frac{c_{f3}}{S}$$

For any given V and ΔV , we can calculate the safety following distance

$$L_1 = V_A t_r + \frac{V_A^2}{2a_A} - \frac{(V_A - \Delta V)^2}{2a_B} + L_s$$

When $S < L_1$, it is of great danger that may occur with the cars, otherwise, the possibility is pretty small. We get the approximate coefficient as shown below

$$\begin{cases} D_f = 0.016V^2 + 0.007\Delta V + \frac{0.056}{S} (S < L_1) \\ D_f = 0.002V^2 + 0.0003\Delta V + \frac{0.014}{S} (S > L_1) \end{cases}$$

In the light traffic, the velocity is high and ΔV has much elasticity, so $S < L_1$ is relatively big. But in this situation the distance is bigger, so we assume $P(S < L_1) = 0.04$ $P(S > L_1) = 0.96$

Then we get the following danger coefficient in light traffic

$$D_{fl} = 0.00256V^2 + 0.000568\Delta V + \frac{0.01568}{S} \quad (14)$$

In the heavy traffic $P(S < L_1) = 0.62$ $P(S > L_1) = 0.58$

Then we get the following danger coefficient in heavy traffic

$$D_{fh} = 0.01108V^2 + 0.004514\Delta V + \frac{0.04284}{S} \quad (15)$$

(2) overtaking danger coefficient

$$D_p = c_{p1}V^2 + c_{p2}\Delta V + \frac{c_{p3}}{S}$$

The safety distance is

$$\begin{cases} S_{sl} = \Delta V t_h + \frac{1}{2} a t_h^2 + l + \frac{1}{2} \omega \sin \theta \\ S_{lt} = -\frac{(V_{LT} - V_A)^2}{2a} + l + \frac{1}{2} \omega \sin \theta \\ S_{ft} = \frac{(V_{FT} - V_A)^2}{2a} + l + \frac{1}{2} \omega \sin \theta \end{cases}$$

Under the rule, the overtaking is more specification, cars would not turn right and turn left to make a mass, thus we can get a smaller coefficient. To make our system more simple and clear, we enact that $D_p^r = 0.87D_p^{nr}$, where D_p^r refers to passing coefficient under the rule, D_p^{nr} means the other.

$$\begin{cases} D_p^r = 0.021V^2 + 0.026\Delta V + \frac{0.0667}{S} (S < S_{pass}) \\ D_p^r = 0.006V^2 + 0.0007\Delta V + \frac{0.035}{S} (S > S_{pass}) \end{cases}$$

In the light traffic, ΔV has much elasticity, so safe distance is relatively high, so we assume

$$P(S < S_{pass}) = 0.17 P(S > S_{pass}) = 0.83$$

$$\text{In the heavy traffic, } P(S < S_{pass}) = 0.79 P(S > S_{pass}) = 0.21 .$$

So, we get the passing danger coefficient in light or heavy traffic with or without the rule.

with the rule

$$\begin{cases} D_{pl}^r = 0.00855V^2 + 0.005\Delta V + \frac{0.04}{S} \\ D_{ph}^r = 0.01785V^2 + 0.021\Delta V + \frac{0.06}{S} \end{cases} \quad (16)$$

without the rule

$$\begin{cases} D_{pl}^{nr} = 0.0098V^2 + 0.022\Delta V + \frac{0.046}{S} \\ D_{ph}^{nr} = 0.0205V^2 + 0.024\Delta V + \frac{0.069}{S} \end{cases} \quad (17)$$

(3) Returning danger coefficient

The danger coefficient is $D_b = c_{b1}V^2 + c_{b2}\Delta V + \frac{c_{b3}}{S}$

Based on the model one and two we get the safety distance.

$$\begin{cases} S_{fi} = \frac{V_{FL}}{V_A} \omega_0 \sin \theta - \omega_0 \cot \theta + l + \frac{1}{2} \omega \sin \theta \\ S_{li} = (\cot \theta - \frac{V_{LI}}{V_A} \sin \theta) \omega_0 \\ S_h = \frac{V_B^2 - V_h^2}{2a'} + \frac{1}{2} \omega \sin \theta \end{cases}$$

With the rule, the car must return its initial lane after passing, and it requirement may add many dangerous, and increase the danger.

$$D_b^r = 1.2D_b^{nr}$$

And calculate the possibilities.

$$\begin{cases} D_b^r = 0.015V^2 + 0.022\Delta V + \frac{0.0552}{S} (S < S_{return}) \\ D_b^r = 0.0034V^2 + 0.0005\Delta V + \frac{0.02}{S} (S > S_{return}) \end{cases}$$

In the light traffic, using the distance of S ft and S lt ,we calculate the

$$P(S < S_{return}) = 0.09 P(S > S_{return}) = 0.91$$

In the heavy traffic, using the three distances, we then calculate that

$$P(S < S_{return}) = 0.85 P(S > S_{return}) = 0.15$$

And we get the returning danger coefficient in different situations

with the rule

$$\begin{cases} D_{bl}^r = 0.00444V^2 + 0.00243\Delta V + \frac{0.023}{S} \\ D_{bh}^r = 0.01326V^2 + 0.0188\Delta V + \frac{0.05}{S} \end{cases} \quad (18)$$

without the rule

$$\begin{cases} D_{bl}^{nr} = 0.0037V^2 + 0.00202\Delta V + \frac{0.019}{S} \\ D_{bh}^{nr} = 0.011V^2 + 0.0156\Delta V + \frac{0.041}{S} \end{cases} \quad (19)$$

In all, we add the three coefficient to get the total rate in under and without rule and light and heavy traffic

In the light traffic

$$\begin{cases} D_l^r = 0.01555V^2 + 0.008\Delta V + \frac{0.079}{S} \\ D_l^{nr} = 0.01606V^2 + 0.0083\Delta V + \frac{0.081}{S} \end{cases} \quad (20)$$

In the heavy traffic

$$\begin{cases} D_h^r = 0.04219V^2 + 0.0044314\Delta V + \frac{0.015282}{S} \\ D_h^{nr} = 0.04258V^2 + 0.004414\Delta V + \frac{0.015284}{S} \end{cases} \quad (21)$$

7.2 conclusion

When the traffic is light, the rule reduced the danger coefficient by about 3%;And when the traffic is heavy, the rule reduced the danger coefficient by about 0.9%;so we can conclude that from the security perspective, the rule perform better when the traffic is light, but the improvement is not obvious.

7.3 model testing

We obtain some data on the website[3]. To see that the rule help saving lives and can turely reduce the accident rate. But we can not find the exact data of accident rate with and without the rule.

8 Model 4 : The traffic flow

8.1 statement

The traffic flow is related to the traffic density and traffic velocity. We now to establish the traffic flow model in light and heavy traffic, and using the model three to analysis the tradeoffs between traffic flow and safety to evaluate the performance of the role under different situations.

(1)The relationships between traffic flow, velocity, and traffic density

We assume that the on a freeway of length L , there are N cars running continuously, and their velocity is v .

The traffic density on the freeway L is $k = \frac{N}{L}$.

The time when these N cars pass a cross-section is $t = \frac{L}{v}$

The traffic flow is $Q = \frac{N}{t}$

So we get $Q = \frac{N}{t} = kv$

We can use the line model to reflect the relationship between velocity and density

$$v = a - bk$$

When $k = 0$,the cars on the road can reach the maximum speed, we note this rate to be v_f

i.e $k = 0, a = v_f$

When $k = K_m$ where K_m means the maximum density of the road, the road get congested and on vehicular moves.

i.e $k = K_m, v = 0$

So we get the linear model in the light traffic

$$v = v_f \left(1 - \frac{k}{K_m}\right)$$

and the traffic flow

$$Q = v_f \left(k - \frac{k^2}{K_m}\right) = K_m \left(v - \frac{v^2}{v_f}\right)$$

Make $\frac{dQ}{dv} = 0$, we know that when $v = \frac{1}{2}v_f, k = \frac{1}{2}K_m$, the traffic flow reach the maximum value, we name this value to be C, and we get $C = \frac{1}{4}v_f K_m$

Suppose that the interval of cars S. In the freeway L, the number of the car is $N = \frac{l}{s} - 1$

So the traffic flow is $Q = \frac{N}{t} = \frac{\frac{l}{s} - 1}{\frac{l}{v}} = \frac{v}{s} - \frac{v}{l}$.

If the free way is infinite, i.e. $l \rightarrow \infty, Q \rightarrow \frac{v}{s}$

(2) The effect of the role on improving the traffic flow

As is shown in the Model 1 testing, we let the distance between two cars remains to be minimum following distance.

Let the velocity in the lane 1,2,3 be v_1, v_2, v_3 , and to reach the largest traffic flow, we assume that each lane, car are at the safe following distance one by one, $\Delta V = 0$.

The danger coefficient would only be following danger coefficient and $S = L_2$.

The traffic flow is $Q = vk \approx \frac{v}{L_2} = \frac{v}{vt_r + v^2 \left(\frac{1}{2a_A} - \frac{1}{2a_B}\right) + L_s}$

However, when the rules carried out, people turns left to exceed cars, they would accelerate and approach the leading car, when the distance is appropriate, they turn. So the distance in the initial lane is smaller than safety following distance, longer than safety overtaking distance (S_{sl}), and the average velocity is higher than the no- overtaking situation.

Without loss of generality, we let $V' = \frac{V_1 + V_2}{2}, S' = \frac{S_{sl} + L_2}{2}$.

The traffic flow in this situation is

$$\begin{aligned} Q' &= \frac{V'}{S'} = \frac{V_1 + V_2}{S_{sl} + L_2} \\ S_{sl} &= (V_A - V_B)t_h + \frac{1}{2}at_h^2 + l + \frac{1}{2}\omega \sin\theta \end{aligned} \quad (22)$$

We assign the speed of each lanes as 80,100,120(m/s^2), and get the table as follows.

The traffic flow under the rule is higher than no-changing situation by 4.85%, which means that the

Table 5

ΔV	80	100	120	Total
Q	0.64086	0.57219	0.51670	1.72974
Q*	1.03236	0.78092	0	1.81328

rule promote better traffic flow but not effectively.

(3) Relationships between traffic flow and safety

Taking the safety evaluation system — model 2.

In the light traffic

$$\begin{cases} D_l^r = 0.0155V^2 + 0.008\Delta V + \frac{0.079}{S} \\ D_l^{rr} = 0.01606V^2 + 0.0083\Delta V + \frac{0.081}{S} \end{cases}$$

In the heavy traffic

$$\begin{cases} D_h^r = 0.04219V^2 + 0.0044314\Delta V + \frac{0.015282}{S} \\ D_h^{nr} = 0.04258V^2 + 0.004414\Delta V + \frac{0.015284}{S} \end{cases}$$

Let $\Delta V = 0$, $S = \frac{V}{Q}$

we get the connection with danger coefficient and traffic flow

$$\begin{cases} D_l^r = 0.0155V^2 + 0.079\frac{Q}{V} \\ D_l^{nr} = 0.01606V^2 + 0.081\frac{Q}{V} \end{cases} \quad (23)$$

$$\begin{cases} D_h^r = 0.04219V^2 + 0.015282\frac{Q}{V} \\ D_h^{nr} = 0.04258V^2 + 0.015284\frac{Q}{V} \end{cases} \quad (24)$$

8.2 Conclusion

On the conclusion of the effect on promote better traffic flow, we know that the rule helps to promote the traffic flow but not effectively, and only increase the flow by 4.85%.

As shown in the relationship of danger coefficient, velocity, and traffic flow above, for any given velocity, the danger coefficient is in the linear relationship with traffic flows, which is consistent with our common sense, the more car in the road the more likely to incur accidents, the higher the danger coefficient is.

As for the traffic velocity, for any fixed traffic flow, the faster the car drives the more accident might happen.

The rule has a better performance in the light traffic to decrease the danger coefficient by approximate 3% when the rule only a slight incline when the traffic is heavy by fewer than 1%. Although the rule did make a change the improvement is not very obvious.

And we can also know that when the traffic is heavy there is almost no change for the relationship with traffic flows and the danger coefficient, which means that when the traffic is busy, whether the rule is carried out or not, the dangerous coefficient is the same.

8.3 Model Testing

There is a little pity that we cannot find any traffic accident data in articles or on the website. So we cannot test our model with objective data.

9 Model Application

9.1 Statement

In this part we will use the model 1 ~ 4 to solve other questions in the task

- The role of under-or-over posted speed limits
- Emergency response and ride comfort
- The performance of The Fixed-Speed-in-The-Different-Lane rule
- Applications for Keep-Left countries
- The improvement made by the intelligent system

9.2 The role of under-or-over posted speed limits

When the speed limits are too low or too high, it will make a difference in the danger coefficient. But it does not change much. We still use the result in model 4 as follows:

$$\begin{cases} D_l^r = 0.0155V^2 + 0.079\frac{Q}{V} \\ D_l^{nr} = 0.01606V^2 + 0.081\frac{Q}{V} \end{cases}$$

$$\begin{cases} D_h^r = 0.04219V^2 + 0.015282\frac{Q}{V} \\ D_h^{nr} = 0.04258V^2 + 0.015284\frac{Q}{V} \end{cases}$$

In light traffic, the cars all at a relatively high speed level, and as we assume before, every driver would not to deliberately have a low speed, so the traffic velocity is close to the speed limit.

When the speed limits are set too high, every vehicle can run at a high speed, and the difference of the danger coefficient is more obvious, that is to say, when in the light traffic and over posted speed limits, it necessary to carry on the KREP Rule.

However, When the speed limits are too low, every vehicle must drive at the speed limits, the drivers can not accelerate to overtaking, because once the accelerate, they would against the rule. The difference of with or without rule is more subtle .

In heavy traffic, when the speed limits are set too high, it has no effect to the cars at all, and when the speed limit are too low, no matter whether the traffic velocity is slower than the limit, the speed limit would have no effect neither.

In conclusion, the role of under-or-over posted speed limits is to examine the necessary to put the KREP rule in to effect, and in the light traffic and over-posted road, there is a necessary to carry out the rule, it would make a relatively obvious improvement of safety. Under other conditions, there is no need.

9.3 Emergency response and ride comfort

When there is an accident in the freeway, such as a car breaks up or an car collision. Suppose that the car slows down and turn to the right lane. without the rule, the passing lane may occupied and the car following the accident car would slow down or passing in unregulated ways, the former choice would decrease the traffic flow, and the latter would increase the danger. However, with the rule, the car can immediately change its way to the right and overtake. Thus from the respective of emergency response, the rule works well.

As for ride comfort, different people have different driving habit. Some may prefer slow and careful driving, while others may enjoy exceeding. Without the rule, the overtaking pattern has a high uncertainty, the road is more dangerous, and passing lane may rarely capable for people to overtaking. Under the rule, passing lane is required for only passing, and people can turn to on this lane more easily. Thus better enjoy their own driving pattern.

9.4 The performance of The Fixed-Speed-in-The-Different- Lane rule

From the discussion above, we get the ideal that the KREP rule do improve the safety and traffic flow to a certain degree. However, the rule make little utilization of passing lane, and if we can make full use of all the lanes and ensure security, then we might get a better result.

Limited speed in different lanes required the cars in different lanes to keep different speed. But as the discussion in Model 4, making full use of three lanes would not help to reach a larger traffic flow.

On the basis of what we' ve done so far, we cannot find a better effective rule to promote traffic flow.

9.5 Applications for Keep-Left countries

In countries where driving automobiles on the left is the norm, they sit on the right and while passing they would change to their right. This overtaking pattern is symmetrical and like specular reflection of that in right norm country.

There might be a little difference about operation proficiency, But it doesn' t make sense. Thus there is no need to add more requirements, just change the turning orientation from right to left. The Keep-Left-Except-To-Pass rule would perform the same in these countries.

9.6 The improvement made by the intelligent system

The application of intelligent system would substantial raise the accuracy of judgments about when to overtake or what's the acceleration to overtake. Also under the reminders of the system, the drivers' sensitivity would develop, the car accident would have a big drop. The danger coefficient may decrease in a large scale.

$$D = c_1 V^2 + c_2 \Delta V + c_3 \frac{Q}{V}$$

So the safety level is high even in fast speed, in this way the KREP rule may have little work to increase the safety thus when intelligent system raises, the rule might have no need to carry out.

$$Q = K_m \left(v - \frac{v^2}{v_f} \right)$$

Besides, when the system installed in the car, the maximum density of the road K_m remains the same, the maximum speed v_f raises, the traffic flow may increase.

9.7 Conclusion

Based on the discussion above we reach the conclusion that the Keep-Right- Except-To-Pass rule has some effect on improve traffic flows and safety, but the effect is not obvious, only by about $3 \approx 5\%$. The rule would have a better performance when the speed limit is too high and the traffic flow is low. Although the rule is not the most effective rule, we may unable to find a better rule to replace it. Considering the function of the rule in emergency reflection and driving comfort, the rule has a good performance in general.

10 Strengths and Weaknesses

10.1 Model 1

Strengths Our model is based on the precise physical process analysis during the following, overtaking, and returning course, so the safety distance is both efficient and objective.

Weakness We getting averaged parameters based on many references, which may cause deviation compared to actual situation.

10.2 Model 2

Strengths We have taken into consideration the difference between light traffic and heavy traffic and revalue the parameters. Besides we combine discussions concerning the physical process with practical process, and add more restrictions.

Weakness The parameter valuing and idealistic movement may cause much deviation.

10.3 Model 3

Strengths We premeditate the danger coefficient in three processes separately, and non-linear relation is more consistent with the actual situation.

Weakness The coefficient lacks some accuracy, and we can improve the accuracy by large data analysis. And there are some other reasons to cause accident besides vehicle and distance.

10.4 Model 4

Strengths We quantify the relationship between safety and traffic flows, and has a good practical value.

Weakness The traffic flow is idealistic and ignore the effect of relative velocity.

11 References

- [1] BAI Wei, LI Cun-jun. 2013. Overtaking Model Based on Different Limiting Speed. *Journal of Transportation Systems Engineering and Information Technology* 13:63-68
- [2] XU Lun-hui, NI Yan-ming, LUO Qiang, HUANG Yan-guo. 2011. Lane-changing Model Based on Minimum Safety Distance. *Journal of Guangxi Normal University: Natural Science Edition* 29:1-6
- [3] KEEP RIGHT *USATM* . <http://www.keeprightusa.com/news.html>



珞珈数学 总二十卷

武汉大学珞珈数学研习会 诚意出品