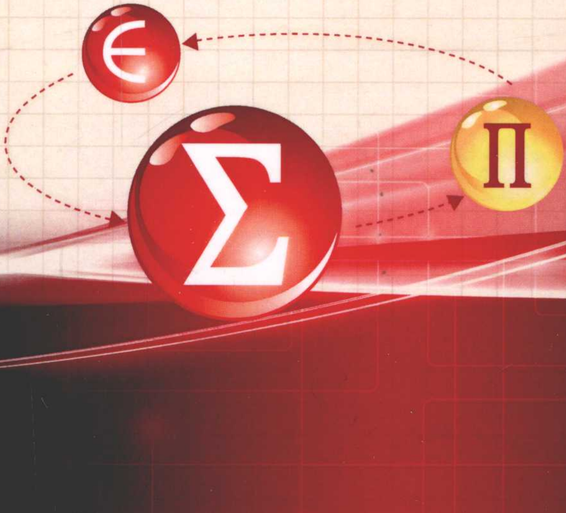
SUIJI GUOCHENG FENXI YU CHULI
—YINGYONG MAERKEFU GUOCHENG YANJIU

随机过程分析与处理

—应用马尔可夫过程研究

沈晋会 著



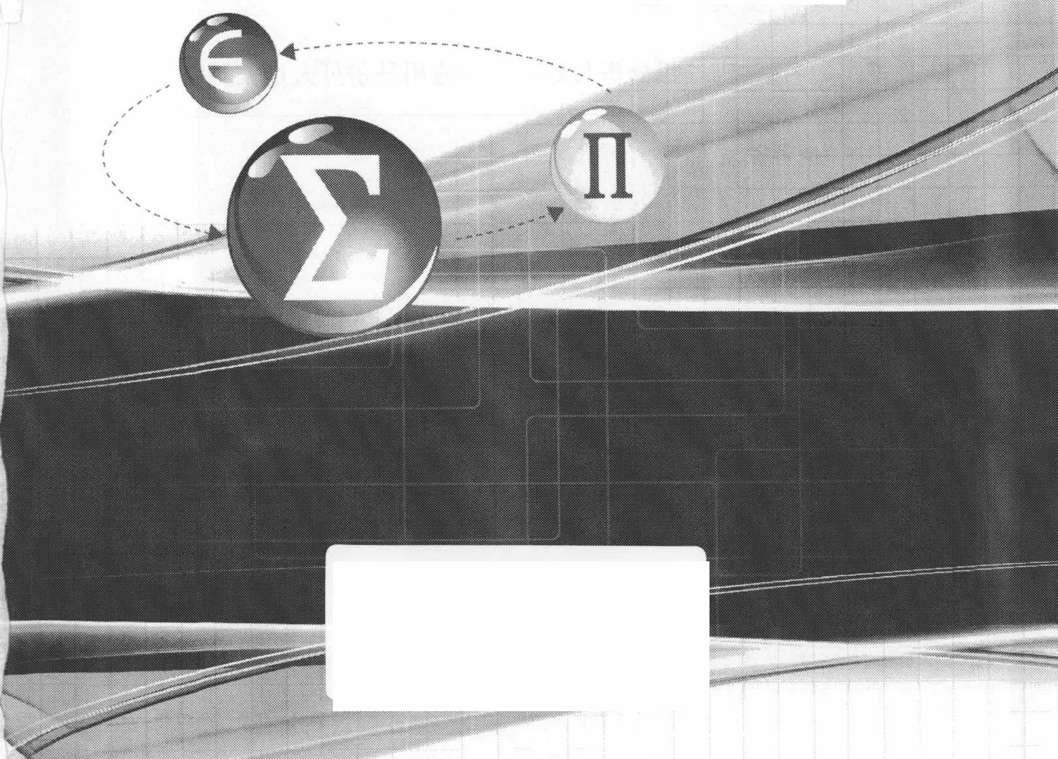
西安地图出版社

分析与处理

SUIJI GUOCHENG FENXI YU CHULI
—YINGYONG MAERKEFU GUOCHENG YANJIU

随机过程分析与处理
—应用马尔可夫过程研究

沈晋会　著



西安地图出版社

图书在版编目(CIP)数据

随机过程分析与处理 ：应用马尔可夫过程研究 / 沈晋会著. -- 西安 ：西安地图出版社，2011.12

ISBN 978-7-80748-802-6

Ⅰ.①随… Ⅱ.①沈… Ⅲ.①马尔柯夫过程—研究 Ⅳ.①O211.62

中国版本图书馆 CIP 数据核字(2011)第 268720 号

著作人及著作方式:沈晋会　著

责任编辑:史英

书　　名:随机过程分析与处理——应用马尔可夫过程研究

出版发行:西安地图出版社

地址邮编:西安市友谊东路 334 号　　710054

印　　刷:三河市铭浩彩色印装有限公司

规格开本:850 毫米×1168 毫米　1/32

印　　张:5.375

字　　数:116 千字

印　　数:0001～1000

版　　次:2011 年 12 月第 1 版　2011 年 12 月第 1 次印刷

书　　号:ISBN 978-7-80748-802-6

定　　价:28.00 元

西安地图出版社通过 ISO9001 国际质量管理体系认证

前　言

马尔可夫链是一种以概率论和随机过程理论为基础，运用随机数学模型来研究现实活动发展变化过程中的数量关系的统计方法．它是由马尔可夫在 1907 年提出的．他的重要发现和研究方法推动了概率论的发展，特别是促进了概率论新分支——随机过程论的发展．后人为了纪念他所作的卓越的贡献，故将他所研究的随机过程称为马尔可夫过程(Markov Process).

马尔可夫链是一个有着广泛应用的随机过程模型，它对一个系统由一种状态转移到另一种状态的现状提出了定量分析．许多经济和社会现象中的动态系统问题都可采用马尔可夫链来描述．近三十年来，在计算机、自动控制、通信、物理、化学、生物、经济和管理等领域有着重要的应用．理论上讲，具有 Markov 性的随机现象只要确定其状态空间和各状态之间的转移规律性，那么，Markov 模型就完全确定了．

本书共 6 章．第 1 章主要介绍了随机过程的一些基本概念与理论；第 2～3 章主要讨论了离散时间与连续时间 Markov 过程的基本理论及其在各个领域的应用，另外在第 3 章结束部分还简单地介绍了时间可逆的 Markov 过程；第 4 章引进

了一种新的统计模型——隐 Markov 模型，并介绍了隐 Markov 模型在语音识别中的应用；第 5 章介绍了 Markov 决策过程，该决策过程在实际中有着广泛应用；第 6 章介绍了基于 Markov 随机场的小波图像建模．

本书在撰写过程中得到了许多同行专家的支持和帮助，在此表示衷心的感谢；撰写时参阅了大量的著作和文献资料，很受教益．在此，谨向有关的作者致谢．

由于作者水平有限，书中难免存在不妥之处，恳请广大读者给予批评指正．

作者

2011.11

目　录

第1章 随机过程概论

通常将随机过程视为概率论的动态部分．概率论中研究的随机现象都是一个或有限多个随机变量的规律性,即使中心极限定理也不过是对随机变量序列的讨论．但在实际问题中,我们还需要研究一些随机现象的发展和变化过程,即随时间不断变化的随机变量,而且所涉及的随机变量通常是无限多个,这就是随机过程的研究对象．本章即对随机过程作简单的概述．

1.1 随机变量及其概率分布

1.1.1 概率空间

概率论的一个基本概念是随机试验,它有如下特征：

(1)试验可在相同条件下重复进行；

(2)每次试验只有一个结果出现且结果事前不能预知；

(3)每次试验所有可能出现的结果已知．

随机试验所有可能出现的结果所组成的集合称为试验的样本空间或基本事件空间,记作 Ω. Ω 中的元素称为样本点或

基本事件,用 ω 表示．而 Ω 中的子集称为事件,常用大写字母 A,B,C 等表示．

在实际问题中,有时并非对样本空间 Ω 的所有子集都感兴趣,只是关心 Ω 的一些子集(事件)及其在一次试验中发生的可能性大小(概率). 为此,引入 σ-域的概念．

定义 1.1.1 设 Ω 是一样本空间,F 是 Ω 的某些子集所组成的集合,满足:

(1)$\Omega\in F$,

(2)若 $A\in F$,则 $\overline{A}\in\Omega-A\in F$,

(3)若 $A_i\in F,i=1,2,\cdots$,则 $\bigcup\limits_{i=1}^{\infty}A_i\in F$,

则称 F 为 Ω 上的 σ-代数或 σ-域.(Ω,F)为可测空间,F 中的元素为事件．

容易验证,若 F 为 σ-域,则 F 可列次交、并、差等运算封闭,即 F 中的任何元素经可列次运算后仍属于 F. 例如,集类 $F_0=\{\varnothing,\Omega\}$,$F_1=\{\varnothing,A,A^C,\Omega\}$以及 $F_2=\{A:\forall A\subset\Omega\}$均是 σ-域,但集类 $A=\{\varnothing,A,\Omega\}$不是 σ-域．

在实际问题中,通常我们只关注包含所要研究对象的最小 σ-域．设 A 为由 Ω 的某些子集构成的集类．一切包含 A 的 σ-域的交,记为 $\sigma(A)$,称 $\sigma(A)$为由 A 生成的 σ-域,或称为包含 A 的最小 σ-域．例如,$A=\{\varnothing,A,\Omega\}$,则 $\sigma(A)=\{\varnothing,A,A^C,\Omega\}$,一维博雷尔(Borel)$\sigma$-域:包含 $\mathbf{R}$ 上所有形如集合$(-\infty,a]$的最小 σ-域称为一维博雷尔 σ-域,即为 B,即 $B=(((-\infty,a],\forall a\in\mathbf{R})$.

定义 1.1.2 设(Ω,F)为可测空间,P 是定义在 F 上的集函数,若满足:

(1)对任意 $A\in F$,有 $0\leqslant P\leqslant 1$,

(2)$P(\Omega)=1$,

(3)若 $A_1,A_2,\cdots,A_n\in F$,且两两互不相容($i\neq j$ 时,$A_iA_j=\varnothing$)总有 $P(\bigcup\limits_{i=1}^{\infty}A_i)=\sum\limits_{i=1}^{\infty}P(A_i)$,

则称 P 为(Ω,F)上的概率,(Ω,F,P)为概率空间,$P(A)$为事件 A 的概率.

事件的概率刻画了事件出现可能性的大小.概率的基本性质如下:

(1)$P(\varnothing)=0,P(\overline{A})=1-P(A)$;

(2)若 $A_1,A_2,\cdots,A_n$ 互不相容,则 $P(\bigcup\limits_{i=1}^{n}A_i)=\sum\limits_{i=1}^{n}P(A_i)$;

(3)若 $A\subset B$,则 $P(A)\leqslant P(B)$;

(4)对任意两个事件 A 与 B,有

$$P(A\cup B)=P(A)+P(B)-P(AB),$$
$$P(A-B)=P(A)-P(AB);$$

(5)(若尔当(Jordan)公式)对任意 $A_1,A_2,\cdots,A_n$ 有

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j\leqslant n}P(A_iA_jA_k)-\cdots+(-1)^{n+1}P(A_1A_2\cdots A_n),$$

特别地,当 $n=2$ 时,有 $P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1A_2)$.

事件的一个重要性质是它具有连续性,为此引入事件列的极限.

已知一事件列$\{A_n,n\geqslant 1\}$,若 $A_n\subset A_{n+1},n\geqslant 1$,则称为单

调增序列，表示为$\lim\limits_{n\to\infty}A_n=\bigcup\limits_{i=1}^{\infty}A_i$；若 $A_n\supset A_{n+1}$，$n\geqslant 1$，则称为单调减序列，表示为$\lim\limits_{n\to\infty}A_n=\bigcap\limits_{i=1}^{\infty}A_i$. 由此得出如下定理．

定理 1.1.1 若$\{A_n,n\geqslant 1\}$是单调增序列（或减序列），则

$$\lim_{n\to\infty}P(A_n)=P(\lim_{n\to\infty}A_n).$$

证 先设$\{A_n,n\geqslant 1\}$是单调增序列，令

$$B_1=A_1,B_n=A_n-\bigcup_{i=1}^{n-1}A_i,n=2,3,\cdots,$$

则 $B_1,B_2,\cdots,B_n\in F$ 且两两互不相容，且$\bigcup\limits_{i=1}^{n}A_i=\bigcup\limits_{i=1}^{n}B_i$，$n=2$，$3,\cdots$. 于是由性质(2)有

$$\begin{aligned}\lim_{n\to\infty}P(A_n)&=\lim_{n\to\infty}P(\bigcup_{i=1}^{n}A_i)=\lim_{n\to\infty}P(\bigcup_{i=1}^{n}B_i)\\&=\lim_{n\to\infty}\sum_{i=1}^{n}P(B_i)=\sum_{n=1}^{\infty}P(B_n)\\&=P(\bigcup_{n=1}^{\infty}B_i)=P(\bigcup_{n=1}^{\infty}A_i)=P(\lim_{n\to\infty}A_n).\end{aligned}$$

若$\{A_n,n\geqslant 1\}$是单调减序列，则$\{\overline{A}_n,n\geqslant 1\}$是单调增序列，于是由上式，得

$$\lim_{n\to\infty}P(\overline{A}_n)=P(\lim_{n\to\infty}\overline{A}_n)=P(\bigcup_{n=1}^{\infty}\overline{A}_n)$$

又因为

$$\lim_{n\to\infty}P(\overline{A}_n)=\lim_{n\to\infty}[1-P(A_n)]=1-\lim_{n\to\infty}P(A_n),$$

$$P(\bigcup_{n=1}^{\infty}\overline{A}_n)=P(\overline{\bigcap_{n=1}^{\infty}A_n})=1-P(\bigcap_{n=1}^{\infty}A_n),$$

因此

$$\lim_{n\to\infty}P(A_n)=P(\bigcap_{n=1}^{\infty}A_n)=P(\lim_{n\to\infty}A_n),$$

证毕 .

下面是著名的 Borel-Cantelli 定理 .

定理 1.1.2　设$\{A_n,n\geqslant 1\}$是一事件序列,若 $\sum\limits_{i=1}^{n}P(A_i)<\infty$,则

$$P(\limsup_{i\to\infty}A_i)=0,$$

其中$\limsup\limits_{i\to\infty}A_i\triangleq\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{i=n}^{\infty}A_i$.

证　显然,$\bigcup\limits_{i=n}^{\infty}A_i$ 是关于 n 的单调减序列,于是由定理 1.1.1 可知

$$\begin{aligned}0\leqslant P(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}A_i)&=P(\lim_{n\to\infty}\bigcup_{i=n}^{\infty}A_i)\\&=\lim_{n\to\infty}P(\bigcup_{i=n}^{\infty}A_i)\leqslant\lim_{n\to\infty}\sum_{i=n}^{\infty}P(A_i)=0.\end{aligned}$$

证毕 .

下面讨论事件间的一种重要关系,即事件的独立性问题 .

定义 1.1.3　设(Ω,F,P)是一概论空间,$A_1,A_2,\cdots,A_n\in F,n\geqslant 2$. 若对任意 $1<k\leqslant n$ 及任意 $1\leqslant i_1<i_2<\cdots<i_k\leqslant n$,有

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}),$$

则称 $A_1,A_2,\cdots,A_n$ 相互独立 .

由定义可知相互独立事件有如下两个性质:

性质 1　若事件 $A_1,A_2,\cdots,A_n$ 相互独立,则 $A_1,A_2,\cdots,A_n$ 中任意 $k(\geqslant 2)$个事件 $A_{i_1},A_{i_2},\cdots,A_{i_k}$ 也相互独立,其中 $1\leqslant i_1<i_2<\cdots<i_k\leqslant n$.

性质 2　若事件 $A_1,A_2,\cdots,A_n$ 相互独立,则事件 $B_1,B_2,\cdots,B_n$ 也相互独立,其中 B_i 或为 A_i,或为$\overline{A_i}$,$i=1,2,\cdots,n$.

定理 1.1.3 若$\{A_n,n\geqslant 1\}$是相互独立的事件序列，且$\sum\limits_{n=1}^{\infty}P(A_n)=\infty$，则有$P(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{i=n}^{\infty}A_i)=1$.

证 由于

$$P(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}A_i)=P(\lim_{n\to\infty}\bigcup_{i=n}^{\infty}A_i)=\lim_{n\to\infty}P(\bigcup_{i=n}^{\infty}A_i)$$
$$=\lim_{n\to\infty}[1-P(\bigcap_{i=n}^{\infty}\overline{A_i})],$$

$$P(\bigcap_{i=n}^{\infty}\overline{A_i})=\prod_{i=n}^{\infty}P(\overline{A_i})=\prod_{i=n}^{\infty}(1-P(A_i))\leqslant\prod_{i=n}^{\infty}e^{-P(A_i)}$$
$$=\exp(-\sum_{i=n}^{\infty}P(A_i))=0,$$

因此$P(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{i=n}^{\infty}A_i)=1$. 证毕.

1.1.2 随机变量与分布函数

定义 1.1.4 设(Ω,F,P)是一概论空间，X是定义在Ω上，取值于实数集$\mathbf{R}$的函数，如果对任意实数$x\in\mathbf{R}$，$\{\omega:X(\omega)\leqslant x\}\in F$，则称$X(\omega)$是$F$上的随机变量，简称为随机变量．函数

$$F(x)=P\{\omega:X(\omega)\leqslant x\},-\infty<x<\infty$$

称为随机变量X的分布函数．

注 在上面的定义中，如果X是广义实值函数，即X可以取∞值，则需要加上条件：X是几乎处处有限的，即$P\{\omega:|X(\omega)|=\infty\}=0$. 否则，会出现按上面定义的分布函数是假分布的情况．

分布函数$F(x)$有如下性质：

性质 1 $F(x)$是单调不减函数，即$\forall x_1,x_2\in\mathbf{R},x_1<x_2$，总

有 $F(x_1)\leqslant F(x_2)$.

性质 2 $F(x)$是右连续函数，即 $\forall x\in\mathbf{R}$，总有 $F(x+0)=\lim\limits_{h\to 0^+}F(x+h)=F(x)$.

性质 3 $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$，$F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$.

定义 1.1.5 两个随机变量 X 与 Y，如果满足 $P\{\omega\in\Omega: X(\omega)\neq Y(\omega)\}=0$，则称它们是等价的.

对两个等价的随机变量，我们视为同一.

在实际中，常用的随机变量有两种类型：离散型随机变量和连续性随机变量.

若随机变量 X 只可能取有限个或可列无穷个(即虽然是无穷个，但可以一个接一个地排列起来)值，则称 X 为离散型随机变量，其分布可用其概率分布(或分布律)

$$p_i=P\{X=x_i\},i=1,2,\cdots$$

来描述. 此时 X 的分布函数为

$$F(x)=\sum_{x_i\leqslant x}p_i,x\in\mathbf{R}.$$

设随机变量 X 的分布函数为 $F(x)$，若存在非负可积函数 $f(x)$，使得

$$F(x)=\int_{-\infty}^{x}f(t)dt,x\in\mathbf{R}$$

则称 X 连续型随机变量，称 $f(x)$为 X 概率密度函数.

定理 1.1.4 若 $F(x_1,\cdots,x_n)$是联合分布函数，则

(1)$F(x_1,\cdots,x_n)$对每个变量都是单调的；

(2)$F(x_1,\cdots,x_n)$对每个变量都是右连续的；

(3) $\lim\limits_{x\to-\infty}F(x_1,\cdots,x_i,\cdots,x_n)=0,(i=1,2,\cdots,n)$，$\lim\limits_{x_1,\cdots,x_n\to\infty}F(x_1,\cdots,x_n)=1$.

(4)任取 $x_i, y_i \in \mathbf{R}, x_i \leqslant y_i, i=1,2,\cdots,n$,总有

$$F(y_1,\cdots,y_n)-\sum_{i=1}^{n}F(y_1,\cdots,y_{i-1},x_i,y_{i+1},\cdots,y_n)+$$

$$\sum_{1\leqslant i<j\leqslant n}F(y_1,\cdots,y_{i-1},x_i,y_{i+1},\cdots,y_{j-1},x_j,y_{j+1},\cdots,y_n)-$$

$$\cdots+(-1)^n F(x_1,\cdots,x_n)\geqslant 0$$

若 $f(x_1,\cdots,x_n)=\dfrac{\partial^n F}{\partial x_1\cdots\partial x_n}$ 对所有的 $(x_1,\cdots,x_n)\in\mathbf{R}^n$ 存在,则称函数 $f(x_1,\cdots,x_n)$ 为 $X=(X_1,\cdots,X_n)$ 或 $F(x_1,\cdots,x_n)$ 的联合密度函数,并且

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x}\cdots\int_{-\infty}^{x}f(x_1,\cdots,x_n)\mathrm{d}t_n\cdots\mathrm{d}t_1.$$

保留 $F(x_1,\cdots,x_n)$ 中的 $k(1\leqslant k<n)$ 个变元,而令其他变元趋于 $+\infty$,便得到 X 的 k 维边际分布函数. 例如,保留 $x_1,\cdots,x_k$,得到 $F(x_1,\cdots,x_n,+\infty,\cdots,+\infty)$. 特别地,当 $k=1$ 时,n 维随机变量 X 的 n 个 1 维边际概率函数和 n 个 1 维边际概率密度函数分别为 $F_{X_1}(x_1),\cdots,F_{X_n}(x_n)$ 与 $f_{X_1}(x_1),\cdots,f_{X_n}(x_n)$.

定义 1.1.6 设 $X=(X_1,\cdots,X_n)$ 是一 n 维连续型随机向量,其联合分布函数及 n 个 1 维边际分布函数分别为 $F(x_1,\cdots,x_n)$ 和 $F_{X_1}(x_1),\cdots,F_{X_n}(x_n)$. 若 $\forall x_1,\cdots,x_n\in\mathbf{R}$,总有

$$F(x_1,\cdots,x_n)=F_{X_1}(x_1),\cdots,F_{X_n}(x_n)$$

则称随机变量 $X_1,\cdots,X_n$ 相互独立.

当 $X=(X_1,\cdots,X_n)$ 是 n 维连续型随机向量时,$X_1,\cdots,X_n$ 相互独立的充要条件是:对几乎所有 $x_1,\cdots,x_n\in\mathbf{R}^n$,有

$$f(x_1,\cdots,x_n)=f_{X_1}(x_1),\cdots,f_{X_n}(x_n).$$

当 $X=(X_1,\cdots,X_n)$ 是 n 维离散型随机向量时，$X_1,\cdots,X_n$ 相互独立的充要条件是：对 X_i 的所有可能取值 $x_i,i=1,2,\cdots,n$，有

$$P\{X_1=x_1,\cdots,X_n=x_n\}=P\{X_1=x_1\}\cdots P\{X_n=x_n\}.$$

1.2　数字特征、矩母函数与特征函数

1.2.1　数字特征

1. 随机变量的数学期望

定义 1.2.1　设 X 为随机变量，$F(x)$ 为 X 的分布函数，若 $\int_{-\infty}^{+\infty}|x|dF(x)$ 存在，则称

$$EX=\int_{\infty}^{+\infty}x\mathrm{d}F(x)$$

为随机变量 X 的数学期望(或称为 X 的均值).

随机变量 X 的数学期望有如下性质：

性质 1　若 $c_i(i=1,2,\cdots,n)$ 为常数，$X_i(i=1,2,\cdots,n)$ 为随机变量，则

$$E(\sum_{i=1}^{n}c_iX_i)=\sum_{i=1}^{n}c_iEX_i.$$

性质 2　设 $g(x)$ 为 x 的函数，$F_X(x)$ 是 X 的分布函数，若 $E[g(X)]$ 存在，则

$$E[g(X)]=\int_{\infty}^{+\infty}g(x)\mathrm{d}F_X(x).$$

性质 3　若 X 为离散型随机变量，即 $P(X=x_n)=P_n(n\in\mathbf{R})$，则