第 3 章

泛函分析及变分法

前面介绍的数学知识是学习图像处理的基础，同时也是大学教育中工科数学的必修内容。如果是仅仅作为数字图像处理学习入门的先修课程基本已经足够。但数字图像处理技术是一门发展非常迅速的学科，一些新方法新理论不断涌现。因此，要想把数字图像处理作为一门学问深入研究，显然仅仅掌握前面的数学知识仍然远远不够。本章主要介绍更进一步的数学知识，这些内容主要围绕泛函分析和变法等主题展开。这些知识与前面的内容相比更加艰深和抽象。对于本章内容的学习，侧重点应该更多地放在有关概念的理解上，而非是深究每一条定理该如何证明。当然本部分内容仍然与前面的内容紧密相连，所以读者务必在牢固掌握之前内容的基础上再进行本章的学习。

3.1 勒贝格积分理论

前面介绍过积分的概念，彼时所讨论的积分首先是由黎曼(Riemann)严格定义的，因此之前所研究的积分通常称为黎曼积分，简称 R 积分。黎曼积分在数学、自然科学或者工程科学中具有非常重要的作用，正如前面所介绍的那样，诸如弧长、面积、体积、做功、通量等概念都可以借助黎曼积分表达。然而，随着现代数学和自然科学的发展，黎曼积分的缺陷也逐渐显现。这时勒贝格(Lebesgue)积分便应运而生了。在介绍勒贝格积分的概念之前，有必要介绍点集的勒贝格测度与可测函数的基本理论，这些内容是建立勒贝格积分的必要前提。

3.1.1 点集的勒贝格测度

点集的测度是区间长度概念的推广。设 E 为直线 R 上任意一个点集，用 mE 表示 E 的测度。如果 E 是直线上的区间(a,b)，或者 $E=[a,b]$、$(a,b]$、$[a,b)$，那么自然会想到可以定义该区间的长度 $b-a$ 为它的测度，即 $mE=b-a$。如果 E 是直线上的开集，那么可以根据

开集构造定理定义它的测度。

定义 设 G 为直线上的有界开集，定义 G 的测度为它的一切构成区间的长度之和。也就是说，若

$$G = \bigcup_{k} (\alpha_k, \beta_k)$$

其中，(α_k, β_k)是 G 的构成区间，则

$$mG = \sum_{k} (\beta_k - \alpha_k)$$

如果 G 的构成区间只有 n 个，那么上式右端是有限项(n 项)之和，即

$$mG = \sum_{k=1}^{n} (\beta_k - \alpha_k)$$

如果 G 的构成区间是可数多个，那么上式右端是一个无穷级数

$$mG = \sum_{k=1}^{+\infty} (\beta_k - \alpha_k)$$

由于 G 是有界开集，因此必然存在开区间(a,b)，使 $G\subset(a,b)$，所以对于任何有限的 n，有

$$\bigcup_{k=1}^{n} (\alpha_k, \beta_k) \subset (a,b)$$

从而有

$$\sum_{k=1}^{n} (\beta_k - \alpha_k) \leqslant b - a$$

令 $n\to+\infty$，得

$$mG = \sum_{k=1}^{+\infty} (\beta_k - \alpha_k) \leqslant b - a < +\infty$$

这表明无穷级数是收敛的，所以上述定义是有意义的。

定义 设 F 为直线上的有界闭集，$F\subset(a,b)$，则 $G=(a,b)-F$ 是有界开集，定义 F 的测度为

$$mF = (b,a) - mG$$

需要说明的是，由属于集 A 但不属于集 B 的元素的全体构成的集合称为 A 与 B 的差集，记为 $A-B$。可以证明，闭集 F 的测度 mF 与区间(a,b)的选择无关。

在直线上，除去开集和闭集之外，还存在大量的既不开也不闭的集合，例如有理数的点集与无理数的点集等。那么又该如何定义它们的测度呢？已知圆的面积既可用其外切正多边形的面积从外面逼近，也可以用其内接正多边形的面积从里面逼近，而且用这两种方法所得的结果也应相等。在做微积分时，也是用这种思想定义任意曲边梯形的面积的。不妨从这个角度定义一般的有界点集的测度。

定义 设 E 为直线上的任意一个有界点集，称所有包含 E 的开集测度的下确界为集 E 的外侧度，记作 m^*E，则

$$m^* E = \inf\{mG \mid G \supset E, G \text{ 为开集}\}$$

而把所有包含于 E 的闭集测度的上确界称为集 E 的内测度，记作 m_*E，则

$$m_* E = \sup\{mF \mid F \subset E, F \text{ 为闭集}\}$$

显然，$m_*E\leqslant m^*E$。事实上，由 $F\subset E\subset G$ 可知

$$m_* E = \sup\{mF \mid F \subset E, F \text{ 为闭集}\} \leqslant mG$$

从而有

$$m_* E \leqslant \inf\{mG \mid G \supset E, G \text{ 为开集}\} = m^* E$$

定义 设 E 为直线上的有界点集，若 $m^* E = m_* E$，则称 E 为勒贝格可测集，简称为 L 可测集，它的外侧度与内测度的共同值称为 E 的勒贝格测度，简称为 E 的 L 测度，记作 mE，则

$$mE = m^* E = m_* E$$

本节后续提及的可测集与测度均为 L 可测集与 L 测度。

直线上的区间，有界开集与有界闭集都是 L 可测的，而且它们的勒贝格测度与前面定义的测度相同。不仅如此，L 可测集还包含更广泛的集类。

点集的测度既然是区间长度概念的推广，那么它理应保持区间长度的一些基本属性。设 $X=[0,1]$ 为基本集，那么它的任意子区间 I 的长度 mI 显然具有下列基本性质。

(1) 非负性：$mI \geqslant 0$；

(2) 有限可加性：设 I_1 和 I_2 是区间 $[a,b]$ 的两个子区间，若 $I_1 \cap I_2 = \varnothing$，则 $m(I_1 \cup I_2) = mI_1 + mI_2$。

在直观上，有限可加性表达了“总量等于各分量之和”这个简单的公理，但这个公理的更完整表述应该是：设 $\{I_n\}$ 是 $X=[a,b]$ 中可列个子区间，$n=1,2,\cdots$，并且 $I_i \cap I_j = \varnothing$，$i \neq j$，则

$$m\left(\bigcup_{n=1}^{+\infty} I_n\right) = \sum_{n=1}^{+\infty} mI_n$$

称该性质为可列可加性(或完全可加性)。

区间长度 mI 还有很多其他的性质，但非负性与可列可加性是其中最基本最重要的定理，称为测度公理。由定义可知，点集的勒贝格测度 mE 是非负的，通过下面的定理可知点集的勒贝格测度同样具有可列可加性。

定理 设 $X=(a,b)$ 为基本集，E、E_1 和 E_2 为 X 的子集。

(1) 若 E 可测，则其补集 $\complement_X E$ 也可测；

(2) 若 E_1 和 E_2 可测，则 $E_1 \cup E_2$、$E_1 \cap E_2$、$E_1 - E_2$ 均可测，又若 $E_1 \cap E_2 = \varnothing$，则有

$$m(E_1 \cup E_2) = mE_1 + mE_2$$

定理

(1) 单调性：若 E_1 和 E_2 可测，且 $E_1 \subset E_2$，则 $mE_1 \leqslant mE_2$；

(2) 可列可加性：若 $\{E_k\}$ 是一个可测集列，$k=1,2,\cdots$，则

$$E = \bigcup_{k=1}^{+\infty} E_k$$

也可测；如果 E_k 两两互不相交，则

$$mE = \sum_{k=1}^{+\infty} mE_k$$

(3) 若 $\{E_k\}$ 是一个可测集列，$k=1,2,\cdots$，则

$$\bigcup_{k=1}^{+\infty} \complement_X E_k$$

也可测。

下面简单证明前两条性质。

证明

（1）由 $E_1 \subset E_2$ 可知 $E_2=(E_2-E_1)\cup E_1$，且 $(E_2-E_1)\cap E_1=\varnothing$，根据前面给出的定理可知 E_2-E_1 可测，且

$$mE_2 = m(E_2-E_1)+mE_1$$

或者

$$m(E_2-E_1) = mE_2 - mE_1$$

而 $m(E_2-E_1)\geqslant 0$，所以 $mE_1\leqslant mE_2$。

（2）假设 E_k 两两互不相交，则同样根据前面给出的定理可知，它们的有限并集也可测，并且

$$m\left(\bigcup_{k=1}^{n} E_k\right)=\sum_{k=1}^{n} mE_k$$

根据内测度的定义以及上确界的意义，对于任意的 $\varepsilon>0$，必然存在闭集

$$F\subset\bigcup_{k=1}^{n} E_k$$

使得

$$mF > m_*\left(\bigcup_{k=1}^{n} E_k\right)-\varepsilon = m\left(\bigcup_{k=1}^{n} E_k\right)-\varepsilon$$

又因为 $F\subset E$，故

$$m_* E \geqslant mF > \sum_{k=1}^{n} mE_k - \varepsilon$$

在上式中，先令 $\varepsilon\to 0$，再令 $n\to+\infty$，得

$$m_* E \geqslant \sum_{k=1}^{+\infty} mE_k$$

类似地，根据外侧度的定义，还可以得到

$$m^* E \leqslant \sum_{k=1}^{+\infty} mE_k$$

于是有

$$m^* E \leqslant \sum_{k=1}^{+\infty} mE_k \leqslant m_* E$$

但是 $m_* E\leqslant m^* E$，于是可得 $m_* E=m^* E$。因此集 E 可测，且

$$mE = \sum_{k=1}^{+\infty} mE_k$$

此外，如果 E_k 中有彼此相交的情况，由

$$\begin{aligned} E = \bigcup_{k=1}^{+\infty} E_k &= E_1 \cup (E_2-E_1)\cup(E_3-(E_1\cup E_2))\cup\cdots \\ &\cup(E_n-(E_1\cup E_2\cup\cdots\cup E_{n-1}))\cup\cdots \end{aligned}$$

即可将 E 分解为互不相交的可测集的并，于是根据上面已经证明的定理，即知 E 可测。

对开集与闭集进行至多可列次的交、并运算所得到的集，通常称为博雷尔(Borel)集。

凡博雷尔集都是勒贝格可测集。因此,勒贝格可测集类是相当广泛的集类,而且通常大多数集合都是勒贝格可测的。但是,也的确有勒贝格不可测集的例子存在,本书对此不做过深涉及。

例如,区间$[0,1]$中的有理点集是L可测的,并且它的测度为0。因为单点集是L可测的,并且测度为0,而有理点集可以看作是可列个单点集的并,所以根据可列可加性就得到上述结论。由此还可以知道,区间$[0,1]$中的无理点集的测度是1。用类似的方法还可以证明,任何可数集的测度都为0,但其逆命题不一定成立。测度为0的集也称为零测集,还可以证明零测集的任何子集都是零测集。

定理　设$X=(a,b)$是基本集,$\{E_k\}$是其中的可测集列。

(1) 若$\{E_k\}$是渐张的,即$E_1\subset E_2\subset\cdots\subset E_k\subset\cdots$,则

$$E=\bigcup_{k=1}^{+\infty} E_k$$

是可测集,并且

$$mE=\lim_{k\to+\infty} mE_k$$

(2) 若$\{E_k\}$是渐缩的,即$E_1\supset E_2\supset\cdots\supset E_k\supset\cdots$,则

$$E=\bigcap_{k=1}^{+\infty} E_k$$

是可测集,并且

$$mE=\lim_{k\to+\infty} mE_k$$

设E是直线上的一个无界点集,如果它与任何开区间的交是可测的,那么称E为可测集,并且定义E的测度为

$$mE=\lim_{\alpha\to+\infty} m[(-\alpha,\alpha)\cap E]$$

需要注意是,无界点集的测度可能是有限值,也可能是无穷大。利用这个定义可以将有界可测集的性质推广到无界可测集。而且仿照上述建立直线点集的测度理论的过程还可以建立平面点集甚至高维空间中点集的勒贝格测度理论。具体过程这里不再赘述。

3.1.2　可测函数及其性质

定义　设E为直线上的可测集(有界或无界),$f(x)$是定义在E上的实值函数。如果对于任何实数α,集合$E(f\geqslant\alpha)=\{x\mid f(x)\geqslant\alpha,x\in E\}$都是勒贝格可测的,那么称$f(x)$是$E$上的勒贝格可测函数,简称可测函数。

定理　函数$f(x)$在可测集E上可测的充要条件是:对于任何实数α和β,集合

$$E(\alpha\leqslant f<\beta)=\{x\mid \alpha\leqslant f(x)<\beta,x\in E\}$$

是勒贝格可测的。

证明　首先证明必要性。设$f(x)$为E上的可测函数,由于

$$E(\alpha\leqslant f<\beta)=E(f\geqslant\alpha)-E(f\geqslant\beta)$$

而$E(f\geqslant\alpha)$与$E(f\geqslant\beta)$都是可测集,所以$E(\alpha\leqslant f<\beta)$也是可测集。

再证明其充分性。假设对于任何实数α和β,$E(\alpha\leqslant f<\beta)$是可测集,而且可以证明

$$E(f \geqslant \alpha) = \bigcup_{n=1}^{\infty} E(\alpha \leqslant f < \alpha + n)$$

并且每个 $E(\alpha \leqslant f < \alpha + n)$ 都是可测集，于是 $E(f \geqslant \alpha)$ 也是可测集，所以 $f(x)$ 为 E 上的可测函数。

定理 函数 $f(x)$ 在可测集 E 上可测的充要条件是下列条件之一成立。

(1) $E(f > \alpha) = \{x \mid f(x) > \alpha, x \in E\}$ 是可测集；

(2) $E(f \leqslant \alpha) = \{x \mid f(x) \leqslant \alpha, x \in E\}$ 是可测集；

(3) $E(f < \alpha) = \{x \mid f(x) < \alpha, x \in E\}$ 是可测集；

(4) 对于直线上的任何开集 G，它的原象 $f^{-1}(x)$ 是可测集，其中，α 是任意实数。

例如，可以证明区间 $[0,1]$ 上的狄利克雷函数

$$D(x) = \begin{cases} 1, & x \text{ 为}[0,1] \text{ 中的有理数} \\ 0, & x \text{ 为}[0,1] \text{ 中的无理数} \end{cases}$$

是可测函数。

事实上，对于任何实数 α，由于

$$E(D \geqslant \alpha) = \begin{cases} \varnothing, & \alpha > 1 \\ [0,1] \text{ 中的有理点集}, & 0 < \alpha \leqslant 1 \\ [0,1] \text{ 中的实数集合}, & \alpha \leqslant 0 \end{cases}$$

是可测集，因此 $D(x)$ 是 $[0,1]$ 上的可测函数。

在集合论中，指示函数(indicator function)，或称特征函数(characteristic function)，是定义在集合 χ 上的函数，它用以表示集合 χ 中的一个元素是否属于 χ 的某一子集 A。如果函数值等于 1，那么表示被考查的元素都在 A 中。反之如果函数值为 0，则表示被考查的元素都在 χ 中，但不在 A 中。例如，设 E 为直线上的任意一点集，而且 E 是可测集，则集 E 的特征函数

$$\chi_E(x) = \begin{cases} 1, & x \in E \\ 0, & x \notin E \end{cases}$$

是 E 上的可测函数，证明方法与前面分析狄利克雷函数的可测性的方法类似。

定理 设 $f(x)$ 与 $g(x)$ 都是可测集 E 上的可测函数，那么 $kf(x)$、$f(x) \pm g(x)$、$f(x) \cdot g(x)$、$f(x)/g(x)$ 以及 $|f(x)|$ 都是 E 上的可测函数。其中，k 是常数，$g(x) \neq 0$。

3.1.3 勒贝格积分的定义

定义 设 $mE < +\infty$，$f(x)$ 是 E 上的有界可测函数，并且 $\alpha < f(x) < \beta$。任意取分点组 $\Delta = \{y_0, y_1, y_2, \cdots, y_n\}$ 分割区间 $[\alpha, \beta]$，则

$$\alpha = y_0 < y_1 < y_2 < \cdots < y_n = \beta$$

令

$$\lambda(\Delta) = \max_{1 \leqslant i \leqslant n}(y_i - y_{i-1}), \quad E_i = E(y_{i-1} \leqslant f < y_i)$$

任意取 $\xi_i \in [y_{i-1}, y_i)$ 做和式

$$\sigma(\Delta)=\sum_{i=1}^{n}\xi_i mE_i$$

如果不论$[\alpha,\beta]$如何分割，不论ξ_i如何选取，当$n\to+\infty$且$\sigma(\Delta)\to 0$时，和式$\sigma(\Delta)$的极限都存在并且相等，则称$f(x)$在E上是勒贝格可积的，而和式$\sigma(\Delta)$的极限值称为$f(x)$在E上的勒贝格积分，简称为L积分，记作

$$\int_E f(x)\mathrm{d}m=\lim_{\lambda(\Delta)\to 0}\sum_{i=1}^{n}\xi_i mE_i$$

下面定理给出了函数勒贝格可积的充分条件。

定理 设$mE<+\infty$，则E上的任何有界可测函数$f(x)$是勒贝格可积的。若$\alpha\leqslant f(x)\leqslant\beta$，则

$$amE\leqslant\int_E f(x)\mathrm{d}m\leqslant\beta mE$$

因此，还可以得到如下推论。

推论 设$mE<+\infty$，$f(x)$为E上的有界可测函数。

(1) 若$f(x)\geqslant 0$，则$\int_E f(x)\mathrm{d}m\geqslant 0$；若$f(x)\leqslant 0$，则$\int_E f(x)\mathrm{d}m\leqslant 0$；

(2) 若$f(x)=\alpha$，α为常数，则$\int_E f(x)\mathrm{d}m=amE$；

(3) 若$mE=0$，则$\int_E f(x)\mathrm{d}m=0$。

勒贝格积分保持了黎曼积分的一些基本性质。

定理 设$mE<+\infty$，$f(x)$与$g(x)$都是E上的有界可测函数。

(1) 线性性：设α与β为常数，则

$$\int_E[\alpha f(x)+\beta g(x)]\mathrm{d}m=\alpha\int_E f(x)\mathrm{d}m+\beta\int_E g(x)\mathrm{d}m$$

(2) 单调性：若几乎处处有$f(x)\leqslant g(x)$，则

$$\int_E f(x)\mathrm{d}m\leqslant\int_E g(x)\mathrm{d}m$$

(3) 有限可加性：设

$$E=\bigcup_{i=1}^{n}E_i$$

E_i均是可测集，并且$E_i\cap E_j=\varnothing$，$i\neq j$，则

$$\int_E f(x)\mathrm{d}m=\sum_{i=0}^{n}\int_{E_i}f(x)\mathrm{d}m$$

推论 若几乎处处有$f(x)=g(x)$，则

$$\int_E f(x)\mathrm{d}m=\int_E g(x)\mathrm{d}m$$

这是一个看似显然但又意味深长的结论。如果$f(x)$与$g(x)$是完全相等的，那么结论是显然成立的。但问题在于该推论的前提是“几乎处处”，也就是表明两个函数不相等的点是有限个。不妨设有$A=E(f\neq g)$，因为这样的点是有限个的，所以$mA=0$，再设有$B=E(f=g)$，则

$$\int_A f(x)\mathrm{d}m = \int_A g(x)\mathrm{d}m = 0$$

又因为

$$\int_B f(x)\mathrm{d}m = \int_B g(x)\mathrm{d}m$$

两式相加即得

$$\int_E f(x)\mathrm{d}m = \int_E g(x)\mathrm{d}m$$

这个结论其实说明，任意改变被积函数 $f(x)$ 在一个零测集上的值，并不影响函数的可积性以及积分的值，即使 $f(x)$ 在此零测集上无意义也未尝不可。因此，对等的两个函数在勒贝格积分理论中可以看成是同一个函数，这是 L 积分与 R 积分的一个显著区别。

接下来就是将勒贝格积分的概念推广到任意可测集 E（mE 可以取无穷值）上的无界可测函数的情形，也就是建立广义勒贝格积分的概念。

首先，设 $mE<+\infty$，$f(x)$ 是 E 上的无界可测函数。

假定 $f(x)$ 是 E 上的非负可测函数，即 $f(x)\geqslant 0$。令

$$[f(x)]_n = \begin{cases} f(x), & f(x)\leqslant n \\ n, & f(x) > n \end{cases}$$

由此对于每一个 $[f(x)]_n$ 而言，它都被控制在了 $[0,n]$ 之间，即 $\{[f(x)]_n\}$ 是 E 上的有界可测函数列，而前面的定理表明“当 $mE<+\infty$ 时，则 E 上的任何有界可测函数 $f(x)$ 是勒贝格可积的”。因此，对于每个 n，都有

$$\int_E [f(x)]_n \mathrm{d}m$$

都存在。又因为 $[f(x)]_1\leqslant[f(x)]_2\leqslant\cdots\leqslant[f(x)]_n\leqslant\cdots$，所以极限

$$\lim_{n\to+\infty}\int_E [f(x)]_n \mathrm{d}m$$

也存在（可以取有限或无限值）。如果极限值是有限的，则称 $f(x)$ 在 E 上勒贝格可积，并且积分值为

$$\int_E f(x)\mathrm{d}m = \lim_{n\to+\infty}\int_E [f(x)]_n \mathrm{d}m$$

如果极限值是无限的，则称 $f(x)$ 在 E 上有积分。

更进一步，假定 $f(x)$ 是 E 上的任意可测函数，那么定义

$$f_+(x) = \begin{cases} f(x), & f(x)\geqslant 0 \\ 0, & f(x) < 0 \end{cases}, \quad f_-(x) = \begin{cases} -f(x), & f(x)\leqslant 0 \\ 0, & f(x) > 0 \end{cases}$$

并分别称它们为 $f(x)$ 的正部和负部，则

$$f(x) = f_+(x) - f_-(x), \quad f_+(x)\geqslant 0, \quad f_-(x)\geqslant 0$$

如若下面两个积分不同时为 $+\infty$，则

$$\int_E f_+(x)\mathrm{d}m, \quad \int_E f_-(x)\mathrm{d}m$$

定义 $f(x)$ 在 E 上的勒贝格积分为

$$\int_E f(x)\mathrm{d}m = \int_E f_+(x)\mathrm{d}m - \int_E f_-(x)\mathrm{d}m$$

若上式右端的两个积分都是有限的，则称 $f(x)$ 在 E 上的勒贝格可积。否则，称 $f(x)$ 在 E 上有积分。

下面再考虑 E 为任意可测集，$f(x)$ 为 E 上的可测函数时的情况。若 $f(x)$ 是全直线 $R=(-\infty,+\infty)$ 上的可测函数，极限

$$\lim_{n\to+\infty}\int_{(-n,n)}|f(x)|\mathrm{d}m$$

存在且有限，则称 $f(x)$ 在实数轴 R 上勒贝格可积，并且定义 $f(x)$ 在 R 上的勒贝格积分为

$$\int_R f(x)\mathrm{d}m=\lim_{n\to+\infty}\int_{(-n,n)}f(x)\mathrm{d}m$$

如果 E 是 R 上的任意可测集（mE 可以为 $+\infty$），则 $f(x)$ 在 E 上的勒贝格积分定义为

$$\int_E f(x)\mathrm{d}m=\int_R f(x)\chi_E(x)\mathrm{d}m$$

其中，$\chi_E(x)$ 是 E 的特征函数。

至此，便完成了勒贝格积分的推广。

需要注意的是，存在勒贝格可积但黎曼不可积的函数。而且还可以证明，在有限区间 $[a,b]$ 上黎曼可积的函数必定勒贝格可积，并且积分值相等。所以，勒贝格可积函数类比黎曼可积函数类广泛得多。而且前面给出的关于有界可测函数勒贝格积分的性质定理及推论对任意可测集上的任意可测函数的勒贝格积分也成立。

定理　（绝对可积性）设 $f(x)$ 是可测集 E 上的可测函数，则 $f(x)$ 在 E 上勒贝格可积的充要条件是 $|f(x)|$ 在 E 上可积，并且有

$$\left|\int_E f(x)\mathrm{d}m\right|\leqslant\int_E|f(x)|\mathrm{d}m$$

定理　（绝对连续性）设 $f(x)$ 是可测集 E 上勒贝格可积，则对于任意的 $\varepsilon>0$，存在 $\delta>0$ 及子集 $e\subset E$，使得当 $me<\delta$ 时，

$$\left|\int_e f(x)\mathrm{d}m\right|<\varepsilon$$

证明　令 $g(x)=|f(x)|$，则由绝对可积性定理即知 $g(x)$ 可积。又由前面给出的

$$\int_E f(x)\mathrm{d}m=\lim_{n\to+\infty}\int_E[f(x)]_n\mathrm{d}m$$

可知对任意的 $\varepsilon>0$，存在自然数 N，使得下面的式子成立：

$$\int_E(g(x)-[g(x)]_N)\mathrm{d}m<\frac{\varepsilon}{2}$$

令 $\delta=\varepsilon/2N$，则当 $e\subset E$，且 $me<\delta$ 时，

$$\left|\int_e f(x)\mathrm{d}m\right|\leqslant\int_E g(x)\mathrm{d}m=\int_E(g(x)-[g(x)]_N)\mathrm{d}m+\int_E[g(x)]_N\mathrm{d}m$$

$$<\frac{\varepsilon}{2}+N\cdot me<\varepsilon$$

所以，定理得证。

定理　（可列可加性）设 $f(x)$ 是可测集 E 上勒贝格可积，且有

$$E=\bigcup_{k=1}^{+\infty}E_k$$

其中，E_k 为互不相交的可测集，则

$$\int_E f(x)\mathrm{d}m = \sum_{k=1}^{+\infty}\int_{E_k} f(x)\mathrm{d}m$$

证明 令

$$R_n = E - \bigcup_{k=1}^{n} E_k$$

则由测度的可列可加性可得当 $n \to +\infty$ 时

$$mR_n = mE - \sum_{k=1}^{n} mE_k \to 0$$

根据积分的有限可加性，有

$$\int_E f(x)\mathrm{d}m - \sum_{k=1}^{n}\int_{E_k} f(x)\mathrm{d}m = \int_{R_n} f(x)\mathrm{d}m$$

因此，再利用积分的绝对连续性即得

$$\lim_{n\to+\infty}\int_{R_n} f(x)\mathrm{d}m = 0$$

至此，便证明了可列可加性。

3.1.4 积分序列极限定理

勒贝格积分的另一个显著优点就是，积分与极限运算交换次序所要求的条件与黎曼积分相比要弱很多，因而使用起来比较灵便。本节介绍几个常用的极限定理。

定理 （勒贝格控制收敛定理）设 $mE < +\infty$，$\{f_n(x)\}$ 是 E 上的可测函数列，并且几乎处处有

$$\lim_{n\to+\infty} f_n(x) = f(x)$$

若存在一个 E 上的勒贝格积分函数 $g(x)$，使得在 E 上几乎处处有

$$|f_n(x)| \leqslant g(x), \quad n = 1,2,\cdots$$

则在 E 上勒贝格可积，并且

$$\int_E f(x)\mathrm{d}m = \lim_{n\to+\infty}\int_E f_n(x)\mathrm{d}m$$

推论 （勒贝格有界收敛定理）在与上述定理相同的条件下，若存在常数 M，使在 E 上几乎处处有

$$|f_n(x)| \leqslant M, \quad n = 1,2,\cdots$$

则 $f(x)$ 在 E 上勒贝格可积，并且

$$\int_E f(x)\mathrm{d}m = \lim_{n\to+\infty}\int_E f_n(x)\mathrm{d}m$$

显然，只要在前面的定理中取 $g(x)=M$ 即可得到此推论。

定理 设 $mE < +\infty$，$f(x)$ 与 $u_n(x)$ 都是 E 上的非负可测函数，$n=1,2,\cdots$，且几乎处处有

$$f(x) = \sum_{n=1}^{+\infty} u_n(x)$$

则

$$\int_E f(x)\mathrm{d}m = \sum_{n=1}^{+\infty}\int_E u_n(x)\mathrm{d}m$$

证明 由于 $f(x)$ 在 E 非负可测，故积分 $\int_E f(x)\mathrm{d}m$ 有意义。又因为对于任意的正整数 N，有

$$\int_E f(x)\mathrm{d}m \geqslant \int_E \sum_{n=1}^{N} u_n(x)\mathrm{d}m = \sum_{n=1}^{N}\int_E u_n(x)\mathrm{d}m$$

所以可得

$$\int_E f(x)\mathrm{d}m \geqslant \sum_{n=1}^{+\infty}\int_E u_n(x)\mathrm{d}m$$

如果上式右端等于无穷大则定理显然成立。现在假设

$$\sum_{n=1}^{+\infty}\int_E u_n(x)\mathrm{d}m < +\infty$$

令

$$S_N(x) = \sum_{n=1}^{N} u_n(x)$$

则对于任意正整数 k，必有

$$\lim_{N\to+\infty}[S_N(x)]_k = [f(x)]_k, \quad x \in E$$

事实上，设 $x_0 \in E$，若 $f(x_0) \leqslant k$，则更有 $S_N(x_0) \leqslant k$，按照 $[S_N(x)]_k$ 的定义，有

$$\lim_{N\to+\infty}[S_N(x_0)]_k = \lim_{N\to+\infty} S_N(x_0) = f(x_0) = [f(x_0)]_k$$

若 $f(x_0) > k$，则存在 N_0，使得 $N > N_0$ 时，$S_N(x_0) > k$，于是当 $N > N_0$ 时，$[S_N(x_0)]_k = k$，从而

$$\lim_{N\to+\infty}[S_N(x_0)]_k = k = [f(x_0)]_k$$

因为 $[S_N(x)]_k \leqslant k$，根据勒贝格有界收敛定理

$$\lim_{N\to+\infty}\int_E [S_N(x)]_k \mathrm{d}m = \int_E [f(x)]_k \mathrm{d}m$$

又因为

$$\int_E [S_N(x)]_k \mathrm{d}m \leqslant \int_E S_N(x)\mathrm{d}m$$

所以

$$\begin{aligned}\sum_{n=1}^{+\infty}\int_E u_n(x)\mathrm{d}m &= \lim_{N\to+\infty}\sum_{n=1}^{N}\int_E u_n(x)\mathrm{d}m = \lim_{N\to+\infty}\int_E S_N(x)\mathrm{d}m \\ &\geqslant \lim_{N\to+\infty}\int_E [S_N(x)]_k \mathrm{d}m = \int_E [f(x)]_k \mathrm{d}m\end{aligned}$$

令 $k \to +\infty$，得

$$\sum_{n=1}^{+\infty}\int_E u_n(x)\mathrm{d}m \geqslant \int_E f(x)\mathrm{d}m$$

综上即得下式，所以结论得证。

$$\int_E f(x)\mathrm{d}m = \sum_{n=1}^{+\infty}\int_E u_n(x)\mathrm{d}m$$

定理 设 $mE < +\infty$，$\{f_n(x)\}$ 是 E 上的非负可测函数列，并且

$$f_1(x) \leqslant f_2(x) \leqslant \cdots \leqslant f_n(x) \leqslant \cdots$$

而函数列$\{f_n(x)\}$逐点收敛于函数 $f(x)$，即

$$\lim_{n\to+\infty} f_n(x) = f(x), \quad \text{a. e.}$$ [1]

则

$$\int_E f(x)\,\mathrm{d}m = \lim_{n\to+\infty}\int_E f_n(x)\,\mathrm{d}m$$

该定理或许是最重要的勒贝格单调收敛定理，又称为莱维(Beppo Levi)定理。

证明 设 $f_0(x)=0$，令 $u_n(x)=f_n(x)-f_{n-1}(x)$，$n=1,2,\cdots$，则 $u_n(x)\geqslant 0$，且

$$\sum_{n=1}^{+\infty} u_n(x) = \sum_{n=1}^{+\infty}[f_n(x)-f_{n-1}(x)] = \lim_{n\to+\infty} f_n(x) = f(x), \quad \text{a. e.}$$

由前面刚刚证明过的定理可得

$$\begin{aligned}\int_E f(x)\,\mathrm{d}m &= \sum_{n=1}^{+\infty}\int_E u_n(x)\,\mathrm{d}m = \sum_{n=1}^{\infty}\int_E [f_n(x)-f_{n-1}(x)]\,\mathrm{d}m \\ &= \lim_{n\to+\infty}\sum_{k=1}^{n}\int_E [f_k(x)-f_{k-1}(x)]\,\mathrm{d}m \\ &= \lim_{n\to+\infty}\int_E \sum_{k=1}^{n}[f_k(x)-f_{k-1}(x)]\,\mathrm{d}m = \lim_{n\to+\infty}\int_E f_n(x)\,\mathrm{d}m\end{aligned}$$

定理得证。

3.2 泛函与抽象空间

牛顿说："把简单的问题看得复杂，可以发现新领域；把复杂的问题看得简单，可以发现新规律。"而从历史的角度看，一个学科的发展也是如此。随着学科的发展，最开始的一个主干方向会不断衍生出各自相对独立的分支，这也就是所谓"把简单的问题看得复杂"的过程。然而，一旦学科发展到一定程度之后，某些分支学科又开始被抽象综合起来，这也就是所谓"把复杂的问题看得简单"的过程。例如，在很长一段时间里，物理学家都把电和磁看成是两种独立的物理现象在研究，当学科研究积累到一定程度时，麦克斯韦就创立了电磁学，从而完成了物理学中的一次大综合。而在数学发展的历史中，几何与代数也曾经在很长的一段时间里是彼此独立的。直到笛卡儿引入了直角坐标系的概念之后，人们才开始建立了一种代数与几何之间的联系，也就是所谓的解析几何。泛函分析也是对以往许多数学问题或者领域进行高度抽象和综合的结果，其主要研究对象之一是抽象空间。其实在学习线性代数的过程中，人们已经建立了一种从矩阵到线性方程组之间的一种联系。而在泛函分析中，实数系、矩阵、多项式以及函数族这些看似关联不大的概念都可以抽成空间。由于泛函分析是一门比较晦涩抽象的学问，应该注意联系以往学习中比较熟悉的一些已知的、具体的概念，从而帮助理解那些全新的、抽象的概念。需要说明的是，本部分内容的重点在于有关

---

[1] a. e. 为 almost everywhere 的缩写。

定义或者概念的介绍，希望能够努力领会这些定义或者概念。

3.2.1　线性空间

线性空间是最基本的一种抽象空间。实数的全体 $\boldsymbol{R}_1$，二维平面向量的全体 $\boldsymbol{R}_2$，三维空间向量的全体 $\boldsymbol{R}_3$，以及所有次数不大于 n 的实系数多项式的全体等，都是线性空间的实例。

定义　设 E 为非空集合，如果对于 E 中任意两个元素 x 和 y，均对应于 E 中的一个元素，称为 x 与 y 之和，记为 $x+y$；对于 E 中任意一个元素 x 和任意一个实数 λ，均对应于 E 中的一个元素，称为 x 与 λ 的数乘，记为 λx；并且上述两种运算满足下列运算规律(x、y、z 为 E 中任意一个元素，λ 与 μ 为任意实数)。

(1) $x+y=y+x$；

(2) $x+(y+z)=(x+y)+z$；

(3) E 中存在唯一的零元素 θ(有时也记为 0)，它满足 $\theta+x=x$，并且对任意 x 均存在唯一的负元素 $-x\in E$，它满足 $x+(-x)=\theta$；

(4) $\lambda(\mu x)=(\lambda\mu)x$；

(5) $1x=x$，$0x=0$；

(6) $\lambda(x+y)=\lambda x+\lambda y$；

(7) $(\lambda+\mu)x=\lambda x+\mu x$。

称 E 是实线性空间。由于本章内容只考虑实数的情况，因此也可以将 E 简称为线性空间。从定义中可见，线性空间的核心思想就在于引入加法和乘法两种代数运算基础上同时保证封闭性。

根据上述定义可以证明下列结论成立。

(1) 所有次数不大于 n 的实系数多项式所构成的结合 P_n 是线性空间。

(2) 所有在区间$[a,b]$上连续的实函数所构成的集合 $C[a,b]$是线性空间。

(3) 所有在区间$[a,b]$上具有连续的 k 阶导数的实函数所构成的集合$C^k[a,b]$是线性空间。

与线性代数中类似，可以在线性空间中引入线性相关、线性无关以及基的概念。设 x_1，x_2，…，x_n 是线性空间 E 中的 n 个元素，其中 $n\geqslant 1$，如果存在不全为零的常数 λ_1，λ_2，…，λ_n，使得

$$\lambda_1 x_1+\lambda_2 x_2+\cdots+\lambda_n x_n=\theta$$

则称 x_1，x_2，…，x_n 是线性相关的。反之，若由 $\lambda_1 x_1+\lambda_2 x_2+\cdots+\lambda_n x_n=\theta$ 的成立可导出 $\lambda_1=\lambda_2=\cdots=\lambda_n=0$，则称 x_1，x_2，…，x_n 是线性无关的。回忆线性代数中关于线性相关的解释，向量组 x_1，x_2，…，x_n 线性相关的充分必要条件是其中至少有一个向量可以由其余 $n-1$ 个向量线性表示。尽管上述结论表明向量组中的线性相关性与其中某一个向量可用其他向量线性表示之间的联系。但是，它并没有断言究竟是哪一个向量可以由其他向量线性表示。关于这个问题可以用下面这个结论来回答。如果向量组 e_1，e_2，…，e_n，x 线性相关，而向量组 e_1，e_2，…，e_n 线性无关，那么向量 x 就可以由向量组 e_1，e_2，…，e_n 线性表示，而且表示形式唯一。

基于上述讨论，便可引出基的概念。如果线性空间 E 中存在 n 个线性无关的元素 e_1，

$e_2,\cdots,e_n$，使得 E 中任意一个元素 x 均可以表示成

$$x=\sum_{i=1}^{n}\xi_i e_i$$

那么，称$\{e_1,e_2,\cdots,e_n\}$为空间 E 的一组基。并且称 n 为空间 E 的维数，记为 $\dim E=n$。而 E 称为有限维（n 维）线性空间。不是有限维的线性空间称为无穷维线性空间。可见，P_n 是有限维的，而 $C[a,b]$和 $C^k[a,b]$都是无穷维的。

3.2.2 距离空间

尽管在线性空间上已经可以完成简单的线性运算，但这仍然不能满足需求。为了保证数学刻画的精确性，还必须引入距离的概念。本章是从极限开始讲起的，它是微积分的必备要素之一，而极限的概念显然也是基于距离上无限接近的角度描述的。

定义 设 X 是非空集合，若对于 X 中任意两个元素 x 和 y，均有一个实数与之对应，此实数记为 $d(x,y)$，满足：

(1) 非负性：$d(x,y)\geqslant 0$；而 $d(x,y)=0$ 的充分条件是 $x=y$；

(2) 对称性：$d(x,y)=d(y,x)$；

(3) 三角不等式：$d(x,y)\leqslant d(x,z)+d(z,y)$。

其中，z 是 X 中的任意元素。称 $d(x,y)$为 x 和 y 的距离，并称 X 是以 d 为距离的距离空间。

例如，通常 n 维向量空间 R_n，其中任意两个元素 $x=[\xi_i]_{i=1}^n$和 $y=[\eta_i]_{i=1}^n$的距离定义为

$$d_2(x,y)=\Big[\sum_{i=1}^{n}|\xi_i-\eta_i|^2\Big]^{\frac{1}{2}}$$

因此，R_n 就是以上式为距离的距离空间。同样，在 R_n 中还可以引入距离

$$d_1(x,y)=\sum_{i=1}^{n}|\xi_i-\eta_i|$$

或

$$d_p(x,y)=\Big[\sum_{i=1}^{n}|\xi_i-\eta_i|^p\Big]^{\frac{1}{p}},\quad p>1$$

或

$$d_{+\infty}(x,y)=\max_{1\leqslant i\leqslant n}|\xi_i-\eta_i|$$

可见，在同一个空间内可以通过不同方式引入距离。而且在同一空间中引入不同的距离后，就认为是得到了不同的距离空间。因此，常用符号(X,d)表示距离空间，如(R_n,d_1)，(R_n,d_p)等。

同样，还可以考虑定义在区间$[a,b]$上的连续函数的全体 $C[a,b]$，其中任意两个元素 $x(t)$与 $y(t)$间的距离可定义为

$$d(x,y)=\max_{a\leqslant t\leqslant b}|x(t)-y(t)|$$

现在思考以上述距离定义为基础的连续函数空间是否是一个距离空间。显然，定义中的前两个条件很容易满足。下面简单地证明。

$$\begin{aligned}|x(t)-y(t)|&\leqslant|x(t)-z(t)|+|z(t)-y(t)|\\&\leqslant\max_{a\leqslant t\leqslant b}|x(t)-z(t)|+\max_{a\leqslant t\leqslant b}|z(t)-y(t)|=d(x,z)+d(z,y)\end{aligned}$$

对所有的 $t\in[a,b]$成立，且上式右端与 t 无关，因此有

$$d(x,y)=\max_{a\leqslant t\leqslant b}|x(t)-y(t)|\leqslant d(x,z)+d(z,y)$$

在文章的最开始讨论过极限的有关内容。现在考虑如何在距离空间中定义极限。设 $\{x_n\}_{n=1}^{+\infty}$是距离空间(X,d)中的元素序列，如果(X,d)中的元素 x 满足

$$\lim_{n\to+\infty}d(x_n,x)=0$$

则称$\{x_n\}$是收敛序列，x 称为它的极限，记作 $x_n\to x$。

而且易得，如果序列$\{x_n\}$有极限，则极限是唯一的。实际上，如果 x 与 y 都是$\{x_n\}$的极限，则在式 $0\leqslant d(x,y)\leqslant d(x,x_n)+d(x_n,y)$中令 $n\to+\infty$，即可得出 $d(x,y)=0$，从而 $x=y$。

可以看出，在 n 维空间 R_n 中，不论距离是 d_1、d_2、$d_p(p>1)$或 $d_{+\infty}$，序列$\{x_n\}$的收敛都是指按(每个)坐标收敛。而连续函数空间 $C[a,b]$中序列$\{x_n(t)\}$的收敛就是前面讲过的一致收敛。

下面再引入球形邻域的概念：设 r 为某一正数，集合

$$S_r(x_0)=\{x\in X;\ d(x,x_0)<r\}$$

称为距离空间(X,d)中的球形邻域，或简称球。x_0 称为 $S_r(x_0)$的中心，r 称为半径。

基于球形邻域的概念就可以定义距离空间中的开集和闭集。

定义 设(X,d)为距离空间，M 是其中的一个子集。$x\in M$。若存在关于 x 的球形邻域 $S_r(x_0)$，它满足 $S_r(x_0)\subset M$，则称 x 是集合 M 的内点。如果集合 M 的元素都是 M 的内点，则称 M 为开集。

定义 设 $M\subset(X,d)$，$x_0\in X$，如果任意一个包含 x_0 的球 $S_r(x_0)$中总含有集合 M 的异于 x_0，则称 x_0 是集合 M 的聚点(或极限点)。

显然，x_0 是集合 M 的聚点的充分必要条件是 M 中存在异于 x_0 的序列$\{x_n\}_{n=1}^{+\infty}$，使得 $x_n\to x_0$。需要说明的是，聚点不一定属于集合 M。例如，在 R_1 中，设集合

$$M=\left\{1,\frac{1}{2},\frac{1}{3},\frac{1}{4}\cdots\right\}$$

则 0 是 M 的聚点，但 $0\notin M$。

定义 记集合 $M\subset(X,d)$的所有聚点所构成的结合为 M'，那么集合 $\overline{M}=M\cup M'$称为集合 M 的闭包。如果集合 M 满足 $M\supset\overline{M}$，称为 M 的闭集。

由此，在距离空间中，可以引入任意逼近的概念，即极限概念。一般来说，一个集合如果能够在其中确切地引入任意逼近的概念，就称为拓扑空间。而距离空间是一种最常用的拓扑空间。

3.2.3 赋范空间

每个实数或复数，都有相对应的绝对值或者模，每一个 n 维矢量，也都可以定义其长度。如果把长度的概念推广到一般抽象空间中的元素上，就可以得到范数这个概念。

定义 设 E 为线性空间，如果对于 E 中的任意个元素 x，都对应于一个实数，它记为 $\|x\|$，且满足：

(1) $\|x\|\geqslant0$，当且仅当 $x=\theta$ 时，$\|x\|=0$；

（2）$\| \lambda x \| = |\lambda| \| x \|$，$\lambda$ 为实数；

（3）$\| x+y \| \leqslant \| x \| + \| y \|$，$x,y \in E$。

则称 $\| x \|$ 为元素 x 的范数。E 称为按范数 $\| \cdot \|$ 的线性赋范空间。

例如，n 维矢量空间 R_n 中的元素 $x=[\xi_i]_{i=1}^n$ 的范数可以定义为如下形式，下面这个范数式也称为欧几里得范数，简称欧氏范数。

$$\| x \|_2 = \left\{ \sum_{i=1}^{n} | \xi_i |^2 \right\}^{\frac{1}{2}}$$

或者可以更一般地定义为（p 为任意不小于 1 的数）

$$\| x \|_p = \left\{ \sum_{i=1}^{n} | \xi_i |^p \right\}^{\frac{1}{p}}$$

还可以定义为

$$\| x \|_{+\infty} = \max_{1 \leqslant i \leqslant n} | \xi_i |$$

很容易证明上述三个定义式都满足范数定义中的三个条件。这里不作具体讨论，但是可以指出的是 $\| \cdot \|_p$ 满足条件（3）可以由闵可夫斯基不等式来证明。

闵可夫斯基（Minkowski）不等式 设 $a_i>0$，$b_i>0$（$i=1,2,\cdots,n$），$p>1$，则

$$\left[\sum_{i=1}^{n} (a_i+b_i)^p \right]^{\frac{1}{p}} \leqslant \left(\sum_{i=1}^{n} a_i^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p \right)^{\frac{1}{p}}$$

证明

$$\sum_{i=1}^{n} (a_i+b_i)^p = \sum_{i=1}^{n} a_i (a_i+b_i)^{p-1} + \sum_{i=1}^{n} b_i (a_i+b_i)^{p-1}$$

对上式右端两个和数分别应用赫尔德不等式，得到

$$\sum_{i=1}^{n} (a_i+b_i)^p \leqslant \left(\sum_{i=1}^{n} a_i^p \right)^{\frac{1}{p}} \left[\sum_{i=1}^{n} (a_i+b_i)^{(p-1)p'} \right]^{\frac{1}{p'}} + \left(\sum_{i=1}^{n} b_i^p \right)^{\frac{1}{p}} \left[\sum_{i=1}^{n} (a_i+b_i)^{(p-1)p'} \right]^{\frac{1}{p'}}$$

$$= \left[\left(\sum_{i=1}^{n} a_i^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p \right)^{\frac{1}{p}} \right] \left[\sum_{i=1}^{n} (a_i+b_i)^{(p-1)p'} \right]^{\frac{1}{p'}}$$

由于 $1/p+1/p'=1$，所以上述不等式可以改写为

$$\sum_{i=1}^{n} (a_i+b_i)^p \leqslant \left[\left(\sum_{i=1}^{n} a_i^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p \right)^{\frac{1}{p}} \right] \left[\sum_{i=1}^{n} (a_i+b_i)^p \right]^{\frac{1}{p'}}$$

然后，用最后一个因式作除式，等式两边同时做除法，即得到欲证明的不等式。

基于前面三个范数的定义，可知空间 R_n 是按范数式 $\| \cdot \|_2$、$\| \cdot \|_p$ 和 $\| \cdot \|_{+\infty}$ 的线性赋范空间。为了区别，通常把这三种线性赋范空分别记为 l_n^2、l_n^p 和 $l_n^{+\infty}$。由此可见，同一线性空间中可以引入多种范数。

连续函数空间 $C[a,b]$ 中元素 $x(t)$ 的范数可以定义为

$$\| x(t) \| = \max_{a \leqslant t \leqslant b} | x(t) |$$

因此，$C[a,b]$ 是按上述范数式的线性赋范空间，仍将它记为 $C[a,b]$。此外，还可以定义 $x(t)$ 的范数表达式为（$p \geqslant 1$）

$$\| x(t) \|_p = \left\{ \int_a^b | x(t) |^p \mathrm{d}t \right\}^{\frac{1}{p}}$$

它称为 p 范数。此时，所对应的线性赋范空间记为 $\widetilde{L}^p[a,b]$。$\| \cdot \|_p$ 可以成为范数的原因

同样是由前面讲过的闵可夫斯基不等式保证,但此时的闵可夫斯基不等式需将原来求和号改为积分符号,即

$$\left\{\int_a^b |x(t)+y(t)|^p \mathrm{d}t\right\}^{\frac{1}{p}} \leqslant \left\{\int_a^b |x(t)|^p \mathrm{d}t\right\}^{\frac{1}{p}} + \left\{\int_a^b |y(t)|^p \mathrm{d}t\right\}^{\frac{1}{p}}$$

可见,线性赋范空间同时也是距离空间,因为可以定义 $d(x,y)=\|x-y\|$。于是,线性赋范空间中的序列$\{x_n\}$收敛于 x 就是指 $\|x_n-x\|\to 0,(n\to+\infty)$。例如,空间 $C[a,b]$的收敛性是一致收敛,而$\widetilde{L}^p[a,b]$中序列 $x_n(t)$收敛于 $x(t)$是 p 幂平均收敛

$$\int_a^b |x_n(t)-x(t)|^p \mathrm{d}t \to 0$$

在线性赋范空间中的收敛性:$\|x_n-x\|\to 0$ 又称为依范数收敛。

设 X_1 和 X_2 都是线性赋范空间。记有次序的元素对$\{x_1,x_2\}$(其中 $x_1\in X_1,x_2\in X_2$)的全体所构成的集合为 $X_1\times X_2$。定义$\{x_1,x_2\}+\{y_1,y_2\}=\{x_1+y_1,x_2+y_2\}$,$\lambda\{x_1,x_2\}=\{\lambda x_1,\lambda x_2\}$及$\|\{x_1,x_2\}\|=\|x_1\|+\|x_2\|$,则 $X_1\times X_2$ 是线性赋范空间。它称为空间 X_1 和 X_2 的乘积空间。

接下来,介绍几条关于范数和依范数收敛的基本性质(这些性质在介绍极限时也有提及):

(1) 范数 $\|x\|$ 关于变元 x 是连续的,即当 $x_n\to x$ 时,$\|x_n\|\to\|x\|$;

(2) 若 $x_n\to x,y_n\to y$,则 $x_n+y_n\to x+y$;

(3) 若 $x_n\to x$,且数列 $a_n\to a$,则 $a_nx_n\to ax$;

(4) 收敛序列必为有界序列,即若 $x_n\to x$,则$\{\|x_n\|\}$是有界序列。

3.2.4　巴拿赫空间

定义　设 X 为线性赋范空间,$\{x_n\}_{n=1}^{+\infty}$是空间 X 中的无穷序列。如果对于任给的 $\varepsilon>0$,总存在自然数 N,使得当 $n>N$ 时,对于任意给定的自然 p,均有 $\|x_{n+p}-x_n\|<\varepsilon$,则称序列$\{x_n\}$是 X 中的基本序列(或称柯西序列)。

显然,X 中的任何收敛序列都是基本序列。为了证明该结论不妨设 $x_n\to x$,即任意给定的 $\varepsilon>0$,总存在自然数 N,使得当 $n>N$ 时,有 $\|x_n-x\|<\varepsilon/2$ 成立。于是对于任意的自然数 p,同时还有 $\|x_{n+p}-x\|<\varepsilon/2$。根据三角不等式,有 $\|x_{n+p}-x_n\|\leqslant\|x_{n+p}-x\|+\|x_n-x\|<\varepsilon$。然而,基本序列却不一定收敛。

定义　如果线性赋范空间 X 中的任何基本序列都收敛于属于 X 的元素,则称 X 为完备的线性赋范空间,或称为巴拿赫(Banach)空间。

下面考虑$\widetilde{L}^2[-1,1]$是不是巴拿赫空间。为此,不妨考查空间$\widetilde{L}^2[-1,1]$中的序列

$$x_n(t)=\begin{cases}-1, & t\in[-1,-1/n]\\ nt, & t\in[-1/n,1/n]\\ +1, & t\in[1/n,1]\end{cases}$$

显然 $x_n(t)$都是连续函数,且$|x_n(t)|\leqslant 1$,因此 $\|x_{n+p}(t)-x_n(t)\|\leqslant 2$,从而当 $n\to+\infty$ 时,则

$$\|x_{n+p}(t)-x_n(t)\|^2=\int_{-1}^{1}|x_{n+p}(t)-x_n(t)|^2\mathrm{d}t$$
$$=\int_{-\frac{1}{n}}^{\frac{1}{n}}|x_{n+p}(t)-x_n(t)|^2\mathrm{d}t\leqslant 4\int_{-\frac{1}{n}}^{\frac{1}{n}}\mathrm{d}t=\frac{8}{n}\to 0$$

这表明$\{x_n(t)\}$是空间$\widetilde{L}^2[-1,1]$中的基本序列。但同时当$n\to+\infty$时，$x_n(t)$的极限函数是间断函数。换言之，$x_n(t)$的极限函数不属于空间$\widetilde{L}^2[-1,1]$。因此，序列$\{x_n(t)\}$是空间$\widetilde{L}^2[-1,1]$中没有极限，或者说$\{x_n(t)\}$不是该空间中的收敛序列。既然线性赋范空间中的存在不收敛于该空间中元素的基本序列，那么空间$\widetilde{L}^2[-1,1]$就不是巴拿赫空间。一般地，$\widetilde{L}^p[a,b]$，其中$p\geqslant 1$，都不是巴拿赫空间。

但是空间$C[a,b]$是巴拿赫空间。为了说明这一点，不妨设$\{x_n(t)\}$是$C[a,b]$中的基本序列，即任意给定的$\varepsilon>0$，总存在自然数N，使得当$n>N$时，对于任意的自然数p均有

$$\max_{a\leqslant t\leqslant b}|x_{n+p}(t)-x_n(t)|<\varepsilon$$

根据前面介绍的函数序列一致收敛的柯西准则可知，$\{x_n(t)\}$是一致收敛序列。由于每个函数$x_n(t)$在$[a,b]$上都连续，因此它的极限函数在$[a,b]$上连续，即该极限函数属于空间$C[a,b]$。类似地，$C^k[a,b]$也是完备的。

关于有限维空间的完备性，有如下一般化结论：任意一个有限维线性赋范空间必为巴拿赫空间。而且由此还可以得到一个推论：任意一个线性赋范空间的有限维子空间都是闭子空间。

于是也得到了无穷维空间与有限维空间的一个重要差别：无穷维空间可以不完备，而有限维空间一定完备。

回忆本章前面关于函数项级数的内容，现在研究巴拿赫空间中的级数。

定理 巴拿赫空间中的级数$\sum\limits_{k=1}^{+\infty}x_k$收敛的充分必要条件是：对于任意给定的$\varepsilon>0$，总存在自然数$N$，当$n>N$时，对任何自然数$p$，均有

$$\left\|\sum_{k=n}^{n+p}x_k\right\|<\varepsilon$$

定义 若数值级数$\sum\limits_{k=1}^{+\infty}\|x_k\|$收敛，则称级数$\sum\limits_{k=1}^{+\infty}x_k$绝对收敛。

回想一下前面介绍过的魏尔斯特拉斯判别法（又称M判别法）：如果函数项级数$\sum\limits_{n=1}^{+\infty}u_n(x)$在区间$I$上满足条件，$\forall x\in I$，$|u_n(x)|\leqslant M_n(n=1,2,\cdots)$，并且正向级数$\sum\limits_{n=1}^{+\infty}M_n$收敛，则函数项级数$\sum\limits_{n=1}^{+\infty}u_n(x)$在区间$I$上一致收敛。

此处便得到了一个更加泛化的表述（只要把其中的$|u_n(x)|\leqslant M_n$替换成$\|f_n(x)\|\leqslant M_n$）：如果函数序列$\{f_n: X\to Y\}$的陪域[①]是一个巴拿赫空间$(Y,\|\cdot\|)$，$\forall x\in X$，存在$\|f_n(x)\|\leqslant M_n(n=1,2,\cdots)$，并且正向级数$\sum\limits_{n=1}^{+\infty}M_n$收敛，即$\sum\limits_{n=1}^{+\infty}M_n<+\infty$，则函数项级数

① 陪域又称上域或到达域，给定一个函数$f: A\to B$，集合B称为是f的陪域。一般来说，值域只是陪域的一个子集。

$\sum_{n=1}^{+\infty} f_n(x)$ 一致收敛。

证明 考虑级数的部分和序列 $s_n = \sum_{i=1}^{n} f_i$，并取任意 $p, n \in \mathbb{N}$，其中 $p \leqslant q$，那么对于任意 $x \in X$，有

$$\| s_q(x) - s_p(x) \| = \left\| \sum_{k=p+1}^{q} f_k(x) \right\| \leqslant \sum_{k=p+1}^{q} \| f_k(x) \| \leqslant \sum_{k=p+1}^{q} M_k$$

因为正向级数 $\sum_{n=1}^{+\infty} M_n$ 收敛，根据数项级数的柯西收敛定理，对于任意给定的 $\varepsilon > 0$，总存在自然数 N，使得当 $p > N$，以及 $x \in X$，有

$$\left\| \sum_{k=p+1}^{q} f_k(x) \right\| \leqslant \sum_{k=p+1}^{q} M_k < \varepsilon$$

而本节前面介绍过的定理也给出了巴拿赫空间中的级数收敛的充分必要条件，由此该定理得证。

从这个证明过程中，还得到了 M 判别法在巴拿赫空间中的一种更简单的表述：若级数 $\sum_{k=1}^{+\infty} x_k$ 是巴拿赫空间中的绝对收敛级数，则 $\sum_{k=1}^{+\infty} x_k$ 收敛。或表述为：当空间是巴拿赫空间时，若其中的级数绝对收敛，则该级数一定收敛。注意，该定理在描述时并没有强调一致收敛，这是因为一致收敛时针对函数项级数而言的，此处所得到的泛化结果是对巴拿赫空间中的元素来说的，即并不要求其中的元素一定是函数，所以也就不再强调一致收敛了。

这是因为，如果级数 $\sum_{k=1}^{+\infty} x_k$ 是巴拿赫空间中的绝对收敛，那么根据定义，就意味着数值级数 $\sum_{k=1}^{+\infty} \| x_k \|$ 收敛。同样根据数项级数的柯西收敛定理，可知对于任意给定的 $\varepsilon > 0$，总存在自然数 N，使得当 $n > N$，对于任何自然数 p，均有

$$\sum_{k=n+1}^{p} \| x_k \| < \varepsilon$$

又因为

$$\left\| \sum_{k=n+1}^{p} x_k \right\| \leqslant \sum_{k=n+1}^{p} \| x_k \|$$

即

$$\left\| \sum_{k=n+1}^{p} x_k \right\| < \varepsilon$$

由此定理得证。

最后，考虑上述定理的逆命题。

定理 如果线性赋范空间 X 中的任意绝对收敛级数都是收敛的，则 X 是巴拿赫空间。

证明 设 $\{x_n\}$ 是 X 中的基本序列，则显然它是有界序列 $\| x_n \| \leqslant c$，且可以选出某一个子序列 $\{x_{n_k}\}$，使得

$$\| x_{n_k} - x_{n_{k+1}} \| < \frac{1}{2^k}, \quad k \geqslant 2$$

于是，级数

$$x_{n_1}+(x_{n_2}-x_{n_1})+\cdots+(x_{n_k}-x_{n_{k+1}})+\cdots$$

绝对收敛。这是因为级数 $c+\sum_{k=1}^{+\infty}(1/2^k)$ 收敛。根据定理假设，上述级数收敛。设其部分和为 s_k，则 $s_k\to x\in X$。但是 $s_k=x_{n_k}$，于是序列 $\{x_n\}$ 有一子序列 $\{x_{n_k}\}$ 收敛于 $x\in X$（当 $k\to+\infty$ 时）。因此，对于任意给定的 $\varepsilon>0$，总存在自然数 N_1，当 $n_k>N_1$ 时，有

$$\| x_{n_k}-x\| < \frac{\varepsilon}{2}$$

又因 $\{x_n\}$ 是基本序列，因此存在自然数 N（不妨设 $N>N_1$），当 n 和 $n_k>N$ 时，有

$$\| x_n-x_{n_k}\| < \frac{\varepsilon}{2}$$

于是，

$$\| x_n-x\| \leqslant \| x_n-x_{n_k}\| + \| x_{n_k}-x\| < \varepsilon$$

这也就表明 $x_n\to x$，定理得证。

通过前面的介绍，应该知道 n 维空间中任意一个元素 x 均可表示为其中某一组基 $\{e_1,e_2,\cdots,e_n\}$ 的线性组合

$$x=\sum_{i=1}^{n}\xi_i \mathrm{e}_i$$

这组基的元素正好是 n 个。现在要在无穷维空间中讨论类似的问题。

以空间 $l_{+\infty}^p(p\geqslant 1)$ 为例，令 $e_k=[0,\cdots,0,1,0,\cdots]$，其第 k 个分量为 1，其余为 0，则显然 $l_{+\infty}^p$ 中任一元素 $x=[\xi_1,\xi_2,\cdots]$ 可唯一地表示为

$$x=\sum_{k=1}^{+\infty}\xi_k e_k$$

因此，元素组 $\{e_k\}_{k=1}^{+\infty}$ 可以作为空间 $l_{+\infty}^p$ 的一组基，但这组基的元素个数不是有限的。一个无穷集合，如果它的全部元素可以安装某种规则与自然数集合 $\{1,2,\cdots\}$ 建立一一对应关系，就称此无穷集合为可数集（或称可列集）。显然有限个可数集的和集仍然是可数集，甚至“可数”个可数集的和集也是可数集。所以，全部有理系数的多项式所构成的集合 P_0 是可数集。

显然，空间 $l_{+\infty}^p$ 的基 $\{e_1,e_2,\cdots\}$ 是一个可数集，这样的基称为可数基。由此，可以进行下面的讨论。

定义　设 M 是线性赋范空间 X 的子集，如果对于任意的元素 $x\in X$ 及正数 ε，均可在 M 中找到一个元素 m。使得 $\| x-m\| <\varepsilon$，则称 M 在 X 中稠密。

稠密性有下列等价定义：

（1）X 中的任一球形邻域内必含有 M 的点；

（2）任取 $x\in X$，则必有序列 $\{x_n\}\subset M$，使得 $x_n\to x$；

（3）M 在 X 中稠密的另一个充分必要条件是 $\overline{M}\supset X$。

定义　如果线性赋范空间 X 中存在可数的稠密子集，则称空间 X 是可分的。

例如，实数集 l_1^p 是可分的，因为所有有理数在其中是稠密的，而有理数集是可数集。进而，n 维空间 $l_n^p(1\leqslant p<+\infty)$ 也是可分的，因为坐标为有理数的点的全体构成其中的一个可数稠密子集。

定理　具有可数集的巴拿赫空间是可分的。

空间的完备性是实数域的基本属性的抽象和推广。完备的线性赋范空间具有许多类似实数域的优良性质，其关键是可以在其中顺利地进行极限运算。而且，不完备的空间也可以在一定的意义下进行完备化。连续函数空间$\widetilde{L}^p[a,b]$的完备化空间记作$L^p[a,b]$，称为p次勒贝格可积函数空间。鉴于本书后续内容中会对此稍有涉及，因此这里需要指出，当$p=1$时，空间$L^1[a,b]$的元素称为勒贝格可积函数。空间$L^1[a,b]$中的元素是“可积”的函数，其积分是关于上限的连续函数；另外，$L^1[a,b]$中两个函数$x(t)$和$y(t)$相等是指

$$\int_a^b |x(t)-y(t)|\,\mathrm{d}t = 0$$

此时，如果$x(t)$和$y(t)$仅在个别点（例如有限个点或者一个可数点集）上取值不等，并不影响上式成立，因此常称$x(t)$和$y(t)$是“几乎处处”相等的。关于空间$L^p[a,b]$，其元素是p次可积的

$$\int_a^b |x(t)|^p\mathrm{d}t < +\infty$$

例如，空间$L^2[a,b]$表示平方可积函数的全体。物理上，平方可积函数可以表示能量有限的信号。此外，有关系$L^1[a,b]\supset L^2[a,b]$，这由下式得知

$$\int_a^b |x(t)|\,\mathrm{d}t = \int_a^b |x(t)|\cdot 1\mathrm{d}t \leqslant \left[\int_a^b |x(t)|^2\mathrm{d}t\right]^{\frac{1}{2}}\left[\int_a^b 1^2\mathrm{d}t\right]^{\frac{1}{2}}$$
$$\leqslant \sqrt{b-a}\left[\int_a^b |x(t)|^2\mathrm{d}t\right]^{\frac{1}{2}}$$

其中用到了赫尔德不等式。一般地，如果$p'<p$，则$L^{p'}[a,b]\supset L^p[a,b]$。

另外还要指出，当$1\leqslant p<+\infty$时，空间$L^p[a,b]$是可分的。因为有理系数多项式集P_0在$\widetilde{L}^p[a,b]$中稠密，而$\widetilde{L}^p[a,b]$在$L^p[a,b]$中稠密。

3.2.5　内积空间

前面已经讨论过关于内积的话题，此处以公理化形式给出内积的定义。

定义　设E为实线性空间，如果对于E中任意两个元素x和y，均有一个实数与之对应，此实数记为(x,y)，且它满足：

(1) $(x,x)\geqslant 0$，当且仅当$x=\theta$时，$(x,x)=0$；

(2) $(x,y)=(y,x)$；

(3) $(\lambda x,y)=\lambda(x,y)$，($\lambda$为任意实数)；

(4) $(x+y,z)=(x,z)+(y,z)$，$(z\in E)$。

则称数(x,y)为x和y的内积，称E为实内积空间（或称欧几里得空间）。

例如，n维实矢量空间R_n中任意两个矢量$x=[\xi_1,\xi_2,\cdots,\xi_n]^{\mathrm{T}}$和$y=[\eta_1,\eta_2,\cdots,\xi_n]^{\mathrm{T}}$的内积定义为

$$(x,y) = \sum_{i=1}^{n}\xi_i\eta_i = x^{\mathrm{T}}y$$

可以验证，这个内积的定义是满足前面介绍过的4个条件的。此外，这种n维实矢量空间通常也称为n维欧几里得空间，记为E_n。

再如，空间 $L^2[a,b]$中的两个函数 $x(t)$和 $y(t)$的内积可以定义为

$$(x,y)=\int_a^b x(t)y(t)\mathrm{d}t$$

很容易验证，这种定义对于上述 4 个条件都是满足的。

定理 （柯西-施瓦茨不等式）内积空间中的任意两个元素 x 和 y 满足不等式

$$|(x,y)|\leqslant\sqrt{(x,x)}\sqrt{(y,y)}$$

当且仅当 $x=\lambda y$ 或 x、y 中只要一个为零时等号成立。

需要指出的是，内积空间是线性赋范空间。只要令

$$\|x\|=\sqrt{(x,x)}$$

此时三角不等式是成立的

$$\begin{aligned}\|x+y\|^2&=|(x+y,x+y)|\leqslant|(x,x)|+|(x,y)|+|(y,x)|+|(y,y)|\\&\leqslant(x,x)+2\sqrt{(x,x)}\sqrt{(y,y)}+(y,y)=(\|x\|+\|y\|)^2\end{aligned}$$

这个证明过程中用到了柯西-施瓦茨不等式。由于内积空间是线性赋范空间，因此线性赋范空间所具有的性质在内积空间中同样成立。

另外，显然柯西-施瓦茨不等式还可以写成

$$|(x,y)|\leqslant\|x\|\cdot\|y\|$$

3.2.6 希尔伯特空间

定义 在由内积所定义的范数意义下完备的内积空间称为希尔伯特(Hilbert)空间。

希尔伯特空间是一类性质非常好的线性赋范空间，在工程上有着非常广泛的应用，而且在希尔伯特空间中最佳逼近问题可以得到比较完满的解决。

定义 设 X 为某一距离空间。设 B 是 X 中的一个集合。$x\in X$ 且 $x\notin B$。现记 $d(x,B)$为点 x 到集合 B 的距离

$$d(x,B)=\inf_{y\in B}d(x,y)$$

如果集合 B 中存在元素$\tilde{x}$，使得

$$d(x,\tilde{x})=d(x,B)$$

则称元素$\tilde{x}\in B$ 是元素 $x\in X$ 在集合 B 中的最佳逼近元，或简称为最佳元。

下面给出关于希尔伯特空间 H 中闭凸子集的最佳元存在的唯一性定理。

定理 设 B 是 H 中的闭凸子集，$x\in H$ 且 $x\notin B$，则存在唯一的$\tilde{x}\in B$，使得

$$\|x-\tilde{x}\|=\inf_{y\in B}\|x-y\|$$

上述定理中提到了有关凸集的概念，其定义如下。

定义 设 M 是线性空间 E 中的一个集合，若对任意 $x,y\in M$ 及满足$\lambda+\mu=1$ 的 $\lambda\geqslant0$，$\mu\geqslant0$，均有 $\lambda x+\mu y\in M$，则称 M 是 E 中凸集。

下面对上述最佳元存在的唯一性定理进行证明。

证明 记

$$d=d(x,B)=\inf_{y\in B}\|x-y\|$$

根据下确界的定义，对于任意自然数 n，必存在 $y_n\in B$，使得

$$d \leqslant \| x - y_n \| < d + \frac{1}{n}$$

下面证明$\{y_n\}$是 H 中的基本序列。为此,对 $x-y_n$ 与 $x-y_m$ 应用平行四边形法则

$$2\| x - y_n \|^2 + 2\| x - y_m \|^2 = \| 2x - y_n - y_m \|^2 + \| y_n - y_m \|^2$$

由 B 的凸性可知$(y_n - y_m)/2 \in B$,从而

$$\| 2x - y_n - y_m \|^2 = 4\left\| x - \frac{y_n - y_m}{2} \right\|^2 \geqslant 4d^2$$

由此便有

$$\begin{aligned}\| y_n - y_m \|^2 &= 2\| x - y_n \|^2 + 2\| x - y_m \|^2 - \| 2x - y_n - y_m \|^2 \\ &\leqslant 2\left(d + \frac{1}{n}\right)^2 + 2\left(d + \frac{1}{m}\right)^2 - 4d^2 \\ &= \frac{4\mathrm{d}}{n} + \frac{4\mathrm{d}}{m} + \frac{2}{n^2} + \frac{2}{m^2}\end{aligned}$$

显然,当 $n \to +\infty, m \to +\infty$时,有 $0 \leqslant \| y_n - y_m \|^2 \leqslant 0$,于是即知$\{y_n\}$是基本序列。对于一个完备的空间而言,其中每个基本序列都收敛,故存在$\tilde{x} \in B$(因为 B 是闭集)使得 $y_n \to \tilde{x}$,即得

$$\| x - \tilde{x} \| = d$$

最后证明$\tilde{x}$的唯一性。设另有$\tilde{x}_1 \in B$ 满足 $\| x - \tilde{x}_1 \| = d$,再次应用平行四边形法则,即有

$$\begin{aligned}4d^2 &= 2\| x - \tilde{x} \|^2 + 2\| x - \tilde{x}_1 \|^2 = \| \tilde{x} - \tilde{x}_1 \|^2 + 4\left\| x - \frac{\tilde{x} + \tilde{x}_1}{2} \right\|^2 \\ &\geqslant \| \tilde{x} - \tilde{x}_1 \|^2 + 4d^2\end{aligned}$$

于是 $\| \tilde{x} - \tilde{x}_1 \|^2 \leqslant 0$,即 $\| \tilde{x} - \tilde{x}_1 \| = 0$,$\tilde{x} = \tilde{x}_1$。定理得证。

定义 如果 $x, y \in H$,且$(x,y)=0$,则称元素 x 与 y 是正交的,并记为 $x \perp y$。设 S 为 H 的子集,而元素 $x \in H$ 与 S 中任意一个元素都正交,则称元素 x 与集合 S 正交,记为 $x \perp S$。

显然,零元素 θ 与任何元素都正交。

定理 (勾股定理)若 $x \perp y$,则 $\| x + y \|^2 = \| x \|^2 + \| y \|^2$。

定理 (投影定理)设 L 是 H 的闭子空间,$x \in H$ 但 $x \notin L$,则$\tilde{l}$是在中的最佳元的充分必要条件是$(x - \tilde{l}) \perp L$。即对任意的 $l \in L$,均有$(x - \tilde{l}, l) = 0$。

证明 设$\tilde{l}$是最佳逼近元。则对于任意的实数 λ 和任意的元素 $l \in L$,有

$$\| x - \tilde{l} \|^2 \leqslant \| x - \tilde{l} + \lambda l \|^2$$

即$(x - \tilde{l}, x - \tilde{l}) \leqslant (x - \tilde{l} + \lambda l, x - \tilde{l} + \lambda l)$,也就是 $2\lambda(x - \tilde{l}, l) + \lambda^2 \| l \|^2 \geqslant 0$。不妨设 $l \neq \theta$,取

$$\lambda = -\frac{(x - \tilde{l}, l)}{\| l \|^2}$$

则原式变为

$$-\frac{(x - \tilde{l}, l)^2}{\| l \|^2} \geqslant 0$$

于是有$(x - \tilde{l}, l) = 0$,即必要性得证。

反之,设$(x - \tilde{l}, l) = 0$ 对任意的元素 $l \in L$ 都成立,则由勾股定理推得

$$\| x-l \|^2 = \| x-\tilde{l}+\tilde{l}-l \|^2 = \| x-\tilde{l} \|^2 + \| \tilde{l}-l \|^2 \geqslant \| x-\tilde{l} \|^2$$

即

$$\inf_{l \in L} \| x-l \|^2 = \| x-\tilde{l} \|^2$$

充分性得证。定理证明完毕。

在此基础上，若 L 是 H 中的有限维子空间，下面的步骤实现了求出 x 到 L 的距离 d 的具体表达式。

定义 设 $x_1,x_2,\cdots,x_k$ 是内积空间中的任意 k 个向量，这些向量的内积所组成的矩阵

$$\boldsymbol{G}(x_1,x_2,\cdots,x_k) = \begin{bmatrix} (x_1,x_1) & \cdots & (x_1,x_k) \\ \vdots & \ddots & \vdots \\ (x_k,x_1) & \cdots & (x_k,x_k) \end{bmatrix}$$

称为 k 个向量 $x_1,x_2,\cdots,x_k$ 的格拉姆(Gram)矩阵，k 阶格拉姆矩阵 $\boldsymbol{G}_k=\boldsymbol{G}(x_1,x_2,\cdots,x_k)$ 的行列式称为格拉姆行列式，通常用 $\Gamma(x_1,x_2,\cdots,x_k)$ 表示，或者记作 $|\boldsymbol{G}(x_1,x_2,\cdots,x_k)|$。

定理 设 $x_1,x_2,\cdots,x_n$ 是内积空间中的一组向量，则格拉姆矩阵 $\boldsymbol{G}_n$ 必定是半正定矩阵，而 $\boldsymbol{G}_n$ 是正定矩阵的充要条件是 $x_1,x_2,\cdots,x_n$ 线性无关。

证明 根据内积空间的定义，对任意的 n 维列矢量 $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots,\lambda_n]^{\mathrm{T}}$，有

$$\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{G}_n\boldsymbol{\lambda} = \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j(x_i,x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}(\lambda_i x_i,\lambda_j x_j) = \left(\sum_{i=1}^{n}\lambda_i x_i,\sum_{j=1}^{n}\lambda_j x_j\right) \geqslant 0$$

因此，n 阶格拉姆矩阵 $\boldsymbol{G}_n=\boldsymbol{G}(x_1,x_2,\cdots,x_n)$ 是半正定对称矩阵。

从上述证明过程中不难得到如下推论：内积空间中的任意 n 个向量 $x_1,x_2,\cdots,x_n$ 的格拉姆行列式恒为非负实数，即 $\Gamma(x_1,x_2,\cdots,x_n)\geqslant 0$，当且仅当 $x_1,x_2,\cdots,x_n$ 线性相关时，等号成立。如果 $x_1,x_2,\cdots,x_n$ 是线性无关的，必然可以推出 $\Gamma(x_1,x_2,\cdots,x_n)\neq 0$。因为如果 $\Gamma(x_1,x_2,\cdots,x_n)=0$，则表明 $\boldsymbol{G}_n$ 的列向量线性相关，即存在不全为零的 $\mu_1,\mu_2,\cdots,\mu_n$，使得

$$\sum_{j=1}^{n}\mu_j(x_i,x_j) = 0, \quad i = 1,2,\cdots,n$$

由此可知（因为内积空间也是线性赋范空间）

$$\left(\sum_{j=1}^{n}\mu_j x_j,\sum_{i=1}^{n}\mu_i x_i\right) = 0 \Rightarrow \left\| \sum_{j=1}^{n}\mu_j x_j \right\|^2 = 0$$

即

$$\sum_{j=1}^{n}\mu_j x_j = \theta$$

这显然与 $x_1,x_2,\cdots,x_n$ 线性无关矛盾。

定理 设 $\{x_1,x_2,\cdots,x_n\}$ 是 H 中的线性无关组。由它们生成的子空间记为 L，其维数为 n。H 中任意一点 x 到 L 的距离 d 为

$$d^2 = \frac{\Gamma(x_1,x_2,\cdots,x_n,x)}{\Gamma(x_1,x_2,\cdots,x_n)}$$

证明 设 x 的最佳元为 $\tilde{l}\in L$，则 $\tilde{l}$ 可表示为（相当于对 $\tilde{l}$ 做线性展开）

$$\tilde{l} = \sum_{j=1}^{n}\lambda_j x_j$$

根据投影定理

$$(x-\tilde{l}) \perp x_i, \quad i = 1,2,\cdots,n$$

这也等价于下列关于 λ_j 的线性方程组

$$\left(x-\sum_{j=1}^{n}\lambda_j x_j, x_i\right)=0 \Rightarrow (x,x_i)-\left(\sum_{j=1}^{n}\lambda_j x_j, x_i\right)=0$$

继续利用内积空间定义中的若干性质便会得到(如下方程组也称为最佳逼近问题的正规方程)

$$\sum_{j=1}^{n}\lambda_j(x_j,x_i)=(x,x_i),\quad i=1,2,\cdots,n$$

其系数行列式 $|G_n|\neq 0$,因此有唯一解。再由内积的定义及投影定理得到

$$d^2=\|x-\tilde{l}\|^2=(x-\tilde{l},x-\tilde{l})=(x,x)-(\tilde{l},x)-(x-\tilde{l},\tilde{l})$$

其中,$(x-\tilde{l},\tilde{l})=0$,即得

$$\sum_{j=1}^{n}\lambda_j(x_j,x)+d^2=(x,x)$$

现在把已经得到的两个和式联立起来,便得到下列关于 $n+1$ 个未知数 $\lambda_1,\lambda_2,\cdots,\lambda_n,d^2$ 的 $n+1$ 个方程:

$$\begin{cases}\lambda_1(x_1,x_1)+\lambda_2(x_2,x_1)+\cdots+\lambda_n(x_n,x_1)+0\cdot d^2=(x,x_1)\\ \lambda_1(x_1,x_2)+\lambda_2(x_2,x_2)+\cdots+\lambda_n(x_n,x_2)+0\cdot d^2=(x,x_2)\\ \qquad\vdots\\ \lambda_1(x_1,x_n)+\lambda_2(x_2,x_n)+\cdots+\lambda_n(x_n,x_n)+0\cdot d^2=(x,x_n)\\ \lambda_1(x_1,x)+\lambda_2(x_2,x)+\cdots+\lambda_n(x_n,x)+1\cdot d^2=(x,x)\end{cases}$$

由线性代数中的克莱姆法则即可求得 d^2 的表达式即为定理中所列出的形式,于是定理得证。

下面给出希尔伯特空间中当逼近集为闭凸集时的最佳元的特征定理。

定理 设 B 是希尔伯特空间 H 中的闭凸子集,$x\in H$,$x\notin B$,则下列命题等价:

(1) $\tilde{x}\in B$ 是 x 的最佳元,即对任意的 $b\in B$,均有 $\|x-\tilde{x}\|\leqslant\|x-b\|$;

(2) $\tilde{x}\in B$ 满足:对任意的 $b\in B$,均有$(x-\tilde{x},b-\tilde{x})\leqslant 0$;

(3) $\tilde{x}\in B$ 满足:对任意的 $b\in B$,均有$(x-b,\tilde{x}-b)\geqslant 0$。

最后,研究希尔伯特空间中的傅里叶级数展开。

定义 设$\{e_1,e_2,\cdots\}$是内积空间 H 中的一组元素,如果对任意的 $i\neq j$,均有$(e_i,e_j)=0$,则称$\{e_1,e_2,\cdots\}$是 H 中的正交系;如果每一个 e_i 的范数为1,则称之为规范正交系。

换言之,$\{e_1,e_2,\cdots\}$是 H 中规范正交系是指

$$(e_i,e_j)=\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$$

其中,δ_{ij} 是克罗内克(Kronecker)函数①。

内积空间中的正交系一定是线性无关组。

现在设$\{x_1,x_2,\cdots\}$是内积空间中的一组线性无关元素。下面讨论的方法实现了由该组

① 克罗内克函数 δ_{ij} 是一个二元函数,得名于数学家克罗内克。克罗内克函数的自变量(输入值)一般是两个整数,如果两者相等,则其输出值为1,否则为0。克罗内克函数的值一般简写为 δ_{ij}。注意,尽管克罗内克函数和狄拉克函数都使用 δ 作为符号,但是克罗内克 δ 函数带两个下标,而狄拉克 δ 函数则只有一个变量。

元素导出一组规范正交系$\{e_1,e_2,\cdots\}$使得e_n是$x_1,x_2,\cdots,x_n$的线性组合。

首先，取$e_1=x_1/\|x_1\|$，再令$u_2=x_2-(x_2,e_1)e_1$，$e_2=u_2/\|u_2\|$，那么显然$\{e_1,e_2\}$是规范正交的。以此类推，若已有规范正交组$\{e_1,e_2,\cdots,e_{n-1}\}$，就再令

$$u_n=x_n-\sum_{i=1}^{n-1}(x_n,e_i)e_i$$

及$e_n=u_n/\|u_n\|$，则显然e_n与$e_1,e_2,\cdots,e_{n-1}$都正交，从而$\{e_1,e_2,\cdots,e_n\}$是规范正交的。如此继续下去就可以得到规范正交系$\{e_1,e_2,\cdots\}$。

上述由线性无关组$\{x_1,x_2,\cdots\}$构造出规范正交系$\{e_1,e_2,\cdots\}$的方法通常称为格拉姆-施密特正交化方法。

例如，显然$\{1,t,t^2,\cdots,t^n,\cdots\}$是空间$L^2[-1,1]$中的线性无关组，但不是正交系。利用格拉姆-施密特方法，可以得到基于内积

$$(x,y)=\int_a^b x(t)y(t)\mathrm{d}t$$

的一个规范正交系。为此，令$x_0=1,x_1=t,\cdots,x_n=t^n,\cdots$，由于$\{x_n\}$是线性无关的，故可取$e_0=x_0/\|x_0\|=1/\sqrt{2}$，$u_1=x_1-(x_1,e_0)e_0=t$，进而取$e_1=u_1/\|u_1\|=\sqrt{3/2}\,t$，类似的有

$$e_2=\sqrt{\frac{5}{2}}\cdot\frac{1}{2}(3t^2-1),\cdots,e_n=\sqrt{\frac{2n+1}{2}}\cdot P_n(t)\quad n=1,2,\cdots$$

其中

$$P_n(t)=\frac{1}{2^n n!}\frac{d^n}{dt^n}(t^2-1)^n$$

称为n阶勒让德(Legendre)多项式。而$\{e_0(t),e_1(t),e_2(t),\cdots,e_n(t),\cdots\}$是$L^2[-1,1]$中的规范正交系。

之前已经推导出了H中任意一点x到其中有限维子空间L的距离公式，下面来考虑无限维子空间的情况。设$\{e_1,e_2,\cdots,e_n\}$是希尔伯特空间H中的规范正交系。现在根据前面讨论过的有限维子空间距离公式求H中任意一个元素x到由$\{e_1,e_2,\cdots,e_n\}$所生成的子空间L的距离d。

显然$\boldsymbol{G}(e_1,e_2,\cdots,e_n)$是单位矩阵，因此$|\boldsymbol{G}(e_1,e_2,\cdots,e_n)|=1$，而

$$|\boldsymbol{G}(e_1,e_2,\cdots,e_n,x)|=\begin{vmatrix}1 & & & (x,e_1)\\ & \ddots & & \vdots\\ & & 1 & (x,e_n)\\ (e_1,x) & \cdots & (e_n,x) & (x,x)\end{vmatrix}=\|x\|^2-\sum_{i=1}^{n}|c_i|^2$$

其中，$c_i=(x,e_i)$。上述化简计算过程中需用到一点线性代数的技巧。根据行列式的性质，把行列式中的某一行(或列)的元素都乘以同一个系数后，再加到另一行(或列)的对应元素上去，则行列式的值不变。于是不妨把第1行乘以$-(e_1,x)$后加到最后一行上，把第2行乘以$-(e_2,x)$后加到最后一行上……最终把原矩阵化成一个上三角矩阵。而上三角矩阵的行列式的值就等于主对角线上所有元素的乘积。基于上述计算结果便可得到

$$d^2=\frac{|\boldsymbol{G}(e_1,e_2,\cdots,e_n,x)|}{|\boldsymbol{G}(e_1,e_2,\cdots,e_n)|}=\|x\|^2-\sum_{i=1}^{n}|c_i|^2$$

而x在L中的最佳逼近元$\tilde{x}$(即元素x在n维子空间L中的投影)可以由一组正交基展开，即

$$\tilde{x}=\sum_{j=1}^{n}c_je_j$$

根据投影定理

$$(x-\tilde{x})\perp e_i,\quad i=1,2,\cdots,n$$

这也等价于下列关于 c_j 的线性方程组

$$\left(x-\sum_{j=1}^{n}c_je_j,e_i\right)=0\Rightarrow(x,e_i)-\left(\sum_{j=1}^{n}c_je_j,e_i\right)=0$$

于是有

$$\sum_{j=1}^{n}c_j(e_j,e_i)=(x,e_i),\quad i=1,2,\cdots,n$$

注意，$\{e_1,e_2,\cdots\}$是规范正交系，所以当 $j\neq i$ 时，$(e_j,e_i)=0$；当 $j=i$ 时，$(e_j,e_i)=1$，有

$$c_i=(x,e_i)$$

进而有

$$\tilde{x}=\sum_{i=1}^{n}c_ie_i=\sum_{i=1}^{n}(x,e_i)e_i$$

其中，系数 $c_i=(x,e_i)$称为元素 x 关于规范正交系$\{e_1,e_2,\cdots,e_n\}$的傅里叶系数。

定义　设内积空间 H 中有一个规范正交系$\{e_1,e_2,\cdots,e_n\}$，则数列$\{(x,e_i)\}(n=1,2,\cdots)$称为 x 关于规范正交系$\{e_1,e_2,\cdots,e_n\}$的傅里叶系数。

事实上，内积空间中的元素关于规范正交系的傅里叶系数就是微积分中的傅里叶系数概念的推广。泛函分析中的理论可以被用来验证或证明之前在微积分中给出的与傅里叶系数有关的许多结论。

定理　设 H 为无穷维希尔伯特空间，$\{e_1,e_2,\cdots\}$为 H 中的一组规范正交系，L 是由$\{e_1,e_2,\cdots\}$张成的一个子空间，即 $L=\mathrm{span}\{e_1,e_2,\cdots,e_n\}$。对于 H 中的任意一个元素 x，则

$$\tilde{x}=\sum_{i=1}^{n}(x,e_i)e_i$$

为元素 x 在 L 上的投影，且

$$\|\tilde{x}\|^2=\sum_{i=1}^{n}|(x,e_i)|^2$$

$$\|x-\tilde{x}\|^2=\|x\|^2-\|\tilde{x}\|^2$$

这个定理根据前面推导而得的结论（H 中任意一点 x 到其中无限维子空间 L 的距离公式）

$$d^2=\|x\|^2-\sum_{i=1}^{n}|c_i|^2$$

可以很容易证明，这里不再赘述。

贝塞尔(Bessel)不等式　设$\{e_n\}$为内积空间 H 中的标准正交系，令 $n\to+\infty$，则

$$\sum_{i=1}^{+\infty}|(x,e_i)|^2\leqslant\|x\|^2$$

这个不等式同样可以根据 H 中任意一点 x 到其中无限维子空间 L 的距离公式推得。当贝塞尔不等式取等号的时候，也就得到了前面曾经讨论过的**帕塞瓦尔等式**

$$\|x\|^2 = \sum_{i=1}^{+\infty} |(x,e_i)|^2$$

定义 设 H 为一内积空间，$\{e_n\}$ 为 H 中的一个标准正交系，若 $x\in H$，$x\perp e_n(n=1,2,\cdots)$，则必有 $x=\theta$。换言之，H 中不再存在非零元素，使它与所有的 e_n 正交，则称 $\{e_n\}$ 为 H 中的完全的标准正交系。

定理 设 $\{e_n\}$ 是希尔伯特空间 H 中的一个标准正交系，且闭子空间 $L=\overline{\mathrm{span}\{e_n|n=1,2,\cdots\}}$，则下述 4 个条件是等价的：

(1) $\{e_n\}$ 为 H 中的完全的标准正交系；

(2) $L=H$；

(3) 对任意 $x\in H$，帕塞瓦尔等式成立；

(4) 对任意 $x\in H$，则

$$x = \sum_{i=1}^{+\infty}(x,e_i)e_i$$

通常把上述定理中的最后一条称为 x 关于完全标准正交系的傅里叶级数（或 x 按 $\{e_n\}$ 展开的傅里叶级数）。该定理把微积分中的傅里叶展开推广到抽象的希尔伯特空间中，并揭示了完全标准正交系、帕塞瓦尔等式以及傅里叶展开之间的本质联系。

证明

(1)⇒(2)，设 $\{e_n\}$ 为完全的标准正交系。若 $L\neq H$，必存在非零元素 $x\in H-L$。由投影定理，存在 $x_0\in L$，$x_1\perp L$，使 $x=x_0+x_1$，因为 $x\neq x_0$，所以有 $x_1=x-x_0\neq\theta$，而 $x_1\perp e_n$，这与 $\{e_n\}$ 的完全性相矛盾。

(2)⇒(3)，若 $L=H$，$x\in H=L$，则 x 可表示为 $\{e_n\}$ 的线性组合的极限。对任意有限个 e_i，如 $e_1,e_2,\cdots,e_n$，有

$$x_n = \sum_{i=1}^{n}(x,e_i)e_i$$

根据前面介绍过的定理，可得

$$\|x-x_n\|^2 = \|x\|^2 - \|x_n\|^2 = \|x\|^2 - \sum_{i=1}^{n}|(x,e_i)|^2$$

另一方面，利用反证法，若由帕塞瓦尔等式不成立，以及贝塞尔不等式，则

$$\|x\|^2 - \sum_{i=1}^{+\infty}|(x,e_i)|^2 = a^2 > 0$$

因而对于任意 n，有

$$\left\|x-\sum_{i=1}^{n}(x,e_i)e_i\right\|^2 = \|x\|^2 - \sum_{i=1}^{+\infty}|(x,e_i)|^2 \geqslant a^2$$

即

$$x \neq \sum_{i=1}^{n}(x,e_i)e_i$$

这与假设条件矛盾，因此命题得证。

(3)⇒(4)，对任意 $x\in H$，帕塞瓦尔等式成立，则由前面介绍过的定理得出

$$\|x-x_n\|^2 = \left\|x-\sum_{i=1}^{n}(x,e_i)e_i\right\|^2 = \|x\|^2 - \sum_{i=1}^{n}|(x,e_i)|^2$$

根据帕塞瓦尔等式，可知

$$\lim_{n\to+\infty}\left\|x-\sum_{i=1}^{n}(x,e_i)e_i\right\|^2=\lim_{n\to+\infty}\left[\|x\|^2-\sum_{i=1}^{n}|(x,e_i)|^2\right]=0$$

即证明了

$$x=\sum_{i=1}^{+\infty}(x,e_i)e_i$$

(4)⇒(1)，对任意 $x\in H$，有

$$x=\sum_{i=1}^{+\infty}(x,e_i)e_i$$

并设 $x\perp e_i(i=1,2,\cdots)$，显然有 $x=\theta$，因此$\{e_n\}$为 H 中的完全的标准正交系。

3.2.7　索伯列夫空间

把区间$[a,b]$上一阶连续可微函数的全体所构成的集合记为$\widetilde{H}^1[a,b]$。显然，在通常的函数加法，乘法意义下，$\widetilde{H}^1[a,b]$是线性空间。对于任意的 $u(t),v(t)\in\widetilde{H}^1[a,b]$，定义其内积为

$$(u,v)=\int_a^b u(t)v(t)\mathrm{d}t+\int_a^b u'(t)v'(t)\mathrm{d}t$$

则不难验证它满足关于内积的四条公理，因而$\widetilde{H}^1[a,b]$是内积空间，相应的范数为

$$\|u\|=\left\{\int_a^b[u(t)]^2\mathrm{d}t+\int_a^b[u'(t)]^2\mathrm{d}t\right\}^{1/2}$$

空间$\widetilde{H}^1[a,b]$在上述范数意义下的完备化空间记为 $H^1(a,b)$，它称为索伯列夫(Sobolev)空间。

设序列$\{u_n(t)\}\subset\widetilde{H}^1[a,b]$是上述范数意义下的基本序列，即当 $n,m\to+\infty$时

$$\|u_n-u_m\|^2=\int_a^b[u_n(t)-u_m(t)]^2\mathrm{d}t+\int_a^b[u'_n(t)-u'_m(t)]^2\mathrm{d}t\to 0$$

如果$\{u_n(t)\}$和$\{\hat{u}_n(t)\}$是$\widetilde{H}^1[a,b]$中的两个基本列，且满足当 $n\to+\infty$时，$\|u_n(t)-\hat{u}_n(t)\|\to 0$，则认为它们属于同一类。上述条件也等价于

$$\int_a^b[u_n(t)-u_m(t)]^2\mathrm{d}t\to 0$$

$$\int_a^b[u'_n(t)-u'_m(t)]^2\mathrm{d}t\to 0$$

根据空间 $L^2[a,b]$的完备性，存在 $u(t)\in L^2[a,b]$及 $w(t)\in L^2[a,b]$，使得当 $n\to+\infty$时，在 L^2 范数的意义下，$u_n(t)\to u(t)$，$u'_n(t)\to w(t)$。对如此所确定的函数 $u(t)$和 $w(t)$，称 $w(t)$是 $u(t)$在索伯列夫意义下的广义导数，并记成 $u'(t)=w(t)$。显然，如果 $u(t),v(t)\in H^1(a,b)$，则 $au(t)+bv(t)\in H^1(a,b)$且$(au+bv)'(t)=au'(t)+bv'(t)$；而常数的广义导数为零。

由广义导数的定义可以看出，这种导数不是关于函数的个别点处局部性质反映，因为它是通过在整个区间上积分的极限确定的，而积分是一种关于函数的整体性质的概念。但也应该指出，广义导数其实是对通常意义下导数概念的推广。如果函数本身是通常意义下可

微的，则其导函数与广义导数是一致的。

类似地，记$\widetilde{H}^2[a,b]$为$[a,b]$上二阶连续可微函数的全体，其内积定义为

$$(u,v)=\int_a^b uv\,\mathrm{d}t+\int_a^b u'v'\,\mathrm{d}t+\int_a^b u''v''\,\mathrm{d}t$$

则$\widetilde{H}^2[a,b]$的完备化空间相应地记为$H^2[a,b]$，也称为索伯列夫空间，空间$H^2[a,b]$中的元素$u(t)$具有一阶和二阶广义导数，且$u'(t),u''(t)\in L^2[a,b]$，即它们都是勒贝格平方可积的。因此，可定义一般的索伯列夫空间$H^k[a,b]$。而且上述这些定义还可以推广到多维的情形，这里不再深究，有兴趣的读者可以参阅泛函分析方面的资料。

3.3 从泛函到变分法

作为数学分析的一个分支，变分法（calculus of variations）在物理学、经济学以及信息技术等诸多领域都有着广泛而重要的应用。变分法是研究依赖于某些未知函数的积分型泛函极值的普遍方法。换句话说，求泛函极值的方法就是变分法。

3.3.1 理解泛函的概念

变分法是现代泛函分析理论的重要组成部分，但变分法却是先于泛函理论建立的。因此，即使不过深地涉及泛函分析的相关内容，也可展开对变分法的学习。而在前面介绍的有关抽象空间的内容上来讨论泛函的概念将是非常方便的。

定义 设X和Y是两个给定的线性赋范空间，并有集合$\mathcal{D}\subset X$。若对于$\mathcal{D}$中的每一个元素x，均对应于Y中的一个确定的元素y，就说这种对应关系确定了一个算子。算子通常用大写字母$T,A,\cdots$表示，记为$y=Tx$或$y=T(x)$。y称为x的象，x称为y的原象。集合$\mathcal{D}$称为算子T的定义域，常记为$\mathcal{D}(T)$；而集合$\mathcal{R}(T)=\{y\in Y;\ y=Tx,x\in\mathcal{D}(T)\}$称为算子$T$的值域。对于算子$T$，常用下述记号$T$：$X\mapsto Y$，读作“$T$是由$X$到$Y$的算子”。但应注意这种表示方法并不意味着$\mathcal{D}(T)=X$及$\mathcal{R}(T)=Y$。

当X和Y都是实数域时，T就是微积分中的函数。因此，算子是函数概念的推广，但是算子这个概念要比函数更抽象，也更复杂。

设X为实（或复）线性赋范空间，则由X到实（或复）数域的算子称为泛函。例如，若$x(t)$是任意一个可积函数$x(t)\in L^2[a,b]$，则其积分

$$f(x)=\int_a^b x(t)\,\mathrm{d}t$$

就是一个定义在$L^1[a,b]$上的泛函，而且是线性的

$$f(\alpha x+\beta y)=\alpha\int_a^b x(t)\,\mathrm{d}t+\beta\int_a^b y(t)\,\mathrm{d}t=\alpha f(x)+\beta f(x)$$

还是有界的

$$|f(x)|\leqslant\int_a^b|x(t)|\,\mathrm{d}t=\|x\|$$

需要说明的是，此处所讨论的仅限于实数范围内的泛函。

如果把上述泛函定义中的线性赋范空间局限于函数空间，那么也可以从另外一个角度来理解此处所要讨论的泛函。

把具有某种共同性质的函数构成的集合称为函数类，记作 F。对于函数类 F 中的每一个函数 $y(x)$，在$\mathbb{R}$中变量$\mathcal{J}$都有一个确定的数值按照一定的规律与之相对应，则$\mathcal{J}$称为函数 $y(x)$的泛函，记作$\mathcal{J}=\mathcal{J}[y(x)]$或者$\mathcal{J}=\mathcal{J}[y]$。函数 $y(x)$称为泛函$\mathcal{J}$的宗量。函数类 F 称为泛函$\mathcal{J}$的定义域。可以这样理解，泛函是以函数类为定义域的实值函数。为了与普通函数相区别，泛函所依赖的函数用方括号括起来。

由泛函的定义可知，泛函的值是数，其自变量是函数，而函数的值与其自变量都是数，所以泛函是变量与函数的对应关系，它是一种广义上的函数。而函数是变量与变量的对应关系，这是泛函与函数的基本区别。此外还应当意识到，泛函的值既不取决于自变量 x 的某个值，也不取决于函数 $y(x)$的某个值，而是取决于函数类 F 中 y 与 x 的函数关系。

由于一元函数在几何上是由曲线来表示的，因此它的泛函也可以称为是曲线函数。类似地，二元函数在几何上的表现形式通常都是曲面，因此它的泛函也可以称为是曲面函数。如果 x 是多维域$(x_1,x_2,\cdots,x_n)$上的变量时，以上定义的泛函也适用。此时，泛函记为$\mathcal{J}=\mathcal{J}[u(x_1,x_2,\cdots,x_n)]$。同时也可以定义依赖于多个未知函数的泛函，记为$\mathcal{J}=\mathcal{J}[y_1(x),y_2(x),\cdots,y_m(x)]$。其中，$y_1(x),y_2(x),\cdots,y_m(x)$都是独立变化的。还有泛函记为$\mathcal{J}=\mathcal{J}[y_1(x_1,x_2,\cdots,x_n),y_2(x_1,x_2,\cdots,x_n),\cdots,y_m(x_1,x_2,\cdots,x_n)]$，同样要求 $y_1(x_1,x_2,\cdots,x_n),y_2(x_1,x_2,\cdots,x_n),\cdots,y_m(x_1,x_2,\cdots,x_n)$也都是独立变化的。这就表示该泛函的定义依赖于多个未知函数，且每个未知函数又依赖于多维变量。

设已知函数 $F(x,y(x),y'(x))$是由定义在区间$[x_0,x_1]$上的三个独立变量 $x,y(x),y'(x)$所共同确定的，并且是二阶连续可微的，则泛函

$$\mathcal{J}[y(x)]=\int_{x_0}^{x_1}F(x,y(x),y'(x))\,\mathrm{d}x$$

称为最简单的积分型泛函，或简称为最简泛函。被积函数 F 称为泛函的核。

同理，还可以定义变量函数为二元函数 $u(x,y)$时的泛函为

$$\mathcal{J}[y]=\iint\limits_{S}F(x,y,u,u_x,u_y)\,\mathrm{d}x\mathrm{d}y$$

其中，$u_x=\partial u/\partial x$，$u_y=\partial u/\partial y$。

此处所讨论的部分主要是古典变分法的内容。它所研究的主要问题可以归结为：在适当的函数类中选择一个函数使得类似于上述形式的积分取得最值。而解决这一问题又归结为求解欧拉-拉格朗日方程。这看起来并非一个多么复杂的问题，而且方法似乎也平常无奇。但依靠这种方法却惊异地发现原来自然世界中许多千差万别的问题居然能够使用统一的数学程序来求解，而且奇妙的变分原理还可以用来解释无数的自然规律。在 3.3.2 节中，将从最简泛函开始导出欧拉-拉格朗日方程。

3.3.2 变分的概念

已知一个函数在某一点处取极值，那么函数在该点处的导数(如果存在)必为零。那么要考虑一个泛函的极值问题，就不妨参照函数求极值的思想引入一个类似的概念，为此需引

入变分的概念，这也是得出欧拉-拉格朗日方程的关键所在。

对于任意定值 $x\in[x_0,x_1]$，可取函数 $y(x)$ 与另一个可取函数 $y_0(x)$ 之差称为函数 $y(x)$ 在 $y_0(x)$ 处的变分，记作 δy，δ 称为变分符号，此时有

$$\delta y = y(x) - y_0(x) = \varepsilon\eta(x)$$

其中，ε 是一个参数，$\eta(x)$ 为 x 的任意函数。由于可取函数都通过区间的端点，即它们在区间的端点值都相等，因此在区间的端点，任意函数 $\eta(x)$ 满足

$$\eta(x_0) = \eta(x_1) = 0$$

因为可取函数 $y(x)$ 是泛函 $\mathcal{J}[y(x)]$ 的宗量，故也可以这样定义变分：泛函的宗量 $y(x)$ 与另一宗量 $y_0(x)$ 之差 $y(x)-y_0(x)$ 称为宗量 $y(x)$ 在 $y_0(x)$ 处的变分。

上述变分的定义也可以推广到多元函数的情形。

显然，函数 $y(x)$ 的变分 δy 是 x 的函数。注意，函数变分 δy 与函数增量 Δy 的区别。函数的变分 δy 是两个不同函数 $y(x)$ 与 $y_0(x)$ 在自变量 x 取固定值时的差 $\alpha\eta(x)$，函数发生了改变；函数的增量 Δy 是由于自变量 x 取了一个增量而使得函数 $y(x)$ 产生的增量，函数仍然是原来的函数。

如果函数 $y(x)$ 与另一函数 $y_0(x)$ 都可导，则函数的变分 δy 有如下性质

$$\delta y' = y'(x) - y_0'(x) = [y(x) - y_0(x)]' = (\delta y)'$$

由此得到变分符号 δ 与导数符号之间的关系

$$\delta\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\delta y$$

即函数导数的变分等于函数变分的导数。换言之，求变分与求导数这两种运算次序可以交换。在进行变分法的推导时要经常用到变分的这个性质。上面这些性质也可推广到高阶导数的变分情形，具体情况这里不再赘述。

上面介绍了函数的变分，下面来考虑泛函的变分。例如，对于泛函

$$\mathcal{J}[y] = \int_a^b y^2(x)\mathrm{d}x$$

的增量，可以表示为

$$\begin{aligned}\Delta\mathcal{J} &= \mathcal{J}[y_1(x)] - \mathcal{J}[y_2(x)] = Q[y(x)+\delta y] - \mathcal{J}[y(x)]\\ &= \int_a^b [y(x)+\delta y]^2\mathrm{d}x - \int_a^b y^2(x)\mathrm{d}x\\ &= \int_a^b [y^2(x) + 2y(x)\delta y + (\delta y)^2]\mathrm{d}x - \int_a^b y^2(x)\mathrm{d}x\\ &= \int_a^b 2y(x)\delta y\mathrm{d}x + \int_a^b (\delta y)^2\mathrm{d}x\end{aligned}$$

其中，$\delta y = y_1(x) - y(x)$。

可见，此泛函 $\mathcal{J}$ 的增量 $\Delta\mathcal{J}$ 由两项相加而得。将第一项记为

$$\int_a^b 2y(x)\delta y\mathrm{d}x = T[y(x),\delta y]$$

当函数 $y(x)$ 固定时，$T[y(x),\delta y]$ 是关于 δy 的线性泛函。这是因为对任何常数 C 而言，有

$$T[y(x),C\delta y] = \int_a^b 2y(x)C\delta y\mathrm{d}x = C\int_a^b 2y(x)\delta y\mathrm{d}x = CT[y(x),\delta y]$$

且

$$\begin{aligned}T[y(x),\delta y_1+\delta y_2]&=\int_a^b 2y(x)(\delta y_1+\delta y_2)\mathrm{d}x\\&=\int_a^b 2y(x)\delta y_1\mathrm{d}x+\int_a^b 2y(x)\delta y_2\mathrm{d}x\\&=T[y(x),\delta y_1]+T[y(x),\delta y_2]\end{aligned}$$

再来考查第二项，此处 $\delta y=y_1(x)-y(x)$，其中 $y(x)$ 是已经给定的函数，$y_1(x)$ 是任意取的函数，$y(x)$ 和 $y_1(x)$ 均属于 $C[a,b]$

若

$$\max_{a\leqslant x\leqslant b}|y_1(x)-y(x)|=\max|\delta y|\to 0$$

由

$$\left|\int_a^b(\delta y)^2\mathrm{d}x\right|\leqslant\max_{a\leqslant x\leqslant b}(\delta y)^2(b-a)$$

可知

$$\frac{\int_a^b(\delta y)^2\mathrm{d}x}{\max|\delta y|}\to 0$$

上式表明，当 $\max|\delta y|\to 0$ 时，分子是比分母更高阶的无穷小量，不妨记为

$$\int_a^b(\delta y)^2\mathrm{d}x=0(\delta y)$$

于是 $\Delta\mathcal{J}=T[y(x),\delta y]+0(\delta y)$。这其实表明，原泛函的增量可以分解为两个部分，第一部分是 δy 的线性泛函，第二部分是比 δy 更高阶的无穷小量。回想函数微分的概念，函数的微分其实是函数增量的线性主要部分。换言之，微分就是当自变量的变化非常小时，用来近似等于因变量的一个量。上述对函数增量及微分关系的分析其实在提示人们，是否可以用泛函增量中的线性主要部分来近似等于泛函的增量。其实这种所谓的泛函增量中的线性主要部分就是下面定义中所给出的泛函的变分。

定义　对于泛函 $\mathcal{J}[y(x)]$，给 $y(x)$ 以增量 δy，即 $y(x)$ 的变分，则泛函 $\mathcal{J}$ 有增量 $\Delta\mathcal{J}=\mathcal{J}[y(x)+\delta y]-\mathcal{J}[y(x)]$。如果 $\Delta\mathcal{J}$ 可以表示为 $\Delta\mathcal{J}=T[(x),\delta y]+\beta[(x),\delta y]$。其中，当 $y(x)$ 给定时，$T[y(x),\delta y]$ 对 δy 来说是线性泛函，而当 $\max|\delta y|\to 0$ 时，有

$$\frac{\beta[(x),\delta y]}{\max|\delta y|}\to 0$$

那么，$T[y(x),\delta y]$ 称为泛函的变分，记作 $\delta\mathcal{J}$。可见，泛函 $\mathcal{J}[y(x)]$ 的变分 $\delta\mathcal{J}$ 本质上来讲就是 $\mathcal{J}$ 的增量的线性主要部分。

3.3.3　变分法的基本方程

导致变分法创立的著名问题是由瑞士数学家约翰·伯努利于 1696 年提出的所谓最速降线(brachistorone)问题。牛顿、莱布尼茨、约翰·伯努利以及他的学生洛必达各自采用不同的方法都成功地解决了这一问题，尽管他们采用的方法各不相同，但最终殊途同归，所得答案都是一致的。后来，欧拉也对最速降线问题进行了研究。1734 年，欧拉给出了更为广泛的最速降线问题的解答。但欧拉对自己当时所采用的方法不甚满意，进而开

始寻求解决这类问题的一种普适方法。而在此过程中，欧拉便建立了变分法。1736 年，欧拉在其著作中给出了变分法中的基本方程，这正是后来变分法所依托的重要基础。欧拉在推导该基本方程时采用的方法非常复杂，而拉格朗日则给出了一个非常简洁的方法，并于 1755 年在信中将该方法告知了欧拉。后来人们便称这个基本方程为欧拉-拉格朗日方程(Euler-Lagrange equation)。

在推导出欧拉-拉格朗日方程之前，先给出一个预备定理，也被称为是变分学引理。

引理 如果函数 $y=f(x)$在$[a,b]$上连续，又

$$\int_a^b f(x)\eta(x)\mathrm{d}x = 0$$

对任何具有如下性质的函数 $\eta(x)$成立，这些性质是：

(1) $\eta(x)$在$[a,b]$上有连续导数；

(2) $\eta(a)=0=\eta(b)$；

(3) $|\eta(x)|<\varepsilon$，其中 ε 是任意给定的正数。

那么，函数 $f(x)$在$[a,b]$上恒为 0。

这里不对该定理进行详细证明，有兴趣的读者可以参阅变分法或数学分析方面的相关资料以了解更多。但同时可以对上述预备定理进行推广，即如果把三个条件中的第一条改为：$\eta(x)$在$[a,b]$上有 n 阶连续导数。其中，n 为任何给定的非负整数，而且规定 $\eta(x)$的零阶导函数就是其本身。那么原命题中的结论仍然成立。特别地，当 $n=1$ 时，所描述的就是原来的预备定理。

至此准备工作已经基本就绪，接下来便可以开始考虑最简泛函的极值问题了。首先，可以利用类似函数极值的概念定义泛函的极值。当变量函数为 $y(x)$时，泛函$\mathcal{J}[y]$取极小值的含义就是：对于极值函数 $y(x)$及其附近的变量函数 $y(x)+\delta y(x)$，恒有

$$\mathcal{J}[y+\delta y] \geqslant \mathcal{J}[y]$$

所谓函数 $y(x)+\delta y(x)$在另一个函数 $y(x)$的附近，指的是：首先，$|\delta y(x)|<\varepsilon$；其次，有时还要求$|(\delta y)'(x)|<\varepsilon$。

接下来，可以仿照函数极值必要条件的导出办法，导出泛函取极值的必要条件。不妨不失普遍性地假定，所考虑的变量函数均通过固定的两个端点 $y(x_0)=a, y(x_1)=b$，即 $\delta y(x_0)=0, \delta y(x_1)=0$。

考虑泛函的差值

$$\mathcal{J}[y+\delta y]-\mathcal{J}[y]=\int_{x_0}^{x_1}F(x,y+\delta y,y'+(\delta y)')\mathrm{d}x-\int_{x_0}^{x_1}F(x,y,y')\mathrm{d}x$$

当函数的变分 $\delta y(x)$足够小时，可以将第一项的被积函数在极值函数的附近进行泰勒展开，于是有

$$F(x,y+\delta y,y'+\delta y')\approx F(x,y,y')+\left[\frac{\partial F}{\partial y}\cdot\delta y+\frac{\partial F}{\partial y'}\cdot(\delta y)'\right]$$

由于舍弃掉了二次项及以上高次项，所以这里用的是约等号。由上式也可推出

$$\mathcal{J}[y+\delta y]-\mathcal{J}[y]=\int_{x_0}^{x_1}\left[\frac{\partial F}{\partial y}\cdot\delta y+\frac{\partial F}{\partial y'}\cdot(\delta y)'\right]\mathrm{d}x$$

上式就称为是$\mathcal{J}[y]$的一阶变分，记为$\delta\mathcal{J}[y]$。泛函$\mathcal{J}[y]$取极值的必要条件是泛函的一阶

变分为 0，即

$$\delta \mathcal{J}[y] \equiv \int_{x_0}^{x_1} \left[\frac{\partial F}{\partial y} \cdot \delta y + \frac{\partial F}{\partial y'} \cdot (\delta y)'\right] \mathrm{d}x = 0$$

应用分部积分，同时代入边界条件，就有

$$\begin{aligned}\delta \mathcal{J}[y] &= \int_{x_0}^{x_1} \frac{\partial F}{\partial y} \cdot \delta y \mathrm{d}x + \int_{x_0}^{x_1} \frac{\partial F}{\partial y'} \cdot (\delta y)' \mathrm{d}x \\ &= \int_{x_0}^{x_1} \frac{\partial F}{\partial y} \cdot \delta y \mathrm{d}x + \frac{\partial F}{\partial y'} \delta y \Big|_{x_0}^{x_1} - \int_{x_0}^{x_1} \delta y \cdot \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) \mathrm{d}x = \int_{x_0}^{x_1} \delta y \cdot \left(\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial F}{\partial y'}\right) \mathrm{d}x\end{aligned}$$

由于 δy 的任意性，结合前面给出的预备定理，就可以得到

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial F}{\partial y'} = 0$$

上述这个方程称为欧拉-拉格朗日方程，而在力学中则被称为拉格朗日方程。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，它是泛函取极小值的必要条件的微分形式。值得指出的是，欧拉-拉格朗日方程只是泛函有极值的必要条件，并不是充分条件。

同理可得二维情况下泛函极值问题的欧拉-拉格朗日方程为

$$\frac{\partial F}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial u_x}\right) - \frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{\partial F}{\partial u_y}\right) = 0$$

定理　设 $F(x,y,y')$ 是三个变量的连续函数，且当点 (x,y) 在平面上的某个有界域 B 内，而 y' 取任何值时，$F(x,y,y')$ 及其直到二阶的偏导数（指对变量 x，y 及 y' 的偏导数）均连续。若满足：

（1）$y(x) \in C^1[a,b]$；

（2）$y(a) = y_0$，$y(b) = y_1$；

（3）$y(x)$ 曲线位于平面上的有界区域 B 内的函数集合中，泛函 $\mathcal{J}[y(x)]$ 在某一条确定的曲线 $y(x)$ 上取极值，且此曲线 $y(x)$ 在 $[a,b]$ 有二阶连续导数，那么函数 $y(x)$ 满足微分方程

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial F}{\partial y'} = 0$$

最后，尝试利用已经得到的欧拉-拉格朗日方程来解决著名的最速降线问题。该问题的描述是这样的：设平面 V 与地面垂直，A 和 B 是此平面上任取的两点，A 点的位置高于 B 点。质点 M 在重力作用下沿着曲线 AB 由 A 点降落到 B 点。现在问 AB 是什么曲线时，总时间最短？设质点在 A 点处的初速度为零，而且 A 点不位于 B 点的正上方。

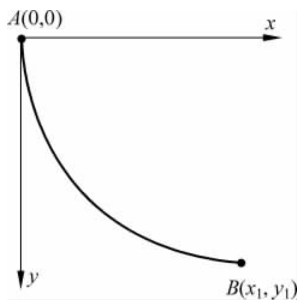

图 3-1　最速降线问题

解　取坐标系如图 3-1 所示，并记质点的质量为 m，速度为 v，又时间为 t，则质点下落时动能的增加就等于势能的减少，则 $mv^2/2 = mgy \Rightarrow v = \sqrt{2gy}$。曲线 $y = y(x)$ 的弧长微分是 $\mathrm{d}S = \sqrt{1 + y'^2}\,\mathrm{d}x$，又有 $v = \mathrm{d}S/\mathrm{d}t$，所以得到

$$\mathrm{d}t = \mathrm{d}S/v = \sqrt{(1+y'^2)/2gy}\,\mathrm{d}x$$

于是得到质点滑落的总时长为

$$T = \int_0^{x_1} \sqrt{(1+y'^2)/2gy}\,\mathrm{d}x = \mathcal{J}[y(x)]/\sqrt{2g}$$

由此可见，只需求出函数 $y=y(x)$，使泛函

$$\mathcal{J}[y(x)] = \int_0^{x_1} \sqrt{(1+y'^2)/y}\,\mathrm{d}x$$

在此曲线 $y(x)$上取得极小值即可。现在设法写出欧拉-拉格朗日方程，因为有

$$F(x,y,y') = \sqrt{(1+y'^2)/y}$$

于是得到

$$\frac{\partial F}{\partial y} = \sqrt{1+y'^2}\left(-\frac{1}{2}\right)y^{-\frac{3}{2}}$$

$$\frac{\partial F}{\partial y'} = \frac{y'}{\sqrt{y(1+y'^2)}}$$

所以得到欧拉-拉格朗日方程方程

$$\sqrt{1+y'^2}\left(-\frac{1}{2}\right)y^{-\frac{3}{2}} = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{y'}{\sqrt{y(1+y'^2)}}\right]$$

下面求解此方程。为了便于更加直观地理解计算过程，不妨将等式右边的$\partial F/\partial y'$用 f 代替。注意，f 是关于 y 和 y'的一个多元复合函数，而 y 和 y' 又分别都是关于 x 的函数。所以，在计算的时候还需用到复合函数的链式求导法则。于是，可得方程的右边为

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\partial f}{\partial y}\cdot\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial f}{\partial y'}\cdot\frac{\mathrm{d}y'}{\mathrm{d}x} = \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial y'}\right)\cdot\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial}{\partial y'}\left(\frac{\partial F}{\partial y'}\right)\cdot\frac{\mathrm{d}y'}{\mathrm{d}x}$$

于是上面得到的欧拉-拉格朗日方程可以写为

$$F_y - F_{y'_y}y' - F_{y'_{y'}}y'' = 0$$

而且上式等价于

$$\frac{\mathrm{d}}{\mathrm{d}x}(F - y'F_{y'}) = 0$$

这是因为

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(F - y'F_{y'}) &= F_y y' + F_{y'}y'' - y''F_{y'} - y'(F_{y'_y}y' - F_{y'_{y'}}y'')\\ &= y'(F_y - F_{y'_y}y' - F_{y'_{y'}}y'') = 0\end{aligned}$$

将

$$\frac{\mathrm{d}}{\mathrm{d}x}(F - y'F_{y'}) = 0$$

做一次积分得到（其中，C 表示任意常数）

$$F - y'F_{y'} = C$$

将 F 的表达式代入上式，得

$$[y(1+y'^2)]^{-\frac{1}{2}} = C$$

即 $y(1+y'^2)=D$，D 为任意常数。

令 $y'=\tan\theta$，则 $y=D/(1+\tan^2\theta)=D\cos^2\theta=D(1+\cos 2\theta)/2$，$\mathrm{d}y=-D\sin 2\theta\mathrm{d}\theta$。又有

$$\mathrm{d}x = \frac{\mathrm{d}y}{y'} = -\frac{-D\sin 2\theta \mathrm{d}\theta}{\tan\theta} = -\frac{-2D\sin\theta\cos\theta\mathrm{d}\theta}{\tan\theta} = -D(1+\cos 2\theta)\mathrm{d}\theta$$

于是有(其中,E 是任意常数)

$$\begin{cases} x = -D\theta - \dfrac{D}{2}\sin 2\theta + E \\ y = \dfrac{D}{2}(1+\cos 2\theta) \end{cases}$$

这就是最速降线问题的欧拉-拉格朗日方程的解。如果令 $2\theta=\pi-\varphi$,则上式化为

$$\begin{cases} x = \dfrac{D}{2}(\varphi - \sin\varphi) - \dfrac{\pi}{2}D + E \\ y = \dfrac{D}{2}(1-\cos\varphi) \end{cases}$$

又当 $\varphi=0$ 时,取 $x=0=y$,于是

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

最终得到,最速降线问题的解是一条旋轮线(也称摆线)。推荐对旋轮线感兴趣的读者参阅文献[11]以了解更多。

最后,讨论其他一些特殊形式变分问题的欧拉方程。

定理 使泛函(其中,F 是具有三阶连续可微的函数,y 是具有四阶连续可微的函数)

$$\mathcal{J}[y(x)] = \int_{x_0}^{x_1} F(x,y,y',y'')\mathrm{d}x$$

取极值且满足固定边界条件 $y(x_0)=y_0, y(x_1)=y_1, y'(x_0)=y'_0, y'(x_1)=y_1$ 的极值曲线 $y=y(x)$ 必满足微分方程

$$F_y - \frac{\mathrm{d}}{\mathrm{d}x}F_{y'} + \frac{\mathrm{d}^2}{\mathrm{d}x^2}F_{y''} = 0$$

上式称为欧拉-泊松方程。

特别地,对含有未知函数的 n 阶导数,或未知函数有两个或两个以上的固定边界变分问题,若被积函数 F 足够光滑,则可得到如下推论。

推论 使依赖于未知函数 $y(x)$ 的 n 阶导数的泛函

$$\mathcal{J}[y(x)] = \int_{x_0}^{x_1} F(x,y,y',\cdots,y^{(n)})\mathrm{d}x$$

取极值且满足固定边界条件

$$y^{(i)}(x_0) = y_0^{(k)}, \quad y^{(i)}(x_1) = y_i^{(k)}, \quad k = 0,1,2,\cdots,n-1$$

的极值曲线 $y=y(x)$ 必满足欧拉-泊松方程

$$F_y - \frac{\mathrm{d}}{\mathrm{d}x}F_{y'} + \frac{\mathrm{d}^2}{\mathrm{d}x^2}F_{y''} - \cdots + (-1)^n \frac{\mathrm{d}^n}{\mathrm{d}x^n}F_{y^{(n)}} = 0$$

其中,F 具有 $n+2$ 阶连续导数,y 具有 $2n$ 阶连续导数,这是 $2n$ 阶微分方程,它的通解中含有 $2n$ 个待定常数,可由 $2n$ 个边界条件来确定。

定理 设 D 是平面区域,$(x,y)\in D, u(x,y)\in C^2(D)$,使泛函

$$\mathcal{J}[u(x,y)]=\iint_D F(x,y,u,u_x,u_y)\mathrm{d}x\mathrm{d}y$$

取极值且在区域 D 的边界 L 上满足边界条件，极值函数 $u=u(x,y)$ 必满足偏微分方程

$$F_u-\frac{\partial}{\partial x}F_{u_x}-\frac{\partial}{\partial y}F_{u_y}=0$$

这个方程称为奥斯特洛格拉茨基方程，简称奥氏方程。它是欧拉方程的进一步发展。

例 3.1 已知 $(x,y)\in D$，求下述泛函的奥氏方程。

$$\mathcal{J}[u(x,y)]=\iint_D\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2\right]\mathrm{d}x\mathrm{d}y$$

根据前面给出的公式，不难写出奥氏方程为

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0$$

这也是二维拉普拉斯方程。

例 3.2 已知 $(x,y)\in D$，写出泛函

$$\mathcal{J}[u(x,y)]=\iint_D\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+2uf(x,y)\right]\mathrm{d}x\mathrm{d}y$$

的奥氏方程。其中，在区域 D 的边界上 u 与 $f(x,y)$ 均为已知。

根据前面给出的公式，不难写出奥氏方程为

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=f(x,y)$$

这就是人们所熟知的泊松方程。

1777 年，拉格朗日研究万有引力作用下的物体运动时指出：在引力体系中，每一质点的质量 m_k 除以它们到任意观察点 P 的距离 r_k，并且把这些商加在一起，其总和

$$\sum_{k=1}^{n}\frac{m_k}{r_k}=V(x,y,z)$$

就是 P 点的势函数，势函数对空间坐标的偏导数正比于在 P 点的质点所受总引力的相应分力。在 1782 年，拉普拉斯证明了引力场的势函数满足偏微分方程

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2}=0$$

该方程叫做势方程，后来通称为拉普拉斯方程。1813 年，泊松撰文指出，如果观察点 P 在充满引力物质的区域内部，则拉普拉斯方程应修改为

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2}=-4\pi\rho$$

该方程叫做泊松方程。其中，ρ 为引力物质的密度。

3.3.4 理解哈密尔顿原理

3.3.3 节从最简泛函开始导出了变分法的基本方程为欧拉-拉格朗日方程。但仍然不禁要问为什么要以形如最简泛函那样的一种表达式来作为问题的开始？事实上，数学中的很多问题都不是凭空而来的，每一个看似高深的数学问题背后往往都有一个具体的实际问题作为支撑。数学问题也仅仅是实际问题抽象化的结果。在这一节中，将从物理问题的角

度阐释变分法的发展与应用。

当牛顿建立了以三大定律及万有引力定律为基础的力学理论之后，无数的自然现象都得到了定量的说明。这部分知识在中学物理中都已经涵盖，大学物理也仅从微积分的角度对这部分内容进行了更为细致的阐述。貌似经典物理学所讨论的内容已经相当完善。然而科学发展的脚步并未因此而停滞。后来，拉格朗日提出了一个变分原理，从这个原理出发，运用变分法，不仅能够十分方便地解决力学问题，而且还能够推导出力学中的主要定律。这些成果后来都收录在他的著作《分析力学》一书中。拉格朗日还创立了拉格朗日运动方程，比牛顿的运动方程适应的范围更广泛，用起来也更加方便。

下面就来导出描写质点运动的拉格朗日方程。先设质点只有一个广义坐标 x。因为，质点的位置由广义坐标 $x(t)$决定，即位置是时间的函数。于是，动能 T 和位能 U 是 x 和 x'（距离对时间的导数其实就是速度）的函数。把 $T-U$ 叫做拉格朗日函数，记为

$$T-U=L=L(t,x,x')$$

于是，质点的作用量定义为

$$S=\int_{t_2}^{t_1}L(t,x,x')\mathrm{d}t$$

根据之前的推导，因为 S 取极值，所以真实轨迹 $x(t)$满足

$$\frac{\partial L}{\partial x}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial x'}=0$$

这就是力学中著名的拉格朗日方程。同样，若质点系的位置由广义坐标 $x_1,x_2,\cdots,x_k$ 决定，且 $x_i(t_1)$及 $x_i(t_2)$均已给定。其中，$i=1,2,\cdots,k$，即在 $t=t_1$ 及 $t=t_2$ 两时刻，体系的位置均已给定。当质点系由 t_1 时刻的位置变到 t_2 时刻的位置时，作用量

$$S=\int_{t_2}^{t_1}L(t,x_1,x_2\cdots,x_k,x'_1,x'_2,\cdots,x'_k)\mathrm{d}t$$

取极值。这种形如

$$\mathcal{J}[y_1(x),y_2(x),\cdots,y_k(x)]=\int_a^b F[x,y_1,y_2,\cdots,y_n,y'_1,y'_2,\cdots,y'_n]\mathrm{d}t$$

的泛函，其对应的欧拉-拉格朗日方程为（具体证明过程略）

$$F_{y_i}-\frac{\mathrm{d}}{\mathrm{d}x}F_{y'_i}=0,\quad i=1,2,\cdots,n$$

由此可知，真实轨迹 $x_i(t)$，$i=1,2,\cdots,k$，满足

$$\frac{\partial L}{\partial x_i}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial x'_i}=0,\quad i=1,2,\cdots,k$$

这就是质点系的拉格朗日方程组。它是在广义坐标系中质点系的运动方程，表达了质点系运动的一般规律。

此后，哈密尔顿又发展了拉格朗日的理论，他在 1834 年提出了一个著名的原理，即哈密尔顿原理，其内容为在质点（甚至是质点系或物体）的一切可能的运动中，真实的运动应当使得积分

$$S=\int_{t_1}^{t_2}(T-U)\mathrm{d}t$$

取极值。其中，T 和 U 分别是动能和位能，t_1 和 t_2 是两个任意取的时刻。

这个原理后来成为了力学中的基本原理。以它为基础，可以导出牛顿三大定律以及能

量、动量和动量矩守恒定律。

哈密尔顿原理的精确表述是：假定在 $t=t_1$ 及 $t=t_2$ 时刻质点的位置已分别确定在 A 点和 B 点，那么质点运动的真实轨道及速度，使积分

$$S=\int_{t_1}^{t_2}(T-U)\mathrm{d}t=\int_{t_1}^{t_2}L\mathrm{d}t$$

取极值，即

$$\delta S=\delta\int_{t_1}^{t_2}(T-U)\mathrm{d}t=\delta\int_{t_1}^{t_2}L\mathrm{d}t=0$$

其中，S 是作用量，而 T 和 U 分别表示质点的动能和位能，$L=T-U$称为拉格朗日函数。

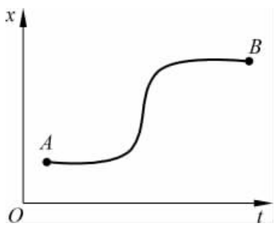

图 3-2 两点间的距离

接下来，尝试利用哈密尔顿原理及变分法来证明欧几里得平面上两点之间直线距离最短这个命题。

解 建立如图 3-2 所示的坐标系。则曲线 AB 的长度可以用弧长积分表示为

$$\mathcal{J}[x(t)]=\int_{t_1}^{t_2}\sqrt{1+x'^2(t)}\,\mathrm{d}t$$

因为 $F(t,x,x')=\sqrt{1+x'^2(t)}$，于是 $F_x=0$，又

$$F_{x'}=\frac{x'}{\sqrt{1+x'^2}}$$

所以得到欧拉-拉格朗日方程为

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x'}{\sqrt{1+x'^2}}\right)=0$$

其中，C 是任意常数

$$\frac{x'}{\sqrt{1+x'^2}}=C$$

由此解得(其中，$C^2\neq1$)

$$x'=\pm\frac{C}{\sqrt{1-C^2}}$$

即 $x'=D$，D 为任意常数。由此式便可看出 $x(t)$是一条直线。结论得证。

3.3.5 等式约束下的变分

在许多极值或最优化问题中，往往要求极值点或最优解满足一定的约束条件。这些所谓的约束条件可能是用等式表示的，也可能是用不等式表示的。这里主要关注采用等式约束的形式。因此，首先介绍著名的拉格朗日乘子法。

定理 (拉格朗日乘子法)设泛函 f 在 $x_0\in X$ 的邻域内连续可微，x_0 是 Φ 的正则点。如果 x_0 是泛函 f 在约束条件 $\Phi(x)=0$ 下的极值点，则存在有界线性泛函 $z_0^*\in Z^*$，使得拉格朗日函数

$$L(x)=f(x)+z_0^*\Phi(x)$$

以 x_0 为驻点，即

$$f'(x_0)+z_0^* \Phi'(x_0)=0$$

其中,上式左端的第 2 项应该理解为两个有界线性算子的复合(乘积)。

例如,在约束条件

$$\begin{cases}\Phi_1(x_1,x_2,\cdots,x_n)=0\\ \Phi_2(x_1,x_2,\cdots,x_n)=0\\ \quad\vdots\\ \Phi_k(x_1,x_2,\cdots,x_n)=0\end{cases}$$

下求 $F=F(x_1,x_2,\cdots,x_n)$的极值,其中 $k<n$。如果要利用拉格朗日乘子法,则设有拉格朗日乘子 $\lambda_1,\lambda_2,\cdots,\lambda_k$,并有

$$F^*=F(x_1,x_2,\cdots,x_n)+\sum_{i=1}^{k}\lambda_i\Phi_i(x_1,x_2,\cdots,x_n)$$

把 F^* 作为 $x_1,x_2,\cdots,x_n$; $\lambda_1,\lambda_2,\cdots,\lambda_k$ 等变量的函数求极值。

$$\mathrm{d}F^*=\sum_{j=1}^{n}\left(\frac{\partial F}{\partial x_j}+\sum_{i=1}^{k}\lambda_i\frac{\partial \Phi_i}{\partial x_j}\right)\mathrm{d}x_j+\sum_{i=1}^{k}\Phi_i(x_1,x_2,\cdots,x_n)\mathrm{d}\lambda_i$$

其中,x_j 和 λ_i 都是独立变量,得

$$\begin{cases}\dfrac{\partial F}{\partial x_j}+\displaystyle\sum_{i=1}^{k}\lambda_i\frac{\partial \Phi_i}{\partial x_j}=0, & j=1,2,\cdots,n\\ \Phi_i(x_1,x_2,\cdots,x_n)=0, & i=1,2,\cdots,k\end{cases}$$

于是便得到了求解 $n+k$ 个变量的 $n+k$ 个方程。

上式也可以通过下面的考虑求得。首先,$F=F(x_1,x_2,\cdots,x_n)$的变分极值要求

$$\mathrm{d}F=\sum_{j=1}^{n}\frac{\partial F}{\partial x_j}\mathrm{d}x_j=0$$

然而,由于问题开始所给出了 k 个约束条件,这些 $\mathrm{d}x_j$ 中只有 $n-k$ 个是独立的。于是从原 k 个约束条件可以求得下列微分条件:

$$\sum_{j=1}^{n}\frac{\partial \Phi_i}{\partial x_j}\mathrm{d}x_j=0,\quad i=1,2,\cdots,k$$

在上式上乘以 λ_i,再加到 $\mathrm{d}F$ 的表达式上,就会得到

$$\mathrm{d}F+\sum_{i=1}^{k}\lambda_i\sum_{j=1}^{n}\frac{\partial \Phi_i}{\partial x_j}\mathrm{d}x_j=\sum_{j=1}^{n}\left(\frac{\partial F}{\partial x_j}+\sum_{i=1}^{k}\lambda_i\frac{\partial \Phi_i}{\partial x_j}\right)\mathrm{d}x_j=0$$

这里的 λ_i 是任选的。其中,$i=1,2,\cdots,k$,如果选择 k 个特定的 λ_i,使 k 个条件

$$\frac{\partial F}{\partial x_j}+\sum_{i=1}^{k}\lambda_i\frac{\partial \Phi_i}{\partial x_j}=0,\quad j=1,2,\cdots,k$$

满足,就可以得到

$$\sum_{j=k+1}^{n}\left(\frac{\partial F}{\partial x_j}+\sum_{i=1}^{k}\lambda_i\frac{\partial \Phi_i}{\partial x_j}\right)\mathrm{d}x_j=0$$

这里只有 $n-k$ 个微分 $\mathrm{d}x_j(j=k+1,k+2,\cdots,n)$,它们是作为独立的微分来处理的。于是

$$\frac{\partial F}{\partial x_j}+\sum_{i=1}^{k}\lambda_i\frac{\partial \Phi_i}{\partial x_j},\quad j=k+1,k+2,\cdots,n$$

综上,得到了同样的求解极值的方程。这也就证明了拉格朗日乘子法。

为了加深对拉格朗日乘子法的理解,这里给出一个例子。

证明算术-几何平均值不等式：设 $x_1,x_2,\cdots,x_n$ 为 n 个正实数，它们的算术平均数是 $A_n=(x_1+x_2+\cdots+x_n)/n$，它们的几何平均数是 $G_n=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n}$。算术-几何平均值不等式表明，对于任意的正实数，总有 $A_n\geqslant G_n$，等号成立当且仅当 $x_1=x_2=\cdots=x_n$。

证明这个不等式的方法有很多，这里采用条件极值的方法来对其进行证明。此时，问题转化为：总和等于常数 C，$C>0$ 的 n 个非负实数，它们的乘积 P 的最大值为多少？

考虑采用拉格朗日乘子法求 n 元函数 $P=x_1x_2\cdots x_n$ 对如下条件的极大值，条件为这 n 个非负实数的和等于 C，即 $x_1+x_2+\cdots+x_n=C$，$x_i\geqslant 0$，$i=1,2,\cdots,n$。于是构造如下函数

$$L=x_1x_2\cdots x_n+\lambda(x_1+x_2+\cdots+x_n-C)$$

其中，λ 是拉格朗日乘子，然后分别对 $x_1,x_2,\cdots,x_n$ 求偏导数，然后令其结果等于 0，构成如下方程组

$$\begin{cases}\dfrac{\partial L}{\partial x_1}=x_2x_3\cdots x_n+\lambda=0\\ \dfrac{\partial L}{\partial x_2}=x_1x_3\cdots x_n+\lambda=0\\ \qquad\qquad\vdots\\ \dfrac{\partial L}{\partial x_n}=x_1x_2\cdots x_{n-1}+\lambda=0\end{cases}$$

求解方程组，可得 $x_1=x_2=\cdots=x_n=C/n$。因为根据题目的描述，P 的极小值是等于 0 的，而当 x_i 满足上述条件时显然 P 是不等于 0 的，所以可知此时函数取极大值，这个极大值就等于

$$P_{\max}=\frac{C}{n}\cdot\frac{C}{n}\cdots\frac{C}{n}=\left(\frac{C}{n}\right)^n$$

即

$$\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)^n\geqslant x_1x_2\cdots x_n$$

对两边同时开根号，显然有下式成立，所以原不等式得证。

$$\frac{x_1+x_2+\cdots+x_n}{n}\geqslant\sqrt[n]{x_1x_2\cdots x_n}$$

下面就参照上述函数条件极值问题的解决思路处理泛函在约束条件 $\Phi_i(x,y_1,y_2,\cdots,y_n)=0$ 作用下的极值问题，其中 $i=1,2,\cdots,k$。

定理 泛函

$$\mathcal{J}=\int_{x_1}^{x_2}F(x,y_1,y_2,\cdots,y_n,y_1',y_2',\cdots,y_n')\mathrm{d}x$$

在约束条件 $\Phi_i(x,y_1,y_2,\cdots,y_n)=0$，$i=1,2,\cdots,k$，$k<n$ 下的变分极值问题所定义的函数 $y_1,y_2,\cdots,y_n$ 必须满足由泛函

$$\mathcal{J}^*=\int_{x_1}^{x_2}\Big[F+\sum_{i=1}^{k}\lambda_i(x)\Phi_i\Big]\mathrm{d}x=\int_{x_1}^{x_2}F^*\,\mathrm{d}x$$

的变分极值问题所确定的欧拉方程

$$\frac{\partial F^*}{\partial y_j}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F^*}{\partial y_j'}\right)=0,\quad j=1,2,\cdots,n$$

其中，$\lambda_i(x)$ 为 k 个拉格朗日乘子。在前面式子的变分中，把 y_j 和 $\lambda_i(x)$ 都看作是泛函 $\mathcal{J}^*$ 的

宗量，所以 $\Phi_i=0$ 同样也可以看作是泛函 $\mathcal{J}^*$ 的欧拉方程。上述欧拉方程也可以写成

$$\frac{\partial F}{\partial y_j}+\sum_{i=1}^{k}\lambda_i(x)\frac{\partial \Phi_i}{\partial y_j}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y_j'}\right)=0,\quad j=1,2,\cdots,n$$

这里不对该定理做详细证明。有兴趣的读者可以参阅变分法方面的资料以了解更多。此处尝试运用该定理解决一个著名的变分问题——短程线问题。设 $\varphi(x,y,z)=0$ 为已知曲面，求曲面上所给两点 A 和 B 间长度最短的曲线。这个最短曲线叫做短程线。位于曲面 $\varphi(x,y,z)=0$ 上的 $A(x_1,y_1,z_1)$ 和 $B(x_2,y_2,z_2)$ 两点间的曲线长度为

$$L=\int_{x_1}^{x_2}\sqrt{1+y'^2+z'^2}\,\mathrm{d}x$$

其中，$y=y(x)$，$z=z(x)$ 满足 $\varphi(x,y,z)=0$ 的条件。

此处把问题描述为：在 $y=y(x)$，$z=z(x)$ 满足 $z=\sqrt{1-x^2}$ 的条件下，从一切 $y=y(x)$，$z=z(x)$ 的函数中，选取一对 $y(x)$ 和 $z(x)$ 使得上述泛函 L 为最小。

用拉格朗日乘子 $\lambda(x)$ 建立泛函

$$L^*=\int_{x_1}^{x_2}\left(\sqrt{1+y'^2+z'^2}+\lambda\varphi\right)\mathrm{d}x$$

其变分（把 y、z 和 λ 当作独立函数）为

$$\delta L^*=\int_{x_1}^{x_2}\left(\frac{y'}{\sqrt{1+y'^2+z'^2}}\delta y'+\frac{z'}{\sqrt{1+y'^2+z'^2}}\delta z'+\lambda\frac{\partial\varphi}{\partial y}\delta y+\lambda\frac{\partial\varphi}{\partial z}\delta z+\varphi\delta\lambda\right\}\mathrm{d}x$$

把积分符号中的首两项做分部积分，得到

$$\delta L^*=\int_{x_1}^{x_2}\left\{\left[-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{y'}{\sqrt{1+y'^2+z'^2}}\right)+\lambda\frac{\partial\varphi}{\partial y}\right]\delta y+\left[-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{z'}{\sqrt{1+y'^2+z'^2}}\right)+\lambda\frac{\partial\varphi}{\partial z}\right]\delta z+\varphi\delta\lambda\right\}\mathrm{d}x$$

根据变分法的预备定理，把 δy，δz 和 $\delta\lambda$ 都看成是独立的函数变分，$\delta L^*=0$ 给出欧拉方程

$$\begin{cases}\lambda\dfrac{\partial\varphi}{\partial y}-\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{y'}{\sqrt{1+y'^2+z'^2}}\right)=0\\[2ex]\lambda\dfrac{\partial\varphi}{\partial z}-\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{z'}{\sqrt{1+y'^2+z'^2}}\right)=0\\[2ex]\varphi(x,y,z)=0\end{cases}$$

这是求解 $y(x)$，$z(x)$ 和 $\lambda(x)$ 的三个微分方程。

现在设所给的约束条件为一个圆柱面 $z=\sqrt{1-x^2}$，于是上述方程组可以写成

$$\begin{cases}\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{y'}{\sqrt{1+y'^2+z'^2}}\right)=0\\[2ex]\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{z'}{\sqrt{1+y'^2+z'^2}}\right)=\lambda(x)\\[2ex]z=\sqrt{1-x^2}\end{cases}$$

第一式和第二式可以积分一次，同时引入弧长 s，则 $\mathrm{d}s=\sqrt{1+y'^2+z'^2}\,\mathrm{d}x$，则积分以后原方

程组可以写成

$$\begin{cases} \mathrm{d}y = a \cdot \mathrm{d}s \\ \mathrm{d}z = \Lambda(x) \cdot \mathrm{d}x \\ z = \sqrt{1-x^2} \end{cases}$$

其中，a 为积分常数，则

$$\Lambda(x) = \int_0^x \lambda(x)\mathrm{d}x + a$$

从方程组中的第二式和第三式，可得

$$\mathrm{d}x = -\frac{\sqrt{1-x^2}}{x}\mathrm{d}z = -\frac{\sqrt{1-x^2}}{x}\Lambda(x)\mathrm{d}s$$

因此，根据 $\mathrm{d}s$ 的定义有

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = \left[\frac{1-x^2}{x^2}\Lambda^2(x) + a^2 + \Lambda^2(x)\right]\mathrm{d}s^2$$

它可以化简为 $\Lambda(x)=\sqrt{1-a^2}\,x$。于是，把上式代入 $\mathrm{d}x$ 的表达式，消去 $\Lambda(x)$，即得

$$-\frac{\mathrm{d}x}{\sqrt{1-x^2}} = \sqrt{1-a^2}\,\mathrm{d}s$$

积分后，得 $\cos^{-1}x=\sqrt{1-a^2}\,s+d$，其中 d 为另一个积分常数，或为

$$x = \cos(\sqrt{1-a^2}\,s + d)$$

并且还可以得到

$$z = \sin(\sqrt{1-a^2}\,s + d)$$

以及 $y=as+b$。其中，b 也为积分常数。于是，便得到了本题的参数解，弧长 s 为参数。积分常数 a,b,d 由起点和终点的坐标决定。这个解就是圆柱面 $z=\sqrt{1-x^2}$ 上的螺旋线。

还可以把原定理加以推广，使得 Φ_i 不仅是 $x,y_1,y_2,\cdots,y_n$ 的函数，而且是 $y'_1,y'_2,\cdots,y'_n$ 的函数的情况，于是有推广后的定理如下。

泛函

$$\mathcal{J}=\int_{x_1}^{x_2} F(x,y_1,y_2,\cdots,y_n,y'_1,y'_2,\cdots,y'_n)\mathrm{d}x$$

在约束条件 $\Phi_i(x,y_1,y_2,\cdots,y_n,y'_1,y'_2,\cdots,y'_n)=0$，$i=1,2,\cdots,k$，$k<n$ 下的变分极值问题所定义的函数 $y_1,y_2,\cdots,y_n$ 必须满足由泛函

$$\mathcal{J}^* = \int_{x_1}^{x_2}\left[F + \sum_{i=1}^{k}\lambda_i(x)\Phi_i\right]\mathrm{d}x = \int_{x_1}^{x_2} F^*\,\mathrm{d}x$$

的变分极值问题所确定的欧拉方程

$$\frac{\partial F^*}{\partial y_j} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F^*}{\partial y'_j}\right) = 0, \quad j = 1,2,\cdots,n$$

或

$$\frac{\partial F}{\partial y_j} - \sum_{i=1}^{k}\lambda_i(x)\frac{\partial \Phi_i}{\partial y_j} - \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial F}{\partial y'_j} + \sum_{i=1}^{k}\lambda_i(x)\frac{\partial \Phi_i}{\partial y'_j}\right] = 0, \quad j = 1,2,\cdots,n$$

在前面式子的变分中，把 y_j 和 $\lambda_i(x)$ 都看作是泛函 $\mathcal{J}^*$ 的宗量，所以 $\Phi_i=0$ 同样也可以看作泛函 $\mathcal{J}^*$ 的欧拉方程。

3.3.6 巴拿赫不动点定理

设 X 为巴拿赫空间，F 为由 X 到 X 的算子，且 $D(F)\cap R(F)$ 非空。如果点 $x^*\in X$ 满足

$$F(x^*)=x^*$$

则称 x^* 为算子 F 的不动点。换句话说，不动点 x^* 是算子方程 $x=F(x)$ 的解。巴拿赫不动点定理，又称为压缩映射原理，它不仅指出了上述算子方程之解的存在性和唯一性，还提供了求出这些近似解的方法及误差估计。

设集合 $Q\subset D(F)$，如果存在常数 $q\in(0,1)$，使得对任意的 $x',x''\in Q$，均有不等式

$$\| F(x')-F(x'')\| \leqslant q\| x'-x''\|$$

则称 F 为集合 Q 上的压缩算子，q 称为压缩系数。

压缩映射原理 设算子 F 映巴拿赫空间 X 中的闭集 Q 为其自身，且 F 为 Q 上的压缩算子，压缩系数为 q，则算法 F 在 Q 内存在唯一的不动点 x^*。若 x_0 为 Q 中任意一点，做序列

$$x_{n+1}=F(x_n),\quad n=0,1,2,\cdots$$

则序列 $\{x_n\}\subset Q$，且 $x_n\to x^*$，并有误差估计

$$\| x_n-x^*\| \leqslant \frac{q^n}{1-q}\| F(x_0)-x_0\|$$

例如，可以利用巴拿赫不动点定理求的 $\sqrt[3]{5}$ 近似值。注意，$\sqrt[3]{5}$ 是方程 $x^3-5=0$ 的实根，构造辅助函数 $f(x)=x^3-5$，则任意给定的 $x\in[1,2]$，都有

$$f'(x)=3x^2\in[3,12]$$

再令

$$g(x)=x-\frac{1}{12}(x^3-5)$$

容易验证，当 $x\in[1,2]$ 时，有 $1\leqslant g(x)\leqslant 2$，以及

$$0\leqslant g'(x)=1-\frac{1}{4}x^2\leqslant\frac{3}{4}$$

所以，g：$[1,2]\to[1,2]$ 是压缩因子 $q=3/4$ 的压缩映射。由于 $[1,2]$ 是 $\mathbb{R}$ 中的有界闭集，因此有 $x^*=\sqrt[3]{5}$ 使得 $g(x^*)=x^*$。进而可用迭代法求得 $\sqrt[3]{5}$ 的近似值。取 $x_0=1$，从而有

$$x_1=g(x_0)=\frac{4}{3},\cdots,x_n=g(x_{n-1})$$

由上述说明可知

$$|x_n-x^*|=|x_n-\sqrt[3]{5}|\leqslant\frac{q^n}{1-q}|x_1-x_0|=\left(\frac{3}{4}\right)^{n-1}$$

从而也可由此求出近似值与精确值之间的误差。

3.3.7 有界变差函数空间

在数学分析中，有界变差(bounded variation)函数，有时也称为 BV 函数，是一个实值函数，它的全变差(total variation)是有界的，即为有限值。首先，讨论最简单的全变差函数定义——单变量的 BV 函数。

定义 一个实值函数 f 定义在区间$[a,b]\subset\mathbb{R}$上的全变差(total variation)，就是如下这样一个量

$$V_a^b(f)=\sup_{p\in P}\sum_{i=0}^{n_P-1}|f(x_{i+1})-f(x_i)|$$

其中，$P=\{x_0,x_1,x_2,\cdots,x_{n_P}\}$是区间$[a,b]$上的一个划分。被考查区间上的所有划分构成一个集合$\mathcal{P}$，而上确界是取遍该集合所得到的。

如果 f 是可微的，并且它的导数是黎曼可积的，那么它的全变差就是

$$V_a^b(f)=\int_a^b|f'(x)|\mathrm{d}x$$

例如，设 $f(x)$是定义在区间$[a,b]$上的有限函数。在$[a,b]$上做分点

$$x_0=a<x_1<x_2<\cdots<x_n=b$$

并且做和

$$V=\sum_{i=0}^{n-1}|f(x_{i+1})-f(x_i)|$$

那么，V 的上确界就是 $f(x)$在$[a,b]$上的全变差，记作 $V_a^b(f)$，有时也会记为 $\overset{b}{\underset{a}{V}}(f)$。本书采用前一种记法。

定义 一个位于实数轴上的实值函数 f 在被选定的区间$[a,b]\subset\mathbb{R}$上被称为是有界变差的(BV 函数)，只需它的全变差是有限的，即

$$f\in \mathrm{BV}([a,b])\Leftrightarrow V_a^b(f)<+\infty$$

换句话说，当 $V_a^b(f)<+\infty$时，称 $f(x)$在$[a,b]$上是有界变差的，或称 $f(x)$在$[a,b]$上具有有界的变差。

定理 单调函数是有界变差的。

本定理，就增函数证明即足矣。设 $f(x)$在定义在$[a,b]$上的一个增函数，那么 $f(x_{i+1})-f(x_i)$不是负的，从

$$V=\sum_{i=0}^{n-1}|f(x_{i+1})-f(x_i)|=f(b)-f(a)$$

即得到定理的证明。

满足利普希茨(R. Lipschitz)条件的函数是有界变差函数的又一个例子。利普希茨条件是一个比一致连续更强的光滑性条件。直观上，利普希茨连续函数限制了函数改变的速度，符合利普希茨条件的函数的斜率，必小于一个称为利普希茨常数的实数(该常数依函数而定)。在微分方程理论中，利普希茨条件是初值条件下解的存在唯一性定理中的一个核心条件。利普希茨条件的一个特殊形式即压缩映射原理，被应用在巴拿赫不动点定理中。

定义 在$[a,b]$上所有定义的有限函数 $f(x)$，如果存在有大于 0 的常数 K 使得不等式

$$|f(x)-f(y)|\leqslant K|x-y|$$

对于$[a,b]$中任何两点 x,y 成立，称 $f(x)$在$[a,b]$上满足利普希茨条件，K 称为利普希茨常数。

若 $f(x)$在区间上满足利普希茨条件，必定有 $f(x)$在此区间上一致连续。假如 $f(x)$在$[a,b]$上的每一点 x 中具有有界的导数 $f'(x)$，那么由拉格朗日中值定理可得

$$f(x)-f(y)=f'(z)(x-y),\quad x<z<y$$

即 $f(x)$是满足利普希茨条件的。

假如 $f(x)$在$[a,b]$上满足利普希茨条件，则

$$|f(x_{i+1})-f(x_i)|\leqslant K(x_{i+1}-x_i)$$

从而有 $V\leqslant K(b-a)$。所以，$f(x)$是有界变差的函数。

连续函数的全变差可以是无穷大的。例如

$$f(x)=x\cos\frac{\pi}{2x},\quad 0<x\leqslant 1,f(0)=0$$

如果在$[0,1]$中采取如下划分方式

$$0<\frac{1}{2n}<\frac{1}{2n-1}<\cdots<\frac{1}{3}<\frac{1}{2}<1$$

那么很容易证明

$$V=1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}$$

这个级数在前面证明过它是发散的。所以有

$$V_0^1(f)\to+\infty$$

定理 有界变差函数是有界的。

证明 对于 $a\leqslant x\leqslant b$，有

$$V=|f(x)-f(a)|+|f(b)-f(x)|\leqslant V_a^b(f)$$

从而得到$|f(x)|\leqslant|f(a)|+V_a^b(f)$，所以结论得证。

关于有界变差函数的性质还有如下一些结论成立，具体证明过程从略。

定理 两个有界变差函数之和、差、积仍然是有界变差的。

定理 设 $f(x)$和 $g(x)$都是有界变差的。若$|g(x)|\geqslant\sigma>0$，则 $f(x)/g(x)$也是有界变差的。

定理 设 $f(x)$是$[a,b]$上的有限函数，又 $a<c<b$，则 $V_a^b(f)=V_a^c(f)+V_c^b(f)$。

推论 设 $a<c<b$，如果 $f(x)$在$[a,b]$上是有界变差的，则 $f(x)$在$[a,c]$及$[c,b]$上也是有界变差的。该命题的逆命题也为真。

推论 若$[a,b]$可分为有限个部分，在每一个部分区间中 $f(x)$成为单调函数，则 $f(x)$在$[a,b]$上是有界变差的。

定理 函数 $f(x)$是有界变差的充分必要条件是 $f(x)$可以表示为两个增函数的差。该定理也称为若尔当(Jordan)分解定理。

证明 其充分性由前面给出的定理很容易推得，此处仅证明其必要性，令

$$\pi(x) = V_a^x(f), \quad a < x \leqslant b$$

$$\pi(a) = 0$$

显然，$\pi(x)$是一个增函数。令 $v(x)=\pi(x)-f(x)$，则可证明 $v(x)$也是增函数。这是因为，当 $a\leqslant x<y\leqslant b$ 时，可得

$$v(y) = \pi(y) - f(y) = \pi(x) + V_x^y(f) - f(y)$$

所以，$v(y)-v(x)=V_x^y(f)-[f(y)-f(x)]$。但是由全变差的定义，可知

$$f(y) - f(x) \leqslant V_x^y(f)$$

即有 $v(y)-v(x)\geqslant 0$，于是 $v(x)$是增函数。而 $f(x)=v(x)-\pi(x)$，即证明了其必要性。

推论 如果 $f(x)$在$[a,b]$上是有界变差的，则 $f'(x)$在$[a,b]$上几乎处处存在且为有限，并且 $f'(x)$在$[a,b]$上是可和的。

推论 有界变差函数的不连续点的全体至多是一个可数集。在每一个不连续点 x_0 存在着两个极限

$$f(x_0 + 0) = \lim_{x\to x_0} f(x), \quad x > x_0$$

$$f(x_0 - 0) = \lim_{x\to x_0} f(x), \quad x < x_0$$

设 $x_1,x_2,\cdots(a<x_n<b)$是 $\pi(x)$或 $v(x)$的不连续点的全体。做跳跃函数

$$s_\pi(x) = [\pi(a+0)-\pi(a)] + \sum_{x_k<x}[\pi(x_k+0)-\pi(x_k-0)] + [\pi(x)-\pi(x-0)], \quad a < x \leqslant b$$

$$s_v(x) = [v(a+0)-v(a)] + \sum_{x_k<x}[v(x_k+0)-v(x_k-0)] + [v(x)-v(x-0)]$$

$$s_\pi(a) = s_v(a) = 0$$

如果 x_k 是$\pi(x)$或 $v(x)$的连续点，那么 x_k 所对应的一项就化为 0。而且还要指出 $v(x)$的不连续点不可能是 $\pi(x)$的连续点，这一点这里不做赘述。

设 $s(x)=s_\pi(x)-s_v(x)$，则

$$s(x) = [f(a+0)-f(a)] + \sum_{x_k<x}[f(x_k+0)-f(x_k-0)] + [f(x)-f(x-0)], \quad a < x \leqslant b$$

$$s(a) = 0$$

$s(x)$也是一个有界变差函数，称为 $f(x)$的跳跃函数。显然，若从 $x_1,x_2,\cdots$中除去 $f(x)$的连续点，则 $s(x)$仍旧没有什么改变。所以，不妨设 $x_1,x_2,\cdots$中的所有点都是 $f(x)$的不连续点。而增函数 $f(x)$与其跳跃函数 $s(x)$的差是一个连续的增函数。因此，$\pi(x)-s_\pi(x)$和 $v(x)-s_v(x)$都是连续的增函数。由此便得到 $\varphi(x)=f(x)-s(x)$是一个连续的有界变差函数。换言之，也证明了如下这个定理。

定理 任意一个有界变差函数可表示为它的跳跃函数与一个连续的有界变差函数的和。

下面讨论更为复杂的情况——多变量的 BV 函数。

定义 令 Ω 是$\mathbb{R}^n$ 的一个开子集。如果存在一个有限的向量拉东(Radon)测度 $Du\in M(\Omega,\mathbb{R}^n)$使得如下等式成立

$$\int_{\Omega} u(x)\operatorname{div}\phi(x)\mathrm{d}x = -\int_{\Omega}\langle \phi, Du(x)\rangle, \quad \forall \phi \in C_C^1(\Omega,\mathbb{R}^n)$$

其中，$u\in L^1(\Omega)$，则函数 u 就是一个有界变差函数，并记作 $u\in \mathrm{BV}(\Omega)$。

也就是说，u 在空间 $C_C^1(\Omega,\mathbb{R}^n)$ 上定义了一个线性泛函，$C_C^1(\Omega,\mathbb{R}^n)$ 表示由在 Ω 中紧支的（compact support）连续可微的向量函数 ϕ 所组成的函数空间。向量测度 Du 因此表示 u 的分布梯度或弱梯度。

上述定义涉及的陌生概念较多，所以给出如下这个等价的定义。

定义 给定一个函数 $u\in L^1(\Omega)$，那么 u 在 Ω 中的全变差就定义为

$$V(u,\Omega):=\sup\left\{\int_{\Omega} u(x)\operatorname{div}\phi(x)\mathrm{d}x:\ \phi \in C_C^1(\Omega,\mathbb{R}^n),\ \|\phi\|_{L^{+\infty}(\Omega)} \leqslant 1\right\}$$

其中，$\|\cdot\|_{L^{+\infty}(\Omega)}$ 是本性上确界的范数。有时，下面的记号也会被使用

$$\int_{\Omega} |Du| = V(u,\Omega)$$

这主要是为了强调 $V(u,\Omega)$ 是 u 分布梯度或弱梯度的全变差。这种记法也提醒人们，如果 u 源自于一个 C^1 空间，即一个连续可微且其一阶导数也连续的函数，那么它的变差就是其梯度的绝对值的积分。

有界变差函数空间可被定义为

$$\mathrm{BV}(\Omega) = \{u \in L^1(\Omega):\ V(u,\Omega) < +\infty\}$$

这两个定义是等价的，因为如果 $V(u,\Omega)<+\infty$，那么

$$\left|\int_{\Omega} u(x)\operatorname{div}\phi(x)\mathrm{d}x\right| \leqslant V(u,\Omega)\|\phi\|_{L^{+\infty}(\Omega)}, \quad \forall \phi \in C_C^1(\Omega,\mathbb{R}^n)$$

因此

$$\int_{\Omega} u(x)\operatorname{div}\phi(x)\mathrm{d}x$$

在空间 $C_C^1(\Omega,\mathbb{R}^n)$ 上定义了一个连续的线性泛函。而且因为 $C_C^1(\Omega,\mathbb{R}^n)\subset C_C^0(\Omega,\mathbb{R}^n)$，作为一个线性子空间，这个连续的线性泛函根据汉恩-巴拿赫定理可以被连续地、线性地扩展到整个 $C^0(\Omega,\mathbb{R}^n)$。即它定义了一个拉东测度（Radon measure）。

接下来，讨论另外一个概念——局部的 BV 函数。如果在前面的定义中所考虑的函数属于一个由局部可积函数组成的空间，即函数属于 $L_{\mathrm{loc}}^1(\Omega)$，而非是来自一个全局可积函数空间，那么如此被定义的函数空间就属于是局部有界变差函数空间。更准确地讲，一个局部变差可以被定义成如下形式：

$$V(u,U):=\sup\left\{\int_{\Omega} u(x)\operatorname{div}\phi(x)\mathrm{d}x:\ \phi \in C_C^1(U,\mathbb{R}^n),\ \|\phi\|_{L^{+\infty}(\Omega)} \leqslant 1\right\}$$

对于每一个集合 $U\in \mathcal{O}_C(\Omega)$，这里 $\mathcal{O}_C(\Omega)$ 表示关于有限维向量空间的标准拓扑 Ω 的所有准紧开子集的集合，那么相应地局部有界变差的函数族被定义成

$$\mathrm{BV}_{\mathrm{loc}}(\Omega) = \{u \in L_{\mathrm{loc}}^1(\Omega):\ V(u,U) < +\infty, \forall U \in \mathcal{O}_c(\Omega)\}$$

通常，用 $\mathrm{BV}(\Omega)$ 表示全局有界变差函数空间，并相对应的用 $\mathrm{BV}_{\mathrm{loc}}(\Omega)$ 表示局部有界变差函数空间；有时，也采用 $\overline{\mathrm{BV}}(\Omega)$ 表示全局有界变差函数空间，并相对应的采用 $\mathrm{BV}(\Omega)$ 表示局部有界变差函数空间。本书中采用第一种记法。

下面讨论一下有界变差函数的基本性质。注意，这里所说的基本性质是指单变量有界变差函数与多变量有界变差函数共有的一些性质。而下面所给出的证明主要是针对多变量

函数进行的，这是因为对于单变量的情况而言，其证明往往是多变量情况的一个简化版。此外，在每个部分还会指出具体的某个性质是否对局部有界变差函数同样适用。

首先，BV 函数仅有跳跃型间断点。对于单变量的情况，这个结论是很显然的：对于函数 u 的定义区间$[a,b]\subset R$ 上的每一点 x_0，下面的两个断言中必有一个是对的（当左右两个极限都存在而且是有限的时）

$$\lim_{x\to x_0-} u(x) = \lim_{x\to x_0+} u(x)$$

$$\lim_{x\to x_0-} u(x) \neq \lim_{x\to x_0+} u(x)$$

对于多变量函数的情况，有一些前提条件需要说明：有一个方向的连续统，沿着这些方向可以逼近属于域 $\Omega\subset\mathbb{R}^n$ 中的一个给定点 x_0。有必要精确地给极限下一个合适的概念。选取一个单位向量$\hat{a}\in\mathbb{R}^n$，它可以将 Ω 划分成两个集合

$$\Omega_{(\hat{a},x_0)} = \Omega \cap \{x\in\mathbb{R}^n \mid \langle x-x_0,\hat{a}\rangle > 0\}$$

$$\Omega_{(-\hat{a},x_0)} = \Omega \cap \{x\in\mathbb{R}^n \mid \langle x-x_0,-\hat{a}\rangle > 0\}$$

那么，对于 BV 函数 u 的定义域 $\Omega\in\mathbb{R}^n$ 中的每一点 x_0，下面的两个断言中仅有一个是正确的

$$\lim_{\substack{x\to x_0\\ x\in\Omega_{(\hat{a},x_0)}}} u(x) = \lim_{\substack{x\to x_0\\ x\in\Omega_{(-\hat{a},x_0)}}} u(x)$$

$$\lim_{\substack{x\to x_0\\ x\in\Omega_{(\hat{a},x_0)}}} u(x) \neq \lim_{\substack{x\to x_0\\ x\in\Omega_{(-\hat{a},x_0)}}} u(x)$$

或者 x_0 属于含有零个 $n-1$ 维的豪斯多夫测度（Hausdorff measure）的 Ω 一个子集。如下的量

$$\lim_{\substack{x\to x_0\\ x\in\Omega_{(\hat{a},x_0)}}} u(x) = u_{\hat{a}}(x_0), \qquad \lim_{\substack{x\to x_0\\ x\in\Omega_{(-\hat{a},x_0)}}} u(x) = u_{-\hat{a}}(x_0)$$

就被称为是 BV 函数 u 在点 x_0 处的近似极限。

其次，$V(\cdot,\Omega)$在 $\mathrm{BV}(\Omega)$上是下半连续的。泛函 $V(\cdot,\Omega)$：$\mathrm{BV}(\Omega)\to\mathbb{R}^+$ 是下半连续的，为了说明这一点，选取一个 BV 函数的柯西序列$\{u_n\}$收敛于 $u\in L^1_{\mathrm{loc}}(\Omega)$，其中 $n\in\mathbb{N}$。因为所有序列中的函数以及它们的极限函数都是可积的，并且根据下限的定义，对于 $\forall\phi\in C^1_C(\Omega,\mathbb{R}^n)$，$\|\phi\|_{L^{+\infty}(\Omega)}\leqslant 1$ 有

$$\liminf_{n\to+\infty} V(u_n,\Omega) \geqslant \liminf_{n\to+\infty}\int_\Omega u_n(x)\mathrm{div}\phi\mathrm{d}x \geqslant \int_\Omega \lim_{n\to+\infty} u_n(x)\mathrm{div}\phi\mathrm{d}x = \int_\Omega u(x)\mathrm{div}\phi\mathrm{d}x$$

现在考虑在函数 $\phi\in C^1_C(\Omega,\mathbb{R}^n)$的集合上的上确界，可知 $\|\phi\|_{L^{+\infty}(\Omega)}\leqslant 1$，那么有下列不等式成立

$$\liminf_{n\to+\infty} V(u_n,\Omega) \geqslant V(u,\Omega)$$

这也就是下半连续的准确定义。

其次，有界变差函数空间 $\mathrm{BV}(\Omega)$是一个巴拿赫空间。根据定义，$\mathrm{BV}(\Omega)$是 $L^1(\Omega)$的一个子集，而线性性质可以从积分的线性属性中得到，即

$$\begin{aligned}\int_{\Omega}[u(x)+v(x)]\operatorname{div}\phi(x)\mathrm{d}x &= \int_{\Omega}u(x)\operatorname{div}\phi(x)\mathrm{d}x+\int_{\Omega}v(x)\operatorname{div}\phi(x)\mathrm{d}x \\ &= -\int_{\Omega}\langle\phi(x),Du(x)\rangle-\int_{\Omega}\langle\phi(x),Dv(x)\rangle \\ &= -\int_{\Omega}\langle\phi(x),[Du(x)+Dv(x)]\rangle\end{aligned}$$

对于所有的 $\phi\in C_C^1(\Omega,\mathbb{R}^n)$成立。因此，对于所有的 $u,v\in \mathrm{BV}(\Omega)$，有 $u+v\in \mathrm{BV}(\Omega)$成立。并且对于所有的 $c\in\mathbb{R}$，还有下式成立

$$\int_{\Omega}c\cdot u(x)\operatorname{div}\phi(x)\mathrm{d}x = c\int_{\Omega}u(x)\operatorname{div}\phi(x)\mathrm{d}x = -c\int_{\Omega}\langle\phi(x),Du(x)\rangle$$

因此，对于所有的 $u\in \mathrm{BV}(\Omega)$，以及 $c\in\mathbb{R}$，有 $cu\in \mathrm{BV}(\Omega)$成立。上述这些被证明的向量空间属性表明 $\mathrm{BV}(\Omega)$是 $L^1(\Omega)$的一个向量子空间。

现在考虑函数 $\|\ \|_{\mathrm{BV}}:\mathrm{BV}(\Omega)\to\mathbb{R}^+$，它的定义形式如下

$$\|u\|_{\mathrm{BV}}:=\|u\|_{L^1}+V(u,\Omega)$$

其中，$\|\ \|_{L^1}$是通常的 $L^1(\Omega)$的范数，很容易证明它是在 $\mathrm{BV}(\Omega)$上的一个范数。为了说明 $\mathrm{BV}(\Omega)$是一个巴拿赫空间，考虑在 $\mathrm{BV}(\Omega)$中的一个柯西序列$\{u_n\}$，其中 $n\in\mathbb{N}$。根据定义它也是 $L^1(\Omega)$中的一个柯西序列，它在 $L^1(\Omega)$中有一个极限 u 存在。因为 u_n 在 $\mathrm{BV}(\Omega)$中对于每一个 n 来说都是有界的，那么 $\|u\|_{\mathrm{BV}}<+\infty$。根据变差 $V(\cdot,\Omega)$的下半连续性，所以 u 是一个 BV 函数。最后，再由下半连续性，选择一个任意小的正数 ε，则有

$$\|u_j-u_k\|_{\mathrm{BV}}<\varepsilon,\forall j,k\geqslant N\in\mathbb{N}\Rightarrow V(u_k-u,\Omega)\leqslant\liminf_{j\to+\infty}V(u_k-u_j)\leqslant\varepsilon$$

此外，$\mathrm{BV}(\Omega)$是不可分的。为了说明这一点，考虑下面这个位于空间 $\mathrm{BV}([0,1])$中的例子，对于每一个 $0<\alpha<1$，定义

$$\chi_\alpha=\chi_{[\alpha,1]}=\begin{cases}0, & x\notin[\alpha,1]\\ 1, & x\in[\alpha,1]\end{cases}$$

为左闭区间$[\alpha,1]$上的指示函数。选取 $\alpha,\beta\in[0,1]$，且 $\alpha\neq\beta$，那么则有下述关系成立

$$\|\chi_\alpha-\chi_\beta\|_{\mathrm{BV}}=2+|\alpha-\beta|$$

现在为了证明 $\mathrm{BV}([0,1])$的每一个稠密子集都不可能是可数的，不妨从下面这个角度考察。对于每一个 $\alpha\in[0,1]$，可以构建一些球

$$B_\alpha=\{\psi\in \mathrm{BV}([0,1]);\ \|\chi_\alpha-\psi\|_{\mathrm{BV}}\leqslant 1\}$$

显然，这些球是两两不相交的，而且它们还是一个集的加标族，其指标集是$[0,1]$。这其实暗示这个族具有连续统的势。如此一来，因为 $\mathrm{BV}([0,1])$的任意稠密子集必须至少有一点在这个族的每个成员里，它的势至少为连续统的势，因此不可能是一个可数集。这个例子可以很显然地扩展到高维的情况，而且因为仅仅涉及局部属性，所以它也表明同样的性质对于 $\mathrm{BV}_{\mathrm{loc}}$也是成立的。

上述描述中涉及一些集合论的内容，在此稍作说明。以集合为元素的集合称为集族

(collection of sets)，记为$\mathcal{A}$。设$\mathcal{A}$是一个非空集族，$\mathcal{A}$的指标函数(indexing function)是从某一个集合J到$\mathcal{A}$的一个满射f，其中J称为指标集(index set)，族$\mathcal{A}$连同指标函数f一起称为一个集的加标族(indexed family of sets)或加标集族。给定$\alpha\in J$，集合$f(\alpha)$记成符号A_α。该加标集族本身则记作$\{A_\alpha\}_{\alpha\in J}$，读作"$\alpha$取遍$J$时，所有$A_\alpha$的族"。当指标集自明时，则简单地记为$\{A_\alpha\}$。

而且，BV(Ω)是一个巴拿赫代数。这个性质从BV(Ω)不仅是一个巴拿赫空间还是一个结合代数(associative algebra)这个事实就可直接得到。结合代数是指一个向量空间，其允许向量有具分配律和结合律的乘法。因此，它是一个特殊的代数。这也暗示如果$\{v_n\}$和$\{u_n\}$是BV函数的柯西序列而且分别收敛到BV(Ω)中的函数v和u，那么

$$\left.\begin{array}{l} vu_n \underset{n\to+\infty}{\longrightarrow} vu \\ v_n u \underset{n\to+\infty}{\longrightarrow} vu \end{array}\right\} \Leftrightarrow vu \in \mathrm{BV}(\Omega)$$

因此，两个函数的普通逐点乘积在空间BV(Ω)中关于每个参数都是连续的。这就使得该函数空间成为一个巴拿赫代数。关于逐点乘积这个概念，此处稍作说明。如果f和g都是函数$f,g: X\to Y$，那么对于每个X中的x，逐点乘积$(f\cdot g): X\to Y$就被定义成$(f\cdot g)(x)=f(x)\cdot g(x)$。前面所说的参数就是指这里的$x$，也就是说$(f\cdot g)(x)$在BV($\Omega$)中是连续的。

索伯列夫空间$W^{1,1}(\Omega)$是BV(Ω)的一个真子集。事实上，对于每个在空间$W^{1,1}(\Omega)$中的u，可以选择一个测度$\mu:=\nabla u\ \mathcal{L}$，其中$\mathcal{L}$是在$\Omega$上的勒贝格测度。如此，即有下列等式成立

$$\int u\mathrm{div}\phi = -\int \phi \mathrm{d}\mu = -\int \phi\ \nabla u, \quad \forall\ \phi \in C_C^1$$

因为它只不过是弱微分的定义，所以等式是成立的。弱微分(weak derivative)是一个函数的微分(强微分)概念的推广，它可以作用于那些勒贝格可积的函数，而不必预设函数的可微性(事实上大部分可以弱微分的函数并不可微)。

很容易找到一个不是$W^{1,1}$的BV函数的例子，在一维情况下，任何带有非平凡跳跃(non-trivial jump)的阶梯函数都是。回忆函数间断点的分类。通常当人们说到函数间断点的类型时，如果按照间断点处的左右极限是否存在来划分，那么可以分为第一类间断点和第二类间断点。其中，如果间断点处的左右极限都存在，这个间断点就是第一类间断点。第一类间断点又分为可去间断点和跳跃间断点两种。如果间断点处的左右极限至少有一个不存在，那么则称该点为函数的第二类间断点。从另外一个角度也可以分成平凡间断点和非平凡间断点。其中，前面提及的可去间断点又称为平凡间断点。当函数在间断点处的极限存在，但此极限不等于该点处的函数值时，这就是一个可去间断点。显然，非平凡间断点包含了跳跃间断点和第二类间断点。如果函数在间断点处的左右极限存在，但是左右极限却不相等，则该间断点就是一个跳跃间断点。如果非平凡间断点特指跳跃间断点，有时也说非平凡跳跃间断点(non-trivial jump discontinuity)。阶梯函数是具有非平凡跳跃间断点的典型例子。

本章参考文献

[1] 钱伟长.变分法及有限元[M].北京：科学出版社，1980.

[2] 龚怀云，寿纪麟，王绵森.应用泛函分析[M].西安：西安交通大学出版社，1985.

[3] 柳重堪.应用泛函分析[M].北京：国防工业出版社，1986.

[4] 刘诗俊.变分法、有限元法和外推法[M].北京：中国铁道出版社，1986.

[5] 老大中.变分法基础[M].2版.北京：国防工业出版社，2007.

[6] 贾正华.Gram矩阵及其行列式[J].安庆师范学院学报(自然科学版).1998(3).

[7] 邓志颖，潘建辉.巴拿赫不动点定理及其应用[J].高等数学研究.2013(4).

[8] 佘守宪，唐莹.浅析物理学中的旋轮线(摆线)[J].大学物理.2001(6).

[9] 克莱鲍尔.数学分析[M].庄亚栋，译.上海：上海科学技术出版社，1981.

[10] 那汤松.实变函数论[M].5版.徐瑞云，译.北京：高等教育出版社，2010.