

处理线性滤波以及预测问题的一种新途径

R. E. Kalman

1 引言

通讯与控制中的理论与实际问题中有很重要的一类具有统计性质。这样的问题有：(1)、随机信号的预测；(2)、从随机噪声中分离随机信号；(3)、在有噪声的情况下探测已知形式的信号（脉冲、正弦波）。

在 Wiener 开拓性的工作中，他证明 [1] 从问题 (1) 和问题 (2) 可导出所谓 Wiener-Hopf 积分方程；他同样给出了解决具有实际重要意义的特殊情况——定态统计和有理数频谱——之积分方程的方法（频谱因式分解）。

在 Wiener 的基础性工作之后出现了许多延伸和推广。Zadeh 与 Ragazzini 给出了有限存储器情况的解 [2]。Bode 和 Shannon[3] 同时独立的给出上述情况的解，并且给出了简化的求解方法。Boonton 讨论了非定态统计 Wiener-Hopf 方程 [4]。这些结果现在都写入了标准教科书中 [5, 6]。最近 Darlington[7] 沿着这些主线给出了一种稍微有些不同的方法。对抽样信号的延伸，参见 Franklin[8] 和 Lees[9] 的工作。基于 Wiener-Hopf 方程（同样应用于非定态问题，尽管前述方法一般并非如此）特征函数的方法由 Davis[10] 开创并被许多人应用，例如 Shinbrot[11], Blum[12], Pugachev[13], Solodovnikov[14]。

在所有这些工作中，目标都是获取一个线性动力系统的明确说明（Wiener 滤波器），由此可以完成预测、分离或者探测随机信号。

现有求解 Wiener 问题的方法受制于若干限制，这样就使得它们的实际用处收到削弱：

1. 最佳的滤波器由其脉冲响应具体指定。由这些数据合成滤波器并非易事。
2. 数值确定最佳的脉冲响应往往十分复杂并且不很适合机器计算。这种情况随着问题复杂度的增加而迅速变得更为糟糕。

3. 重要的推广（如增长存储器滤波器、非定态预测）需要新的推导过程，经常给非专业人士带来相当大的困难。
4. 这些推导过程的数学部分并不透明。基本假设及其后果趋于模糊。

本文回避上述困难，提出看待这些问题的整个集合的新方式。以下是本文的亮点：

5. 最佳估计和正交投影。Wiener 问题是以条件分布与期望的观点处理的。这样，Wiener 理论的基本事实可以迅速获取；结果的范围以及基本假设可以清楚的显现出来。可以看到所有的统计计算以及结果都基于一阶和二阶平均；不需要其他的统计数据。这样一来困难（4）便被排除。这种方法在概率论中为人们所熟知（见 Doob[15] 第 148 至 155 页以及 Loève[16] 第 455 至 464 页），但在工程上还没有大量的应用。
6. 随机过程模型。继前人之后，尤其是 Bode 与 Shannon[3]，任意随机信号可以被表示（直到二阶平均统计性质）为线性动力系统受独立或不相关随机信号（“白噪声”）激励后的输出。这是工程上应用 Wiener 理论的标准手法 [2, 3, 4, 5, 6, 7]。这里用到的方法与传统方法相比只在线性动力系统的描述方法上不同。我们将强调状态以及状态过渡；换言之，线性系统将以一阶差分（或微分）方程组来刻画。为了利用（5）中提到的简化，这种观点是自然的，也是必要的。
7. 求解 Wiener 问题。使用状态——过渡方法，单独一次推导即覆盖多种问题：增长与有限存储器滤波器、定态与非定态统计等等；（3）中的困难消失了。正确猜测出估计问题的“状态”后，接下来就是最佳估计误差协方差矩阵的非线性差分（或微分）方程。这个方程与 Wiener-Hopf 方程有些相似。对方称的求解开始于观测开始的 t_0 时刻；随后每个时刻 t 方程的解都代表给定区间 (t_0, t) 上观测的最佳预测误差协方差。从 t 时刻的协方差矩阵我们立刻可以获得刻画最佳线性滤波器的系数，而无需进一步的计算。
8. 对偶问题。对 Wiener 问题的新的公式化使其接触到基于“状态”观点的成长中的控制系统新理论 [17, 18, 19, 20, 21, 22, 23, 24]。令人惊讶的是，Wiener 问题是无噪声最佳调整器问题的对偶，而此问题已经被本文的作者利用状态——过渡方法解决 [18, 23, 24]。两个问题的

数学背景完全一致——这一点一直以来都被人们所怀疑，但直到现在两者的类比才被明确指出。

9. 应用。新方法的威力在理论调研与复杂实际问题的数值解答中大多显而易见。在后面的案例中，最好借助机器计算。这种类型的例子将在后面讨论。为了给应用提供一些感觉，包含了两个非定态预测的标准例子；在这些例子中，(7) 中提到的非线性差分方程甚至可以得到近似形式的解。

为了参考方便，主要结果都用定理的形式显示。只有定理 3 和定理 4 是原创的。下一个章节以及附录主要服务于用适用于当前目的的形式来回顾人们熟知的资料。

2 符号约定

贯穿本文，我们主要与离散（或者抽样）动力系统打交道；换句话说，信号将在等间距的时刻（抽样瞬间）被观测到。选择合适的时间尺度，相邻两次抽样瞬间的时间间隔常数（抽样周期）可以被选择为单位时间。如此一来，表示时间的变量如 t, t_0, τ, T 等将一直是整数。对离散动力系统施加这样的约束条件并不是必需的（至少从工程的角度来看是这样）；使用这样的离散性，我们可以保有严密的、基础的数学。矢量将用小写粗体字母如 $\mathbf{a}, \mathbf{b}, \dots, \mathbf{x}, \mathbf{y}, \dots$ 表示。矢量，或者更精确的说， n 维矢量是 n 个数 x_1, \dots, x_n 的集合； x_i 是矢量 \mathbf{x} 的坐标或分量。

矩阵将使用大写粗体字母 $\mathbf{A}, \mathbf{B}, \mathbf{Q}, \mathbf{\Phi}, \mathbf{\Psi}, \dots$ 表示；它们是元素 $a_{ij}, b_{ij}, q_{ij}, \dots$ 的 $m \times n$ 维数列。矩阵的转置（交换行与列）用一撇来表示。使用公式时，为求方便，视矩阵为只有一列元素的矩阵。

使用传统的矩阵乘法定义，我们将两个 n 维矢量 \mathbf{x}, \mathbf{y} 的标量积写成

$$\mathbf{x}'\mathbf{y} = \sum_{i=1}^n x_i y_i = \mathbf{y}'\mathbf{x}$$

标量积显然是标量，也就是说并非矢量。类似的，关于 $n \times n$ 维矩阵 \mathbf{Q} 的二次型是，

$$\mathbf{x}'\mathbf{Q}\mathbf{y} = \sum_{i,j=1}^n x_i q_{ij} x_j$$

我们定义表达式 \mathbf{xy}' (式中 \mathbf{x}' 是 m 维矢量, \mathbf{y} 是 n 维矢量) 为 $m \times n$ 维矩阵, 矩阵元为 $x_i y_j$.

我们将随机矢量 \mathbf{x} 的期望值记为 $E(\mathbf{x}) = E\mathbf{x}$ (见附录)。为方便起见, 通常省略 E 后面的括号。因为常数与 E 算符对易, 故这种省略在简单情况中并不会引起混淆。从而, $E\mathbf{xy}'$ 是矩阵元为 $E(x_i y_j)$ 的矩阵; $E\mathbf{x}E\mathbf{y}$ 是以 $E(x_i)E(y_j)$ 为矩阵元的矩阵。

为方便参考, 下面给出基本符号表:

最佳估计

- t 时间, 当前时间。
- t_0 观测开始时刻。
- $x_1(t), x_2(t)$ 基本随机变量。
- $y(t)$ 观测到的随机变量。
- $x_1^*(t_1|t)$ 给定 $y(t_0), \dots, y(t)$ 之后对 $x_1(t_1)$ 的最佳估计。
- L 损耗函数 (是其自变量的非随机函数)。
- ϵ 估计误差 (随机变量)。

正交投影

- $\mathcal{Y}(t)$ 随机变量 $y(t_0), \dots, y(t)$ 生成的线性流形。
- $\bar{x}(t_1|t)$ $x(t_1)$ 在 $\mathcal{Y}(t)$ 上的正交投影。
- $\tilde{x}(t_1|t)$ $x(t_1)$ 正交于 $\mathcal{Y}(t)$ 的分量。

随机过程模型

- $\Phi(t+1; t)$ 过渡矩阵。
- $\mathbf{Q}(t)$ 随机激励的协方差。

求解 Wiener 问题

- $\mathbf{x}(t)$ 基本随机变量。
- $\mathbf{y}(t)$ 观测到的随机变量。
- $\mathcal{Y}(t)$ 由 $\mathbf{y}(t_0), \dots, \mathbf{y}(t)$ 生成的线性流形。
- $\mathcal{Z}(t)$ 由 $\tilde{\mathbf{y}}(t|t-1)$ 生成的线性流形。
- $\mathbf{x}^*(t_1|t)$ 给定 $\mathcal{Y}(t)$ 之后对 $\mathbf{x}(t_1)$ 的最佳估计。
- $\tilde{\mathbf{x}}(t_1|t)$ 给定 $\mathcal{Y}(t)$ 之后对 $\mathbf{x}(t_1)$ 最佳估计的误差。

3 最佳估计

为具体描述将要研究问题的类型，需考虑以下情况。已知信号 $x_1(t)$ 和噪声 $x_2(t)$ 。只能观察到和 $y(t) = x_1(t) + x_2(t)$ 。假定我们已经观测并确切的知道 $y(t_0), \dots, y(t)$ 的值，关于 $t = t_1$ (t_1 可能小于、等于或者大于 t) 处的值（非观测量）我们可以从已知信息中推断出什么？如果 $t_1 < t$ ，这是数据平滑（插值）问题。如果 $t_1 = t$ ，这称为滤波。如果 $t_1 > t$ ，称为预测问题。既然我们有足够一般性的方法来处理以上以及类似的问题，以下我们将使用共同的术语估计。

正如 Wiener 指出 [1] 的那样，估计问题的天然背景属于概率论和统计学的范畴。因此信号、噪声以及它们的和都是随机变量，进而它们可被视为随机过程。从随机过程的概率论描述中我们可以确定特定信号或者噪声抽样发生的概率。对于随机变量 $y(t)$ 的任意给定的测量值 $\eta(t_0), \dots, \eta(t)$ 原则上也可以确定随机变量 $x_1(t_1)$ 于同一时刻取不同值 $\xi_1(t)$ 的概率。这就是条件概率分布函数

$$Pr[x_1(t_1) \leq \xi_1 | y(t_0) = \eta(t_0), \dots, y(t) = \eta(t)] = F(\xi_1) \quad (1)$$

显然， $F(\xi_1)$ 代表了随机变量 $y(t_0), \dots, y(t)$ 测量结果传递的关于随机变量 $x_1(t_1)$ 所有信息。随机变量 $x_1(t_1)$ 的任何统计估计都是上述分布的某种函数，因而是随机变量 $y(t_0), \dots, y(t)$ 的（非随机）函数。该统计估计记为 $X_1(t_1|t)$ ，如果观测到的随机变量集合或者待估计时间在上下文中是明确的，也可记为 $X_1(t_1)$ 或者 X_1 。

假定 X_1 以随机变量 $y(t_0), \dots, y(t)$ 的固定函数的形式给出。那么 X_1 本身就是一个随机变量，只要 $y(t_0), \dots, y(t)$ 的实际值已知，即可知 X_1 的实际值。一般来说， $X_1(t_1)$ 的实际值与 $x_1(t_1)$ 的（未知）实际值是不同的。为了取得确定 X_1 的合理方法，自然要为不正确的估计指定罚函数或损耗函数。确切的说，损耗函数应当（1）非负，（2）是估计误差 $\epsilon = x_1(t_1) - X_1(t_1)$ 的单调不递减函数。故此，定义损耗函数

$$L(0) = 0$$

$$L(\epsilon_2) \leq L(\epsilon_1) \leq 0 \text{ when } \epsilon_2 \leq \epsilon_1 \leq 0 \quad (2)$$

$$L(\epsilon) = L(-\epsilon)$$

常见的损耗函数有： $L(\epsilon) = a\epsilon^2, a\epsilon^4, a|\epsilon|, a[1 - \exp(-\epsilon^2)]$ 等等，其中 a 是大于零的常数。

一种（但并非唯一的）自然而然的选择随机变量 X_1 的方法是令选取的值最小化损耗或风险的平均值

$$E\{L[x_1(t_1) - X_1(t_1)]\} = E[E\{L[x(t_1) - X_1(t_1)]|y(t_0), \dots, y(t)\}] \quad (3)$$

既然式3右边第一个期望值不依赖于 X_1 的选择，而是由 $y(t_0), \dots, y(t)$ 唯一决定，所以最小化 refeq3 等价于最小化

$$E\{L[x_1(t_1) - X_1(t_1)]|y(t_0), \dots, y(t)\} \quad (4)$$

在少量附加的假设之下，最佳估计就可以用简单的方法刻画出来。

定理 1. 假定 L 如式2且由式1定义的条件分布函数 $F(\xi)$ ：

A 关于均值 $\bar{\xi}$ 对称：

$$F(\xi - \bar{\xi}) = 1 - F(\bar{\xi} - \xi)$$

B 对 $\xi \leq \bar{\xi}$ 是凸的：

$$F(\lambda\xi_1 + (1-\lambda)\xi_2) \leq \lambda F(\xi_1) + (1-\lambda)F(\xi_2) \\ \text{for all } \xi_1, \xi_2 \leq \bar{\xi} \text{ and } 0 \leq \lambda \leq 1$$

则最小化损耗（式3）的随机变量 $x_1^*(t_1|t)$ 是条件期望

$$x_1^*(t_1|t) = E[x_1(t_1)|y(t_0), \dots, y(t)] \quad (5)$$

证明： 正如 Sherman 最近所指出 [25] 的，该定理是概率论中一个著名引理的直接结论。

推论. 如果随机过程 $x_1(t), x_2(t)$ 和 $y(t)$ 是高斯的，则定理1成立。

证明： 由定理3（见附录），高斯随机过程的条件概率仍然是高斯的。因而总是满足定理1的要求。

在控制系统的文献中，上面的定理以某种程度上更为受限而换言之也更为一般的形式出现：

定理 1-A. 如果 $L(\epsilon) = \epsilon^2$ ，那么无需假设 A 和 B 定理1即成立。

证明： 展开条件期望（式4）：

$$E[x_1^2(t_1)|y(t_0), \dots, y(t)] - 2X_1(t_1)E[x_1(t_1)|y(t_0), \dots, y(t)] + X_1^2(t_1)$$

然后对 $X_1(t_1)$ 求导。这并非严密的论证；若求简单严密的论证可参考 Doob 专著的第 77 至 78 页。

评论。（一）、就作者所知，人们并不知道条件分布函数满足定理1要求的随机过程 $x_1(t), x_2(t)$ 属于何种最为一般的类型。

（二）、除了 Sherman 的记录，定理1似乎从来没有在控制系统的文献中明确出现过。事实上，一般类型（如式2）的损耗函数在数学上无法方便的处理，对这种影响的陈述随处可见。

（三）、后文中，我们将主要应付矢量值随机变量。相应的，估计问题表述为：给定矢量值随机过程 $\mathbf{x}(t)$ 和观测随机变量 $\mathbf{y}(t_0), \dots, \mathbf{y}(t)$ ，其中 $\mathbf{y}(t) = \mathbf{M}\mathbf{x}(t)$ （ \mathbf{M} 是奇异矩阵；换句话说，并非 $\mathbf{x}(t)$ 的所有坐标都能被观测到），试求可以最小化损耗期望 $E[L(\|\mathbf{x}(t_1) - \mathbf{X}(t_1)\|)]$ 的估计 $\mathbf{X}(t_1)$ ， $\|\cdot\|$ 是矢量的范数。

定理1在矢量情况下依然成立，只要矢量 $\mathbf{x}(t_1)$ 的 n 个分量的条件分布函数

$$Pr[x_1(t_1) \leq \xi_1, \dots, x_n(t_1) \leq \xi_n | \mathbf{y}(t_0), \dots, \mathbf{y}(t)] = F(\xi_1, \dots, \xi_n)$$

关于 n 个变量 $\xi_1 - \bar{\xi}_1, \dots, \xi_n - \bar{\xi}_n$ 对称且在所有这些变量取负值的区域是凸的。

4 正交投影

一般来说，作为观测变量的函数准确计算最佳估计是不可能的。但这里有个重要的例外：过程 $x_1(t)$ 和 $x_2(t)$ 是高斯的。

另一方面，如果令损耗函数 $L(\epsilon) = \epsilon^2$ 且令估计是观测随机变量的线性函数，我们将得到与高斯情形相同的最佳估计，而高斯情形则无需线性性质以及二次型损耗函数的假设。这证明非线性估计仅在以下情况能得到优于线性估计的结果：（一）、随机过程不是高斯的（依定理3中 C 的观点），且（二）、要至少考虑三阶的概率分布函数。

在上面提到的特殊情况中，显式求解估计问题在几何图像的帮助下大致上可以容易理解一些。这正是本部分的主旨。

考虑 (实值) 随机变量 $y(t_0), \dots, y(t)$. 系数为实数的这些随机变量的所有线性组合的集合

$$\sum_{i=t_0}^t a_i y(i) \quad (6)$$

构成一个矢量空间 (线性流形), 记为 $\mathcal{Y}(t)$. 我们将所有形如式6的表达式抽象的视为 $\mathcal{Y}(t)$ 上的 “点” 或 “矢量”. 当然, 此处使用的 “矢量” 当不与随机矢量中的 “矢量” 或者其它地方的 “矢量” 相混淆. 由于我们并不想限定 t 的值 (可能的观测的总数), 故 $\mathcal{Y}(t)$ 应该被视为所有可能的观测空间的有限维子空间。

任意给定 $\mathcal{Y}(t)$ 中的两个矢量 u, v (即可由式6表达的随机变量), 如果 $Euv = 0$ 我们就说 u, v 是正交的. 使用 Schmidt 正交化过程, 正如 Doob[15] (第 151 页) 或 Loève[16] (第 459 页) 借助实例描绘的, 很容易就可以找出 $\mathcal{Y}(t)$ 的一组正交基, 亦即, $\mathcal{Y}(t)$ 的一组正交矢量 e_{t_0}, \dots, e_t , 使用这组矢量, $\mathcal{Y}(t)$ 中的任意矢量都可以被唯一的表示为 e_{t_0}, \dots, e_t 的线性组合, 且

$$\left. \begin{aligned} Ee_i e_j &= \delta_{ij} = 1 \text{ 如果 } i = j \\ &= 0 \text{ 如果 } i \neq j \end{aligned} \right\} \quad (i, j = t_0, \dots, t) \quad (7)$$

故 $\mathcal{Y}(t)$ 中的任意矢量可写成

$$\bar{x} = \sum_{i=t_0}^t a_i e_i$$

系数 a_i 也可借由7立刻得出

$$E\bar{x}e_j = E\left[\sum_{i=t_0}^t a_i e_i\right]e_j = \sum_{i=t_0}^t a_i Ee_i e_j = \sum_{i=t_0}^t \delta_{ij} a_i = a_j \quad (8)$$

进一步的, 所有随机变量 x (并不一定是 $\mathcal{Y}(t)$ 中的) 可以唯一的被分解成两部分: 一部分为 \bar{x} 在 $\mathcal{Y}(t)$ 中, 另一部分同 $\mathcal{Y}(t)$ 正交 (即同 $\mathcal{Y}(t)$ 中所有矢量都正交). 事实上, 我们可以将其写为

$$x = \bar{x} + \tilde{x} = \sum_{i=t_0}^t (Exe_i) e_i + \tilde{x} \quad (9)$$

故 \bar{x} 可以由式9唯一确定, 且显然是 $\mathcal{Y}(t)$ 中的矢量. 这样 \tilde{x} 也被唯一的确定了; 接下来检验其是否与 $\mathcal{Y}(t)$ 正交:

$$E\tilde{x}e_i = E(x - \bar{x})e_i = Exe_i - E\bar{x}e_i$$

\bar{x} 关于基 e_{t_0}, \dots, e_t 的坐标或者如式8以 $E\bar{x}e_i$ 的形式给出, 或者如式9以 $E\bar{x}e_i$ 的形式给出。既然坐标是唯一的, 故 $E\bar{x}e_i = E\bar{x}e_i$ ($i = t_0, \dots, t$); 因此 $E\tilde{x}e_i = 0$, \tilde{x} 与每个基矢量 e_j 都正交; 也就是说与 $\mathcal{Y}(t)$ 正交。我们称 \tilde{x} 为 x 在 $\mathcal{Y}(t)$ 上的正交投影。

这里还有另外一种表征正交投影的方式: \bar{x} 是 $\mathcal{Y}(t)$ 上最小化二次型损耗函数的矢量 (即随机变量 $y(t_0), \dots, y(t)$ 的线性组合)。事实上, 如果 \bar{w} 是 $\mathcal{Y}(t)$ 上的任意另一矢量, 有

$$E(x - \bar{w})^2 = E(\tilde{x} + \bar{x} - \bar{w})^2 = E[(x - \bar{x}) + (\bar{x} - \bar{w})]^2$$

既然 \tilde{x} 与 $\mathcal{Y}(t)$ 上的所有矢量, 特别的, 与 $\bar{x} - \bar{w}$ 正交, 有

$$E(x - \bar{w})^2 = E(x - \bar{x})^2 + E(\bar{x} - \bar{w})^2 \geq E(x - \bar{x})^2 \quad (10)$$

恰好证明, 如果 \bar{w} 也最小化二次型损耗函数, 则必有 $E(\bar{x} - \bar{w})^2 = 0$, 即, 随机变量 \bar{x} 和 \bar{w} 相等 (除非对于概率为零的一组事件)。

以上结果摘要如下:

定理 2. 令 $\{x(t)\}, \{y(t)\}$ 为零均值的随机过程 (即, 对于一切 t 均有 $Ex(t) = Ey(t) = 0$)。观测 $y(t_0), \dots, y(t)$ 。

如果有

A 随机过程 $\{x(t)\}, \{y(t)\}$ 是高斯的; 或者

B 将最佳估计限定为观测随机变量的线性函数且 $L(\epsilon) = \epsilon^2$;

那么

$$\begin{aligned} x^*(t_1|t) &= \text{给定 } y(t_0), \dots, y(t) \text{ 对 } x(t_1) \text{ 的最佳估计} \\ &= x(t_1) \text{ 在 } \mathcal{Y}(t) \text{ 上的正交投影} \end{aligned} \quad (11)$$

以上结果为人们所熟知, 尽管这在控制系统文献中并不容易获取。见 Doob[15] 第 75 至 78 页, 或着 Pugachev[26]。有时为求方便, 将正交投影记为

$$\bar{x}(t_1|t) \equiv x^*(t_1|t) = \hat{E}[x(t_1)|\mathcal{Y}(t)]$$

使用记号 \hat{E} 的目的是: 如果讨论的随机过程是高斯的, 则正交投影盒条件期望事实上是一样的。

证明:

A 关于式10评论的直接结果。

B 既然 $x(t), y(t)$ 都是零均值的随机变量, 从式9可知, $x(t_1)$ 关于 $\mathcal{Y}(t)$ 的正交部分 $\tilde{x}(t_1|t)$ 也是零均值的随机变量。零均值随机变量是不相关的; 如果同时他们也是高斯的 (由于定理3B 部分), 则他们是独立的。所以

$$\begin{aligned} 0 &= E\tilde{x}(t_1|t) = E[\tilde{x}(t_1|t)|y(t_0), \dots, y(t)] \\ &= E[x(t_1) - \bar{x}(t_1|t)|y(t_0), \dots, y(t)] \\ &= E[x(t_1)|y(t_0), \dots, y(t)] - \bar{x}(t_1|t) = 0 \end{aligned}$$

评论。 (四)、 $t \rightarrow \infty$ 时本部分内容的严格公式化需要希尔伯特空间的一些基本概念。见 Doob[15] 与 Loève[16]。

(五)、定理2的物理学阐述大体上。如果我们不担心高斯性质的假设, A 部分证明正交投影是所有合理损耗函数的最佳估计。如果我们确实担心高斯性质的假设, 甚至我们只考虑线性估计, 对于多数合理的损耗函数而言, 正交投影都不是最佳估计。由于事实上有一个有物理来源的随机过程在多大程度上是高斯的很难把握, 定理2究竟具有很广泛的还是很有限的重要性也很难判断。

(六)、直接将定理2推广为矢量值随机变量的情况。事实上, 定义 $\mathbf{y}(t_0), \dots, \mathbf{y}(t)$ 生成的线性流形 $\mathcal{Y}(t)$ 为随机矢量 $\mathbf{y}(t_0), \dots, \mathbf{y}(t)$ 中每一个的所有 m 个分量的所有线性组合

$$\sum_{i=t_0}^t \sum_{j=1}^m a_{ij} y_j(i)$$

的集合。随后的部分可仿效前面的部分。

(七)、定理2有效的说明了, 条件 A 或 B 下最佳估计是所有既往观测的线性组合。换言之, 最佳估计可考虑成线性滤波器的输出, 滤波器的输入是可观测随机变量实际发生的数值; 定理2为计算最佳滤波器的脉冲响应提供了方法。正如前面指出的, 对脉冲响应的认识并不是问题的完整的解; 出于这个原因, 不会给出计算脉冲响应的显式的公式。

5 随机过程模型

在同物理现象打交道时, 仅仅给出经验性的描述是不够的, 还必须对潜在的原因有一定了解。如果不能在某种意义上区分原因和影响, 亦即,

如果没有因果关系的假设，那么就几乎无法期待有用的结果。

通常人们都接受这样一个事实，即随机现象的主要宏观来源是独立的高斯过程。一个著名的例子是电阻中由热扰动造成的噪声电压。在大多数情况下，观测到的随机现象不能通过独立的随机变量来描述。通常，不同时刻观测到的随机信号之间的统计依赖性（相关性）是由主要随机来源于观测者之间存在动力学系统来解释的。因此以时间为自变量的随机函数可以考虑为接受独立高斯随机过程激励的动力学系统的输出。

高斯随机信号的一个重要属性是，当它们通过线性系统之后，它们仍然是高斯的（定理3的 A 部分）。假定有独立的、高斯的主要随机源，如果观测到的随机信号也是高斯的，我们即可假定观测者与主要源之间的动力学系统是线性的。我们之所以必须接受这条结论，也是因为对观测到的随机信号的统计属性缺乏细致的了解：给定已知一阶、二阶平均的任意随机过程，我们可以找到具有相同属性的高斯随机过程（定理3的 C 部分）。因此，高斯分布和线性动力学性质是自然而然的、互为印证的假设，特别是当统计数据贫乏的时候。

一个动力学系统（线性的或非线性的）是如何被描述的？基本思想是状态的概念。这意味着，直观的说，一些定量的信息（数的集合、函数，等等）。如果要预测系统的未来行为，这些信息就是必须知道的有关系统过去行为的最少量数据。这样，动力学性质就用术语状态过渡来描述，即，必须指出随着时间的流逝，状态是如何过渡成另一个状态的。

线性动力学系统一般可以用矢量微分方程

$$\left. \begin{aligned} & \frac{d\mathbf{x}}{dt} = \mathbf{F}(t)\mathbf{x} + \mathbf{D}(t)\mathbf{u}(t) \\ & \text{和} \\ & \mathbf{y}(t) = \mathbf{M}(t)\mathbf{x}(t) \end{aligned} \right\} \quad (12)$$

表示，其中 \mathbf{x} 是 n 维矢量，系统的状态（ \mathbf{x} 的分量 x_i 称为系统的状态变量）； $\mathbf{u}(t)$ 是 m 维（ $m \leq n$ ）矢量，表示系统的输入； $\mathbf{F}(t)$ 和 $\mathbf{D}(t)$ 分别是 $n \times n$ 和 $n \times m$ 维矩阵。如果 $\mathbf{F}(t), \mathbf{D}(t), \mathbf{M}(t)$ 的所有系数都是常数，我们就说动力学系统（式12）是时不变或定态的。最后， $\mathbf{y}(t)$ 是 p 维矢量，表示系统的输出； $\mathbf{M}(t)$ 是 $n \times p$ 维的矩阵； $p \leq n$ 。

式12的物理阐释已经在别的文献中详细讨论过 [18, 20, 23]。图1或许有所帮助。这是一副矩阵框图（图中箭头指示信号流向）。图1中的积分号事实上表示 n 个积分器，每个积分器的输出都是标量变量； $\mathbf{F}(t)$ 表明积分器的输出如何反馈到积分器的输入。故 $f_{ij}(t)$ 是第 j 积分器的输出反馈到 i

积分器的输入的系数。将这种重视形式的方法与更为传统的线性系统分析方法联系起来并不困难。

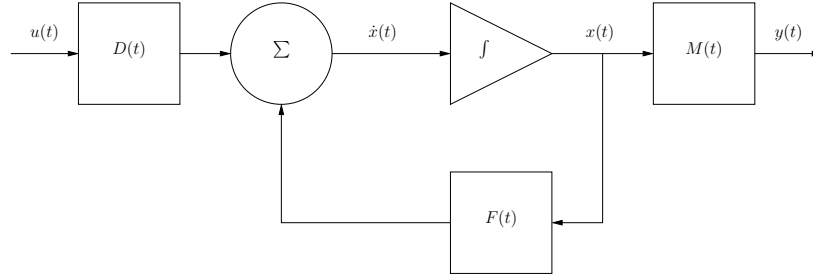


图 1: 一般线性连续动力学系统的矩阵框图

如果我们假定如式12的系统是定态的, $\mathbf{u}(t)$ 在每个抽样周期都不变, 即

$$\mathbf{u}(t + \tau) = \mathbf{u}(t); 0 \leq \tau < 1, t = 0, 1, \dots \quad (13)$$

那么式12立即可以变形成为更方便的离散形式。

$$\mathbf{x}(t + 1) = \Phi(1)\mathbf{x}(t) + \Delta(1)\mathbf{u}(t); t = 0, 1, \dots$$

其中 [18, 20]

$$\Phi(1) = \exp \mathbf{F} = \sum_{i=0}^{\infty} \mathbf{F}^i / i! \quad (\mathbf{F}^0 = \text{unit matrix})$$

且

$$\Delta(1) = \left(\int_0^1 \exp \mathbf{F} \tau d\tau \right) \mathbf{D}$$

见图2。也可以通过拉普拉斯变换的方法 [18, 20, 22, 24] 将 $\mathbf{F}\tau$ 表达成闭合的形式。如果 $\mathbf{u}(t)$ 满足式13但系统 (式12非定态, 类似的

$$\left. \begin{aligned} \mathbf{x}(t + 1) &= \Phi(t + 1; t) + \Delta(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{M}(t)\mathbf{x}(t) \end{aligned} \right\} \quad t = 0, 1, \dots \quad (14)$$

但是显然现在无法用闭合形式表示 $\Phi(t + 1; t), \Delta(t)$. 形如式14的方程也经常在研究复杂的抽样数据系统中遇到 [22], 见图2.

$\Phi(t + 1; t)$ 是系统 (如式12或式14) 的过渡矩阵。记号 $\Phi(t_2; t_1)$ 意味着从时刻 t_1 到时刻 t_2 的过渡。显然, $\Phi(t; t) = \mathbf{I}$ = 单位矩阵。如果系统

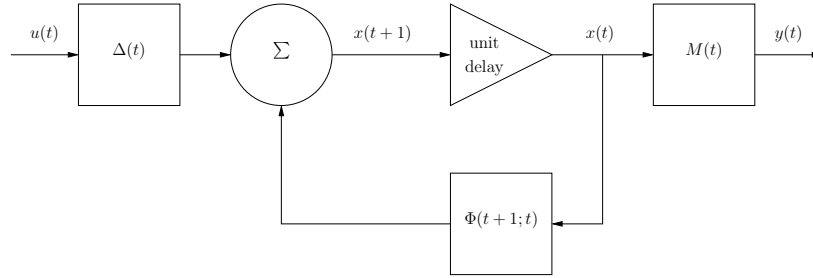


图 2: 一般线性离散动力学系统的矩阵框图

(如式12) 是定态的, 那么 $\Phi(t+1; t) = \Phi(t+1-t) = \Phi(1) = \text{常数}$ 。注意乘法规则: $\Phi(t; s)\Phi(s; r) = \Phi(t; r)$ 和倒数规则 $\Phi^{-1}(t; s) = \Phi(s; t)$, 其中 t, s, r 都是整数。定态系统中, $\Phi(t; \tau) = \exp \mathbf{F}(t - \tau)$ 。

作为前面讨论的结果, 我们将用

$$\mathbf{x}(t+1) = \Phi(t+1; t)\mathbf{x}(t) + \mathbf{u}(t) \quad (15)$$

模型来代表随机性现象, 其中 $\{\mathbf{u}(t)\}$ 是矢量值独立高斯随机过程, 均值为零, 完全由

$$E\mathbf{u}(t) = 0 \text{ 对于一切 } t;$$

$$E\mathbf{u}(t)\mathbf{u}'(s) = 0 \text{ 如果 } t \neq s$$

$$E\mathbf{u}(t)\mathbf{u}'(t) = \mathbf{G}(t).$$

描述 (依照定理3C 的观点)。当然 (定理3A), $\mathbf{x}(t)$ 也是零均值的高斯随机过程, 但不是独立的。事实上, 如果我们认为式15处于稳定状态 (假设是稳定系统), 换句话说, 如果我们忽略初始状态 $\mathbf{x}(t_0)$, 那么

$$\mathbf{x}(t) = \sum_{r=-\infty}^{t-1} \Phi(t; r+1)\mathbf{u}(r).$$

因此如果 $t \geq s$ 有

$$E\mathbf{x}(t)\mathbf{x}'(s) = \sum_{r=-\infty}^{s-1} \Phi(t; r+1)\mathbf{Q}'(r)\Phi'(s; r+1).$$

那么如果我们设想一个线性动力学系统并且知道高斯随机激励的统计学属性, 则可以简单的找到相应的高斯随机 $\{\mathbf{x}(t)\}$ 过程属性。

但是在现实生活中，情况通常是相反的。给定协方差矩阵 $E\mathbf{x}(t)\mathbf{x}'(s)$ (或者甚至要试着从有限的统计数据中估计该矩阵) 问题是得到式15和 $\mathbf{u}(t)$ 的统计学属性。这是实验与数据处理中很微妙且当前大部分悬而未决的问题。正如在大多数 Wiener 问题的工程学著作中那样，我们将发现从式15的模型开始，将获得模型本身视为单独的问题是很方便的。的确，如果可能应该将这两个问题联合优化；但是据作者所知并没有联合优化问题方面的研究。

作为小结，有关随机过程的以下假设已经作出：

物理随机现象可认为是源于主要随机来源激励的动力学系统。假定主要随机来源为独立的高斯随机过程，均值为零；动力学系统是线性的。随机过程因此由模型（如式15）描述。不考虑如何从实验数据获取具体数值以表示模型。

6 求解 Wiener 问题

让我们先来定义本文的主要问题。

问题一。 考虑动力学模型

$$\mathbf{x}(t+1) = \Phi(t+1; t)\mathbf{x}(t) + \mathbf{u}(t) \quad (16)$$

$$\mathbf{y}(t) = \mathbf{M}(t)\mathbf{x}(t) \quad (17)$$

其中 $\mathbf{u}(t)$ 是一个 n 维向量零均值独立高斯随机过程， $\mathbf{y}(t)$ 是一个 p 维向量 ($p \leq n$)， $\Phi(t+1; t)$, $\mathbf{M}(t)$ 分别是 $n \times n$, $p \times n$ 维矩阵，矩阵元是时间的非随机函数。

给定 $\mathbf{y}(t_0), \dots, \mathbf{y}(t)$ 的观测值，找到 $\mathbf{x}(t_1)$ 的估计 $\mathbf{x}^*(t_1|t)$ 以最小化期望损耗。（见图2, 其中 $\Delta(t) = \mathbf{I}$ 。）

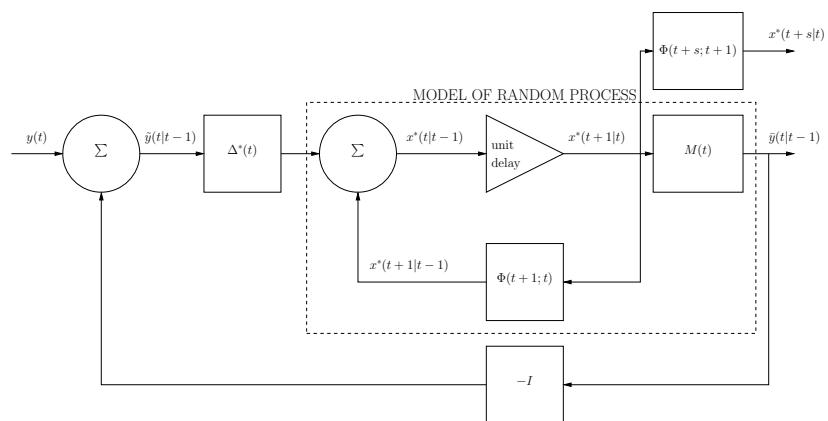


图 3: 最优滤波器的矩阵框图

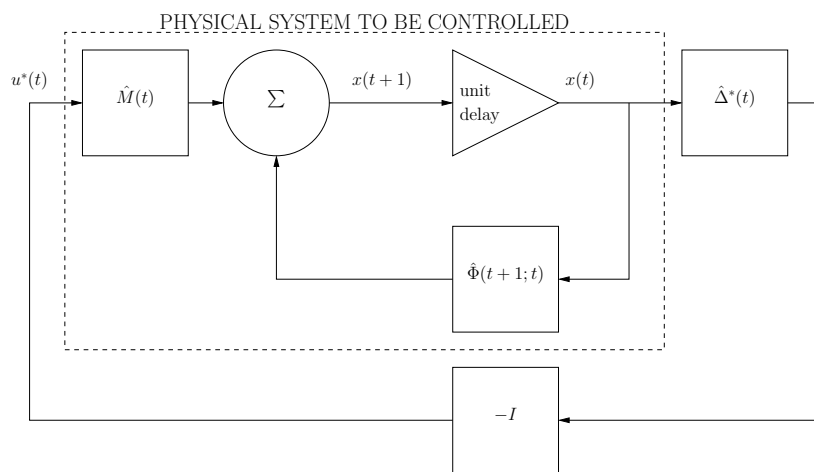


图 4: 最优控制器的矩阵框图

7 对偶问题

8 应用

9 结论

本文从“状态”观点公式化并解决了 Wiener 问题。一方面，这引领我们找到一种常规的方法来处理问题（包括用其他方法将碰上困难的问题）。另一方面，这也证明 Wiener 问题与控制论中的其他问题紧密相连。为了探索这其中的联系，还需要更进一步的工作。

参考文献

- [1] N. Wiener, “The Extrapolation, Interpolation and Smoothing of Stationary Time Series,” John Wiley & Sons, Inc., New York, N. Y., 1949.
- [2] L. A. Zadeh and J. R. Ragazzini, “An Extension of Wiener’s Theory of Prediction,” *Journal of Applied Physics*, vol. 21, 1950. pp. 645-655.
- [3] H. W. Bode and C. E. Shannon, “A Simplified Derivation of Linear Least-Squares Smoothing and Prediction Theory,” *Proceedings IRE*, vol. 38, 1950, pp. 417-425.
- [4] R. C. Booton, “An Optimization Theory for Time-Varying Linear Systems with Nonstationary Statistical Inputs,” *Proceedings IRE*, vol. 40, 1952, pp. 977-981.
- [5] J. H. Laning and R. H. Battin, “Random Processes in Automatic Control,” McGraw-Hill Book Company, Inc., New York, N. Y., 1956.
- [6] W. H. Davenport, Jr., and W. L. Root, “An Introduction to the Theory of Random Signals and Noise,” McGraw-Hill Book Company, Inc., New York, N. Y., 1958.
- [7] S. Darlington, “Linear Least-Squares Smoothing and Prediction, With Applications,” *Bell System Tech. Journal*, vol. 37, 1958, pp. 1221-1294.

- [8] G. Franklin, "The Optimum Synthesis of Sampled-Data Systems" Doctoral dissertation, Dept. of Elect. Engr., Columbia University, 1955.
- [9] A. B. Lees, "Interpolation and Extrapolation of Sampled Data," *Trans. IRE Prof. Group on Information Theory*, IT-2, 1956, pp. 173-175.
- [10] R. C. Davis, "On the Theory of Prediction of Nonstationary Stochastic Processes," *Journal of Applied Physics*, vol. 23, 1952, pp. 1047-1053.
- [11] M. Shibrot, "Optimization of Time-Varying Linear Systems With Nonstationary Inputs," *TRANS. ASME*, vol. 80, 1958, pp. 457-462.
- [12] M. Blum, "Recursion Formulas for Growing Memory Digital Filters," *Trans. IRE Prof. Group on Information Theory*, IT-4, 1958, pp. 24-30.
- [13] V. S. Pugachev, "The Use of Canonical Expansions of Random Functions in Determining an Optimum Linear Systems," *Automatics and Remote Control(USSR)*, vol. 17, 1956, pp. 489-499; translation pp. 545-556.
- [14] V. V. Solodovnikov and A. M. Batkov, "On the Theory of Self-Optimizing Systems (in German and Russian)," *Proc. Heidelberg Conference on Automatic Control*, 1956, pp. 308-323.
- [15] J. L. Doob, "Stochastic Processes," John Wiley & Sons, Inc., New York, N. Y., 1955.
- [16] M. Loève, "Probability Theory," Van Nostrand Company, Inc., New York, N. Y., 1955.
- [17] R. E. Bellman, I. Glicksberg, and O. A. Gross, "Some Aspects of the Mathematical Theory of Control Processes," *RAND Report R-313*, 1958, 244 pp.
- [18] R. E. Kalman and R. W. Koepcke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," *TRANS. ASME*, vol. 80, 1958, pp. 1820-1826.
- [19] J. E. Bertram, "Effect of Quantization in Sampled-Feedback Systems," *Trans. AIEE*, vol. 77, II, 1958, pp. 177-182.

- [20] R. E. Kalman and J. E. Bertram, "General Synthesis Procedure for Computer Control of Single and Multi-Loop Linear Systems," *Trans. AIEE*, vol. 77, II, 1958, pp. 602-609.
- [21] C. W. Merriam, III, "A Class of Optimum Control Systems," *Journal of the Franklin Institute*, vol. 267, 1959, pp. 267-281.
- [22] R. E. Kalman and J. E. Bertram, "A Unified Approach to the Theory of Sampling Systems," *Journal of the Franklin Institute*, vol. 267, 1959, pp. 405-436.
- [23] R. E. Kalman and R. W. Koepcke, "The Role of Digital Computers in the Dynamic Optimization of Chemical Reactors," Proc. Western Joint Computer Conference, 1959, pp. 107-116.
- [24] R. E. Kalman, "Dynamic Optimization of Linear Control Systems, I. Theory," to appear.
- [25] S. Sherman, "Non-Mean-Square Error Criteria," *Trans. IRE Prof. Group on Information Theory*, IT-4, 1958, pp. 125-126.
- [26] V. S. Pugachev, "On a Possible General Solution of the Problem of Determining Optimum Dynamic Systems," *Automatics and Remote Control*(USSR), vol. 17, 1956, pp. 585-589.
- [27] G. C. Newton, Jr., L. A. Gould, and J. F. Kaiser, "Analytical Design of Linear Feedback Controls," John Wiley & Sons, Inc., New York, N. Y., 1957.
- [28] O. J. M. Smith, "Feedback Control Systems," McGraw-Hill Book Company, Inc., New York, N. Y., 1958.
- [29] R. E. Kalman, "On the General Theory of Control Systems," Proceedings First International Conference on Automatic Control, Moscow, USSR, 1960.

A 随机过程：基本概念

为方便读者，我们在此对有关概率论以及随机过程的一些基本定义以及事实进行回顾。一切都用尽可能简明的方式展现；如需深入或延伸，可查阅 Laning 与 Battin 或 Doob 的工作。

随机变量是取值依赖于偶然性事件之结果的函数。随机变量的值可以是任意方便的数学量；实数或复数，矢量，等等。为了简明，我们在此将仅考虑实值随机变量，但这并非现实的约束。记随机变量为 x, y, \dots ，并记其值为 ξ, η, \dots 。随机变量的和、积、函数仍然是随机变量。

随机变量 x 可通过说明其小于或等于某个常数 ξ 的概率来显式的定义。可以用符号表示为

$$Pr(x \leq \xi) = F_x(\xi); F_x(-\infty) = 0, F_x(+\infty) = 1 \quad (18)$$

$F_x(\xi)$ 称为随机变量 x 的概率分布函数。当 $F_x(\xi)$ 对 ξ 可导时 $f_x(\xi) = \frac{dF_x(\xi)}{d\xi}$ 称为 x 的概率密度函数。

随机变量 x 的任意非随机函数 $g(x)$ 的期望值（数学期望、统计平均、系综平均、平均值都是常见的同义词）定义为

$$Eg(x) = E[g(x)] = \int_{-\infty}^{\infty} g(\xi) dF_x(\xi) = \int_{-\infty}^{\infty} g(\xi) f_x(\xi) d\xi \quad (19)$$

正如上文指出的，通常为了方便将会省略 E 后面的括号。随机变量（有限的或无限的）的序列

$$\{x(t)\} = \dots, x(-1), x(0), x(1), \dots \quad (20)$$

叫做离散（或离散参数）随机（或概率性，此处原文为 *stochastic*）过程。式20表示的随机过程的一组特定的观测值

$$\dots, \xi(-1), \xi(0), \xi(1), \dots$$

称为该过程的实现（或抽样函数）。直觉上，随机过程是一组编排好索引从而有了时间概念的随机变量。

称随机过程是不相关的，如果

$$Ex(t)x(s) = Ex(t)Ex(s) \quad (t \neq s)$$

如果，更进一步的，

$$Ex(t)x(s) = 0 \quad (t \neq s)$$

则该随机过程是正交的。将 $x(t)$ 替换为 $x'(t) = x(t) - Ex(t)$ 则任意不相关的随机过程都可以变成正交的随机过程，因为

$$Ex'(t)x'(s) = E[x(t) - Ex(t)] \cdot [x(s) - Ex(s)] = Ex(t)x(s) - Ex(t)Ex(s) = 0$$

记住这个实用的结论，如果某随机过程是正交的，那么

$$E[x(t_1) + x(t_2) + \dots]^2 = Ex^2(t_1) + Ex^2(t_2) + \dots \quad (t_1 \neq t_2 \neq \dots)$$

如果 \mathbf{x} 是以 x_1, \dots, x_n （当然也是随机变量）为分量的矢量值随机变量，矩阵

$$[E(x_i - Ex_i)(x_j - Ex_j)] = E(\mathbf{x} - E\mathbf{x})(\mathbf{x}' - E\mathbf{x}') = \text{cov } \mathbf{x} \quad (21)$$

称为 \mathbf{x} 的协方差矩阵。

随机过程可以通过说明任意有限数量的如下类型的事例同时发生的概率来显式的明确说明

$$x(t_1) \leq \xi_1, \dots, x(t_n) \leq \xi_n; (t_1 \neq \dots \neq t_n), \text{ i.e.,}$$

$$Pr[(x(t_1) \leq \xi_1, \dots, x(t_n) \leq \xi_n)] = F_{x(t_1), \dots, x(t_n)}(\xi_1, \dots, \xi_n) \quad (22)$$

其中 $F_{x(t_1), \dots, x(t_n)}$ 叫做随机变量 $x(t_1), \dots, x(t_n)$ 的联合概率分布函数。这样，联合概率密度函数即为

$$f_{x(t_1), \dots, x(t_n)}(\xi_1, \dots, \xi_n) = \frac{\partial^n F_{x(t_1), \dots, x(t_n)}}{\partial \xi_1 \dots \partial \xi_n}$$

，假如需要的导数存在。 n 个随机变量的任意非随机函数的期望值 $Eg[x(t_1), \dots, x(t_n)]$ 由类似式19的 n 重积分定义。

如果对于任意有限个 $t_1 \neq \dots \neq t_n$ 式22等于一阶分布的积

$$Pr[x(t_1) \leq \xi_1] \dots Pr[x(t_n) \leq \xi_n]$$

则该随机过程是独立的。如果一组随机变量是独立的，显然它们也是不相关的。但是一般来说，反过来则并不成立。对于多于两个随机变量的一个集合，任意两个互相独立并不足以使整个集合是独立的。

给定随机变量 $x(t_1), \dots, x(t_n)$ 出现的实际值 ξ_1, \dots, ξ_n 来考虑随机变量 $x(t_{n+1})$ 的概率分布常常是人们关注的。记此分布为

$$\begin{aligned} Pr[x(t_{n+1}) \leq \xi_{n+1} | x(t_1) = \xi_1, \dots, x(t_n) = \xi_n] \\ = \frac{\int_{-\infty}^{\xi_{n+1}} f_{x(t_1), \dots, x(t_{n+1})}(\xi_1, \dots, \xi_{n+1}) d\xi_{n+1}}{f_{x(t_1), \dots, x(t_n)}(\xi_1, \dots, \xi_n)} \end{aligned} \quad (23)$$

并称之为给定 $x(t_1), \dots, x(t_n)$ 的 $x(t_{n+1})$ 的条件概率分布函数。条件期望

$$E\{g[x(t_{n+1})]|x(t_1), \dots, x(t_n)\}$$

的定义类似式19. 条件期望是一个随机变量；它服从

$$E[E\{g[x(t_{n+1})]|x(t_1), \dots, x(t_n)\}] = E\{g[x(t_{n+1})]\}$$

在本文所关注的所有情况中，如式19或者式23的积分并不需要显式的计算出来，只有期望值的概念是需要的。

随机变量 x 是高斯的（或者正态分布的），如果

$$f_x(\xi) = \frac{1}{[2\pi E(x - Ex)^2]^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \frac{(\xi - Ex)^2}{E(x - Ex)^2}\right]$$

亦即人们熟知的钟状曲线。类似的，随机矢量 \mathbf{x} 是高斯的，如果

$$f_{\mathbf{x}}(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \mathbf{C})^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (\xi - E\mathbf{x})' \mathbf{C}^{-1} (\xi - E\mathbf{x})\right]$$

其中 \mathbf{C}^{-1} 是 \mathbf{x} 协方差矩阵（式21）的逆。类似的可以定义高斯随机过程。

高斯随机变量与随机过程的重要性大体上是由于下列事实：

定理 3. A 高斯随机过程的线性函数（因而有条件期望）是高斯随机变量。

B 正交高斯随机变量是独立的。

C 给定任意均值为 $Ex(t)$ ，协方差为 $Ex(t)x(s)$ 的随机过程，存在唯一的高斯随机过程具备同样的均值与协方差。