高等学校教材

最优化理论和方法

秦寿康 编著

電子工業出版社

最优化理论和方法

秦寿康　编著

電子工業出版社

内容简介

本书旨在对本世纪四十年代末至七十年代最优化的基本理论和方法作一概括，它以泛函分析和凸分析为数学工具，统一处理静态和动态最优化问题，即规划和控制的最优化问题。大体上包含三部分内容：第一部分为数学基础，第二部分为有限维约束极值的理论和方法，第三部分为无限维约束极值的理论和方法。在选材方面，力图体现三个面向，反映最优化理论和方法的当前水平。

本书可作为自动控制、系统工程、经济管理等专业大学高年级学生的选修教材和研究生教材，以及应用数学专业大学生有关课程的教学参考书，也可供上述领域的科学技术工作者和大学教师阅读。

最优化理论和方法

秦寿康　编著

责任编辑　王昌铭

•

电子工业出版社出版（北京市万寿路）

北京密云卫新综合印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

•

开本：787×1092毫米 1/16　印张：19.125　字数：465千字

1986年11月第1版　　1986年11月第1次印刷

印数：0,001—7000册　　定价：3.15 元

统一书号：15290·412

出 版 说 明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校工科电子类专业课教材的编审、出版的组织工作。从一九七七年底到一九八二年初，由于各有关院校，特别是参与编审工作的广大教师的努力和有关出版社的紧密配合，共编审出版了教材 159 种。

为了使工科电子类专业教材能更好地适应社会主义现代化建设培养人才的需要，反映国内外电子科学技术水平，达到“打好基础、精选内容、逐步更新、利于教学”的要求，在总结第一轮教材编审出版工作经验的基础上，电子工业部于一九八二年先后成立了高等学校《无线电技术与信息系统》、《电磁场与微波技术》、《电子材料与固体器件》、《电子物理与器件》、《电子机械》、《计算机与自动控制》、中等专业学校《电子类专业》、《电子机械类专业》共八个教材编审委员会，作为教材工作方面的一个经常性的业务指导机构，并制定了一九八二～一九八五年教材编审出版规划。列入规划的教材、教学参考书、实验指导书等共 217 种选题。在努力提高教材质量，适当增加教材品种的思想指导下，这一批教材的编审工作由编审委员会直接组织进行。

这一批教材的书稿，主要是从通过教学实践、师生反映较好的讲义中评选择优和从第一轮较好的教材中修编产生出来的。广大编审者，各编审委员会和有关出版社都为保证和提高教材质量作出了努力。

这一批教材，分别由电子工业出版社、国防工业出版社、上海科学技术出版社、西北电讯工程学院出版社、湖南科学技术出版社、江苏科学技术出版社、黑龙江科学技术出版社和天津科学技术出版社承担出版工作。

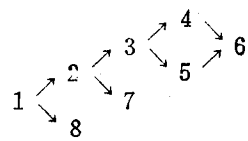
限于水平和经验，这一批教材的编审出版工作肯定还会有许多缺点和不足之处，希望使用教材的单位、广大教师和同学积极提出批评建议，共同为提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前　　言

本书旨在对本世纪四十年代末至七十年代确定型系统中最优化的基本理论和方法作一概括，并为后续的研究打下一个基础，具体说来就是：以泛函分析和凸分析为数学工具，统一处理静态和动态最优化问题，即规划和控制的最优化问题。

本书内容大体上分成三部分，第一部分（第一、二章）为数学基础，即泛函分析和凸分析；第二部分（第三至六章及第八章的一部分）为有限维空间中约束极值的理论和方法，其中包括线性规划、非线性规划、离散系统极值控制和离散型动态规划；第三部分（第七章和第八章的另一部分）为无限维空间中约束极值的理论和方法，即连续系统极值控制和连续型动态规划。各章的相互关系见附图。书末附录为 Banach 空间微分学。



在选材方面，力图体现三个面向（面向现代化，面向世界，面向未来），反映最优化理论和方法的当前水平。主要表现在：第一，有目标地加强数学基础，即以约束极值统一理论为主要线索，将泛函分析和凸分析融成一体；第二，着眼于揭示各种典型方法的实质及其内部联系；第三，表述方式尽可能简明易懂，如加强与微积分的联系，自行设计一些图形等。

北京工业学院吴沧浦教授审阅了全稿，提出了许多宝贵意见；中国科学院学部委员、上海交通大学张钟俊教授给予了多方面的关怀和帮助，深致谢忱！

由于作者学识不足，加诸编写时间急促，书中必有错漏之处，敬请读者不吝赐教，作者将竭诚欢迎！

编著者谨识　一九八五年八月于武汉

目　录

绪论……1

第一章　泛函分析导论……5
§1　度量空间的基本概念……5
§2　度量空间中的点集……9
1　开集……9
2　闭集……10
3　稠密性……12
4　点集间的距离……13
§3　连续映射……14
1　映射的概念和基本性质……14
2　连续映射……15
§4　度量空间的完备性……16
§5　列紧集……18
1　列紧集和完全有界集……19
2　紧集……23
3　紧集上连续映射和连续映射序列的重要性质……25
§6　赋范线性空间……26
1　赋范线性空间的概念……26
2　有界线性算子……28
§7　内积空间……32
1　内积空间的概念……32
2　直交投影……34
3　内积空间中的直交系……37
4　黎茨表示定理……41
§8　拓扑空间……42
1　拓扑空间……42
2　线性拓扑空间……44

第二章　凸集和凸泛函基本理论……48
§1　线性流形……48
1　超平面……48
2　线性流形……49
§2　凸集……51

1 凸集及其性质 ……51
2 集的凸包 ……54
§3 Hahn Banach 定理 ……58
1 半范数 ……58
2 Hahn-Banach定理 ……60
§4 凸集分离定理 ……65
1 E^n空间中的凸集分离定理 ……66
2 无限维空间中的凸集分离定理 ……69
§5 凸锥与极锥 ……71
1 凸锥及其性质 ……71
2 极锥及其性质 ……73
3 凸锥分离定理 ……75
§6 凸泛函 ……76
1 凸泛函及其性质 ……76
2 E^n空间中可微凸函数的性质 ……79
3 半连续泛函 ……82
4 凸泛函的基本定理 ……84
§7 不动点定理 ……87
1 Brouwer不动点定理 ……87
2 Schauder不动点定理 ……92

第三章 有限维空间中约束极值的统一理论 ……96
§1 导锥 ……96
1 等式约束集的导锥 ……97
2 不等式约束集的导锥 ……99
3 凸集的导锥 ……102
§2 E^n 空间中约束极值的必要条件 ……103
§3 E^n 空间中约束极值的充分条件 ……110
§4 Lagrange 乘子理论 ……113

第四章 线性规划 ……116
§1 线性规划模型的标准形式及其解的性质 ……116
1 线性规划模型的标准形式 ……116
2 线性规划问题解的基本性质 ……118
§2 单纯形法 ……120
1 单纯形法 ……120
2 应用举例 ……124
3 求解退化线性规划问题的摄动法 ……132
§3 线性规划的对偶理论 ……133

1 原问题与对偶问题的关系及其基本性质 ……133
2 对偶单纯形法 ……138
§4 有界变数的线性规划 ……141
§5 用约束极值的统一理论验证线性规划模型的最优解 ……143
1 用Lagrange乘子法验证线性规划问题的最优解 ……143
2 用约束极值统一理论验证线性规划问题的最优解 ……145

第五章 非线性规划 ……146
§1 非线性规划模型的最优性条件 ……146
1 非线性规划模型的一阶最优性条件 ……146
2 非线性规划模型的二阶最优性条件 ……150
§2 非线性规划的对偶理论 ……151
1 Lagrange对偶问题及其涵义 ……151
2 对偶性定理 ……152
3 鞍点定理 ……155
§3 算法的概念 ……157
1 算法的基本概念 ……157
2 收敛定理 ……158
3 评价算法的几个要素 ……159
§4 无约束极值问题的算法 ……160
1 一维搜索 ……161
2 多维搜索 ……165
3 共轭梯度法 ……167
4 变尺度法 ……170
§5 惩罚函数法和障碍函数法 ……177
1 惩罚函数法 ……177
2 障碍函数法 ……182
§6 可行方向法 ……187
1 可行方向法 ……187
2 梯度投影法 ……190

第六章 离散系统极值控制 ……195
§1 离散系统极值控制的一般理论 ……195
§2 线性二次型最优控制系统 ……203
1 无约束控制问题 ……203
2 约束控制问题 ……205
§3 最短时间控制系统 ……210

第七章 连续系统极值控制 ……216

§1 泛函约束极值的必要条件 ……216
1 泛函约束极值的一般性必要条件 ……216
2 泛函正则性约束极值的基本定理 ……222
§2 最大值原理 ……227
1 连续函数线性空间的拓扑及本性有界可测控制 ……227
2 一阶凸近似及微分 ……228
3 极值控制问题与最大值原理 ……230
§3 时间极值控制系统 ……238
1 一类非线性系统的时间极值控制 ……238
2 双积分装置的最短时间控制系统 ……241
3 简谐振荡器的最短时间控制系统 ……243
§4 燃料极值控制系统 ……247
1 一类非线性系统的燃料极值控制 ……247
2 双积分装置的时间和燃料综合最优控制系统 ……250
§5 线性二次型最优控制系统 ……253
1 状态调节器问题 ……254
2 输出调节器问题 ……258
3 跟踪问题 ……259

第八章 动态规划 ……262
§1 基本概念 ……262
§2 最优性原理和逆序递推法 ……266
§3 哈密顿-雅谷比-贝尔曼方程 ……270
§4 离散极值控制系统 ……274
§5 资源分配问题 ……278
1 单种资源分配问题 ……278
2 多种资源分配问题 ……279

附录 Banach 空间微分学 ……283
1 连续映射的导数 ……283
2 偏导数 ……288
3 高阶导数 ……290
主要参考文献 ……293
名词索引 ……297

绪　论

最优化理论和方法，是现代科学技术发展的综合产物，而归根结蒂，是生产力发展到大工业生产后的产物。

微分学的创立，从根本上推动了极值问题的研究。设 $f^0(\cdot)$ 是在E^n空间中某开集X上定义的连续可微函数，$\hat{x}$是f^0的一个极值点，则按 f^0在$\hat{x}$的驻点性质，可得极值解的必要条件

$$\frac{\partial f^0(\hat{x})}{\partial x} = \nabla f^0(\hat{x}) = 0 \tag{1}$$

于是极值问题就归结为解方程，而求极值解的必要条件也就成为最优化理论的一个中心课题。

极值问题的进一步研究导致等式约束极值问题：

$$\min f^0(x) \tag{2}$$

约束条件为

$$f(x) = 0 \tag{3}$$

其中$f^0 \in E^1$和$f \in E^l$都是在开集$X \subset E^n$上定义的连续可微函数，且$l < n$。这一问题可以用拉格朗日(Lagrange)乘子法求解：设$\hat{x}$是等式约束极值问题的一个极小值点，则必存在$1+l$个不全为零的实数 λ^0，$\lambda=(\lambda^1, \cdots, \lambda^l)$，使得拉格朗日函数

$$L(x, \lambda^0, \lambda) = \lambda^0 f^0(x) + \lambda f(x)$$

有一驻点$\hat{x}$，即成立

$$\frac{\partial L(\hat{x})}{\partial x} = 0 \tag{4}$$

变分学起源于一些实际问题（如最速降线问题）[78] 它研究的是如下的泛函极值问题：

$$\min J(x) = \int_{t_0}^{t_f} f^0(x(t), \dot{x}(t), t)dt \tag{5}$$

边界条件为

$$x(t_0) = x_0, \quad x(t_f) = x_f \tag{6}$$

其中 $x \in E^n$ 和 $f^0 \in E^1$对其变元两次连续可微。变分学研究的基本方法，是微分法的直接推广：设$\hat{x}(t)$是经x_0和x_f的极值轨线，又

$$x(t) = \hat{x}(t) + \delta x(t)$$

是$\hat{x}(t)$的邻近轨线，其中 $\delta x(t)$ ($\delta x(t_0)=0$，$\delta x(t_f)=0$)是轨线 $\hat{x}(t)$ 的一阶变分，通过一些简易运算，便可获得泛函达极值的必要条件：

$$\frac{\partial f^0}{\partial x} - \frac{d}{dt}\frac{\partial f^0}{\partial \dot{x}} = 0 \tag{7}$$

这就是著名的欧拉-拉格朗日(Euler-Lagrange)方程。

由等周问题[78]引起对微分方程约束下泛函极值问题，即拉格朗日问题的研究：

$$\min J(x)=\int_{t_0}^{t_f} f^0(x(t),\dot{x}(t),t)dt$$
$$x(t_0)=x_0,\quad x(t_f)=x_f \tag{8}$$

约束条件为

$$\varphi(x(t),\dot{x}(t),t)=0 \tag{9}$$

其中$\varphi\in E^l$，$l<n$。这一问题的结果归结为下面的乘子法则：设 $\hat{x}(t)$ 是拉格朗日问题的极值轨线，则存在实数λ^0和l维函数 $\lambda(t)=(\lambda^1(t),\ \cdots,\ \lambda^l(t))$，$|\lambda^0|+\|\lambda(t)\|\neq 0$，使得拉格朗日函数

$$L(x(t),\dot{x}(t),\ t,\ \lambda^0,\ \lambda(t))=\lambda^0 f^0(x(t),\dot{x}(t),t)+\lambda(t)\varphi(x(t),\dot{x}(t),t)$$

在轨线$\hat{x}(t)$上满足欧拉-拉格朗日方程

$$\frac{\partial L}{\partial x}-\frac{d}{dt}\ \frac{\partial L}{\partial \dot{x}}=0 \tag{10}$$

随后，又出现了维尔斯特拉斯(Weierstrass)和马克申（Mcshane）的工作。维尔斯特拉斯针对拉格朗日问题乘子法则中$\lambda^0=1$的情形，推导出维尔斯特拉斯条件：设 $\hat{x}(t)$ 是拉格朗日问题当$\lambda^0=1$时的极值轨线，则维尔斯特拉斯E函数非负：

$$E(x(t),\dot{x}(t),\ \dot{\hat{x}}(t),\ t,\ \lambda(t))=L(x(t),\dot{x}(t),\ t,\ \lambda(t))$$
$$-L(x(t),\ \dot{\hat{x}}(t),t,\lambda(t))-(\dot{x}(t)-\dot{\hat{x}}(t))\frac{\partial L(x(t).\ \dot{\hat{x}}(t),t,\lambda(t))}{\partial \dot{\hat{x}}}\geqslant 0 \tag{11}$$

E 函数判别式（11）把拉格朗日问题从弱极值推进到强极值的情形[58]。马克申在他的著作中[79]，首次运用了对现代最优控制理论有着决定意义的方法初步，他借助于以“针状变分”著称的新型变分和凸锥分离定理，证明了对任意拉格朗日问题的解的维尔斯特拉斯条件。

以上就是截至本世纪三十年代末最优化理论和方法的一个概貌。

四十年代末期至六十年代中期，是各种最优化理论和方法蓬勃兴起的年代，它们大体上是循着规划（即静态最优化）和控制（即动态最优化）两个方向进行的。

三十至四十年代，由于生产、流通和军事上的迫切需要，提出了两类数学规划问题，一类是线性规划问题：

$$\min f^0(x)=cx \tag{12}$$

约束条件为

$$Ax=b,\ \ x\geqslant 0 \tag{13}$$

其中 $x\in E^n$，$c\in E^n$，$b\in E^m$，A为$m\times n$矩阵；另一类是非线性规划问题：

$$\min f^0(x) \tag{14}$$

约束条件为

$$f(x)=0$$
$$g(x)\leqslant 0 \tag{15}$$

其中$x\in E^n$，$f^0: E^n\to E^1$，$f: E^n\to E^l$，$g: E^n\to E^m$ 都是连续可微函数。这两类问题与等式约束极值问题的本质差异，在于这里函数f^0的约束集为闭集，因而不能用拉格朗日乘子法求解。于是从丹捷克(Dantzig)和库恩-塔克(Kuhn-Tucker)开始，提出了一系列极值解的必要条件，并导致数学规划某些更细分支（如凸规划、二次规划等）的建立。

电子计算机的问世，在数学与现代科学技术之间架起了一座桥梁，对最优化方法的发展有着本质的影响，而基于迭代原理的各种数值方法，如单纯形法、共轭方向法、惩罚函数法等，就象雨后春笋般地产生了。

第二次世界大战后，由于航天事业的迫切需求，提出了一类极值控制问题。这类问题的一个共同特点，是它们的控制集或决策集为闭集，因而不能用经典变分法求解，动态规划和最大值原理就是为着解决这类问题而创立的。

动态规划是研究决策过程最优化的一种理论和方法，它最初应用于时间离散问题，即多阶段决策问题，随后又推广到了时间连续问题。五十年代初，贝尔曼(Bellman)针对一种无后效的时间离散问题，提出了一条朴素的原理——最优性原理，并将这一原理运用于上述问题，推导出贝尔曼递推方程，尔后，贝尔曼又将最优性原理应用于时间连续问题，把哈密顿-雅谷比(Hamilton-Jacobi)理论推进了一步，导致哈密顿-雅谷比-贝尔曼方程的建立。

与动态规划相反，最大值原理最初应用于时间连续系统，随后又推广到了时间离散系统。连续系统的最大值原理，研究的是如下一类极值控制问题：设系统的运动方程是

$$\dot{x}(t)=f(x(t),u(t)) \tag{16}$$

其中$x\in E^n$为状态，$u\in U\subset E^r$为控制，$f:E^n\times E^r\to E^n$对u连续，对 x 为C^1类函数。寻找容许控制集U上逐段连续的极值控制 $\hat{u}(t)$，将状态从初态 $x(t_0)=x_0$转移到终态$x(t_f)=x_f$，并使泛函

$$J(u)=\int_{t_0}^{t_f} f^0((t),\ u(t))dt \tag{17}$$

极小。庞特里雅金（Понтлягин）等人采用与经典变分不同的方式，即针状变分和凸锥分离定理，在增广维数的有限维空间中证明了最大值条件[63]：

$$H(\hat{\psi}(t),\ \hat{x}(t),\ \hat{u}(t))=\max_{u(t)\in U} H(\hat{\psi}(t),\ \hat{x}(t),\ u(t)) \tag{18}$$

这一条件的极值具有整体特征，它不仅适用于U为开集的情形，而且也适用于U为闭集的情形，从而把变分学从经典变分阶段推进到了现代变分阶段。

综上所述，第二次世界大战后的二十年，是最优化理论和方法大发展的二十年。这一时期的基本特征，在理论上表现为凸分析的发现和采用，解决了约束集为闭集的种种极值问题；在方法上表现为基于迭代原理的各种数值方法的建立，解决了最优化理论到实际应用的问题。

六十年代中期至七十年代初期，是最优化理论和方法向纵深发展的时期。这一时期的主要成就，是两种并行的最优化统一理论——有限维空间中约束极值的统一理论和无限维空间中约束极值的统一理论——的建立。由前一种统一理论，可以推出有限维规划和离散系统极值控制的主要理论成果；而由后一种统一理论，可以推出无限维规划和连续系统极值控制的主要理论成果。统一理论的基本思想，是按照问题的需要，构造出若干个空间，并对每一空间中的约束集，在极值点的邻域构成各个凸锥，作为对原约束集的线性近似，然后运用凸集理论，依据各种凸锥中向量之间的相互关系，推出极值解的必要条件和充分条件。

1970年，米萨洛维奇(Mesarovic)等人仿照久已存在于人类社会乃至生物界的递阶组织结构形式，提出了一种新型系统——递阶系统，这类系统的构成原理乃是分解-协调原理，而其数学基础则是非线性规划对偶理论和最优控制理论。整个七十年代，人们不仅建立了多种递阶最优化算法[72]，较好地解决了大系统计算中出现的计算机存储容量及解题速度等方面

的矛盾，而且构成了一些实际的递阶控制系统[81]，从而把最优化理论和方法推进到了大系统阶段。

今天，最优化理论和方法已发展到一个空前的高度，并广泛应用于人类社会的众多领域。随着人们对客观世界认识的不断深化，势将把最优化理论和方法推向新的更高的阶段。

第一章　泛函分析导论

泛函分析是现代数学中一个较新的重要分支，它是研究无限维线性空间上的泛函数与算子的理论的一门分析数学。无限维线性空间是描述具无限多自由度的物理系统的数学工具。因此，泛函分析是定量地研究诸如连续介质力学、电磁场理论、连续控制系统等具无限多自由度的物理系统的有力工具。泛函分析起源于经典数学物理中的某些变分问题与边值问题，到本世纪三十年代即已初具规模，成为数学中一门独立的学科，并在四十至五十年代得到了大发展。今天，泛函分析的概念和方法已渗透到现代纯粹与应用数学、理论物理学和技术科学的许多分支，促进了这些学科的发展。

本章的目的是为研究最优化理论提供一个必要的数学背景，主要介绍度量空间和拓扑空间中的一些基本概念和常用的定理。

§1　度量空间的基本概念

距离是日常生活中最基本的概念之一。它具备三个简单属性：

（i）　任意两点之间的距离非负。

(ii）　两点间的距离与以哪一点作为计量的起点无关，即距离是标量函数。

(iii)　两点间的最短距离是连接这两点的直线段的长度。

回顾数学分析中常见的一些基本概念，如数列的极限，函数列的一致收敛等，都包含着某种距离渐趋于零的涵义。为了从更高的角度来概括并发展这些初看起来似乎各不相同的概念，结合距离三属性，我们引进抽象的距离和度量空间（也叫距离空间）的定义。

定义1　设R是一个非空集。在$R\times R$上定义一个实函数d，满足如下三条件，即对任意的x，y，$z\in R$，成立

（i）　非负性：$d(x,\ y)\geqslant 0$，且$d(x,\ y)=0$的充要条件是 $x=y$。

(ii）　对称性：$d(x,\ y)=d(y,\ x)$。

(iii)　三点不等式：$d(x,\ z)\leqslant d(x,\ y)+d(y,\ z)$。

$d(\cdot,\cdot)$ 称为R上的距离，而这种赋有距离意义的空间R，就叫做度量空间或距离空间，记作$(R,\ d)$，或简记为R①。

在引入了距离之后，就可进而建立度量空间中极限的概念。

定义2　度量空间的点列$\{x_n\}$叫做按距离$d(\cdot,\cdot)$收敛于点x_0，或称点列$\{x_n\}$以点 x_0为极限，是指当 $n\to\infty$时，有

$$d(x_n,\ x_0)\to 0$$

简记成

$$\lim_{n\to\infty} x_n=x_0\quad 或\quad x_n\to x_0\ (n\to\infty)$$

①今后，凡度量空间中的距离，都是指平移不变的距离，即具备性质$d(x,y)=d(x-y,0)$。

定理1 在度量空间中，任一点列$\{x_n\}$至多收敛于一点。如点列$\{x_n\}$收敛于点x_0，它的每个子列也收敛于x_0。

证 假若点列$\{x_n\}$既收敛于点x_0，又收敛于点y_0，则对任给的 $\varepsilon>0$，存在正整数 N_1 和 N_2，使当 $n\geqslant N_1$ 时，成立

$$d(x_n, x_0)<\frac{\varepsilon}{2}$$

而当$n\geqslant N_2$时，成立

$$d(x_n, y_0)<\frac{\varepsilon}{2}$$

令 $n\geqslant \max(N_1, N_2)$，依据三点不等式，可得

$$d(x_0, y_0)\leqslant d(x_0, x_n)+d(x_n, y_0)<\varepsilon$$

由ε的任意性，推出 $d(x_0, y_0)=0$，即$x_0=y_0$。

设$\{x_n\}$收敛于x_0，而$\{x_{n_k}\}$是$\{x_n\}$的某个子列，则对任给的 $\varepsilon>0$，存在正整数 N，使当 $n\geqslant N$ 时，有

$$d(x_n, x_0)<\varepsilon$$

相应地，存在某个下标k_0，使得$n_{k_0}>N$；而当下标$k>k_0$时，必有 $n_k>n_{k_0}>N$，并推出

$$d(x_{n_k}, x_0)<\varepsilon$$

即$x_{n_k}\to x_0$ $(n\to\infty)$。

定理2 距离$d(x, y)$是度量空间中两点 x，y 的连续函数，即若 $x_n\to x_0$，$y_n\to y_0$，则 $d(x_n, y_n)\to d(x_0, y_0)$。

证 依据三点不等式，我们有

$$d(x_n, y_n)\leqslant d(x_n, x_0)+d(x_0, y_0)+d(y_0, y_n)$$
$$d(x_0, y_0)\leqslant d(x_0, x_n)+d(x_n, y_n)+d(y_n, y_0)$$

综合上列二式，可得

$$|d(x_n, y_n)-d(x_0, y_0)|\leqslant d(x_n, x_0)+d(y_n, y_0)$$

在上式中令 $n\to\infty$，便得所需结果。

下面举几个重要的例子。

例1 E^n空间 在n维欧几里得空间E^n中，两点

$$x=(x^1, x^2, \cdots, x^n),\quad y=(y^1, y^2, \cdots, y^n)$$

的距离可定义成

$$d(x, y)=\sqrt{\sum_{i=1}^{n}(x^i-y^i)^2} \tag{1}$$

并称之为欧几里得距离。

现证 $d(x, y)$ 满足距离三条件。条件 (i)，(ii) 在这里显然成立。下证条件 (iii)。应用柯西 (Cauchy) 不等式

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2\leqslant\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

可推出

$$\sum_{i=1}^{n}(a_i+b_i)^2=\sum_{i=1}^{n}a_i^2+2\sum_{i=1}^{n}a_ib_i+\sum_{i=1}^{n}b_i^2$$

$$\leqslant\sum_{i=1}^{n}a_i^2+2\sqrt{\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right)}+\sum_{i=1}^{n}b_i^2$$

$$=\left(\sqrt{\sum_{i=1}^{n}a_i^2}+\sqrt{\sum_{i=1}^{n}b_i^2}\right)^2$$

令 $a_i=z^i-x^i$，$b_i=y^i-z^i$，其中 $z=(z^1, z^2, \cdots, z^n)$，并把这些关系代入上式，便知三点不等式成立。

还可以在E^n中给出距离的另一定义：

$$d_1(x, y)=\max_{1\leqslant i\leqslant n}|x^i-y^i| \tag{2}$$

不难证明，$d_1(x, y)$ 也满足距离三条件。

由此可见，即使是同一个空间，也可通过赋予不同的距离，来构成不同的度量空间。但在本例中，无论是(E^n, d)空间，还是(E^n, d_1)空间，从按距离收敛的角度来看，它们都等价于按每个坐标收敛。

例2 $C[a,b]$ 空间 用$C[a,b]$表示闭区间$[a, b]$上的连续函数全体。对$x(t)$，$y(t)\in C[a,b]$，定义距离

$$d(x, y)=\max_{t\in[a,b]}|x(t)-y(t)| \tag{3}$$

它满足距离三条件。

若函数列$\{x_n(t)\}\in C[a,b]$，则$\{x_n(t)\}$一致收敛于函数$x(t)\in C[a,b]$的充要条件，是

$$d(x_n, x)\to 0 \quad (n\to\infty) \tag{4}$$

这是因为，如果我们把$x_n(t)$和$x(t)$都看成是$C[a, b]$空间中运动着的点，那末(3)、(4)两式表明，在整个运动过程中，x_n和x之间的最大距离，将随着$n\to\infty$而趋于零，从而保证了每一瞬刻都有$\lim\limits_{n\to\infty}x_n(t)=x(t)$，而这正好表述了一致收敛的性质。按照传统的说法就是：对任意的$\varepsilon>0$，存在正整数N，使当$n>N$时，就有

$$\max_{t\in[a,b]}|x_n(t)-x(t)|<\varepsilon$$

这意味着对所有的$t\in[a,b]$，存在同一个正整数N，只要$n>N$，就成立$|x_n(t)-x(t)|<\varepsilon$。

例3① **L^p空间** 用$L^p[a,b]$表示闭区间$[a,b]$上勒贝格p方$(p\geqslant 1)$可积函数全体组成的空间，即对$x(t)\in L^p[a, b]$，成立

$$\int_a^b|x(t)|^p dt<\infty$$

对$x(t)$，$y(t)\in L^p[a, b]$，定义距离

$$d(x, y)=\left(\int_a^b|x(t)-y(t)|^p dt\right)^{\frac{1}{p}} \tag{5}$$

易证它满足距离三条件。下面验证第三个条件：由闵可夫斯基(Minkowski)不等式

---

① 例3、4详见[1]第四章§3。

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

立即推出

$$\begin{aligned}
d(x,z) &= \left(\int_a^b |x(t)-z(t)|^p\,dt\right)^{\frac{1}{p}} \\
&= \left(\int_a^b |[x(t)-y(t)]+[y(t)-z(t)]|^p\,dt\right)^{\frac{1}{p}} \\
&\leqslant \left(\int_a^b |x(t)-y(t)|^p\,dt\right)^{\frac{1}{p}} + \left(\int_a^b |y(t)-z(t)|^p\,dt\right)^{\frac{1}{p}} \\
&= d(x,y)+d(y,z)
\end{aligned}$$

因此，L^p按(5)定义的距离成为度量空间。

例 4　l^p空间　$l^p(p\geqslant 1)$是指由满足

$$\sum_{k=1}^{\infty} |x_k|^p < \infty$$

的实（或复）数列 $x=\{x_k\}$的全体组成的空间。类比L^p空间，定义l^p中的距离

$$d(x,y)=\left(\sum_{k=1}^{\infty} |x_k-y_k|^p\right)^{\frac{1}{p}} \tag{6}$$

则l^p 构成一度量空间。

例 5　s空间　设s为实数列$\{x_i\}$的全体组成的空间。对 $x=\{x_i\},y=\{y_i\}\in s$,定义距离

$$d(x,y)=\sum_{i=1}^{\infty} \frac{1}{2^i}\,\frac{|x_i-y_i|}{1+|x_i-y_i|} \tag{7}$$

它满足距离三条件。以下证三点不等式成立。从两个简单的事实，即函数

$$\frac{a}{1+a}=1-\frac{1}{1+a}$$

当 $a\in[0,\infty)$时是单调增函数，及 $|a+b|\leqslant |a|+|b|$，可以证明下列不等式：

$$\begin{aligned}
\frac{|a+b|}{1+|a+b|} &\leqslant \frac{|a|+|b|}{1+|a|+|b|} \\
&= \frac{|a|}{1+|a|+|b|}+\frac{|b|}{1+|a|+|b|} \\
&\leqslant \frac{|a|}{1+|a|}+\frac{|b|}{1+|b|}
\end{aligned}$$

在上式中令 $a=x_i-z_i$，$b=z_i-y_i$，就有

$$\frac{|x_i-y_i|}{1+|x_i-y_i|} \leqslant \frac{|x_i-z_i|}{1+|x_i-z_i|}+\frac{|z_i-y_i|}{1+|z_i-y_i|}$$

把这一关系式代入（7），便得三点不等式

$$d(x,y)\leqslant d(x,z)+d(z,y)$$

可以证明，s 空间中点列按距离收敛等价于按每个坐标收敛。

还可以举出许多度量空间的例子。但仅从上列数例，就足以看出距离的多样性。至于应当采用什么样的距离函数，主要取决于两个因素，一个是描述实际系统的函数类别，另一个是