

本章讨论从统计意义上的最优滤波问题或波形的最优线性估计问题。首先讨论对于平稳随机信号的维纳滤波器的设计,紧接着讨论一种特殊的维纳滤波器,最优一步线性预测,通过前向线性预测和后向线性预测的对称关系,导出了求解 Yule-Walker 方程的快速递推算法,并由此导出格型滤波器结构。接着讨论在有限数据集条件下的最小二乘滤波器(LS),这是一种易于实现的有效滤波器并具有统计上的最优性,并介绍求解 LS 问题的奇异值分解(SVD)算法。

3.1 维纳滤波

维纳滤波器是统计意义上的最优滤波,或者等价地说是波形的最优线性估计,它要求输入信号是宽平稳随机信号。本章针对离散信号情况,详细讨论 FIR 结构和 IIR 结构的维纳滤波器。

由信号当前值及它的各阶位移 $\{x(n-k)\}_{k=-\infty}^{+\infty}$, 估计一个期望信号 $d(n)$, 输入信号 $x(n)$ 是宽平稳的, $x(n)$ 和 $d(n)$ 是联合宽平稳的, 要求这个估计的均方误差最小, 这就是维纳滤波器, 它实质上是一个波形估计问题。为了方便, 本章假设所涉及的信号都是零均值。

3.1.1 实际问题中的维纳滤波

初看起来, 维纳滤波器有些抽象, 用一个输入信号估计一个期望响应, 期望响应是什么? 通过如下几个来自于实际的应用实例来理解维纳滤波器是对许多不同应用问题的抽象, 从而构成了最优滤波器的基础框架。

1. 通信的信道均衡器

在通信系统中, 为了在接收器端补偿信道传输引入的各种畸变, 在对接收信号进行检测之前, 通过一个滤波器对信道失真进行校正, 这个滤波器称为信道均衡器。为了说明均衡器问题, 这里只简单地讨论基带传输, 忽略通信系统中的调制解调和各种编解码过程, 设通信系统的发射端发送序列 $s(n)$, 通过信道传输, 在接收端的滤波器输入端得到因信道不理想而发生了失真的信号 $x(n)$, $x(n)$ 可能包含了信道不理想产生的畸变和加性噪声, 例如, 无线通信系统中的多径效应和环境噪声。 $x(n)$ 作为维纳滤波器的输入信号, 通过确定滤波器的权系数, 使得滤波器的输出 $s'(n)$ 尽可能逼近一个期望信号 $d(n)$, 也就是发送序列的延迟 $s(n-k)$ 。采用均方误差最小准则设计滤波器系数, 使估计误差 $e(n) = d(n) - s'(n)$ 的



视频讲解

均方误差值最小,见图 3.1.1。

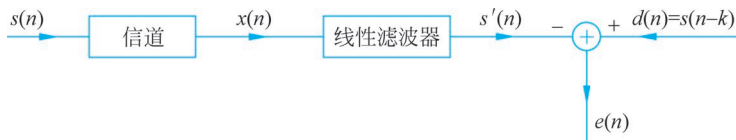


图 3.1.1 信道均衡器的结构示意图

2. 系统辨识

系统辨识的问题是: 有一个系统是未知的或需要用一个 LTI 系统进行模型化, 设计一个线性滤波器尽可能精确地逼近这个待处理系统。维纳滤波器实现一个从统计意义上最优的对未知系统的线性逼近。图 3.1.2 示出一个实现系统辨识的原理框图。使维纳滤波器和未知系统使用同一个输入信号 $x(n)$, 未知系统的输出作为维纳滤波器的期望响应 $d(n)$, 设计滤波器系数使得滤波器输出 $y(n)$ 与期望响应 $d(n)$ 之间的估计误差的均方值最小。

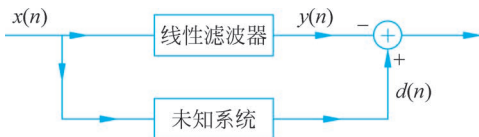


图 3.1.2 线性系统辨识的结构

3. 最优线性预测

通过一个随机信号已存在的数据 $\{x(n-1), x(n-2), \dots, x(n-M)\}$ 来预测一个新值 $x(n)$, 这是一步前向线性预测问题, 由 $\{x(n-1), x(n-2), \dots, x(n-M)\}$ 的线性组合得到对 $x(n)$ 的最优估计, 相当于设计一个 FIR 滤波器对 $\{x(n-1), x(n-2), \dots, x(n-M)\}$ 进行线性运算, 来估计期望响应 $d(n) = x(n)$, 维纳滤波器是用于设计均方误差最小的最优预测器。

通过如上几个例子, 可以抽象出一种有用的滤波器结构, 这就是维纳滤波器。维纳滤波器的一般结构示于图 3.1.3, 滤波器自身是一个 FIR 或 IIR 滤波器, 滤波器输入信号 $x(n)$, 输出 $y(n)$, 有一个待估计的期望响应 $d(n)$, 滤波器系数的设计准则是使得滤波器的输出 $y(n)$ (或写成 $\hat{d}(n)$) 是均方意义上对期望响应的最优线性估计。

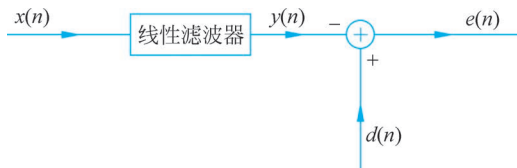


图 3.1.3 维纳滤波器的一般结构

维纳滤波器的目的是求最优滤波器系数 $\mathbf{w}_o = [\dots, w_{o,0}, w_{o,1}, \dots, w_{o,k}, \dots]^T$, 使

$$J(n) = E[|e(n)|^2] = E[|d(n) - \hat{d}(n)|^2] \quad (3.1.1)$$

最小。当滤波器系数有无穷多个(即单位抽样响应无限长)时, 对应 IIR 结构的维纳滤波器,

当滤波器系数为有限个时,对应 FIR 结构的维纳滤波器。FIR 结构的维纳滤波器的滤波部分的示意图如图 3.1.4 所示,在信号处理的文献中,也常称这个结构为横向滤波器。

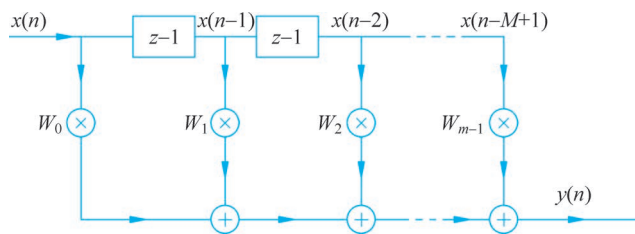


图 3.1.4 维纳滤波的横向滤波器

3.1.2 从估计理论观点导出维纳滤波

将存在有限个观测数据情况下的线性贝叶斯估计的结论,用于波形的估计,可以直接得到 FIR 结构的维纳滤波器,这表明维纳滤波器是波形的线性贝叶斯估计。

为简单起见,先讨论实信号的情况。将线性贝叶斯估计推广到波形估计的情况。在 n 时刻,由输入 $x(n)$ 及其延迟量 $\{x(n-1), x(n-2), \dots, x(n-M+1)\}$ 作为观测值,估计期望信号的波形在 n 时刻的值 $d(n)$,确定权系数 $\mathbf{w} = [w_0, w_1, \dots, w_{M-1}]^T$,使估计误差均方值最小,即在权系数取 $\mathbf{w}_0$ 时

$$J(n) = E[(d(n) - \hat{d}(n))^2] \quad (3.1.2)$$

达到最小,这里波形估计 $\hat{d}(n)$ 写为

$$\hat{d}(n) = \sum_{i=0}^{M-1} w_{0i} x(n-i) = \mathbf{w}_0^T \mathbf{x}(n) \quad (3.1.3)$$

其中

$$\mathbf{w}_0 = [w_{00}, w_{01}, \dots, w_{0(M-1)}]^T \quad (3.1.4)$$

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T \quad (3.1.5)$$

对一个固定时间 n ,波形的瞬时估计与随机参数的线性贝叶斯估计是完全相同的问题,利用式(2.5.10),时刻 n 的波形估计器的系数满足如下方程:

$$\mathbf{R}(n) \mathbf{w}_0 = \mathbf{r}_{xd}(n) \quad (3.1.6)$$

这里

$$\begin{aligned} \mathbf{R}(n) &= E[\mathbf{x}(n) \mathbf{x}^T(n)] \\ &= \begin{bmatrix} r(n, n) & r(n, n-1) & \cdots & r(n, n-M+2) & r(n, n-M+1) \\ r(n-1, n) & r(n-1, n-1) & \cdots & r(n-1, n-M+2) & r(n-1, n-M+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(n-M+2, n) & r(n-M+2, n-1) & \cdots & r(n-M+2, n-M+2) & r(n-M+2, n-M+1) \\ r(n-M+1, n) & r(n-M+1, n-1) & \cdots & r(n-M+1, n-M+2) & r(n-M+1, n-M+1) \end{bmatrix} \end{aligned} \quad (3.1.7)$$

$$\mathbf{r}_{xd}(n) = E[\mathbf{x}(n) d(n)] = [r_{xd}(n, n), r_{xd}(n-1, n), \dots, r_{xd}(n-M+1, n)]^T \quad (3.1.8)$$

在一般情况下,对于不同时刻 n ,权系数不同,因此,需要对每个时刻重新计算权系数。

当输入信号 $x(n)$ 是一个宽平稳随机信号, 并且 $x(n)$ 和 $d(n)$ 是联合平稳时, 上述自相关矩阵和互相关向量简化为

$$\mathbf{R}(n) = \mathbf{R} = \begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(M-1) \\ r_x(-1) & r_x(0) & \cdots & r_x(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(-M+1) & r_x(-M+2) & \cdots & r_x(0) \end{bmatrix} \quad (3.1.9)$$

和

$$\mathbf{r}_{xd}(n) = \mathbf{r}_{xd} = [r_{xd}(0), r_{xd}(-1), \cdots, r_{xd}(-M+1)]^T \quad (3.1.10)$$

权系数满足的方程简化为

$$\mathbf{R}\mathbf{w}_o = \mathbf{r}_{xd} \quad (3.1.11)$$

注意, 权系数是个确定性的量, 不是随机量。对每个时刻 n , $d(n)$ 的估计值 $\hat{d}(n)$ 相当于图 3.1.4 线性时不变滤波器的输出。最小均方估计误差为

$$J_{\min} = \sigma_d^2 - \mathbf{r}_{xd}^T \mathbf{R}^{-1} \mathbf{r}_{xd} \quad (3.1.12)$$

从维纳滤波器是线性贝叶斯波形估计的观点, 需注意如下几点。

(1) 在均方误差意义上, 维纳滤波器是线性 FIR 滤波器中的最优滤波器, 但可能存在一些非线性滤波器能达到更好结果。

(2) 在 $x(n)$ 和 $d(n)$ 是联合高斯分布条件下, 维纳滤波也是总体最优的, 不存在非线性滤波器能达到更好的结果。

(3) 从线性贝叶斯估计推导过程知, 在滤波器系数取非最优的任意权系数 $\mathbf{w}$ 时, 其误差性能表达式为

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{w}^T \mathbf{r}_{xd} - \mathbf{r}_{xd}^T \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (3.1.13)$$

它是 $\mathbf{w}$ 的超二次曲面, 当自相关矩阵 $\mathbf{R}$ 是正定矩阵时, 该曲面只有一个最小点, 当 $\mathbf{w} = \mathbf{w}_o$ 时 $J(\mathbf{w}) = J_{\min}$ 。

3.1.3 维纳滤波器-正交原理

维纳滤波器是一个最优线性滤波器, 图 3.1.3 是一个一般表示框图, 滤波器核是 IIR 或 FIR 的, 在实信号情况下, 已经导出了求解 FIR 型维纳滤波器的方程。在第 2 章讨论了线性最优估计的正交性原理, 第 2 章的正交原理是由最优线性估计方程导出的。在最优线性滤波器理论中, 正交原理是一个基本分析工具, 由正交原理出发, 很容易导出线性最优估计和维纳滤波器的方程式。由于正交原理应用的广泛性和简洁性, 并且贯穿于平稳、非平稳和有限数据等多种情况, 在本节, 对复信号的一般情况, 重新导出正交原理的一般形式, 并利用正交原理, 重新推导复信号情况下维纳滤波器的一般方程。先推导适应于 IIR 和 FIR 的一般结论, 然后分别讨论 FIR 和 IIR。

将一般的复数形式维纳滤波器的问题重新描述如下。

设输入随机过程 $x(n)$ 为复信号, 由 $\{x(n-k)\}_{k=-\infty}^{+\infty}$ 估计期望响应 $d(n)$, 求复数权系数 $\cdots, w_{o(-1)}, w_{o0}, w_{o1}, w_{o2}, \cdots$, 使输出误差 $e(n) = d(n) - y(n)$ 的均方值最小。

如下推导一般的正交性原理, 实数据是特例。在复信号的一般情况下, 滤波器输出和估计误差分别写为

$$y(n) = \sum_{k=-\infty}^{+\infty} w_k^* x(n-k) \quad (3.1.14)$$

$$e(n) = d(n) - y(n) \quad (3.1.15)$$

均方误差表示为

$$J(\mathbf{w}) = E\{|e(n)|^2\} = E\{e(n)e^*(n)\} \quad (3.1.16)$$

设滤波器权系数是如下复数形式:

$$w_k = a_k + jb_k \quad (3.1.17)$$

定义针对滤波器系数向量的梯度算子为

$$\nabla = [\cdots, \nabla_0, \nabla_1, \cdots, \nabla_k, \cdots]^T \quad (3.1.18)$$

其中, 针对参数 w_k 的梯度算子为

$$\nabla_k = \frac{\partial}{\partial w_k^*} = \frac{\partial}{\partial a_k} + j \frac{\partial}{\partial b_k} \quad (3.1.19)$$

为求解最优滤波器的权系数, 将梯度算子作用于均方误差 $J(\mathbf{w})$, 并令其为零。其中, 第 k 项为

$$\nabla_k J(\mathbf{w}) = \frac{\partial J(\mathbf{w})}{\partial a_k} + j \frac{\partial J(\mathbf{w})}{\partial b_k} \quad (3.1.20)$$

求参数 $\cdots, w_{-1}, w_0, w_1, \cdots$ 的值, 使得 $J(\mathbf{w})$ 最小, 即使得

$$\nabla J(\mathbf{w}) = \mathbf{0} \quad (3.1.21)$$

或等价地每个分量

$$\nabla_k J(\mathbf{w}) = 0, \quad k = \cdots, -1, 0, 1, 2, \cdots \quad (3.1.22)$$

将 $J(\mathbf{w}) = E\{e(n)e^*(n)\}$ 代入梯度算子定义式(3.1.20), 得到

$$\nabla_k J(\mathbf{w}) = E \left[\frac{\partial e(n)}{\partial a_k} e^*(n) + \frac{\partial e^*(n)}{\partial a_k} e(n) + \frac{\partial e(n)}{\partial b_k} j e^*(n) + \frac{\partial e^*(n)}{\partial b_k} j e(n) \right] \quad (3.1.23)$$

由

$$e(n) = d(n) - \sum_{k=-\infty}^{+\infty} w_k^* x(n-k) \quad (3.1.24)$$

得到

$$\frac{\partial e(n)}{\partial a_k} = -x(n-k) \quad (3.1.25)$$

$$\frac{\partial e(n)}{\partial b_k} = jx(n-k) \quad (3.1.26)$$

$$\frac{\partial e^*(n)}{\partial a_k} = -x^*(n-k) \quad (3.1.27)$$

$$\frac{\partial e^*(n)}{\partial b_k} = -jx^*(n-k) \quad (3.1.28)$$

将如上各项代入 $\nabla_k J(\mathbf{w})$ 表达式(3.1.23), 整理得

$$\nabla_k J(\mathbf{w}) = -2E[x[n-k]e^*[n]], \quad k = \cdots, -1, 0, 1, \cdots \quad (3.1.29)$$

当

$$\nabla_k J(\mathbf{w}) = 0, \quad k = \cdots, -1, 0, 1, \cdots \quad (3.1.30)$$

时, $J(\mathbf{w})$ 达到最小。设 $J(\mathbf{w})$ 达到最小时, 用 $w_o, e_o(n)$ 表示权系数和误差 $e(n)$, 即当

$$J(\mathbf{w}) = J_{\min} \quad (3.1.31)$$

时,满足

$$E[x(n-k)e_o^*(n)] = 0, \quad k = \dots, -1, 0, 1, \dots \quad (3.1.32)$$

或等价地

$$E[x^*(n-k)e_o(n)] = 0, \quad k = \dots, -1, 0, 1, 2, \dots \quad (3.1.33)$$

以上两式为一般的正交性原理:滤波器达到最优时,误差和输入信号正交。正交性原理的一个直接推论是,滤波器达到最优时,输出和误差也是正交的,即

$$E[y_o(n)e_o^*(n)] = 0 \quad (3.1.34)$$

注意到,在正交性原理式(3.1.32)中,输入信号各阶位移 $x(n-k)$ 与 $e_o(n)$ 正交,式(3.1.32)所表示的正交方程数目与滤波器权系数 $\dots, w_{-1}, w_0, w_1, \dots$ 数目是一致的,即维纳滤波器有多少个权系数,式(3.1.32)就包含多少个正交方程式。

现在通过正交性原理,导出更一般的维纳滤波器的设计方程,即维纳-霍普夫(Wiener-Hopf)方程。由正交性原理

$$E[x(n-k)e_o^*(n)] = 0, \quad k = \dots, -1, 0, 1, 2, \dots \quad (3.1.35)$$

代入误差的表达式(3.1.24)到式(1.1.35),得

$$E\left[x(n-k)\left(d^*(n) - \sum_{i=-\infty}^{+\infty} w_{oi}x^*(n-i)\right)\right] = 0, \quad k = \dots, -1, 0, 1, 2, \dots \quad (3.1.36)$$

交换求期望和求和次序,将如上方程展开为

$$\sum_{i=-\infty}^{+\infty} w_{oi}E[x(n-k)x^*(n-i)] = E[x(n-k)d^*(n)], \quad k = \dots, -1, 0, 1, 2, \dots \quad (3.1.37)$$

由自相关和互相关的定义

$$r_x(i-k) = E[x(n-k)x^*(n-i)] \quad (3.1.38)$$

$$r_{xd}(-k) = E[x(n-k)d^*(n)] \quad (3.1.39)$$

代入式(3.1.37),得到

$$\sum_{i=-\infty}^{+\infty} w_{oi}r_x(i-k) = r_{xd}(-k), \quad k = \dots, -1, 0, 1, 2, \dots \quad (3.1.40)$$

式(3.1.40)称为维纳-霍普夫方程,解此方程,可得到最优权系数 $\{w_{oi}, i = \dots, -1, 0, 1, 2, \dots\}$ 。式(3.1.40)是维纳滤波器的一般方程,根据滤波器权系数是有限个还是无限个可以分别设计 FIR 和 IIR 型维纳滤波器。

3.1.4 FIR 维纳滤波器

对于 M 个系数的 FIR 滤波器(横向滤波器),只有 M 个权系数 $\{w_{oi}, i = 0, 1, \dots, M-1\}$ 非零,维纳-霍普夫方程简化为有限个未知量的线性方程组,即式(3.1.40)简化为

$$\sum_{i=0}^{M-1} w_{oi}r_x(i-k) = r_{xd}(-k), \quad k = 0, 1, \dots, M-1 \quad (3.1.41)$$

更方便的是写成矩阵形式。令

$$\begin{aligned} \mathbf{R} &= E[\mathbf{x}(n)\mathbf{x}^H(n)] \\ &= \begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(M-1) \\ r_x^*(1) & r_x(0) & \cdots & r_x(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x^*(M-1) & r_x^*(M-2) & \cdots & r_x(0) \end{bmatrix} \end{aligned} \quad (3.1.42)$$

和

$$\mathbf{r}_{xd} = E[\mathbf{x}(n)d^*(n)] = [r_{xd}(0), r_{xd}(-1), \dots, r_{xd}(1-M)]^T \quad (3.1.43)$$

维纳-霍普夫方程的矩阵形式为

$$\mathbf{R}\mathbf{w}_o = \mathbf{r}_{xd} \quad (3.1.44)$$

解方程求得

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{r}_{xd} \quad (3.1.45)$$

这个关系式与通过线性贝叶斯波形估计得到的结论是相同的。

下面用正交性原理,导出一般复信号情况下维纳滤波器的最小均方误差公式。在达最优时, $y_o(n)$ 也写成 $\hat{d}(n|X_n)$,意指由 $x(n), x(n-1), \dots, x(n-M+1)$ 张成的线性空间 X_n 对 $d(n)$ 的最优估计。估计误差为

$$e_o(n) = d(n) - y_o(n) = d(n) - \hat{d}(n|X_n) \quad (3.1.46)$$

也可以写成

$$d(n) = e_o(n) + \hat{d}(n|X_n) \quad (3.1.47)$$

由正交性原理的推论, $\hat{d}(n|X_n)$ 和 $e_o(n)$ 是正交的,因此

$$\sigma_d^2 = E[|e_o(n)|^2] + \sigma_{\hat{d}}^2 = J_{\min} + \sigma_{\hat{d}}^2 \quad (3.1.48)$$

即

$$J_{\min} = \sigma_d^2 - \sigma_{\hat{d}}^2 \quad (3.1.49)$$

由

$$\hat{d}(n|X_n) = \sum_{k=0}^{M-1} \tau_{ok}^* x(n-k) = \mathbf{w}_o^H \mathbf{x}(n) \quad (3.1.50)$$

得

$$\begin{aligned} \sigma_{\hat{d}}^2 &= E[\hat{d}(n|X_n)\hat{d}^*(n|X_n)] = E[\mathbf{w}_o^H \mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}_o] \\ &= \mathbf{w}_o^H E[\mathbf{x}(n)\mathbf{x}^H(n)]\mathbf{w}_o \\ &= \mathbf{w}_o^H \mathbf{R}\mathbf{w}_o = \mathbf{w}_o^H \mathbf{r}_{xd} = \mathbf{r}_{xd}^H \mathbf{w}_o = \mathbf{r}_{xd}^H \mathbf{R}^{-1}\mathbf{r}_{xd} \end{aligned} \quad (3.1.51)$$

将式(3.1.51)代入式(3.1.49),得

$$J_{\min} = \sigma_d^2 - \sigma_{\hat{d}}^2 = \sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{w}_o = \sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{R}^{-1}\mathbf{r}_{xd} \quad (3.1.52)$$

由正交原理导出的维纳滤波器和从线性贝叶斯估计得到的维纳滤波器,分别得到一些有益的启示。从估计理论看,维纳滤波器只是一个线性贝叶斯估计,它是最优估计的一个线性逼近,只有在高斯情况下,它才是真正的最优滤波器,在其他分布情况下,非线性滤波器可能达到比线性最优滤波器更优的结果。从一般线性最优滤波器的正交原理出发,我们容易忽视这些限制。

为了后续导出自适应滤波器的需要(第5章),导出在复信号的一般情况下的误差性能表面。上面的讨论集中在最优滤波器的权系数设计和达到最优时均方误差的结果。讨论在滤波器权系数未取最优值时,滤波器的均方误差怎样随滤波器系数的不同取值而变化的规律是有意义的,对导出自适应滤波算法有指导作用。假设滤波器取一组任意权系数,对期望响应的估计误差为

$$e(n) = d(n) - \sum_{k=0}^{M-1} \tau_k^* x(n-k) \quad (3.1.53)$$

直接代入

$$J(\mathbf{w}) = E[e(n)e^*(n)] \quad (3.1.54)$$

整理得

$$J(\mathbf{w}) = \sigma_d^2 - \sum_{k=0}^{M-1} w_k^* r_{xd}(-k) - \sum_{k=0}^{M-1} w_k r_{xd}^*(-k) + \sum_{k=0}^{M-1} \sum_{i=0}^{M-1} w_k^* w_i r_x(i-k) \quad (3.1.55)$$

由上式可以看出, $J(\mathbf{w})$ 是 w_k 的二次曲面, 由于 2 次项的系数是一个半正定序列, 因此二次曲面是一个碗状曲面, 碗口向上, 若自相关矩阵是正定的, 则 $J_{\min}$ 在碗底且唯一, 其实, 由上式直接对各 w_k 求导, 并令其为 0, 得到一组方程, 正是维纳-霍普夫方程。

上式也可以直接写成矩阵形式:

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{w}^H \mathbf{r}_{xd} - \mathbf{r}_{xd}^H \mathbf{w} + \mathbf{w}^H \mathbf{R} \mathbf{w} \quad (3.1.56)$$

它可以整理成如下形式:

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{R}^{-1} \mathbf{r}_{xd} + (\mathbf{w} - \mathbf{R}^{-1} \mathbf{r}_{xd})^H \mathbf{R} (\mathbf{w} - \mathbf{R}^{-1} \mathbf{r}_{xd}) \quad (3.1.57)$$

上式在 $\mathbf{w}_o = \mathbf{R}^{-1} \mathbf{r}_{xd}$ 时, 达到最小, 且最小值为

$$J_{\min} = \min_{\mathbf{w}} J(\mathbf{w}) = \sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{R}^{-1} \mathbf{r}_{xd} \quad (3.1.58)$$

式(3.1.58)与式(3.1.52)的结果是一致的。性能表面 $J(\mathbf{w})$ 也可以写成

$$J(\mathbf{w}) = J_{\min} + (\mathbf{w} - \mathbf{w}_o)^H \mathbf{R} (\mathbf{w} - \mathbf{w}_o) \quad (3.1.59)$$

利用自相关矩阵的特征值分解式

$$\mathbf{R} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H \quad (3.1.60)$$

得到

$$J(\mathbf{w}) = J_{\min} + (\mathbf{w} - \mathbf{w}_o)^H \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H (\mathbf{w} - \mathbf{w}_o) \quad (3.1.61)$$

令

$$\mathbf{v} = \mathbf{Q}^H (\mathbf{w} - \mathbf{w}_o) \quad (3.1.62)$$

上式可写成

$$J(\mathbf{w}) = J_{\min} + \mathbf{v}^H \mathbf{\Lambda} \mathbf{v} = J_{\min} + \sum_{k=1}^M \lambda_k v_k v_k^* = J_{\min} + \sum_{k=1}^N \lambda_k |v_k|^2 \quad (3.1.63)$$

这里, $\mathbf{v}$ 是 $\mathbf{w}$ 的坐标变换, 由 $\mathbf{w}$ 通过原点移位和坐标旋转得到 $\mathbf{v}$ 的坐标系, 通过坐标变换, 得到以 $\mathbf{v}$ 为变量的均方误差的如上的规范形式, 对于一个给定 $J \neq J_{\min}$, 有

$$J - J_{\min} = \sum_{k=1}^M \frac{|v_k|^2}{\frac{1}{\lambda_k}} \quad (3.1.64)$$

这是超椭圆, $\frac{1}{\lambda_k}$ 为其一个轴。 $J - J_{\min}$ 取一个固定值所对应的 $\mathbf{v}$ (等价于 $\mathbf{w}$) 的所有取值称为性能表面的等高线, 式(3.1.64)表明线性滤波器性能表面的等高线是超平面上的椭圆。

例 3.1.1 有一信号 $s(n)$, 它的自相关序列为 $r_s(k) = \frac{10}{27} \left(\frac{1}{2}\right)^{|k|}$, 被一加法性白噪声所污染, 噪声方差为 $2/3$, 白噪声与信号不相关。被污染信号 $x[n]$ 作为维纳滤波器的输入, 求两个系数 FIR 滤波器使输出信号是 $s(n)$ 的尽可能的恢复。

解: 本题中, 显然维纳滤波器的输入 $x(n) = s(n) + v(n)$, 期望响应 $d(n) = s(n)$ 。故

$$r_x(k) = r_s(k) + r_v(k) = \frac{10}{27} \left(\frac{1}{2}\right)^{|k|} + \frac{2}{3} \delta(n)$$

$$r_{xd}(k) = E\{x(n)d(n-k)\} = r_s(k) = \frac{10}{27}\left(\frac{1}{2}\right)^{|k|}$$

由于只需要 2 阶滤波器设计, 因此

$$\mathbf{R} = \begin{bmatrix} \frac{10}{27} + \frac{2}{3} & \frac{10}{27} \times \frac{1}{2} \\ \frac{10}{27} \times \frac{1}{2} & \frac{10}{27} + \frac{2}{3} \end{bmatrix}, \quad \mathbf{r}_{xd} = \begin{bmatrix} \frac{10}{27} \\ \frac{10}{27} \times \frac{1}{2} \end{bmatrix}$$

代入维纳-霍普夫方程解, 得

$$\mathbf{w}_0 = \mathbf{R}^{-1} \mathbf{r}_{xd} = [0.3359, 0.1186]^T$$

$$J_{\min} = \sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{w}_0 = \frac{10}{27} - \begin{bmatrix} \frac{10}{27} \\ \frac{10}{27} \times \frac{1}{2} \end{bmatrix}^T \begin{bmatrix} 0.3359 \\ 0.1186 \end{bmatrix} = 0.2240$$

例 3.1.2 设期望响应 $d(n)$ 是一个 AR(1) 过程, 参数 $a(1) = 0.8458$, 激励白噪声 $v_1(n)$ 的方差 $\sigma_1^2 = 0.27$, 由白噪声驱动的该过程的系统函数为

$$H_1(z) = \frac{1}{1 + 0.8458z^{-1}}$$

$d(n)$ 经过了一个通信信道, 信道的系统函数为

$H_2(z)$, 在信道输出端加入了白噪声 $\sigma_2^2 = 0.1$, 通道模型如图 3.1.5 所示, 系统函数

$$H_2(z) = \frac{1}{1 - 0.9458z^{-1}}$$

信道输出 $x(n) = s(n) + v_2(n)$ 。

在接收端设计一个二系数 FIR 结构的维纳滤波器, 目的是由 $x[n]$ 恢复 $d[n]$, 并写出等高线方程^[117]。

解: $d(n)$ 是一个 AR(1) 过程, $A_1(z) = 1 + a_1 z^{-1}$, $\sigma_1^2 = 0.27$, 如例 1.6.3 所求, 期望响应的方差为

$$\sigma_d^2 = \frac{\sigma_1^2}{1 - a_1^2} = \frac{0.27}{1 - (0.8458)^2} = 0.9486$$

在 $x(n) = s(n) + v_2(n)$ 中, $s(n)$ 是一个 2 阶 AR(2) 过程, 由白噪声产生 $s(n)$ 的系统函数相当于 $H(z) = H_1(z)H_2(z)$, 因此

$$A(z) = (1 + 0.8458z^{-1})(1 - 0.9458z^{-1}) = 1 - 0.1z^{-1} - 0.8z^{-2}$$

2 阶 AR(2) 过程的参数为 $a_{2,1} = -0.1$, $a_{2,2} = -0.8$, 方差仍为 $\sigma_1^2 = 0.27$ 。由 2 阶 AR(2) 参数, 可以确定 $r_s(k)$, 由 Yule-Walker 方程

$$\begin{bmatrix} r_s(0) & r_s(1) \\ r_s(1) & r_s(0) \end{bmatrix} \begin{bmatrix} a_{2,1} \\ a_{2,2} \end{bmatrix} = - \begin{bmatrix} r_s(1) \\ r_s(2) \end{bmatrix}$$

$$\sigma_1^2 = r_s(0) + a_{2,1}r_s(1) + a_{2,2}r_s(2)$$

反解 $r_s(0)$ 、 $r_s(1)$, 得

$$r_s(0) = \left(\frac{1 + a_{2,2}}{1 - a_{2,2}} \right) \frac{\sigma_1^2}{[(1 + a_{2,2})^2 - a_{2,1}^2]} = \frac{1 - 0.8}{1 + 0.8} \frac{0.27}{[(1 - 0.8)^2 - 0.1^2]} = 1$$

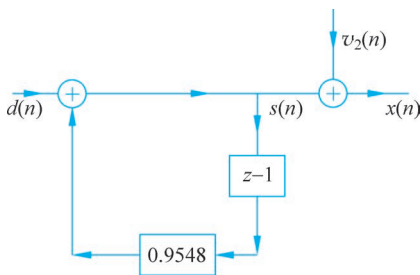


图 3.1.5 通道模型

$$r_s(1) = \frac{-a_{2,1}r_s(0)}{1+a_{2,2}} = \frac{0.1 \times 1}{1-0.8} = 0.5$$

由上确定 $s(n)$ 的自相关矩阵为 $\mathbf{R}_s = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$, 进而有

$$\mathbf{R}_x = \mathbf{R}_s + \sigma_v^2 \mathbf{I} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times 0.1 = \begin{bmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{bmatrix}$$

接下来, 求 $r_{xd}(-k)$:

$$r_{xd}(-k) = E[x(n-k)d(n)]$$

把 $s(n) - 0.9458s(n-1) = d(n)$ 和 $x(n) = s(n) + v_2(n)$ 代入上式, 得

$$r_{xd}(-k) = r_s(k) - 0.9458r_s(k-1)$$

故

$$r_{xd}(0) = r_s(0) - 0.9458 \times r_s(-1) = 0.5272$$

$$r_{xd}(-1) = r_s(1) - 0.9458 \times r_s(0) = -0.4458$$

$$\mathbf{r}_{xd} = \begin{bmatrix} 0.5272 \\ -0.4458 \end{bmatrix}$$

代入维纳-霍普夫方程, 解得最优权系数为

$$\mathbf{w}_o = \mathbf{R}_x^{-1} \mathbf{r}_{xd} = \begin{bmatrix} 0.8368 \\ -0.7853 \end{bmatrix}$$

最小均方误差为

$$J_{\min} = \sigma_d^2 - \mathbf{r}_{xd}^H \mathbf{w}_o = 0.9486 - [0.5272, -0.4458] \begin{bmatrix} 0.8368 \\ -0.7853 \end{bmatrix} = 0.1579$$

对于任意取的权系数, 性能表面为

$$\begin{aligned} J(\mathbf{w}_0, \mathbf{w}_1) &= \sigma_d^2 - \mathbf{w}^H \mathbf{r}_{xd} - \mathbf{r}_{xd}^H \mathbf{w} + \mathbf{w}^H \mathbf{R}_x \mathbf{w} = \sigma_d^2 - 2\mathbf{w}^T \mathbf{r}_{xd} + \mathbf{w}^T \mathbf{R}_x \mathbf{w} \\ &= 0.9486 - 2[0.5272, -0.4458] \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} + [w_0, w_1] \begin{bmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \\ &= 0.9486 - 1.0544w_0 + 0.8916w_1 + w_0w_1 + 1.1(w_0^2 + w_1^2) \end{aligned}$$

为求规范化性能表面, 解 $|R - \lambda I| = 0$, 即

$$\begin{vmatrix} 1.1 - \lambda & 0.5 \\ 0.5 & 1.1 - \lambda \end{vmatrix} = 0 \Rightarrow (1.1 - \lambda)^2 - (0.5)^2 = 0$$

得到两个特征值分别为 $\lambda_1 = 1.6, \lambda_2 = 0.6$, 规范化性能表面为

$$J(v_1, v_2) = J_{\min} + 1.6v_1^2 + 0.6v_2^2$$

对于给定的一个均方误差值, 性能表面的等高线由如下椭圆方程描述:

$$1 = \frac{v_1^2}{\frac{(J - J_{\min})}{\lambda_1}} + \frac{v_2^2}{\frac{(J - J_{\min})}{\lambda_2}}$$

这是一个椭圆, 主轴 $\left(\frac{J - J_{\min}}{\lambda_2}\right)^{\frac{1}{2}}$, 副轴 $\left(\frac{J - J_{\min}}{\lambda_1}\right)^{\frac{1}{2}}$ 。

* 3.1.5 IIR 维纳滤波器

考虑维纳-霍普夫方程在 IIR 滤波器时的情况, 式(3.1.40)是针对非因果 IIR 滤波器的

一般方程,可以直接转换成 z 变换关系式。为了简便,先讨论无因果性要求的最一般的 IIR 滤波器的设计式。对 IIR 情况,我们仅考虑实信号和实滤波器系数的情况。

在非因果条件下,式(3.1.40)重写为

$$\sum_{i=-\infty}^{+\infty} w_{oi} r_x(i-k) = r_{xd}(-k) \quad k = -\infty, \dots, -1, 0, 1, \dots, \infty \quad (3.1.65)$$

注意到,式(3.1.65)左侧是两个序列 $w(k)=w_{ok}$ 和 $r_x(k)$ 的卷积表达式,两边取 z 变换,并由 $r_{dx}(k)=r_{xd}(-k)$,得

$$W(z)S_x(z) = S_{dx}(z) \quad (3.1.66)$$

或

$$W(z) = \frac{S_{dx}(z)}{S_x(z)} \quad (3.1.67)$$

这里,系统函数 $W(z)$ 是滤波器权系数序列 $w(k)=w_{ok}$ 的 z 变换, $S_x(z)$ 是 $r_x(k)$ 的 z 变换, $S_{dx}(z)$ 是 $r_{dx}(k)$ 的 z 变换。

最小均方误差为

$$J_{\min} = \sigma_d^2 - \sum_{l=-\infty}^{+\infty} w_{ol} r_{dx}(l) \quad (3.1.68)$$

例 3.1.3 有一信号 $s(n)$,它的自相关序列为 $r_s(k) = \frac{10}{27} \left(\frac{1}{2}\right)^{|k|}$,被一加法性白噪声所污染,噪声方差为 $2/3$,噪声和信号不相关,被污染信号 $x(n)$ 作为维纳滤波器的输入,求 IIR 滤波器使输出信号是 $s(n)$ 的尽可能的恢复。

解: 本题中取 $x(n)=s(n)+v(n)$, $d(n)=s(n)$ 。显然

$$r_x(k) = r_s(k) + r_v(k) = \frac{10}{27} \left(\frac{1}{2}\right)^{|k|} + \frac{2}{3} \delta(n)$$

$$r_{dx}(k) = E\{x(n-k)d(n)\} = r_s(k) = \frac{10}{27} \left(\frac{1}{2}\right)^{|k|}$$

如上两式直接取 z 变换,得

$$S_x(z) = \frac{5/18}{\left(1 - \frac{1}{2}z^{-1}\right)\left(1 - \frac{1}{2}z\right)} + \frac{2}{3} = \frac{20 - 6z - 6z^{-1}}{18\left(1 - \frac{1}{2}z^{-1}\right)\left(1 - \frac{1}{2}z\right)}$$

$$S_{dx}(z) = \frac{5/18}{\left(1 - \frac{1}{2}z^{-1}\right)\left(1 - \frac{1}{2}z\right)}$$

将其代入式(3.1.67),得

$$W(z) = \frac{S_{dx}(z)}{S_x(z)} = \frac{5}{20 - 6z - 6z^{-1}} = \frac{5}{18\left(1 - \frac{1}{3}z^{-1}\right)\left(1 - \frac{1}{3}z\right)}$$

求 $W(z)$ 的反 z 变换,得到

$$w_{ok} = \frac{5}{16} \left(\frac{1}{3}\right)^{|k|}$$

最小均方误差

$$J_{\min} = \sigma_d^2 - \sum_{l=-\infty}^{+\infty} w_{ol} r_{dx}(l) = \frac{10}{27} - \sum_{l=-\infty}^{+\infty} \frac{5}{16} \left(\frac{1}{3}\right)^{|l|} \left(\frac{10}{27}\right) \left(\frac{1}{2}\right)^{|l|} = \frac{5}{24} \approx 0.2083$$

从滤波器输入的噪声功率为 $2/3$, 下降到滤波器输出的误差功率为 $5/24$ 。

现在考虑因果 IIR 维纳滤波器设计。它的推导比非因果情况复杂一些, 这里先给出结果。因果 IIR 维纳滤波器的系统函数为

$$W(z) = \frac{1}{S_x^+(z)} \left[\frac{S_{dx}(z)}{S_x^-(z)} \right]_+ \quad (3.1.69)$$

上式中, $S_x^+(z)$ 是由 $S_x(z)$ 中位于单位圆内的极点和零点组成; $S_x^-(z)$ 是由 $S_x(z)$ 中位于单位圆外的极点和零点组成; $\left[\frac{S_{dx}(z)}{S_x^-(z)} \right]_+$ 是对应于 $\frac{S_{dx}(z)}{S_x^-(z)}$ 的反变换中的因果序列部分的 z 变换。最小均方误差为

$$J_{\min} = \sigma_d^2 - \sum_{l=0}^{+\infty} w_{ol} r_{dx}(l) \quad (3.1.70)$$

下面通过一个白化滤波器导出因果 IIR 维纳滤波器的结果。一个规则随机过程可以通过一个白化滤波器产生其新息表示(白噪声), 因此, 将因果 IIR 维纳滤波器分成两步, 第一步将输入信号白化, 将新息作为维纳滤波器的输入对期望响应进行估计, 如图 3.1.6 所示。



图 3.1.6 因果 IIR 维纳滤波器的分级形式

在 1.6.1 节的谱分解定理指出, 一个实规则随机信号的功率谱可以分解为

$$S(z) = \sigma^2 Q(z) Q(1/z) \quad (3.1.71)$$

通过一个系统函数为 $\Gamma(z) = 1/Q(z)$ 的白化滤波器, 输出其新息序列, 这里为了表示方便, 定义

$$\frac{1}{S_x^+(z)} = \frac{1}{\sqrt{\sigma^2} Q(z)} \quad (3.1.72)$$

为白化滤波器的系统函数, 由此产生的新息序列是方差为 1 的白噪声。

先求新息作为输入的因果 IIR 维纳滤波器的单位抽样响应 $h_2(n)$, 因为

$$e(n) = d(n) - y(n) = d(n) - \sum_{k=0}^{+\infty} h_2(k) v(n-k) \quad (3.1.73)$$

利用正交原理

$$E[v(n-l)e(n)] = E[v(n-l)d(n)] - E\left[v(n-l) \sum_{k=0}^{+\infty} h_2(k) v(n-k)\right] = 0, \quad l = 0, 1, \dots, \infty \quad (3.1.74)$$

由于 $v(n)$ 是方差为 1 的白噪声, 上式整理为

$$r_{dv}(l) = \sum_{k=0}^{+\infty} h_2(k) \delta(k-l) = h_2(l), \quad l \geq 0 \quad (3.1.75)$$

由于 $v(n)$ 是中间量, 需要将 $r_{dv}(l)$ 进一步表示为

$$\begin{aligned} r_{dv}(l) &= E[d(n)v(n-l)] = E\left[\sum_{k=0}^{+\infty} \gamma(k)x(n-l-k)d(n)\right] \\ &= \sum_{k=0}^{+\infty} \gamma(k)r_{dx}(l+k) = \sum_{m=0}^{-\infty} \gamma(-m)r_{dx}(l-m) = \gamma(-l) * r_{dx}(l) \end{aligned} \quad (3.1.76)$$

上式中, $\gamma(k)$ 是白化滤波器的单位抽样响应, 上式两侧取 z 变换为

$$S_{dv}(z) = \frac{1}{\sqrt{\sigma^2} Q(1/z)} S_{dx}(z) = \frac{S_{dx}(z)}{S_x^-(z)} \quad (3.1.77)$$

注意到, $h_2(l)$ 只取 $r_{dv}(l)$ 的因果部分, 因此

$$H_2(z) = [S_{dv}(z)]_+ = \left[\frac{1}{\sqrt{\sigma^2} Q(1/z)} S_{dx}(z) \right]_+ = \left[\frac{S_{dx}(z)}{S_x^-(z)} \right]_+$$

最终的因果 IIR 维纳滤波器是 $H_2(z)$ 与白化滤波器的级联, 因此因果 IIR 维纳滤波器的系统函数为式(3.1.69)。

例 3.1.4 用因果 IIR 滤波器实现例 3.1.3 的相同问题。

解: 由

$$S_x(z) = \frac{20 - 6z - 6z^{-1}}{18\left(1 - \frac{1}{2}z^{-1}\right)\left(1 - \frac{1}{2}z\right)} = \frac{\left(1 - \frac{1}{3}z^{-1}\right)\left(1 - \frac{1}{3}z\right)}{\left(1 - \frac{1}{2}z^{-1}\right)\left(1 - \frac{1}{2}z\right)}$$

得到

$$S_x^+(z) = \frac{\left(1 - \frac{1}{3}z^{-1}\right)}{\left(1 - \frac{1}{2}z^{-1}\right)}, \quad S_x^-(z) = \frac{\left(1 - \frac{1}{3}z\right)}{\left(1 - \frac{1}{2}z\right)}$$

另外, 令

$$\Gamma(z) = \frac{S_{dx}(z)}{S_x^-(z)} = \frac{5}{18\left(1 - \frac{1}{2}z^{-1}\right)\left(1 - \frac{1}{3}z\right)} = \frac{\frac{1}{6}z^{-1}}{\left(1 - \frac{1}{2}z^{-1}\right)} + \frac{1/3}{\left(1 - \frac{1}{3}z\right)}$$

由反变换得

$$\gamma(n) = \frac{1}{6}\left(\frac{1}{2}\right)^{n-1} u(n-1) + \frac{1}{3}\left(\frac{1}{3}\right)^{-n} u(-n)$$

上式中的 $u(n)$ 代表阶跃序列。取 $\gamma(n)$ 的因果部分为 $\gamma_+(n) = \frac{1}{3}\left(\frac{1}{2}\right)^n u(n)$, 因此

$$[\Gamma(z)]_+ = \left[\frac{S_{dx}(z)}{S_x^-(z)} \right]_+ = \frac{1/3}{1 - \frac{1}{2}z^{-1}}$$

将上式代入式(3.1.69), 得

$$W(z) = \frac{1}{S_x^+(z)} \left[\frac{S_{dx}(z)}{S_x^-(z)} \right]_+ = \frac{\left(1 - \frac{1}{2}z^{-1}\right)}{\left(1 - \frac{1}{3}z^{-1}\right)} \frac{1/3}{1 - \frac{1}{2}z^{-1}} = \frac{1/3}{1 - \frac{1}{3}z^{-1}}$$

取反 z 变换得到滤波器单位抽样响应为

$$w_{ok} = \frac{1}{3}\left(\frac{1}{3}\right)^k, \quad k \geq 0$$

$$J_{\min} = \sigma_d^2 - \sum_{l=0}^{+\infty} w_{ol} r_{dx}[l] = \frac{10}{27} - \sum_{l=-\infty}^{+\infty} \frac{1}{3}\left(\frac{1}{3}\right)^{|l|} \left(\frac{10}{27}\right)\left(\frac{1}{2}\right)^{|l|} = \frac{2}{9} \approx 0.2222$$

因果的 IIR 维纳滤波器比非因果的剩余误差要略大。

注意到,例 3.1.3、例 3.1.4 和例 3.1.1 是同一个问题分别用非因果 IIR、因果 IIR 和 2 系数 FIR 维纳滤波器进行处理,得到输出最小均方误差分别为 0.2083、0.2222 和 0.2240。虽然非因果 IIR 的误差最小,但是不可实现的,可实现的因果 IIR 和 2 系数 FIR 的误差很接近。在这个例子中,由于输入信号的极点远离单位圆,故用 2 系数的 FIR 滤波器即可取得接近于 IIR 滤波器的性能,若输入信号更复杂可能需要的 FIR 系数更多,但该例子给出一个直观说明,对于一个给定问题,选择适当阶数的 FIR 滤波器可能得到与因果 IIR 滤波器非常接近的性能。由于 FIR 滤波器不存在数值稳定性问题,容易实现和集成,所以实际中更易使用。

* 3.1.6 应用实例——通信系统的最佳线性均衡器

在通信系统中,发射机发送的信号通过信道传输到接收机,传输信道有不同的媒体,主要分成有线和无线两类。以无线为例,发射信号通过无线信道传播,经过多径到达接收机而被接收,这些多径是由于信道中的信号经反射、折射和衍射形成的。多径传播产生的接收信号是由多个延迟衰落的发射信号叠加而成的。因此接收机接收到的信号中存在着串扰和畸变,直接进行检测会产生较大的误码。

一种改善信号检测性能的装置是信道均衡器,它的目标是补偿信道造成的串扰和畸变,为了叙述简单,我们采用图 3.1.7 的基带模型讨论均衡器的设计问题。

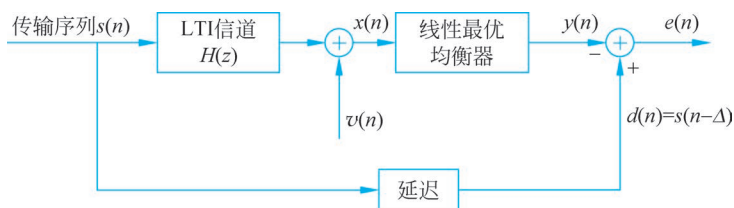


图 3.1.7 通信系统线性均衡器示意图

设发射机发送的信号序列是 $s(n)$, 用一个线性时不变(LTI)滤波器模拟一个信道的传输性质(对于慢时变系统,这个 LTI 模型是合理的), 假定接收机和发射机用相同的采样频率, 接收到的信号表示为

$$x(n) = \sum_k h(k)s(n-k) + v(n)$$

其中, $v(n)$ 表示加性干扰白噪声, $h(n)$ 表示信道的单位抽样响应, 为了降低加性噪声和信道传输的影响, 设计一个线性滤波器作为均衡器, 均衡器的单位抽样响应记为 $w(n)$, 均衡器输出为

$$y(n) = \sum_k w(k)x(n-k)$$

线性均衡器的设计目标是使 $y(n)$ 是 $s(n)$ 的尽可能精确的估计, 即

$$y(n) = \sum_k w(k)x(n-k) \approx s(n-\Delta)$$

其中, Δ 表示一个整数延迟量。一种设计线性最优均衡器的准则是最小均方准则, 这实际是设计一个维纳滤波器, 滤波器的输入为 $x(n)$, 期望响应为 $d(n)=s(n-\Delta)$ 。既可以设计 FIR 结构的维纳滤波器, 也可以设计 IIR 结构的维纳滤波器, 并且只需要 $x(n)$ 的自相关和 $x(n)$ 与 $s(n)$ 的互相关函数。如下以非因果 IIR 结构为例再进一步做说明, 非因果 IIR 结构

可以取得理论上最好的结果。非因果 IIR 型维纳滤波器的系统函数为

$$W(z) = \frac{S_{dx}(z)}{S_x(z)}$$

容易求得

$$\begin{aligned} S_{dx}(z) &= H\left(\frac{1}{z}\right) S_s(z) z^{-\Delta} \\ S_x(z) &= H(z) H\left(\frac{1}{z}\right) S_s(z) + \sigma_v^2 \end{aligned}$$

因此,系统函数进一步写为

$$W(z) = \frac{H\left(\frac{1}{z}\right) S_s(z) z^{-\Delta}}{H(z) H\left(\frac{1}{z}\right) S_s(z) + \sigma_v^2} \quad (3.1.78)$$

频率响应为

$$W(e^{j\omega}) = \frac{H(e^{-j\omega}) S_s(\omega) e^{-j\omega\Delta}}{|H(e^{j\omega})|^2 S_s(\omega) + \sigma_v^2} \quad (3.1.79)$$

由以上两式可以看到,一般情况下,维纳滤波器的频率响应与发射信号的功率谱和信道的频率响应有关。讨论两个特例。

(1) 加性噪声不存在的情况,即 $\sigma_v^2=0$,线性最优均衡器的系统函数简化为

$$W(z) = \frac{z^{-\Delta}}{H(z)} \quad (3.1.80)$$

此时,最优线性均衡器是信道模型的逆系统。 $y(n)=s(n-\Delta)$ 达到理想均衡。

(2) 发射信号是方差为 1 的白噪声,均衡器的系统函数简化为

$$W(z) = \frac{H\left(\frac{1}{z}\right) z^{-\Delta}}{H(z) H\left(\frac{1}{z}\right) + \sigma_v^2} \quad (3.1.81)$$

此时,得到的均衡器是所有线性系统中,使 $y(n)$ 与 $s(n-\Delta)$ 均方误差最小的系统,但达不到理想均衡条件 $y(n)=s(n-\Delta)$ 。

在实际系统中,信道性质可能是时变的,可以通过自适应滤波技术达到最优线性均衡和跟踪信道的时变性,有关用自适应滤波算法实现自适应均衡器的讨论见第 5 章。

3.2 最优线性预测

最优线性预测是维纳滤波器的一种特殊类型,将特殊的输入向量和期望响应代入维纳-霍普夫方程,得到线性最优预测的解。线性最优预测分为前向预测和后向预测两种情况,将会发现解前向线性预测系数的方程与解 AR 模型系数的方程是完全等价的。通过讨论前向和后向线性预测滤波器结构,导出快速算法: Levinson-Durbin 递推算法,用于快速求解线性预测或 AR 模型问题。通过 Levinson-Durbin 递推算法也导出了一种重要的滤波器结构,即格型滤波器。本节如不加特别说明,均指一步预测。



视频讲解

3.2.1 前向线性预测

由 $\{x(n-i)\}_{i=1}^M$ 张成的空间记为 X_{n-1} , 在 X_{n-1} 中预测 $x(n)$ 的值, 这是一步前向预测问题。在所有可能的线性预测器中, 存在一个使预测误差的均方值最小的预测器, 称为最优前向线性预测, 设最优预测器的系数为 $w_{f,1}, w_{f,2}, \dots, w_{f,M}$ 。记最优预测为

$$\hat{x}(n | X_{n-1}) = \sum_{k=1}^M w_{f,k}^* x(n-k) \quad (3.2.1)$$

显然, 最优前向线性预测相当于一个维纳滤波器, 对应于维纳滤波的术语, 期望响应为

$$d(n) = x(n) \quad (3.2.2)$$

预测误差为

$$f_M(n) = x(n) - \hat{x}(n | X_{n-1}) \quad (3.2.3)$$

最小预测误差功率为

$$p_M = E[|f_M(n)|^2] \quad (3.2.4)$$

为使用维纳滤波器的结论, 各变量和参数分别描述如下, 输入向量为

$$\mathbf{x}(n-1) = [x(n-1), x(n-2), \dots, x(n-M)]^T \quad (3.2.5)$$

自相关矩阵为

$$\mathbf{R}_x = E[\mathbf{x}(n-1)\mathbf{x}^H(n-1)] = E[\mathbf{x}(n)\mathbf{x}^H(n)] \quad (3.2.6)$$

式(3.2.6)的第二个等号成立是假定了输入信号是平稳的。输入信号与期望响应的互相关向量为

$$\mathbf{r}_{xd} = E[\mathbf{x}(n-1)x^*(n)] = \begin{bmatrix} r_x(-1) \\ r_x(-2) \\ \vdots \\ r_x(-M) \end{bmatrix} = \mathbf{r} \quad (3.2.7)$$

相应的求解前向最优线性预测系数的维纳-霍普夫方程为

$$\mathbf{R}_x \mathbf{w}_f = \mathbf{r} \quad (3.2.8)$$

这里, $\mathbf{w}_f = [w_{f,1}, w_{f,2}, \dots, w_{f,M}]^T$, 注意到, 这个方程与求解 AR 模型系数的 Yule-Walker 方程是一致的, 只是系数相差一个负号。

当采用最优预测系数时, 预测误差均方值达到最小, 即最小预测误差功率为

$$p_M = r_x(0) - \mathbf{r}^H \mathbf{w}_f \quad (3.2.9)$$

前向预测和预测误差输出可用图 3.2.1 表示。

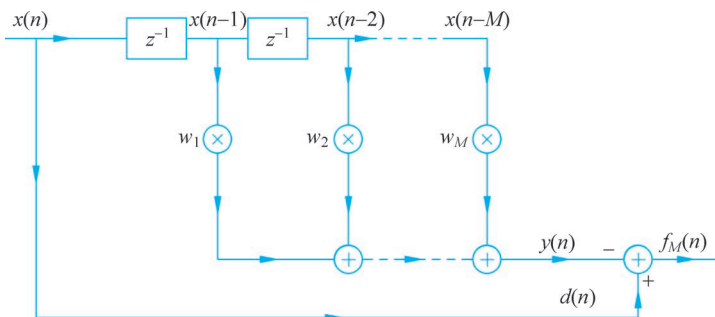


图 3.2.1 前向线性预测器及预测误差输出

1. 前向线性预测误差滤波器

如果将预测误差作为一个滤波器的输出,由输入 $x(n)$ 到输出 $f_M(n)$ 构成前向线性预测误差滤波器,相应的滤波器输入输出关系表示为

$$\begin{aligned} f_M(n) &= x(n) - \sum_{k=1}^M w_{f,k}^* x(n-k) \\ &= \sum_{k=0}^M a_{M,k}^* x(n-k) = \mathbf{a}_M^H \mathbf{x}_{M+1}(n) \end{aligned} \quad (3.2.10)$$

这里,相应的前向线性预测误差滤波器系数 $a_{M,k}$ 与最优预测器系数 $w_{f,k}$ 的关系为

$$a_{M,k} = \begin{cases} 1, & k=0 \\ -w_{f,k}, & k=1, 2, \dots, M \end{cases} \quad (3.2.11)$$

并使用向量符号为

$$\mathbf{a}_M = [1, a_{M,1}, a_{M,2}, \dots, a_{M,M}]^T \quad (3.2.12)$$

$$\mathbf{x}_{M+1}(n) = [x(n), x(n-1), \dots, x(n-M+1), x(n-M)]^T$$

2. 增广维纳-霍普夫方程

将前向预测的维纳-霍普夫方程式(3.2.8)和预测误差功率 p_M 公式(3.2.9)写在一起,有

$$\begin{bmatrix} r_x(0) & \mathbf{r}^H \\ \mathbf{r} & \mathbf{R}_x \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{w}_f \end{bmatrix} = \begin{bmatrix} p_M \\ \mathbf{0} \end{bmatrix} \quad (3.2.13)$$

或等价

$$\mathbf{R}_{M+1} \mathbf{a}_M = \begin{bmatrix} p_M \\ \mathbf{0} \end{bmatrix} \quad (3.2.14)$$

式(3.2.13)或式(3.2.14)是线性预测的增广维纳-霍普夫方程,从增广方程看,前向线性预测误差滤波器和 AR 模型所服从的方程组也是一致的。

3. 前向线性预测误差滤波器与 AR 模型的关系

前文已经提到,在已知一组自相关序列值 $\{r_x(k), k=0, \dots, M\}$ 的条件下,前向线性预测误差滤波器和 AR 模型的求解方程是一致的。为了便于比较,将 AR(M)模型的差分方程表达式写为如下形式:

$$\sum_{k=0}^M a_k^* x(n-k) = v(n) \quad (3.2.15)$$

比较式(3.2.10)和式(3.2.15)发现,AR 模型与前向线性预测误差滤波器满足相同的方程(但等号左右两侧内容是相互调换的),只需要进行如下的对应关系:

$$a_k^* \leftrightarrow a_{M,k}^* \quad (3.2.16)$$

$$v(n) \leftrightarrow f_M(n) \quad (3.2.17)$$

注意,这种对应只是数学形式的对应,从物理意义上,前向线性预测误差是分析器,即 $x(n)$ 是输入信号,通过滤波器产生输出 $f_M(n)$; AR 模型是合成器, $v(n)$ 是激励信号,通过系统产生合成的信号 $x(n)$ 。两者分别用于解决不同的问题,但遵从相同的数学关系。因此,在已知一组自相关序列值 $\{r_x(k), k=0, 1, \dots, M\}$ 的条件下,两者的求解过程和所求的解相同,即两者的解等价。

3.2.2 后向线性预测

初看起来,后向预测似乎没有实质用处,但其可与前向预测构成互补结构,从而得到有效的系统结构。由 $\{x(n-i)\}_{i=0}^{M-1}$ 构成的空间 X_n 预测 $x(n-M)$,称为一步后向预测,简称后向预测。最优后向线性预测器系数记为 $w_{b,1}, w_{b,2}, \dots, w_{b,M}$,预测值记为

$$\hat{x}(n-M | X_n) = \sum_{k=1}^M w_{b,k}^* x(n-k+1) \quad (3.2.18)$$

与前向预测类似,后向最优线性预测器也是维纳滤波器的特殊类型,对应于维纳滤波器的术语,分别有期望响应为

$$d(n) = x(n-M) \quad (3.2.19)$$

预测误差为

$$b_M(n) = x(n-M) - \hat{x}(n-M | X_n) \quad (3.2.20)$$

预测误差功率为

$$p_M = E[|b_M(n)|^2] \quad (3.2.21)$$

与标准维纳滤波器相比,后向预测器的有关向量分别写为,输入向量

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T$$

自相关矩阵为

$$\mathbf{R}_x = E[\mathbf{x}(n)\mathbf{x}^H(n)]$$

互相关向量为

$$\mathbf{r}_{xd} = E[\mathbf{x}(n)x^*(n-M)] = [r(M), r(M-1), \dots, r(1)]^T = \mathbf{r}^{B*}$$

这里上标符号 B 指向量反向排序。

求后向预测器系数的维纳-霍普夫方程为

$$\mathbf{R}_x \mathbf{w}_b = \mathbf{r}^{B*} \quad (3.2.22)$$

这里, $\mathbf{w}_b = [w_{b,1}, w_{b,2}, \dots, w_{b,M}]^T$ 。预测误差功率为

$$p_M = r(0) - \mathbf{r}^{BT} \mathbf{w}_b \quad (3.2.23)$$

可以证明(留作习题),前向和后向最优预测系数之间满足如下关系:

$$\mathbf{w}_b^{B*} = \mathbf{w}_f \quad (3.2.24)$$

或写成如下分量形式:

$$w_{b,M-k+1}^* = w_{f,k} \quad k=1,2,\dots,M \quad (3.2.25)$$

$$w_{b,k} = w_{f,M-k+1}^* \quad k=1,2,\dots,M \quad (3.2.26)$$

由此也可证明前向最优线性预测误差功率与后向最优线性预测误差功率也相等,因此都用了同一个符号 p_M 表示。

1. 后向预测误差滤波器

与前向预测类似,定义后向预测误差滤波器

$$\begin{aligned} b_M(n) &= x(n-M) - \sum_{k=1}^M w_{b,k}^* x(n-k+1) \\ &= \sum_{k=0}^M C_{M,k}^* x(n-k) \end{aligned} \quad (3.3.27)$$

这里

$$C_{M,k}^* = \begin{cases} -w_{b,k+1}^*, & k = 0, 1, \dots, M-1 \\ 1, & k = M \end{cases}$$

容易验证

$$C_{M,k} = a_{M,M-k}^* \quad k = 0, 1, \dots, M \quad (3.2.28)$$

故

$$b_M(n) = \sum_{k=0}^M a_{M,M-k} x(n-k) = \mathbf{a}_M^{BT} \mathbf{x}_{M+1}(n) \quad (3.2.29)$$

2. 增广维纳-霍普夫方程

后向预测问题也可以写成一个增广方程：

$$\begin{bmatrix} \mathbf{R}_x & \mathbf{r}^{B*} \\ \mathbf{r}^{BT} & r_x(0) \end{bmatrix} \begin{bmatrix} -\mathbf{w}_b \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ p_M \end{bmatrix} \quad (3.2.30)$$

或

$$\mathbf{R}_{M+1} \mathbf{a}_M^{B*} = \begin{bmatrix} \mathbf{0} \\ p_M \end{bmatrix} \quad (3.2.31)$$

3.2.3 Levinson-Durbin 算法

在实际应用中,不管是解 AR 模型问题或解线性最优预测问题,都需要求解增广的 Yule-Walker 方程(或线性预测的增广维纳-霍普夫方程),但实际中,往往并不知道模型或预测的最好阶数。一些学者研究了阶数的确定准则;另一种方法就是对一个范围内的若干阶数分别进行求解,当预测误差随阶数增加而明显减少时,就可以得到相应的模型阶。分别求解若干阶的增广 Yule-Walker 方程,是比较费时的。用高斯消元法解一个 M 阶方程需要 $O(M^3)$ 的运算量,随着 M 增加,运算量增加很快。一种好的方法是:如果已知 $M-1$ 阶的系数和最小均方误差,通过一个简单的递推关系得到 M 阶的系数与均方误差值,这就是 Levinson-Durbin 递推算法。

Levinson-Durbin 递推算法实际是利用相关矩阵 $\mathbf{R}$ 的 Toeplitz 性质,由 Toeplitz 性得到了自相关矩阵的增广形式,再利用前向和后向预测的增广 Yule-Walker 方程的结构,导出了由 $m-1$ 阶的解 $\mathbf{a}_{m-1}$ 、 p_{m-1} 递推求解 $\mathbf{a}_m$ 、 p_m 的快速算法。如下推导这个算法。

从 $m-1$ 阶出发,对前向预测有增广方程

$$\mathbf{R}_m \mathbf{a}_{m-1} = \begin{bmatrix} p_{m-1} \\ \mathbf{0}_{m-1} \end{bmatrix} \quad (3.2.32)$$

注意,为了推导过程清晰,自相关矩阵 $\mathbf{R}_m$ 的下标表示 $m \times m$ 增广矩阵。将式(3.2.32)的系数矩阵增加阶数,将系数自相关矩阵增广为 $(m+1) \times (m+1)$ 矩阵,为保持方程成立,系数向量尾部增加一个 0,得到

$$\begin{bmatrix} \mathbf{R}_m & \mathbf{r}_m^{B*} \\ \mathbf{r}_m^{BT} & r_x(0) \end{bmatrix} \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} = \begin{bmatrix} p_{m-1} \\ \mathbf{0}_{m-1} \\ \mathbf{r}_m^{BT} \mathbf{a}_{m-1} \end{bmatrix} \triangleq \begin{bmatrix} p_{m-1} \\ \mathbf{0}_{m-1} \\ \Delta_{m-1} \end{bmatrix} \quad (3.2.33)$$

这里, Δ_{m-1} 只是引入的一个缩写符号,即

$$\Delta_{m-1} = \mathbf{r}_m^{BT} \mathbf{a}_{m-1} = \sum_{\ell=0}^{m-1} r_x(\ell-m) a_{m-1,\ell} \quad (3.2.34)$$

实际上 Δ_{m-1} 有其明确的物理意义, 它的物理意义稍后再讨论。

类似地, 后向 $m-1$ 阶预测的增广方程为

$$\mathbf{R}_m \mathbf{a}_{m-1}^{B*} = \begin{bmatrix} \mathbf{0}_{m-1} \\ p_{m-1} \end{bmatrix} \quad (3.2.35)$$

同样对式(3.2.35)的系数矩阵做 1 阶增广, 得如下方程:

$$\begin{bmatrix} r_x(0) & \mathbf{r}_m^H \\ \mathbf{r}_m & \mathbf{R}_m \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_m^H \mathbf{a}_{m-1}^{B*} \\ \mathbf{0}_{m-1} \\ p_{m-1} \end{bmatrix} = \begin{bmatrix} \Delta_{m-1}^* \\ \mathbf{0}_{m-1} \\ p_{m-1} \end{bmatrix} \quad (3.2.36)$$

观察式(3.2.33)和式(3.2.36)发现, 它们都是以 $\mathbf{R}_{m+1}$ 为系数矩阵的方程, 为了与 m 阶前向线性预测的增广方程进行比较, 构造一个方程式为式(3.2.33)+式(3.2.36) $\times k_m$, 即

$$\mathbf{R}_{m+1} \left\{ \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} + k_m \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} \right\} = \begin{bmatrix} p_{m-1} \\ \mathbf{0}_{m-1} \\ \Delta_{m-1} \end{bmatrix} + k_m \begin{bmatrix} \Delta_{m-1}^* \\ \mathbf{0}_{m-1} \\ p_{m-1} \end{bmatrix} \quad (3.2.37)$$

式(3.2.37)中, k_m 是一个自由参数。现在可以比较式(3.2.37)和 m 阶的前向预测的增广方程, 为了比较方便, 将 m 阶的前向预测的增广方程写在如下:

$$\mathbf{R}_{m+1} \mathbf{a}_m = \begin{bmatrix} p_m \\ \mathbf{0}_m \end{bmatrix} \quad (3.2.38)$$

比较式(3.2.37)和式(3.2.38)发现, 如果能够确定 k_m 的值, 使式(3.2.37)和式(3.2.38)两式等价, 即式(3.2.37)和式(3.2.38)方程两侧的各向量相等, 也就是

$$\mathbf{a}_m = \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} + k_m \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} \quad (3.2.39)$$

和

$$\begin{bmatrix} p_m \\ \mathbf{0}_m \end{bmatrix} = \begin{bmatrix} p_{m-1} \\ \mathbf{0}_{m-1} \\ \Delta_{m-1} \end{bmatrix} + k_m \begin{bmatrix} \Delta_{m-1}^* \\ \mathbf{0}_{m-1} \\ p_{m-1} \end{bmatrix} \quad (3.2.40)$$

是有唯一解的。由于式(3.2.37)和式(3.2.38)的系数都是 $\mathbf{R}_{m+1}$, 如果式(3.2.39)和式(3.2.40)有解, 那么由式(3.2.39)就得到预测滤波系数的递推公式, 式(3.2.39)可以写成标量形式:

$$a_{m,\ell} = a_{m-1,\ell} + k_m a_{m-1,m-\ell}^* \quad \ell = 0, 1, \dots, m \quad (3.2.41)$$

在使用上式时, 应注意到 $a_{m-1,m} = 0$ 和 $a_{m-1,0} = 1$ 。

解式(3.2.40), 确定 k_m 参数和 p_m 的公式, 注意到式(3.2.40)中间 $m-2$ 行恒为零, 因此只有如下两个方程:

$$\begin{cases} p_m = p_{m-1} + k_m \Delta_{m-1}^* \\ 0 = \Delta_{m-1} + k_m p_{m-1} \end{cases}$$

解之, 得到

$$k_m = -\frac{\Delta_{m-1}}{p_{m-1}} \quad (3.2.42)$$

$$p_m = p_{m-1}(1 - |k_m|^2) \quad (3.2.43)$$

可以看到, 只要按式(3.2.42)取 k_m 的值, 就使式(3.2.39)和式(3.2.40)等式成立, 由

式(3.2.41)和式(3.2.43)得到由 $\mathbf{a}_{m-1}$ 、 p_{m-1} 递推求解 $\mathbf{a}_m$ 、 p_m 的公式。

不难验证,与式(3.2.39)对应的另一个形式的递推公式是:

$$\mathbf{a}_m^{B*} = \begin{bmatrix} 0 \\ \mathbf{a}_{m-1}^{B*} \end{bmatrix} + k_m^* \begin{bmatrix} \mathbf{a}_{m-1} \\ 0 \end{bmatrix} \quad (3.2.44)$$

或写成单项的形式:

$$a_{m,m-\ell}^* = a_{m-1,m-\ell}^* + k_m^* a_{m-1,\ell} \quad (3.2.45)$$

为了更全面地理解 Levinson-Durbin 算法,以下对几个相关问题作一些更详细地讨论。

1. Δ_{m-1} 和 k_m 的解释

在上述的推导过程中, Δ_{m-1} 和 k_m 是在推导过程中引入的参数,其中 Δ_{m-1} 由式(3.2.34)定义, k_m 由式(3.2.42)可以求解,其实这两个参数有明确的物理意义,可以证明(留作习题)

$$\Delta_{m-1} = E[b_{m-1}(n-1)f_{m-1}^*(n)] \quad (3.2.46)$$

称为偏相关系数。而

$$\begin{aligned} k_m &= -\frac{E[b_{m-1}(n-1)f_{m-1}^*(n)]}{E[|f_{m-1}(n)|^2]} \\ &= -\frac{E[b_{m-1}(n-1)f_{m-1}^*(n)]}{E[|f_{m-1}(n)|^2]^{\frac{1}{2}}E[|b_{m-1}(n-1)|^2]^{\frac{1}{2}}} \end{aligned} \quad (3.2.47)$$

称为反射系数,由施瓦茨不等式知, $|k_m| \leq 1$ 。

2. Levinson-Durbin 算法小结

目的是由 $m-1$ 阶的解 $\mathbf{a}_{m-1}$ 、 p_{m-1} , 递推求解 $\mathbf{a}_m$ 、 p_m 。需进行如下计算:

- (1) $\Delta_{m-1} = \mathbf{r}_m^{BT} \mathbf{a}_{m-1} = \sum_{\ell=0}^{m-1} r_x(\ell-m)a_{m-1,\ell}$
- (2) $k_m = -\frac{\Delta_{m-1}}{p_{m-1}}$
- (3) $a_{m,\ell} = a_{m-1,\ell} + k_m a_{m-1,m-\ell}^*, \quad \ell=0,1,\dots,m$
- (4) $p_m = p_{m-1}(1-|k_m|^2)$

如果需要初始时从最低阶开始进行递推,需要确定其初始条件。

3. Levinson-Durbin 算法初始化条件

根据预测误差滤波器的定义,可以得到如下一组初始条件:

$$\begin{cases} f_0(n) = b_0(n) = x(n) \\ p_0 = r_x(0) \\ \Delta_0 = E[b_0(n-1)f_0^*(n)] = E[x(n-1)x^*(n)] = r_x^*(1) \\ a_{0,0} = 1 \end{cases} \quad (3.2.48)$$

注意到上式中,因为零阶预测的全部预测系数为零,因此预测误差等于信号本身,由此导出各初始值。

如果用高斯消元法直接求解一个 M 阶的最优预测器系数,所需运算量为 $O(M^3)$,用 Levinson-Durbin 算法为 $O(M^2)$ 。但是 Levinson-Durbin 算法可以得到从 1 阶到 M 阶的各阶系数和最小均方误差值。

由 Levinson-Durbin 算法的 4 个计算公式可以看到,如果给出随机信号自相关函数的

$M+1$ 个值 $\{r_x(0), r_x(1), \dots, r_x(M)\}$, 递推得到各阶反射系数 $\{k_m, m=1, 2, \dots, M\}$ 和各阶预测误差滤波器系数 $\{a_{m,k}, m=1, 2, \dots, M, k=1, 2, \dots, m-1, m\}$, 最终确定了第 M 阶系数, $\{a_{M,k}, k=0, 1, \dots, M\}$ 。进一步注意到, 在每增加 1 阶的递推过程中, 除计算 Δ_{m-1} 外, 其他参数计算不需要自相关函数的值, 如果可以用其他方法获得反射系数, 仅由 M 个反射系数 k_m , 可以递推得到 M 阶预测误差滤波器的系数。

例 3.2.1 写出实信号的用 k_m 表示的 Levinson-Durbin 算法前 3 阶的系数递推公式。

解: 由 $a_{0,0}=1$ 作为初始条件, 各阶系数的递推公式为

$$1 \text{ 阶: } a_{1,1} = k_1$$

$$a_{2,1} = a_{1,1} + k_2 a_{1,1} = k_1 + k_2 k_1$$

$$2 \text{ 阶: } a_{2,2} = k_2$$

$$a_{3,1} = a_{2,1} + k_3 a_{2,2} = k_1 + k_2 k_1 + k_3 k_2$$

$$3 \text{ 阶: } a_{3,2} = a_{2,2} + k_3 a_{2,1} = k_2 + k_3 (k_1 + k_2 k_1)$$

$$a_{3,3} = k_3$$

例 3.2.2 设随机信号是满足 1 阶 AR 模型的, 即满足 $x(n) = -a_1 x(n-1) + v(n)$, 其中 $v(n)$ 是方差为 σ_v^2 的白噪声, 用 Levinson-Durbin 算法求解对 $x(n)$ 的各阶最优线性预测误差滤波器的系数。

解: 由例 1.6.3 已求得 $x(n)$ 的自相关序列为

$$r_x(k) = \frac{\sigma_v^2}{1 - a_1^2} (-a_1)^{|k|}$$

用 Levinson-Durbin 算法可以获得各阶预测误差滤波器的系数。

初始的零阶预测误差滤波器的参数为

$$a_{0,0} = 1, \quad p_0 = r_x(0) = \frac{\sigma_v^2}{1 - a_1^2}, \quad \Delta_0 = r_x(1) = \frac{-a_1 \sigma_v^2}{1 - a_1^2}$$

1 阶参数为

$$k_1 = -\frac{\Delta_0}{p_0} = -\frac{r_x(1)}{r_x(0)} = a_1$$

$$a_{1,0} = 1$$

$$a_{1,1} = k_1 = a_1$$

$$p_1 = p_0 (1 - |k_1|^2) = \sigma_v^2$$

2 阶参数为

$$\Delta_1 = r_x(2) + a_{1,1} r_x(1) = 0$$

$$k_2 = 0$$

$$p_2 = p_1 = \sigma_v^2$$

$$a_{2,0} = 1, \quad a_{2,1} = a_{1,1} = a_1, \quad a_{2,2} = 0$$

类似地, 容易验证 M 阶预测误差滤波器系数为

$$a_{M,0} = 1, \quad a_{M,1} = a_1$$

$$a_{M,k} = 0, \quad k > 1$$

对于 1 阶 AR 模型, 只需要 1 阶线性预测器可以达到使预测误差为白噪声, 且预测误差

功率为 σ_v^2 , 更高阶预测不能改善预测性能, 这个结论可以推广到任意阶 AR 模型, 即对于 AR(p) 模型产生的信号, 只需要 p 阶预测, 更高阶预测不再改善预测性能。

4. 反 Levinson-Durbin 算法

正如上所述, 由 M 个反射系数 $\{k_m, m=1, \dots, M\}$, 可以递推得到 M 阶预测误差滤波器的系数。反射系数 k_m 是一个重要的量, 3.2.4 节将看到, 由反射系数可以构成预测误差滤波器另一种规范的实现结构: 格型结构。这里, 讨论一个反问题, 如果给出了预测误差滤波器的 M 阶系数, $\{a_{M,k}, k=0, 1, \dots, M\}$, 也可以递推得到各阶反射系数 k_m 。

将式(3.2.41)和式(3.2.45)联立, 得到如下联立方程:

$$\begin{cases} a_{m,k} = a_{m-1,k} + k_m a_{m-1,m-k}^* \\ a_{m,m-k}^* = a_{m-1,m-k}^* + k_m a_{m-1,k} \end{cases} \quad (3.2.49)$$

且注意到, 在式(3.2.41)中, 取 $k=m$, 并利用 $a_{m-1,m}=0$ 和 $a_{m-1,0}=1$, 得到

$$k_m = a_{m,m} \quad (3.2.50)$$

将式(3.2.50)代入式(3.2.49), 以 $a_{m-1,k}$ 为未知量, 解得

$$a_{m-1,k} = \frac{a_{m,k} - a_{m,m} a_{m,m-k}^*}{1 - |a_{m,m}|^2}, \quad k=0, 1, \dots, m-1 \quad (3.2.51)$$

由此得到 $k_{m-1} = a_{m-1,m-1}$ 。由 $m=M$ 开始, 连续用式(3.2.50)和式(3.2.51), 每次使 m 减 1, 依次得到 $k_M, k_{M-1}, \dots, k_1$ 。这个过程称为反 Levinson-Durbin 算法, 它是由 M 阶预测误差滤波器系数 $\{a_{M,k}, k=0, 1, \dots, M\}$, 递推得到各阶反射系数 $k_M, k_{M-1}, \dots, k_1$ 。

3.2.4 格型预测误差滤波器

由 m 阶前向预测误差滤波器的输出 $f_m(n) = \mathbf{a}_m^H \mathbf{x}_{m+1}(n)$, 和 m 阶后向预测误差滤波器输出 $b_m(n) = \mathbf{a}_m^{BT} \mathbf{x}_{m+1}(n)$, 利用 Levinson-Durbin 递推公式, 整理得到

$$f_m(n) = f_{m-1}(n) + k_m^* b_{m-1}(n-1) \quad (3.2.52)$$

$$b_m(n) = b_{m-1}(n-1) + k_m f_{m-1}(n) \quad (3.2.53)$$

注意到

$$b_{m-1}(n-1) = Z^{-1}[b_{m-1}(n)] \quad (3.2.54)$$

表示后向预测误差的 1 阶延迟, 并且注意到 0 阶预测误差滤波器输出为

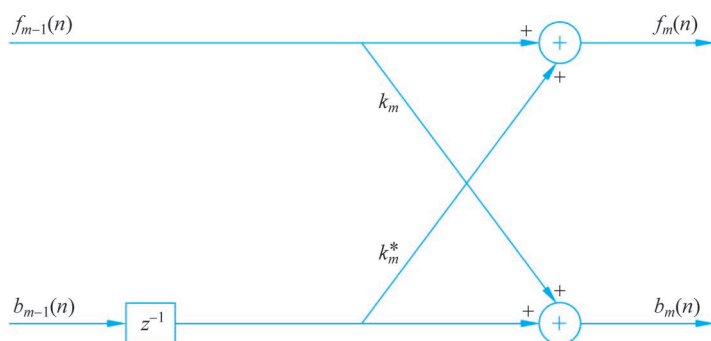
$$f_0(n) = b_0(n) = x(n) \quad (3.2.55)$$

由式(3.2.52)~式(3.2.55)得到格型预测误差滤波器结构如图 3.2.2 所示。图 3.2.2(a) 是 1 阶格型模块, 实现了式(3.2.52)和式(3.2.53)的运算关系, 通过 M 级级联, 同时实现了 M 阶前向预测误差和后向预测误差滤波器, 第 m 级格型单元的反射系数为 k_m 。

如下简单证明前向预测误差滤波器的递推关系式(3.2.52), 后向预测误差滤波器的递推关系式(3.2.53)的证明很相似, 此处从略。由于

$$\begin{aligned} f_m(n) &= \sum_{k=0}^m a_{m,k}^* x(n-k) \\ b_m(n) &= \sum_{k=0}^m a_{m,m-k} x(n-k) \end{aligned}$$

在如上第 1 式代入 Levinson-Durbin 递推公式(3.2.41), 得到



(a) 单级格型模块

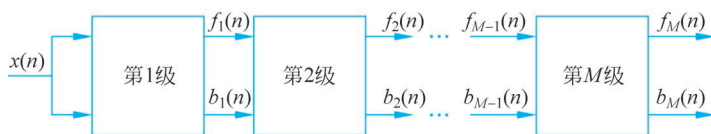
(b) M 阶格型预测误差滤波器

图 3.2.2 预测误差滤波器的格型结构

$$\begin{aligned}
 f_m(n) &= \sum_{k=0}^m a_{m,k}^* x(n-k) = \sum_{k=0}^m (a_{m-1,k} + k_m a_{m-1,m-k}^*) x(n-k) \\
 &= \sum_{k=0}^m (a_{m-1,k})^* x(n-k) + k_m^* \sum_{k=0}^m a_{m-1,m-k} x(n-k)
 \end{aligned}$$

利用 $a_{m-1,m}=0$, 代入上式, 得到

$$\begin{aligned}
 f_m(n) &= \sum_{k=0}^{m-1} a_{m-1,k}^* x(n-k) + k_m^* \sum_{k=1}^m a_{m-1,m-k} x(n-k) \\
 &= \sum_{k=0}^{m-1} a_{m-1,k}^* x(n-k) + k_m^* \sum_{k=0}^{m-1} a_{m-1,m-1-k} x(n-1-k) \\
 &= f_{m-1}(n) + k_m^* b_{m-1}(n-1)
 \end{aligned}$$

预测误差滤波器的格型结构实现有几个明确的特点。

(1) 模块化结构, 格型滤波器的每一级结构完全相同, 只有唯一的参数即反射系数的取值不同, 这利于硬件实现时模块的复用。

(2) 增加阶数后不改变前级的参数, 对于格型实现, 由 M 阶增加到 $M+1$ 阶只需要在最后增加一个反射系数为 k_{M+1} 模块, 不需要改变前 M 级的参数。

(3) 格型结构可以同时计算前向和后向预测误差。

例 3.2.3 一个 WSS, 已知它的 4 个自相关值分别为

$$r_x(0)=2, \quad r_x(1)=1, \quad r_x(2)=1, \quad r_x(3)=0.5$$

求它的 3 阶前向预测误差滤波器的格型结构和横向结构的参数, 分别画出结构图。

解: 用 Levinson-Durbin 递推算法, 初始条件: $p_0=r_x(0)=2, \Delta_0=r_x(1)=1, a_{0,0}=1$ 第 1 阶递推:

$$k_1 = -\frac{\Delta_0}{p_0} = -\frac{1}{2}, \quad p_1 = p_0(1 - k_1^2) = 1.5$$

$$a_{1,0} = 1$$

$$a_{1,1} = a_{0,1} + k_1 a_{0,0} = k_1 = -1/2$$

第2阶递推:

$$\Delta_1 = r_x(2) + a_{1,1} r_x(1) = 1/2$$

$$k_2 = -\frac{\Delta_1}{p_1} = -\frac{1}{3}$$

$$p_2 = p_1(1 - k_2^2) = 4/3$$

$$a_{2,0} = 1$$

$$a_{2,1} = a_{1,1} + k_2 a_{1,1} = -1/3$$

$$a_{2,2} = k_2 = -1/3$$

第3阶递推:

$$\Delta_2 = r_x(3) + a_{2,1} r_x(2) + a_{2,2} r_x(1) = -1/6$$

$$k_3 = -\frac{\Delta_2}{p_2} = \frac{1}{8}$$

$$p_3 = p_2(1 - k_3^2) = 21/16$$

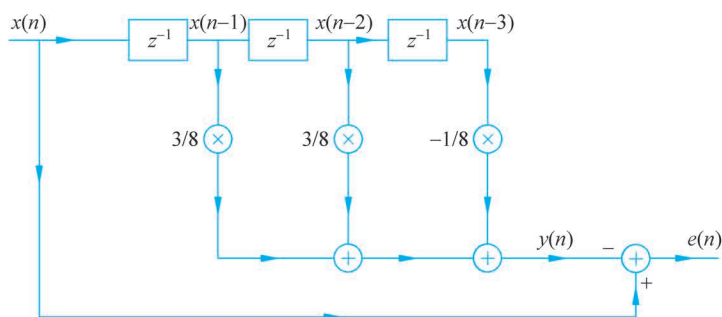
$$a_{3,0} = 1$$

$$a_{3,1} = a_{2,1} + k_3 a_{2,2} = -3/8$$

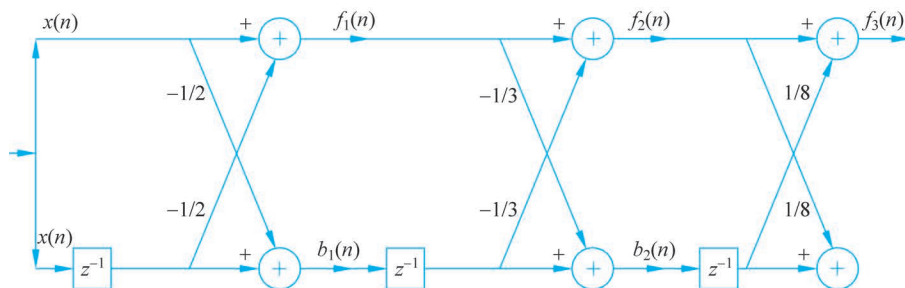
$$a_{3,2} = a_{2,2} + k_3 a_{2,1} = -3/8$$

$$a_{3,3} = k_3 = 1/8$$

由以上递推的参数,分别画出横向和格型结构图如图3.2.3所示。



(a) 前向预测误差滤波器的横向结构



(b) 前向预测误差滤波器的格型结构

图 3.2.3 横向与格型滤波器结构

3.2.5 预测误差滤波器的性质

预测误差滤波器有一组性质,由这些性质可能导出一些有用的结论。

性质 1 m 阶前向预测误差滤波器系统函数和 m 阶后向预测误差滤波器系统函数之间满足

$$H_{b,m}(z) = z^{-m} H_{f,m}^*(z^{-1}) \quad (3.2.56)$$

m 阶前向预测误差滤波器系统函数满足递推关系式

$$H_{f,m}(z) = H_{f,m-1}(z) + k_m^* z^{-1} H_{b,m-1}(z) \quad (3.2.57)$$

这里, $H_{f,m}(z)$ 是 m 阶前向预测误差滤波器系统函数, $H_{b,m}(z)$ 是 m 阶后向预测误差滤波器系统函数。由式(3.2.10)可得 $H_{f,m}(z)$ 的表达式为

$$H_{f,m}(z) = 1 + \sum_{k=1}^m a_{m,k} z^{-k} \quad (3.2.58)$$

利用式(3.2.58)直接可以证明式(3.2.56),将 Levinson-Durbin 递推式(3.2.41)代入式(3.2.58),直接可以证明式(3.2.57),推导过程类似于式(3.2.52)的证明,证明的详细步骤留作习题。

性质 2 前向预测误差滤波器是最小相位的,后向预测误差滤波器是最大相位的。如果各阶反射系数 $|k_m| < 1$, 对应有 $|z_{m,i}| < 1$, 这里 $z_{m,i}$ 表示 m 阶前向预测误差滤波器的第 i 个零点。

用数学归纳法来证明性质 2,为简单假设滤波器系数是实的。对 1 阶前向预测误差滤波器,由 $H_{f,1}(z) = 1 + k_1 z^{-1}$,得到的零点是 $z_{1,1} = k_1$,因此 $|z_{1,1}| = |k_1| < 1$ 。假设 $m-1$ 阶前向预测误差滤波器的各零点都位于单位圆内,定义

$$H_{a,m-1}(z) = \frac{z^{-m} H_{f,m-1}(1/z)}{H_{f,m-1}(z)}$$

显然, $H_{a,m-1}(z)$ 是一个全部极点在单位圆内的全通系统。设 $z_{m,i}$ 是 m 阶前向预测误差滤波器的零点,因此

$$H_{f,m}(z_{m,i}) = H_{f,m-1}(z_{m,i}) + k_m^* z_{m,i}^{-1} H_{b,m-1}(z_{m,i}) = 0$$

利用上式和式(3.2.56),得

$$|H_{a,m-1}(z_{m,i})| = \frac{1}{|k_m|} > 1$$

利用全通系统的一个性质:如果全通系统 $H_{ap}(z)$ 的全部极点在单位圆内,则有

$$H_{ap}(z) \begin{cases} > 1, & |z| < 1 \\ = 1, & |z| = 1 \\ < 1, & |z| > 1 \end{cases}$$

由此性质,得 $|z_{m,i}| < 1$,最小相位性质得证。

由式(3.2.56)得到后向预测误差滤波器的零点位于单位圆外,是最大相位的。

性质 3 如果随机信号是 $AR(p)$ 过程,它通过 p 阶前向预测误差滤波器后的输出是白噪声。

由 $AR(p)$ 模型和 p 阶前向预测误差滤波器解的等价性,这个性质是显然的,由例 3.2.2 的推广得到该性质的证明。

性质 4 预测误差滤波器的正交或功率关系。

预测误差滤波器存在许多正交和功率关系,几个较重要的关系分述如下:

$$\begin{aligned} (1) \quad & E[f_m(n)x^*(n-k)] = 0, 1 \leq k \leq m \\ & E[b_m(n)x^*(n-k)] = 0, 0 \leq k \leq m-1 \end{aligned} \quad (3.2.59)$$

这个关系式就是最优滤波器的正交性原理在前向线性预测和后向线性预测的具体形式。

$$(2) \quad E[f_m(n)x^*(n)] = [b_m(n)x^*(n-m)] = p_m \quad (3.2.60)$$

式(3.2.60)可由(3.2.59)直接推出,即

$$\begin{aligned} E[f_m(n)x^*(n)] &= E\left[f_m(n)\left[x^*(n) - \sum_{k=1}^m a_{m,k}x^*(n-k)\right]\right] = E[f_m(n)f_m^*(n)] = p_m \\ (3) \quad E[b_m(n)b_i^*(n)] &= \begin{cases} p_m, & i=m \\ 0, & i \neq m \end{cases} \end{aligned} \quad (3.2.61)$$

式(3.2.61)说明,各不同阶后向预测误差滤波器在同时刻的输出互相正交。也就是图 3.2.2(b)中,底下一行的各级输出是相互正交的。

$$(4) \quad E[f_m(n)f_i^*(n)] = p_i, l = \max(m, i) \quad (3.2.62)$$

前向预测误差滤波器没有与后向预测误差滤波器相同的性质,即图 3.2.2(b)中,顶上一行的各级输出在同一时刻不是相互正交的。

$$(5) \quad E[f_i(n)f_j^*(n-k)] = 0, \begin{cases} 1 \leq k \leq i-j, & i > j \\ -1 \geq k \geq i-j, & i < j \end{cases} \quad (3.2.63)$$

$$E[b_i(n)b_j^*(n-k)] = 0, \begin{cases} 0 \leq k \leq i-j-1, & i > j \\ 0 \geq k \geq i-j+1, & i < j \end{cases} \quad (3.2.64)$$

式(3.2.63)和式(3.2.64)给出了前向预测误差和后向预测误差在不同阶和不同时刻的正交关系,作为一个特例,对于前向预测误差,取 $i=n, k=i-j, j < i$,即等价的 $i=n, j=i-k, k=1, 2, \dots, n-1$,前向预测误差关系的特例是

$$E[f_n(n)f_{n-k}^*(n-k)] = 0, \quad k = 1, 2, \dots, n-1 \quad (3.2.65)$$

或写成另一种形式

$$E[f_n(n)f_m^*(m)] = 0, \quad m < n \quad (3.2.66)$$

这个关系式清楚地说明, m 阶前向预测误差滤波器在 m 时刻的输出与 n 阶前向预测误差滤波器在 n 时刻的输出是正交的(当 $n \neq m$)。

式(3.2.61)至式(3.2.64)的证明留作习题。

3.3 最小二乘滤波

本章前几节采用均方误差准则意义下的最优滤波,通过一个滤波器输出估计一个期望响应。为使滤波器输出是期望响应的最优估计,求滤波器系数使估计误差的均方值最小,由此得到了维纳滤波器。维纳滤波器的求解需要输入信号向量的自相关矩阵和输入与期望响应的互相关向量,在许多实际问题中,这两个统计量并不存在,实际存在的仅是一段时间内的输入信号和期望响应的波形。在这段有限时间内滤波器对期望响应进行估计,使估计误



视频讲解

差的平方和最小得到的滤波器权系数,就是最小二乘(LS)意义下的最优滤波器。

需要注意,也有文献把维纳滤波这样的按均方意义的统计最优问题,统称为 LS 问题,本书中,我们用 LS 专指“有限误差平方和最小”这种情况,以区别维纳滤波。在 2.6 节已讨论了线性模型参数估计情况下的 LS 技术,本节讨论 LS 的滤波问题。作为线性系统的表示,讨论复数据和复系统的最一般情况,实数据和实系统简单地作为一种特例,通过去掉共轭符号或将共轭转置符“H”变为转置符“T”而得到其解。

由于 LS 滤波器仅讨论有限误差和,其滤波器工作的持续时间有限,因此,对 LS 滤波器仅讨论 FIR 结构。由输入信号 $x(i), x(i-1), \dots, x(i-M+1)$ 估计期望响应 $d(i)$, 估计误差 $e(i)$, 则有

$$\begin{aligned} e^*(i) &= d^*(i) - y^*(i) \\ &= d^*(i) - \sum_{k=0}^{M-1} w_k x^*(i-k) = d^*(i) - \mathbf{x}^H(i) \mathbf{w} \end{aligned} \quad (3.3.1)$$

注意到,为使待求系数 w_k 上没有共轭符号,对误差和信号项加上共轭符号。误差表示的向量形式为

$$\mathbf{e} = \mathbf{d} - \mathbf{A} \mathbf{w} \quad (3.3.2)$$

对期望响应向量的估计为

$$\hat{\mathbf{d}} = \mathbf{y} = \mathbf{A} \mathbf{w} \quad (3.3.3)$$

这里,

$$\begin{aligned} \mathbf{e} &= [e^*(i_1), e^*(i_1+1), \dots, e^*(i_2)]^T \\ \mathbf{d} &= [d^*(i_1), d^*(i_1+1), \dots, d^*(i_2)]^T \\ \mathbf{x}(i) &= [x(i), x(i-1), \dots, x(i-M+1)]^T \\ \mathbf{w} &= [w_0, w_1, \dots, w_{M-1}]^T \end{aligned} \quad (3.3.4)$$

和

$$\mathbf{A} = \begin{bmatrix} \mathbf{x}^H(i_1) \\ \mathbf{x}^H(i_1+1) \\ \vdots \\ \mathbf{x}^H(i_2) \end{bmatrix} \quad (3.3.5)$$

i_1, i_2 表示处理问题所考虑的起始时间和终止时间,即所谓 LS 滤波所考虑的误差取值范围为 $i_1 \leq n \leq i_2$, $\mathbf{A}$ 是数据矩阵,稍后会进一步讨论考虑有限区间边界问题时 $\mathbf{A}$ 的取值问题。

LS 滤波问题中,需确定 w_k 的值使

$$\begin{aligned} \xi(w_0, w_1, \dots, w_{M-1}) &= \sum_{i=i_1}^{i_2} |e(i)|^2 = \sum_{i=i_1}^{i_2} \left| d(i) - \sum_{k=0}^{M-1} w_k^* x(i-k) \right|^2 \\ &= (\mathbf{d} - \mathbf{A} \mathbf{w})^H (\mathbf{d} - \mathbf{A} \mathbf{w}) \end{aligned} \quad (3.3.6)$$

最小。

从如上的讨论中不难看出,LS 滤波器是一个“现实”的最优滤波器。维纳滤波器的设计目标是使得滤波器误差模平方的期望值最小,在各态历经的情况下,相当于在 $(-\infty, +\infty)$ 区间上误差模平方的均值最小,因此,滤波器的求解需要输入信号的自相关矩阵和输入信号与期望响应的互相关向量,这在一些实际应用中可能难以做到。LS 滤波器的目标是现实

的,使用得到的观测数据及其持续时间,在这个持续时间内令误差模平方和最小(也等价于均值最小)而设计出滤波器系数。由于在 2.6 节已给出了 LS 解的形式,并且在 2.6.3 节给出了复数据和系统情况下 LS 的解,对滤波问题 LS 解重写为如下几个方程。LS 滤波器系数向量满足方程式

$$\mathbf{A}^H \mathbf{A} \mathbf{w}_{LS} = \mathbf{A}^H \mathbf{d} \quad (3.3.7)$$

用 $\mathbf{w}_{LS}$ 表示 LS 滤波器的解,若 $\mathbf{A}^H \mathbf{A}$ 非奇异,则解为

$$\mathbf{w}_{LS} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d} \quad (3.3.8)$$

最小误差平方和为

$$J_{\min} = J(\mathbf{w}_{LS}) = \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{A} \mathbf{w}_{LS} \quad (3.3.9)$$

LS 滤波器的解只用到输入数据和期望响应的值。

3.3.1 LS 滤波的边界问题

在 LS 方法中,因为数据是有限长的,因此,存在对边界的不同处理方式,有几种(i_1, i_2)的不同取法,对应着对数据窗或数据矩阵的不同取法。

在实际滤波和参数估计问题中,可以将起始时间定义为 0,如果需要考虑终止时间,将终止时间表示为 $N-1$ 。即假设输入数据区间为 $[0, N-1]$,在不同应用中,对输入数据的处理方式有多种选择,以下是 4 种基本方式。

1. 协方差方法

在产生滤波器输出或对期望响应进行估计时,避免对数据窗外的信号取值做任何假设。因此,如果 FIR 滤波器的非 0 系数为 M 个,考虑滤波器输出误差时,必须取 $i_1 = M-1, i_2 = N-1$,这样在产生滤波器输出误差时才不会使用到 $[0, N-1]$ 区间之外的数据。这种方法不需要对数据窗之外的信号取值做任何假设,相当于不做任何“加窗”处理。考虑到式(3.3.8)对数据计算的要求, $\mathbf{A}$ 和 $\mathbf{A}^H$ 起的作用一样重要,为书写方便,直接给出转置形式的数据矩阵定义为

$$\begin{aligned} \mathbf{A}^H &= [\mathbf{x}(M-1), \mathbf{x}(M), \dots, \mathbf{x}(N-1)] \\ &= \begin{bmatrix} x(M-1) & x(M) & \cdots & x(N-1) \\ x(M-2) & x(M-1) & \cdots & x(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ x(0) & x(1) & \cdots & x(N-M) \end{bmatrix} \end{aligned} \quad (3.3.10)$$

这里, $\mathbf{x}(i) = [x(i), x(i-1), \dots, x(i-M+1)]^T$ 是计算每个时刻滤波器输出所需要的输入信号向量。这种取数据的方法称为协方差方法,需要注意,LS 的协方差方法的名称是一种习惯的叫法,它与随机信号的“协方差”的定义并无严格的联系,是一种因习惯而保留下来的名词。

2. 自相关方法

对一个有 M 个非 0 系数的 FIR 滤波器,如果输入数据范围为 $[0, N-1]$,滤波器输出范围可扩展到 $[0, N+M-2]$ 。取所有可能存在的滤波器输出值,得到相应的估计误差,即取滤波器输出误差的求和范围为 $i_1 = 0, i_2 = N+M-2$ 。这种情况下,当需要数据观测窗 $[0, N-1]$ 之外的数据时,假设 $x(i)$ 在 $0 \leq i \leq N-1$ 的窗外取值为 0,由此,该方法可以计入所有不为 0

的滤波器输出。但该方法相当于对输入数据进行了“加窗”预处理,假设窗外数据为 0。数据矩阵的共轭转置形式为

$$\begin{aligned} \mathbf{A}^H &= [\mathbf{x}(0), \mathbf{x}(2), \dots, \mathbf{x}(N-1), \dots, \mathbf{x}(N+M-2)] \\ &= \begin{bmatrix} x(0) & x(1) & \cdots & x(M-1) & \cdots & x(N-1) & 0 & \cdots & 0 \\ 0 & x(0) & \cdots & x(M-2) & \cdots & x(N-2) & x(N-1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x(0) & \cdots & x(N-M) & x(N-M-1) & \cdots & x(N-1) \end{bmatrix} \end{aligned} \quad (3.3.11)$$

3. 预加窗法

令 $i_1=0, i_2=N-1$, 相当于假设 $i < 0$ 时, $x(i)=0$ 。不考虑 $i \geq N$ 之后的数据。相当于加了前窗,即窗的起始时间是 0,但没有考虑终止时间端的情况,因此称为预加窗。数据矩阵的转置形式为

$$\begin{aligned} \mathbf{A}^H &= [\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(N-1)] \\ &= \begin{bmatrix} x(0) & x(1) & \cdots & x(M-1) & \cdots & x(N-1) \\ 0 & x(0) & \cdots & x(M-2) & \cdots & x(N-2) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x(0) & \cdots & x(N-M) \end{bmatrix} \end{aligned} \quad (3.3.12)$$

4. 后加窗法

令 $i_1=M-1, i_2=N+M-2$ 相当于假设 $i > N$ 时, $x(i)=0$,但避开对 $i \leq 0$ 时的数据做假设。数据矩阵为

$$\begin{aligned} \mathbf{A}^H &= [\mathbf{x}(M-1), \dots, \mathbf{x}(N-1), \dots, \mathbf{x}(N+M-2)] \\ &= \begin{bmatrix} x(M-1) & \cdots & x(N-1) & 0 & \cdots & 0 \\ x(M-2) & \cdots & x(N-2) & x(N-1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x(0) & \cdots & x(N-M) & x(N-M-1) & \cdots & x(N-1) \end{bmatrix} \end{aligned} \quad (3.3.13)$$

LS 方法在很多不同的信号处理问题中得以应用,在不同的应用中,数据矩阵有不同的取法。例如,在谱估计的应用时,由于加窗会引起谱分辨率的下降,最常用的数据矩阵是按协方差方法取的;在自适应滤波应用时,往往从接收到第一个数据就开始算法迭代,因此常用预加窗法。

以上式(3.3.9)~式(3.3.13)是在假设输入数据范围为 $[0, N-1]$,在标准滤波问题时数据矩阵的形式,对于各种不同应用,只需写出式(3.3.2)形式的误差向量表达式,即可得到相应的数据矩阵和期望响应向量,并不必要按式(3.3.9)~式(3.3.13)套用公式。

3.3.2 LS 的正交性原理

尽管由式(3.3.8)已经得到了 LS 滤波器的解,以下还是要导出 LS 滤波器的另外一种解的形式,这来自于与维纳滤波器相对应的 LS 的正交性原理,这可以使得我们用向量空间角度进一步理解 LS 滤波问题。

LS 的正交性原理的推导思路与维纳滤波情况一致,只是将取期望运算变成有限求和运

算。令

$$J(\mathbf{w}) = \sum_{i=i_1}^{i_2} |e(i)|^2 = \sum_{i=i_1}^{i_2} e(i)e^*(i) \quad (3.3.14)$$

将

$$e(i) = d(i) - \sum_{k=0}^{M-1} w_k^* x(i-k)$$

代入上式,并令

$$\nabla J(\mathbf{w}) = \mathbf{0} \quad (3.3.15)$$

或标量形式

$$\nabla_k J = \frac{\partial J}{\partial w_k} = 0, \quad k=0,1,\dots,M-1 \quad (3.3.16)$$

在满足式(3.3.16)时,LS 误差 $e(i)$ 达到最小值 $e_o(i)$,由式(3.3.16)得到

$$\sum_{i=i_1}^{i_2} x(i-k)e_o^*(i) = 0, \quad k=0,1,\dots,M-1 \quad (3.3.17)$$

并由

$$\hat{d}(i) = y_o(i) = \sum_{k=0}^{M-1} w_{o,k}^* x(i-k) \quad (3.3.18)$$

进一步得到

$$\sum_{i=i_1}^{i_2} \hat{d}(i)e_o^*(i) = 0 \quad (3.3.19)$$

式(3.3.17)、式(3.3.19)是 LS 意义下的正交原理及其推论。

1. 正则方程与最小误差平方和公式

与维纳滤波器类似,在 LS 意义下达到最优时,滤波器系数、期望响应的估计、误差和满足的关系式,均可以由正交性原理推导出。由

$$e_o(i) = d(i) - \sum_{t=0}^{M-1} w_{o,t}^* x(i-t) \quad (3.3.20)$$

两边取共轭后,同乘 $x(i-k)$ 并从 $i=i_1$ 到 i_2 求和,得

$$\sum_{t=0}^{M-1} w_{o,t} \sum_{i=i_1}^{i_2} x(i-k)x^*(i-t) = \sum_{i=i_1}^{i_2} x(i-k)d^*(i), \quad k=0,1,\dots,M-1 \quad (3.3.21)$$

令

$$\Phi(t,k) = \sum_{i=i_1}^{i_2} x(i-t)x^*(i-k), \quad 0 \leq (t,k) \leq M-1 \quad (3.3.22)$$

$$z(-k) = \sum_{i=i_1}^{i_2} x(i-k)d^*(i), \quad 0 \leq k \leq M-1 \quad (3.3.23)$$

式(3.3.21)简化为

$$\sum_{t=0}^{M-1} w_{o,t} \Phi(t,k) = z(-k), \quad k=0,1,\dots,M-1 \quad (3.3.24)$$

式(3.3.24)可称为 LS 意义下的维纳-霍普夫方程,或 LS 的正则方程。写成矩阵形式:

$$\Phi \mathbf{w}_o = \mathbf{z} \quad (3.3.25)$$

这里,

$$[\Phi]_{ij} = \Phi(j, i) \\ \mathbf{z} = [z(0), z(-1), \dots, z(-M+1)]^T$$

若 Φ 可逆,式(3.3.25)的解为

$$\mathbf{w}_o = \Phi^{-1} \mathbf{z} \quad (3.3.26)$$

Φ 的几个性质,由 Φ 的定义可以验证它的几个性质。

- (1) $\Phi^H = \Phi$,但一般情况下, Φ 不再满足 Toeplitz 性,除非数据矩阵采用自相关方法。
- (2) 对任意向量 $\mathbf{x}$,满足 $\mathbf{x}^H \Phi \mathbf{x} \geq 0$,也就是说 Φ 是半正定的。
- (3) Φ 的特征值是实的和非负的。
- (4) Φ 可以分解为

$$\Phi = \mathbf{A}^H \mathbf{A} = \sum_{i=i_1}^{i_2} \mathbf{x}(i) \mathbf{x}^H(i) \quad (3.3.27)$$

$\mathbf{z}$ 可以表示成

$$\mathbf{z} = \mathbf{A}^H \mathbf{d} = \sum_{i=i_1}^{i_2} \mathbf{x}(i) d(i) \quad (3.3.28)$$

将式(3.3.27)和式(3.3.28)代入式(3.3.25),得到 LS 的正则方程的矩阵形式

$$\mathbf{A}^H \mathbf{A} \mathbf{w}_o = \mathbf{A}^H \mathbf{d} \quad (3.3.29)$$

如果 Φ 可逆,滤波器系数为

$$\mathbf{w}_o = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d} \quad (3.3.30)$$

对期望响应的 LS 估计为

$$\hat{\mathbf{d}} = \mathbf{A} \mathbf{w}_o = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d} \quad (3.3.31)$$

由正交原理得到的式(3.3.29)和式(3.3.30)与前面用矩阵运算得到的式(3.3.7)和式(3.3.8)是一致的,即式(3.3.30)的解 $\mathbf{w}_o$ 和式(3.3.8)的解 $\mathbf{w}_{LS}$ 相等。

与维纳滤波类似,由正交原理同样验证在 LS 意义下,各变量和的关系为

$$\xi_d = \sum_{i=i_1}^{i_2} |d(i)|^2 = \sum_{i=i_1}^{i_2} |e_o(i)|^2 + \sum_{i=i_1}^{i_2} |\hat{d}(i)|^2 = J_{\min} + \xi_{est}$$

故最小误差平方和为

$$J_{\min} = \xi_d - \xi_{est} \quad (3.3.32)$$

由式(3.3.3)得

$$\xi_{est} = \mathbf{w}_o^H \Phi \mathbf{w}_o = \mathbf{w}_o^H \mathbf{z} = \mathbf{z}^H \mathbf{w}_o = \mathbf{z}^H \Phi^{-1} \mathbf{z}$$

最小误差平方和为

$$J_{\min} = \xi_d - \mathbf{w}_o^H \mathbf{z} = \mathbf{d}^H \mathbf{d} - \mathbf{z}^H \Phi^{-1} \mathbf{z} \quad (3.3.33)$$

2. 投影算子

如果将矩阵 $\mathbf{A}$ 的列向量张成线性空间,在这个线性空间上估计向量 $\mathbf{d}$,得到 $\mathbf{d}$ 在该空间上的投影是 $\mathbf{d}$ 在该空间上的最优估计,这个估计由式(3.3.31)表示。由此,可以定义

$$\mathbf{P} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \quad (3.3.34)$$

为对 $\mathbf{A}$ 的列向量张成的空间的投影算子,任意向量 $\mathbf{d}$ 在该空间的投影为

$$\hat{\mathbf{d}} = \mathbf{P}\mathbf{d} = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{d} \quad (3.3.35)$$

很自然地定义正交投影算子

$$\mathbf{P}^\perp = \mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H \quad (3.3.36)$$

它作用于 $\mathbf{d}$ 得到的向量是 $\mathbf{d}$ 与 $\mathbf{d}$ 在 $\mathbf{A}$ 的列空间的投影的向量差,该向量是与 $\mathbf{A}$ 的列空间垂直的向量,或者说是最小误差向量,即

$$\mathbf{e} = \mathbf{d} - \hat{\mathbf{d}} = \mathbf{P}^\perp\mathbf{d} \quad (3.3.37)$$

3.3.3 最小二乘滤波的几个性质

为了更直观地讨论 LS 的性质,假设用 LS 估计一个自回归过程(AR)的参数。假设一个自回归过程为

$$x(n) = \sum_{k=1}^M w_{0,k}x(n-k) + v(n)$$

将 AR 过程在时间范围 $[0, N-1]$ 的取值写成向量形式:

$$\mathbf{x} = \mathbf{A}\mathbf{w}_0 + \mathbf{v}$$

由测量向量 $\mathbf{x} = \mathbf{d}$ 估计自回归过程的参数 $\mathbf{w}_0$, 令 $\mathbf{v} = \mathbf{e}_0$, 方程写为

$$\mathbf{d} = \mathbf{A}\mathbf{w}_0 + \mathbf{e}_0 \quad (3.3.38)$$

这里, $\mathbf{e}_0$ 等价于测量噪声,假设它是均值为 0 的白噪声,方差为 σ_e^2 。可以分析得到 $\mathbf{w}_0$ 的 LS 估计 $\hat{\mathbf{w}}_0$ 的性质。首先, $\hat{\mathbf{w}}_0$ 满足正则方程

$$\mathbf{A}^H\mathbf{A}\hat{\mathbf{w}}_0 = \mathbf{A}^H\mathbf{d} \quad (3.3.39)$$

将式(3.3.38)代入式(3.3.39),得

$$\mathbf{A}^H\mathbf{A}\hat{\mathbf{w}}_0 = \mathbf{A}^H\mathbf{d} = \mathbf{A}^H(\mathbf{A}\mathbf{w}_0 + \mathbf{e}_0)$$

因此 $\mathbf{w}_0$ 的 LS 估计为

$$\hat{\mathbf{w}}_0 = \mathbf{w}_0 + (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{e}_0 \quad (3.3.40)$$

由式(3.3.40)不难验证 LS 估计的如下统计性质(1)、(2),性质(3)、(4)的证明是第 2 章习题 4 的结论。

(1) LS 估计 $\hat{\mathbf{w}}_0$ 是无偏估计, $E[\hat{\mathbf{w}}_0] = \mathbf{w}_0$ 。

(2) $\hat{\mathbf{w}}_0$ 的协方差矩阵是 $\mathbf{C}_w = \sigma_e^2(\mathbf{A}^H\mathbf{A})^{-1} = \sigma_e^2\Phi^{-1}$ 。

(3) $\hat{\mathbf{w}}_0$ 是最小方差无偏估计器。

(4) 如果 $\mathbf{e}_0$ 是高斯白噪声, $\hat{\mathbf{w}}_0$ 是最小方差无偏估计,且克拉美-罗下界可达,同时它也是最大似然估计。

比较一下维纳滤波器和 LS 滤波器是有意义的。重写 FIR 结构维纳滤波器的维纳-霍普夫方程如下:

$$\mathbf{R}\mathbf{w}_0 = \mathbf{r}_{xd}$$

比较 LS 的正则方程

$$\Phi\mathbf{w}_0 = \mathbf{z}$$

为了比较方便,考虑 LS 的输入信号取值范围为 $[0, N-1]$, 令 $\Phi' = \frac{1}{N}\Phi$ 和 $\mathbf{z}' = \frac{1}{N}\mathbf{z}$, 正则方

程重写为

$$\Phi' w_o = z'$$

3.3.1 节讨论了 LS 滤波的 4 种边界处理方法, 相应于 4 种数据矩阵的构成方式, 4 种方法对应的矩阵 Φ' 不同, 可以证明(留作习题), 若使用自相关方法处理边界, 则 Φ' 满足 Toeplitz 性, 实际上 $\Phi' = \hat{R}_x$ 和 $z' = \hat{r}_{xd}$, 这里, $\hat{R}_x$ 和 $\hat{r}_{xd}$ 相当于用 $[0, N-1]$ 范围的观测数据估计 $r_x(k)$ 和 $r_{xd}(k)$, 然后用估计得到的自相关序列和互相关序列构成的自相关矩阵和互相关向量, 因此, 自相关方法构成数据矩阵的 LS 滤波器是维纳滤波器在有限样本情况下的近似。需要注意, 若利用其他边界处理方式, 例如协方差方法构成的数据矩阵, 则 Φ' 满足半正定性但不满足 Toeplitz 性。当观测数据量变大, 边界效应减弱, 当 $N \rightarrow \infty$ 时, 若信号满足平稳性和遍历性, 则 $\Phi' \rightarrow R_x$ 和 $z' \rightarrow r_{xd}$, LS 滤波和维纳滤波趋于一致。

在观测数据量有限, 尤其 N 较小时, LS 的不同边界处理方式, 可能对结果有明显的影响, 这时对一类应用选择合适的边界处理会使 LS 取得更好的效果。尽管自相关方法使得 Φ' 具有 Toeplitz 性, 但自相关方法往往不是好的选择, 例如, 在信号的现代谱估计中协方差方法得到的估计性能更好, 而在自适应滤波时更常用预加窗方法。LS 边界处理问题在第 5 章和第 6 章会进一步讨论。

3.3.4 最小二乘的线性预测

与最优线性预测是维纳滤波器的一个特例一样, LS 意义下的最优预测问题也是 LS 滤波的特例, 假设已观测到随机信号的一组数据 $\{x(0), x(1), \dots, x(N-1)\}$, 求 M 阶线性 LS 预测器, 目标是求解最优线性预测系数。同 LS 滤波问题一样, 可以有 4 种数据窗的取法, 这里仅以预加窗法为例进行说明。

很显然, 期望响应向量为

$$d = [x^*(0), x^*(1), \dots, x^*(N-1)]^T$$

利用预加窗法, 得到对期望响应的预测为

$$\hat{d} = \begin{bmatrix} \hat{x}^*(0) \\ \hat{x}^*(1) \\ \vdots \\ \hat{x}^*(N-1) \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ x^*(0) & 0 & \cdots & 0 & 0 \\ x^*(1) & x^*(0) & 0 & \vdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x^*(M-1) & x^*(M-2) & x^*(M-3) & \cdots & x^*(0) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x^*(N-2) & x^*(N-3) & \cdots & x^*(N-M) & x^*(n-M-1) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix}$$

数据矩阵为

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ x^*(0) & 0 & \cdots & 0 & 0 \\ x^*(1) & x^*(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x^*(M-1) & x^*(M-2) & x^*(M-3) & \cdots & x^*(0) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x^*(N-2) & x^*(N-3) & \cdots & x^*(N-M) & x^*(n-M-1) \end{bmatrix}$$

为了与前面 LS 滤波的数据矩阵比较,也可以写成数据矩阵的共轭转置:

$$\begin{aligned} \mathbf{A}^H &= [\mathbf{x}(-1), \mathbf{x}(0), \mathbf{x}(1), \cdots, \mathbf{x}(N-2)] \\ &= \begin{bmatrix} 0 & x(0) & x(1) & \cdots & x(N-3) & x(n-2) \\ 0 & 0 & x(0) & \cdots & x(n-4) & x(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x(n-M) & x(n-M-1) \end{bmatrix} \end{aligned}$$

例 3.3.1 LS 预测,设一个随机信号的前几个数据是 $x(n) = \{1, 2, 3, 3, 4, 5\}$,用 LS 方法求一个前向预测器,预测器是两系数的,用预加窗法。

解: 对这个预测问题,列出如下期望响应和预测器估计方程:

$$\mathbf{d} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \\ 4 \\ 5 \end{bmatrix} \quad \hat{\mathbf{d}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 2 & 1 \\ 3 & 2 \\ 3 & 3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \mathbf{A} \mathbf{w}$$

得到数据矩阵为

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 2 & 1 \\ 3 & 2 \\ 3 & 3 \\ 4 & 3 \end{bmatrix}$$

LS 解的预测系数为

$$\mathbf{w} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d} = \begin{bmatrix} 1.48 \\ -0.3036 \end{bmatrix}$$

因此,对期望响应向量的预测值向量为

$$\hat{\mathbf{d}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 2 & 1 \\ 3 & 2 \\ 3 & 3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \mathbf{A} \mathbf{w} = \begin{bmatrix} 0 \\ 1.4821 \\ 2.6607 \\ 3.8393 \\ 3.5357 \\ 5.0179 \end{bmatrix}$$

预测误差向量为

$$\mathbf{e} = \mathbf{d} - \hat{\mathbf{d}} = [1.0000 \quad 0.5179 \quad 0.3393 \quad -0.8393 \quad 0.4643 \quad -0.0179]^T$$

对此问题,也可以直接求投影矩阵,并由投影矩阵得到对期望响应向量的估计,即

$$\mathbf{P} = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4107 & 0.3036 & 0.1964 & -0.3214 & 0.0893 \\ 0 & 0.3036 & 0.2679 & 0.2321 & -0.1071 & 0.1964 \\ 0 & 0.1964 & 0.2321 & 0.2679 & 0.1071 & 0.3036 \\ 0 & -0.3214 & -0.1071 & 0.1071 & 0.6429 & 0.3214 \\ 0 & 0.0893 & 0.1964 & 0.3036 & 0.3214 & 0.4107 \end{bmatrix}$$

$$\hat{\mathbf{d}} = \mathbf{P}\mathbf{d} = [0 \quad 1.4821 \quad 2.6607 \quad 3.8393 \quad 3.5357 \quad 5.0179]^T$$

3.3.5 正则最小二乘滤波

与 2.6 节讨论线性模型 LS 估计遇到的问题一样,在 LS 滤波应用时,若数据矩阵 $\mathbf{A}$ 非满秩,则 $\Phi = \mathbf{A}^H \mathbf{A}$ 是奇异的,式(3.3.8)所表示的解不存在。即使 Φ 满秩但其条件数很大时, Φ 的逆的数值解结果不稳定。在应用中, Φ 不可逆的情况很少出现,但 Φ 条件数很大的情况更常遇到。对于解决这类问题,发展了正则化最小二乘(Regularized Least Squares)方法,本节对 2.6 节讨论的正则 LS 稍加一般化,讨论复数据和复系统的更一般形式的正则 LS。

所谓复正则化 LS 是指在式(3.3.6)用误差和表示的目标函数中增加一项约束滤波器权系数向量自身的量,常用的约束量选择为滤波器权系数向量的范数平方,即 $\|\mathbf{w}\|^2 = \mathbf{w}^H \mathbf{w}$,因此加了正则化约束的目标函数为

$$\begin{aligned} \xi(\mathbf{w}) &= \sum_{i=i_1}^{i_2} |e(i)|^2 + \lambda \sum_{k=0}^{M-1} |\mathbf{w}_k|^2 \\ &= \sum_{i=i_1}^{i_2} \left| d(i) - \sum_{k=0}^{M-1} \mathbf{w}_k^* x(i-k) \right|^2 + \lambda \|\mathbf{w}\|^2 \\ &= (\mathbf{d} - \mathbf{A}\mathbf{w})^H (\mathbf{d} - \mathbf{A}\mathbf{w}) + \lambda \mathbf{w}^H \mathbf{w} \end{aligned} \quad (3.3.41)$$

这里, λ 是一个可选择的参数,用于控制误差和项与参数向量范数约束项的作用。为求使得式(3.3.41)最小的 $\mathbf{w}$ 值。计算

$$\begin{aligned} \frac{\partial \xi(\mathbf{w})}{\partial \mathbf{w}^*} &= \frac{\partial (\mathbf{d} - \mathbf{A}\mathbf{w})^H (\mathbf{d} - \mathbf{A}\mathbf{w})}{\partial \mathbf{w}^*} + \frac{\partial \lambda \mathbf{w}^H \mathbf{w}}{\partial \mathbf{w}^*} \\ &= -2\mathbf{A}^H \mathbf{d} + 2\mathbf{A}^H \mathbf{A} \mathbf{w} + 2\lambda \mathbf{w} \end{aligned}$$

令 $\mathbf{w} = \mathbf{w}_{\text{LS}}$ 时上式为 0,得

$$(\mathbf{A}^H \mathbf{A} + \lambda \mathbf{I}) \mathbf{w}_{\text{LS}} = \mathbf{A}^H \mathbf{d} \quad (3.3.42)$$

求得滤波器权系数向量的正则化 LS 解为

$$\mathbf{w}_{\text{LS}} = (\mathbf{A}^H \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^H \mathbf{d} \quad (3.3.43)$$

正则化最小二乘是一般性正则理论的一个特例;Tikhonov 正则化理论的泛函由两部分组成:一项是经验代价函数,如式(3.3.41)中的误差和是一种经验代价函数,另一项是正则化项,它是约束系统结构的,在滤波器设计中用于约束滤波器权系数向量的范数。在实际

应用中,参数 λ 是一个经验性参数。

* 3.3.6 基于非线性函数的最小二乘滤波

到目前为止,所讨论的 LS 滤波器是一个线性滤波器,输出是输入的线性函数,即

$$y(n) = \sum_{k=0}^{M-1} w_k^* x(n-k) \quad (3.3.44)$$

可以通过一组非线性映射函数

$$\phi_i(\mathbf{x}(n)), \quad i=1,2,\dots,K \quad (3.3.45)$$

将滤波器输出与输入之间建立起非线性关系,但是与滤波器权系数向量仍是线性关系。注意,这里

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T$$

$$\mathbf{w}_\psi = [\tau w_{\psi,1}, \tau w_{\psi,1}, \dots, \tau w_{\psi,K}]^T$$

注意输入信号向量 $\mathbf{x}(n)$ 为 M 维向量,权系数向量 $\mathbf{w}$ 为 K 维向量与映射函数集 ϕ_i 数目相等。利用映射函数集定义滤波器输出为

$$y(n) = \sum_{k=1}^K \tau w_{\psi,k}^* \phi_k(\mathbf{x}(n)) = \mathbf{w}_\psi^H \boldsymbol{\Psi}(\mathbf{x}(n)) \quad (3.3.46)$$

这里

$$\boldsymbol{\Psi}(\mathbf{x}(n)) = [\phi_1(\mathbf{x}(n)), \phi_2(\mathbf{x}(n)), \dots, \phi_K(\mathbf{x}(n))]^T \quad (3.3.47)$$

在此基础上讨论 LS 滤波,滤波器的结构如式(3.3.46),在有限区间 $i_1 \leq n \leq i_2$ 范围内通过滤波器估计期望响应 $d(n)$,与线性 LS 滤波类似,定义向量

$$\mathbf{e} = [e^*(i_1), e^*(i_1+1), \dots, e^*(i_2)]^T$$

$$\mathbf{d} = [d^*(i_1), d^*(i_1+1), \dots, d^*(i_2)]^T$$

则

$$\mathbf{e} = \mathbf{d} - \mathbf{A}_\psi \mathbf{w}_\psi$$

$$= \begin{bmatrix} d^*(i_1) \\ d^*(i_1+1) \\ \vdots \\ d^*(i_2) \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Psi}^H(\mathbf{x}(i_1)) \\ \boldsymbol{\Psi}^H(\mathbf{x}(i_1+1)) \\ \vdots \\ \boldsymbol{\Psi}^H(\mathbf{x}(i_2)) \end{bmatrix} \mathbf{w}_\psi \quad (3.3.48)$$

这里

$$\mathbf{A}_\psi = \begin{bmatrix} \boldsymbol{\Psi}^H(\mathbf{x}(i_1)) \\ \boldsymbol{\Psi}^H(\mathbf{x}(i_1+1)) \\ \vdots \\ \boldsymbol{\Psi}^H(\mathbf{x}(i_2)) \end{bmatrix}$$

$$= \begin{bmatrix} \phi_1^*(\mathbf{x}(i_1)) & \phi_2^*(\mathbf{x}(i_1)) & \cdots & \phi_K^*(\mathbf{x}(i_1)) \\ \phi_1^*(\mathbf{x}(i_1+1)) & \phi_2^*(\mathbf{x}(i_1+1)) & \cdots & \phi_K^*(\mathbf{x}(i_1+1)) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1^*(\mathbf{x}(i_2)) & \phi_2^*(\mathbf{x}(i_2)) & \cdots & \phi_K^*(\mathbf{x}(i_2)) \end{bmatrix} \quad (3.3.49)$$

注意到,与式(3.3.6)的线性 LS 问题相比,这里除了数据矩阵 $\mathbf{A}_\psi$ 的定义由式(3.3.49)通过

映射函数进行计算外,一旦数据矩阵 $\mathbf{A}_\psi$ 确定了,由于待求量 $\mathbf{w}_\psi$ 仍保持线性关系,求解的问题是一致的,故滤波器权系数的解为

$$\mathbf{w}_\psi = (\mathbf{A}_\psi^H \mathbf{A}_\psi)^{-1} \mathbf{A}_\psi^H \mathbf{d} \quad (3.3.50)$$

注意到,与线性 LS 滤波的不同主要表现在数据矩阵 $\mathbf{A}_\psi$ 中,对于线性 LS 滤波,若输入信号向量 $\mathbf{x}(n)$ 是 M 维的,则数据矩阵 $\mathbf{A}$ 是 $(i_2 - i_1 + 1) \times M$ 维矩阵,且矩阵的每个元素是输入信号 $x(n)$ 在某一时刻的取值。对于基于非线性映射函数的 LS 滤波器(简称 NM-LS),数据矩阵 $\mathbf{A}_\psi$ 是 $(i_2 - i_1 + 1) \times K$ 维矩阵,即数据矩阵的列数为 K ,由映射函数数目确定,且数据矩阵 $\mathbf{A}_\psi$ 的每一个元素需要通过相应映射函数计算得到,一旦数据矩阵 $\mathbf{A}_\psi$ 计算得到了,则 NM-LS 的求解问题与线性 LS 是一致的。

例 3.3.2 讨论一个基于非线性映射函数的 LS 滤波器,考虑滤波器的记忆性为两个单位延迟,故取 $\mathbf{x}(n) = [x(n), x(n-1), x(n-2)]^T$,为了简单假设信号和系统都是实的。设映射函数向量为

$$\Psi(\mathbf{x}(n)) = \begin{bmatrix} \psi_1(\mathbf{x}(n)) \\ \psi_2(\mathbf{x}(n)) \\ \vdots \\ \psi_K(\mathbf{x}(n)) \end{bmatrix} = \begin{bmatrix} x(n) \\ x(n-1) \\ x(n-2) \\ x^2(n) \\ x^2(n-1) \\ x^2(n-2) \\ x(n)x(n-1) \\ x(n)x(n-2) \\ x(n-1)x(n-2) \end{bmatrix}$$

这里, $K=9$,故权系数向量记为

$$\mathbf{w}_\psi = [\mathbf{w}_{\psi,1}, \mathbf{w}_{\psi,2}, \dots, \mathbf{w}_{\psi,9}]^T$$

滤波器输出为

$$\begin{aligned} y(n) &= \sum_{k=1}^K \mathbf{w}_{\psi,k} \psi_k(\mathbf{x}(n)) \\ &= \mathbf{w}_{\psi,1} x(n) + \mathbf{w}_{\psi,2} x(n-1) + \mathbf{w}_{\psi,3} x(n-2) + \mathbf{w}_{\psi,4} x^2(n) + \\ &\quad \mathbf{w}_{\psi,5} x^2(n-1) + \mathbf{w}_{\psi,6} x^2(n-2) + \mathbf{w}_{\psi,7} x(n)x(n-1) + \\ &\quad \mathbf{w}_{\psi,8} x(n)x(n-2) + \mathbf{w}_{\psi,9} x(n-1)x(n-2) \end{aligned}$$

假设考虑的误差范围为 $i_1=2, i_2=50$, 则

$$\mathbf{d} = [d(2), d(3), \dots, d(50)]^T$$

数据矩阵 $\mathbf{A}_\psi$ 为

$$\mathbf{A}_\psi = \begin{bmatrix} x(2) & x(1) & x(0) & x^2(2) & x^2(1) & x^2(0) & x(2)x(1) & x(2)x(0) & x(1)x(0) \\ x(3) & x(2) & x(1) & x^2(3) & x^2(2) & x^2(1) & x(3)x(2) & x(3)x(1) & x(2)x(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x(50) & x(49) & x(48) & x^2(50) & x^2(49) & x^2(48) & x(50)x(49) & x(50)x(48) & x(49)x(48) \end{bmatrix}$$

这里 $\mathbf{A}_\psi$ 是一个 49×9 的数据矩阵,计算 $(\mathbf{A}_\psi^H \mathbf{A}_\psi)^{-1}$ 需要求 9×9 方阵的逆矩阵。

注意到,对此问题若采用线性 LS 滤波器,则输出写为

$$y(n) = w_1 x(n) + w_2 x(n-1) + w_3 x(n-2)$$

数据矩阵 $\mathbf{A}$ 是 49×3 的矩阵, 则 $(\mathbf{A}^H \mathbf{A})^{-1}$ 的计算只需求 3×3 方阵的逆矩阵。另外也需注意到, 写出 $\mathbf{A}_\psi$ 需要一定的计算量, 尤其当 $\psi_K(\mathbf{x}(n))$ 中存在复杂非线性函数时, 附加运算量可能是相当可观的, 而写出 $\mathbf{A}$ 不需要附加计算量。

对于许多实际应用, 怎样选择合适的非线性函数 $\Psi(\mathbf{x}(n))$ 是一个很重要的问题。很多情况下, 非线性函数的选择与所处理的问题密切相关。有一些构造非线性函数集的方法, 例如 Volterra 级数, 例 3.3.2 给出的这组函数是 Volterra 级数的一个例子。近年来, 人们利用核函数的思想研究非线性逼近取得了一些成果, 核函数的思想可用于结合 LS 技术处理非线性问题, 有关核函数的问题, 限于篇幅本书不展开讨论, 可参考相关文献。

3.4 奇异值分解计算 LS 问题

由于 LS 应用的广泛性, LS 问题的求解又主要是矩阵运算问题, 因此对 LS 的计算已成为“矩阵计算”领域的一个专门问题, 已有多种成熟算法, 例如 Householder 变换、Givens 旋转、QR 分解和奇异值分解等, 对这些算法的详细介绍不是本书目标, 有兴趣的读者可参考有关矩阵计算方面的专著, 例如 G. H. 戈卢布的著作^[290]。在 LS 求解的诸多方法中, 由于奇异值分解问题在信号处理领域有多方面的应用, 本节以 LS 求解为例, 介绍奇异值分解问题 (Singular-Value Decomposition, SVD)。

解 LS 问题, 关键是求

$$\mathbf{A}^+ = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \quad (3.4.1)$$

$\mathbf{A}^+$ 称为矩阵 $\mathbf{A}$ 的伪逆 (Pseudo inverse), 当伪逆求得后, LS 滤波器系数的解为

$$\mathbf{w}_{\text{LS}} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{d} = \mathbf{A}^+ \mathbf{d} \quad (3.4.2)$$

如果 $(\mathbf{A}^H \mathbf{A})$ 是满秩的, 它是可逆的, 式 (3.4.1) 定义的伪逆存在, LS 解是唯一的。当 $(\mathbf{A}^H \mathbf{A})$ 不是满秩的, $(\mathbf{A}^H \mathbf{A})$ 不可逆, 由式 (3.4.1) 定义的伪逆不存在, 使 LS 误差最小化的解不是唯一的。下面介绍奇异值分解 (SVD) 定理, 利用 SVD 定理, 不管 $(\mathbf{A}^H \mathbf{A})$ 是否满秩, 总可以得到一个 LS 解, 并且解的数值稳定性是良好的。

1. SVD 定理

定理 3.4.1 (SVD 定理) 给定数据矩阵 $\mathbf{A}$ 是 $N \times M$ 维矩阵, 存在两个酉矩阵 $\mathbf{V}$ 和 $\mathbf{U}$, 使得

$$\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.4.3)$$

这里, $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_W\}$, 且 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_W > 0$, $W \leq \min\{N, M\}$ (注意到, $\mathbf{U}$ 是 $N \times N$ 矩阵, $\mathbf{V}$ 是 $M \times M$ 矩阵)。

证明: 这里给出一种构造式的证明, 主要证明过确定情况 $N > M$, 对欠确定情况的证明是非常相似的。

由于 $\mathbf{A}^H \mathbf{A}$ 是共轭对称的和非负定的, 它的特征值为实的和非负的, 按从大到小顺序排列为 $\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2$, 这里设 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_W > 0$ 和 $\sigma_{W+1} = \dots = \sigma_M = 0$, 并设 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$ 是与各 σ_i^2 对应的特征向量, 互相正交且取其模为 1。因此 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M]$ 是 $M \times M$ 酉矩阵。故

$$\mathbf{V}^H \mathbf{A}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \boldsymbol{\Sigma}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.4.4)$$

令

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_W : \mathbf{v}_{W+1}, \dots, \mathbf{v}_M] = [\mathbf{V}_1, \mathbf{V}_2]$$

这里, $\mathbf{V}_1 = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_W]$, $\mathbf{V}_2 = [\mathbf{v}_{W+1}, \dots, \mathbf{v}_M]$, $\mathbf{V}_1$ 是 $M \times W$ 矩阵, $\mathbf{V}_2$ 是 $M \times (M-W)$ 矩阵, 且

$$\mathbf{V}_1^H \mathbf{V}_2 = \mathbf{0}$$

将 $\mathbf{V}$ 的分块 $\mathbf{V}_1, \mathbf{V}_2$ 代入式(3.4.4), 得 $\mathbf{V}_1^H \mathbf{A}^H \mathbf{A} \mathbf{V}_1 = \boldsymbol{\Sigma}^2$, 即

$$\boldsymbol{\Sigma}^{-1} \mathbf{V}_1^H \mathbf{A}^H \mathbf{A} \mathbf{V}_1 \boldsymbol{\Sigma}^{-1} = \mathbf{I} \quad (3.4.5)$$

和

$$\mathbf{V}_2^H \mathbf{A}^H \mathbf{A} \mathbf{V}_2 = \mathbf{0} \quad (3.4.6)$$

即

$$\mathbf{A} \mathbf{V}_2 = \mathbf{0} \quad (3.4.7)$$

令

$$\mathbf{U}_1 = \mathbf{A} \mathbf{V}_1 \boldsymbol{\Sigma}^{-1} \quad (3.4.8)$$

这里, $\mathbf{U}_1$ 是 $N \times W$ 矩阵, 由式(3.4.5)得

$$\mathbf{U}_1^H \mathbf{U}_1 = \mathbf{I}$$

以 $\mathbf{U}_1$ 为基础, 构造一个酉矩阵 $\mathbf{U}$, $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, 且 $\mathbf{U}_1^H \mathbf{U}_2 = \mathbf{0}$, 由以上关系式, 得到

$$\begin{aligned} \mathbf{U}^H \mathbf{A} \mathbf{V} &= \begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix} \mathbf{A} (\mathbf{V}_1, \mathbf{V}_2) = \begin{bmatrix} \mathbf{U}_1^H \mathbf{A} \mathbf{V}_1 & \mathbf{U}_1^H \mathbf{A} \mathbf{V}_2 \\ \mathbf{U}_2^H \mathbf{A} \mathbf{V}_1 & \mathbf{U}_2^H \mathbf{A} \mathbf{V}_2 \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{A} \mathbf{V}_1 \boldsymbol{\Sigma}^{-1})^H \mathbf{A} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{U}_2^H \mathbf{U}_1 \boldsymbol{\Sigma} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{aligned}$$

最后一行左上角用了式(3.4.5)。

对于欠确定情况 $N < M$, 令 $\mathbf{U}$ 是 $\mathbf{A} \mathbf{A}^H$ 的特征向量矩阵, 是酉的, 其他步骤类似, 此处从略。

SVD 定理的证明过程也同时给出了两个酉矩阵 $\mathbf{V}$ 和 $\mathbf{U}$ 的定义, 其中 $\mathbf{V}$ 是 $\mathbf{A}^H \mathbf{A}$ 的特征向量构成的矩阵, 可通过矩阵特征分解算法求得。在过确定情况下, $\mathbf{A}^H \mathbf{A}$ 是 $M \times M$ 矩阵, M 是滤波器非零系数的个数, 一般远比可用的数据记录数 N 为小, $\mathbf{A}^H \mathbf{A}$ 的特征分解运算量尚可接受。类似 $\mathbf{U}$ 是 $\mathbf{A} \mathbf{A}^H$ 的特征向量构成的矩阵。后面会看到, 在实际利用 SVD 求解 LS 问题时, 一般只要求一个矩阵的特征分解, 例如在过确定性问题时, 只需求 $\mathbf{A}^H \mathbf{A}$ 的特征分解。

SVD 定理也可以写成如下形式:

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^H = \mathbf{U}_1 \boldsymbol{\Sigma} \mathbf{V}_1^H = \sum_{i=1}^W \sigma_i \mathbf{u}_i \mathbf{v}_i^H \quad (3.4.9)$$

可以看到, 用 SVD 定理可以确定数据矩阵 $\mathbf{A}$ 和 $\boldsymbol{\Phi} = \mathbf{A}^H \mathbf{A}$ 的秩为 W 。

2. 伪逆矩阵

由奇异值分解定理, 定义一个更一般的伪逆矩阵为

$$\mathbf{A}^+ = \mathbf{V} \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^H = \sum_{i=1}^W \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^H \quad (3.4.10)$$

对存在的以下三种情况分别讨论, 这里只叙述有关结论, 详细证明留作习题。

(1) 在过确定系统中 $N > M$, 若 $W = M$ 即 $(\mathbf{A}^H \mathbf{A})^{-1}$ 存在, 那么有

$$\mathbf{A}^+ = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \mathbf{V} [\mathbf{\Sigma}^{-1} \quad \mathbf{0}_{M \times (N-M)}] \mathbf{U}^H \quad (3.4.11)$$

注意, 上式中用 $\mathbf{0}$ 的下标更清晰地表示了这个全 0 矩阵的维数。

(2) 在欠确定系统中 $N < M$, 若 $W = N$, 即 $(\mathbf{A} \mathbf{A}^H)^{-1}$ 存在, 伪逆阵为

$$\mathbf{A}^+ = \mathbf{A}^H (\mathbf{A} \mathbf{A}^H)^{-1} = \mathbf{V} \begin{bmatrix} \mathbf{\Sigma}^{-1} \\ \mathbf{0}_{(M-N) \times N} \end{bmatrix} \mathbf{U}^H \quad (3.4.12)$$

(3) 在一般情况下, 即 $\mathbf{A}^H \mathbf{A}$ 或 $\mathbf{A} \mathbf{A}^H$ 的秩小于 $\min\{N, M\}$ 的情况下, $(\mathbf{A}^H \mathbf{A})^{-1}$ 或 $(\mathbf{A} \mathbf{A}^H)^{-1}$ 不存在, 由

$$\mathbf{A}^+ = \mathbf{V} \begin{bmatrix} \mathbf{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^H = \mathbf{V}_1 \mathbf{\Sigma}^{-1} \mathbf{U}_1^H = \sum_{i=1}^W \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^H \quad (3.4.13)$$

定义的伪逆矩阵得到的 LS 解

$$\mathbf{w}_{\text{LS}} = \mathbf{V} \begin{bmatrix} \mathbf{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^H \mathbf{d} = \mathbf{V}_1 \mathbf{\Sigma}^{-1} \mathbf{U}_1^H \mathbf{d} \quad (3.4.14)$$

是最小范数解, 即它一方面产生最小平均误差 $J_{\min}$, 另一方面, 它的滤波器系数的模是所有可能解中最小的。

进一步, 在过确定情况下, 将 $\mathbf{U}_1 = \mathbf{A} \mathbf{V}_1 \mathbf{\Sigma}^{-1}$ 代入式 (3.4.13) 和式 (3.4.14), 解进一步简化为

$$\mathbf{A}^+ = \mathbf{V} \begin{bmatrix} \mathbf{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^H = \mathbf{V}_1 \mathbf{\Sigma}^{-2} \mathbf{V}_1^H \mathbf{A}^H = \sum_{i=1}^W \frac{1}{\sigma_i^2} \mathbf{v}_i \mathbf{v}_i^H \mathbf{A}^H \quad (3.4.15)$$

和

$$\mathbf{w}_o = \sum_{i=1}^W \frac{\mathbf{v}_i}{\sigma_i^2} \mathbf{v}_i^H \mathbf{A}^H \mathbf{d} = \sum_{i=1}^W \frac{(\mathbf{v}_i^H \mathbf{A}^H \mathbf{d})}{\sigma_i^2} \mathbf{v}_i \quad (3.4.16)$$

在欠确定情况下, LS 滤波器权系数向量的解为

$$\mathbf{w}_o = \sum_{i=1}^W \frac{(\mathbf{u}_i^H \mathbf{d})}{\sigma_i^2} \mathbf{A}^H \mathbf{u}_i \quad (3.4.17)$$

注意在式 (3.4.16) 和式 (3.4.17) 中, 括号内的乘积是标量。在式 (3.4.16) 中只需要对 $\mathbf{A}^H \mathbf{A}$ 做特征分解, 在式 (3.4.17) 中只需要对 $\mathbf{A} \mathbf{A}^H$ 做特征分解。

LS 的 SVD 解有好的性质, 对过确定情况, 不管 $\mathbf{A}^H \mathbf{A}$ 是否满秩, 都可以用式 (3.4.16) 得到 LS 解, $\mathbf{A}^H \mathbf{A}$ 满秩时, LS 解是唯一的, 式 (3.4.16) 和式 (3.4.2) 在理论上是相等的, 在存在计算误差的情况下, SVD 解的数值稳定性良好。在 $\mathbf{A}^H \mathbf{A}$ 不满秩的情况下, $\mathbf{A}^H \mathbf{A}$ 的逆不存在, 式 (3.4.2) 不可直接使用, 这种情况下存在多个解都可以使 LS 误差达到最小 $J_{\min}$, 即解不是唯一的, 但在所有可能的解中, 模值最小的解是唯一的, 这就是 SVD 解。在 $\mathbf{A}^H \mathbf{A}$ 不满秩的情况下, 尽管存在多个解, 即有多种滤波器系数的取值都可以达到最小的误差, 但从数值稳定性的角度考虑, 模值最小的解一般是最好的。

* 3.5 总体最小二乘 (TLS)

本节仅简要叙述总体最小二乘 (Total Least Squares, TLS) 的定义, 并给出一个基于 SVD 的求解算法, 关于 TLS 的详细材料, 可参考文献 [258]。

LS方法求解 $\mathbf{Ax}=\mathbf{b}$ 的解,目标是求得一个解 $\mathbf{x}$ 使 $\|\mathbf{Ax}-\mathbf{b}\|^2$ 最小,在LS方法中,这可以理解为向量 $\mathbf{b}$ 被一个向量 $\Delta\mathbf{b}$ 所扰动,求一个范数最小的扰动向量 $\Delta\mathbf{b}$,即 $\min\{\|\Delta\mathbf{b}\|\}$,使解 $\mathbf{x}$ 满足方程式 $\mathbf{Ax}=\mathbf{b}+\Delta\mathbf{b}$,这里的 $\Delta\mathbf{b}$ 相当于3.4节中的误差 $\mathbf{e}$ 的另一种解释。将这个思路进一步扩展,假设系数矩阵 $\mathbf{A}$ 也被 $\Delta\mathbf{A}$ 扰动,这就是TLS问题。可以找到 $\Delta\mathbf{A}$ 和 $\Delta\mathbf{b}$,使得方程组

$$(\mathbf{A} + \Delta\mathbf{A})\mathbf{x} = \mathbf{b} + \Delta\mathbf{b} \quad (3.5.1)$$

有解,但使得式(3.5.1)有解的 $\Delta\mathbf{A}$ 和 $\Delta\mathbf{b}$ 可能有很多,为此,构成一个扩展矩阵

$$[\Delta\mathbf{A} \quad \Delta\mathbf{b}]$$

找到使式(3.5.1)有解且使范数 $\|[\Delta\mathbf{A} \quad \Delta\mathbf{b}]\|$ 最小的 $\Delta\mathbf{A}_0$ 和 $\Delta\mathbf{b}_0$,即TLS解 $\mathbf{x}_{\text{TLS}}$ 满足

$$(\mathbf{A} + \Delta\mathbf{A}_0)\mathbf{x}_{\text{TLS}} = \mathbf{b} + \Delta\mathbf{b}_0 \quad (3.5.2)$$

如下把问题稍加扩充,然后给出一种求TLS的方法。考虑更一般性的方程组的解,这里有

$$\mathbf{AX} = \mathbf{B} \quad (3.5.3)$$

设 $\mathbf{A}$ 是 $N \times M$ 矩阵, $\mathbf{B}$ 是 $N \times P$ 矩阵, $\mathbf{X}$ 是 $M \times P$ 个未知量组成的矩阵。方程组 $\mathbf{Ax}=\mathbf{b}$ 是 $P=1$ 的特例。式(3.5.3)的TLS解的定义如下。

定义 3.5.1 TLS解的定义是,找到范数最小的扰动矩阵 $\Delta\mathbf{A}_0$ 和 $\Delta\mathbf{B}_0$,使如下方程式成立:

$$(\mathbf{A} + \Delta\mathbf{A}_0)\mathbf{X}_{\text{TLS}} = \mathbf{B} + \Delta\mathbf{B}_0 \quad (3.5.4)$$

即 $\Delta\mathbf{A}_0$ 和 $\Delta\mathbf{B}_0$ 是使式(3.5.4)成立,且满足优化关系 $\min\{\|[\Delta\mathbf{A}, \Delta\mathbf{B}]\|\}$ 的矩阵。方程组(3.5.3)的TLS解 $\mathbf{X}_{\text{TLS}}$ 是针对这样的扰动矩阵 $\Delta\mathbf{A}_0$ 和 $\Delta\mathbf{B}_0$ 所构成的方程组(3.5.4)的解。

利用SVD可得到一种TLS解。讨论过确定性问题,即 $N > M + P$,对于扩展矩阵 $[\mathbf{A} \quad \mathbf{B}]$ 进行SVD分解表示为

$$\mathbf{U}^H [\mathbf{A} \quad \mathbf{B}] \mathbf{V} = \Sigma \quad (3.5.5)$$

对各矩阵写成分块形式:

$$\mathbf{U} = [\mathbf{U}_1 \quad \mathbf{U}_2] \quad (3.5.6)$$

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \quad (3.5.7)$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \quad (3.5.8)$$

最重要的是矩阵 $\mathbf{V}$ 的分块形式,其他矩阵的分块形式可与其对应。这里 $\mathbf{V}_{11}$ 是 $M \times M$ 矩阵, $\mathbf{V}_{12}$ 是 $M \times P$ 矩阵, $\mathbf{V}_{21}$ 是 $P \times M$ 矩阵, $\mathbf{V}_{22}$ 是 $P \times P$ 矩阵。如果 $\mathbf{V}_{22}^{-1}$ 存在,TLS解为

$$\mathbf{X}_{\text{TLS}} = -\mathbf{V}_{12}\mathbf{V}_{22}^{-1} \quad (3.5.9)$$

注意,在针对标准方程组和LS滤波器等常用情况时 $P=1$, $\mathbf{V}_{22}^{-1}$ 只是一个标量值,而 $\mathbf{V}_{12}$ 是一个列向量。

需要注意到,式(3.5.9)作为TLS的唯一解有一个前提条件。设以 $\mathbf{A}$ 作为数据矩阵所对应的第 M 个特征值为 σ_M ,以 $[\mathbf{A} \quad \mathbf{B}]$ 为数据矩阵的所对应的第 $M+1$ 个特征值为 $\sigma_{AB,M+1}$,若 $\sigma_M > \sigma_{AB,M+1}$,则式(3.5.9)是TLS的唯一解。关于式(3.5.9)的推导过程请参考文献[290]。

3.6 本章小结和进一步阅读

本章的核心内容是统计意义下最优的滤波器理论,这些结果构成了实际线性滤波器可能达到的最优结果。其中,在平稳条件下,最优滤波器就是维纳滤波器,在采用 FIR 结构时由维纳-霍普夫方程表示了维纳滤波器的解,而在 IIR 结构时,要求出的是滤波器的系统函数。最优线性预测可以看作维纳滤波器的一个特例,并且观察到前向线性最优预测与 AR 模型具有相同的解的方程,都可以采用 Levenson-Durbin 递推算法进行快速计算,Levenson-Durbin 算法也导出了滤波器的格型实现结构。

本章将 LS 滤波的概念限定在有限误差和最小意义下的最优滤波器。可以看到,在很多情况下,LS 是在存在有限数据的情况下,对维纳滤波器的逼近实现。可以证明,在用自相关方法构成数据矩阵时,对于最优预测或最优滤波的 LS 实现,相当于用观测的有限数据集,首先利用第 2 章给出的自相关和互相关的估计公式估计这些统计量,然后代入维纳-霍普夫方程所得到的解。很显然,在各态历经的情况下,当存在的数据量趋于无穷时,LS 解趋于维纳滤波器。

关于最优滤波器理论几乎所有涉及随机信号处理和统计信号处理的著作均有详细介绍,例如 Hayes(1996)^[114]、Haykin(2002)^[117] 和 Kailath(1990)^[135] 等,也可以参考维纳的原始论文^[271]。

存在大量方法用于解 LS 方程。用高斯消元和 Cholesky 分解可以求解正则方程,QR 分解、Givens 旋转、Householder 变换可以用于更加有效的解 LS 方程,奇异值分解也是有效和稳定的方法,但这些算法的细节,超出本书的范围,我们仅简要介绍了奇异值分解技术,对其他方法的详细讨论,请参考矩阵计算的专门著作,例如戈卢布(2001)^[290]。

习题

1. 有一个零均值信号 $x(n)$, 它的自相关序列的前两个值为 $r_x(0)=10, r_x(1)=5$, 该信号在传输中混入了一个均值为 0、方差为 5 的加性高斯白噪声, 该噪声与信号是不相关的, 设计一个 2 系数 FIR 型维纳滤波器, 使滤波器输出尽可能以均方意义逼近原信号 $x(n)$ 。

(1) 求滤波器系数。

(2) 滤波器输出与 $x(n)$ 之间均方误差。

2. 考虑如题图 3.1 所示的系统, 用于估计期望响应 $d(n)$, 其中零延时支路固定系数为 1, 求系数 w_1 使估计的均方误差最小, 并求此最小均方误差。已知参数为

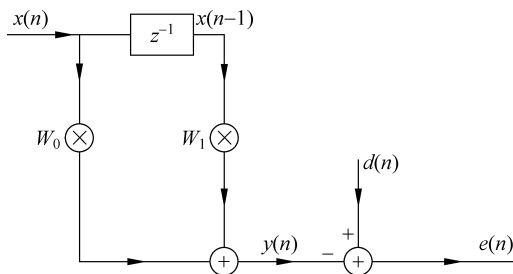
$$\sigma_d^2 = 4, \quad r_x(0) = 1.0, \quad r_x(1) = 0.5, \quad r_x(2) = 0.25, \\ r_{xd}(0) = -1.0, \quad r_{xd}(-1) = 1.0$$

3. 维纳滤波器的输入为 $x(n)=s(n)+v(n)$, $s(n)$ 和 $v(n)$ 相互独立, 设计一个非因果 IIR 滤波器, 希望尽可能降低噪声 $v(n)$, 证明

(1) 最优滤波器的输出均方误差可表示为

$$\epsilon = E[|e_0|^2] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_v(e^{j\omega}) W(e^{j\omega}) d\omega$$

式中, $S_v(e^{j\omega})$ 是噪声 $v(n)$ 的功率谱密度函数, $W(e^{j\omega})$ 是滤波器的频率响应。



题图 3.1

(2) 若 $v(n)$ 是方差为 σ_v^2 的白噪声, 证明上式简化为 $\epsilon = \sigma_v^2 w_{o0}$, 这里 w_{o0} 是最优滤波器 0 序号的权系数。

4. 源信号 $s(n)$ 是满足 $s(n) = \alpha s(n-1) + v(n)$ 的 AR(1) 信号, $v(n)$ 的方差为 σ_v^2 , 实际测量信号 $x(n) = s(n) + w(n)$ 混入了白噪声 $w(n)$, 其方差为 σ_w^2 , 设计一个维纳滤波器, 以 $x(n)$ 为输入, 输出尽可能近似 $s(n+K)$, K 是固定整数。

(1) 给出 2 系数 FIR 滤波器权系数的一般解的方程组。

(2) 在(1)中, 假设 $\alpha = 0.8$, $\sigma_w^2 = \sigma_v^2 = 1$, 分别求出 $K=0$ 和 $K=1$ 时滤波器的权系数和最小均方误差。

(3) 求出非因果 IIR 滤波器系统函数的表达式。

(4) 在(3)中, 假设 $\alpha = 0.8$, $\sigma_w^2 = \sigma_v^2 = 1$, 分别求出 $K=0$ 和 $K=1$ 时滤波器的权系数和最小均方误差。

5. 利用 FIR 结构的维纳滤波器, 实现一个陷波器, 即当输入为 $e^{j\omega_0 n}$ 时, 输出为 0, 输入为其他频率信号时, 无损通过滤波器(这是指理想要求, 实际不可实现, 实际中是近似无损), 导出这个最优滤波器系数 $\mathbf{W} = [w_0, w_1, \dots, w_{M-1}]^T$ 的设计公式(输入信号的 M 阶自相关矩阵用 $\mathbf{R}_x$ 表示)。

6. 计算
$$\left[\frac{\left(z + \frac{1}{3}\right)\left(z^{-1} + \frac{1}{3}\right)}{\left(z + \frac{1}{2}\right)\left(z^{-1} + \frac{1}{2}\right)} \right]_+$$
。

7. 证明前向最优线性预测器和后向最优线性预测器系数之间满足: $w_b^{B*} = w_f$, 并且满足前向最优线性预测误差与后向最优线性预测误差的均方值相等。

8. 一个平稳随机信号的自相关序列为 $r_x(k) = \frac{1}{2}\delta(k+1) + \frac{5}{4}\delta(k) + \frac{1}{2}\delta(k-1)$ 。

(1) 设计该信号的 1 阶、2 阶最优前向一步预测器。

(2) 利用 Levinson-Durbin 算法, 计算 3 阶前向预测误差滤波器对应的反射系数, 画出格型结构图。

9. 如题图 3.2 所示两个线性因果系统级联, 其中, $H_1(z) = \frac{1}{1-0.8z^{-1}}$, $H_2(z) = \frac{1}{1-0.5z^{-1}}$, 输入 $w(n)$ 是



题图 3.2

白噪声,均值为0,方差 $\sigma_w^2=1$,系统输出是 $x(n)$ 。

(1) 求 $x(n)$ 的自相关序列值 $r_x(0)$ 、 $r_x(1)$ 和 $r_x(2)$;

(2) 如果对 $x(n)$ 进行最优预测,设计 $x(n)$ 的最优格型预测误差滤波器,求出格型预测误差滤波器的合适的阶数、各反射系数,画出实现结构图。

10. 设由输入信号 $x(n)$ 通过一个非因果 IIR 维纳滤波器对期望响应 $d(n)$ 进行估计,估计误差 $e(n)$,证明:估计误差功率可由下式计算:

$$E\{|e(n)|^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - |C_{dx}(\omega)|^2) S_d(\omega) d\omega$$

其中, $S_d(\omega)$ 是期望响应的功率谱密度, $C_{dx}(\omega)$ 称为 $d(n)$ 和 $x(n)$ 的复相关,定义为下式:

$$C_{dx}(\omega) = \frac{S_{dx}(\omega)}{[S_x(\omega)S_d(\omega)]^{1/2}}$$

11. Levinson-Durbin 递推算法中,定义了标量: $\Delta_{m-1} = \sum_{k=0}^{m-1} r(k-m)a_{m-1,k}$,这里 $a_{m-1,k}$ 是 $m-1$ 阶前向线性最优预测滤波器的第 k 个权系数,用 $f_{m-1}(n)$ 和 $b_{m-1}(n)$ 分别表示前向和后向线性最优预测滤波器的预测误差,信号和滤波器系数都是实的,证明:

$$\Delta_{m-1} = E\{b_{m-1}(n-1)f_{m-1}(n)\}$$

12. 用 Levinson-Durbin 算法证明:一个 $AR(p)$ 信号输入到格型预测误差滤波器,仅有 p 个非零的反射系数,阶次高于 p 的反射系数均为零。

13. 证明:前向预测误差滤波器和后向预测误差滤波器输出满足如下的关系式。

$$(1) E[b_m(n)b_i^*(n)] = \begin{cases} p_m, & i=m \\ 0, & i \neq m \end{cases}$$

$$(2) E[f_m(n)f_i^*(n)] = p_l, l = \max(m, i)$$

$$(3) E[f_i(n)f_j^*(n-k)] = 0, \begin{cases} 1 \leq k \leq i-j, & i > j \\ -1 \geq k \geq i-j, & i < j \end{cases}$$

$$E[b_i(n)b_j^*(n-k)] = 0, \begin{cases} 0 \leq k \leq i-j-1, & i > j \\ 0 \geq k \geq i-j+1, & i < j \end{cases}$$

14. 证明:前向预测误差滤波器系统函数和后向预测误差滤波器系统函数满足如下关系:

$$H_{f,m}(z) = H_{f,m-1}(z) + k_m^* z^{-1} H_{b,m-1}(z)$$

15. 由测量的数据 $\{x(1), x(2), \dots, x(N)\}$,设计 M 阶前向一步 LS 线性预测器,分别按协方差方法和自相关方法使用数据,写出其数据矩阵、期望响应向量。

16. 如果要求设计后向一步 LS 线性预测器,重做上题。

17. 设有记录数据 $\{x(1), x(2), \dots, x(6)\}$,用3阶前向线性预测,采用协方差方法,写出求解预测系数的正则方程,确定 Φ 、 z 的各项。

18. 用自相关方法重解上题,比较两种方法得到的 Φ 、 z 各项的异同,并检查哪种方法得到的 Φ 满足 Toeplitz 性。

19. 用奇异值分解方法重新求解例题 3.3.1。

20. 证明:在过确定系统中($N > M$),若 $W=M$ 即 $(A^H A)^{-1}$ 存在,那么有

$$\mathbf{A}^+ = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \mathbf{V} [\mathbf{\Sigma}^{-1} \quad \mathbf{0}_{M \times (K-M)}] \mathbf{U}^H$$

* 21. 一个随机信号是由一个 AR(4)模型产生的,模型的极点分别位于 $p_{1,2} = 0.8e^{\pm j\pi/6}$, $p_{3,4} = 0.9e^{\pm j5\pi/8}$,激励白噪声方差为 1。用此模型产生 100 个数据,作为数据集,设计一个 LS 前向预测器,设预测器阶数为 4,分别使用自相关方法和协方差方法构成数据矩阵,估计预测系数并与 AR 模型的实际参数进行比较。

* 22. 一个随机信号是由一个 ARMA(4,2)模型产生的,模型极点位于 $p_{1,2} = 0.8e^{\pm j\pi/6}$, $p_{3,4} = 0.9e^{\pm j5\pi/8}$,零点位于 $z_{1,2} = 0.8e^{\pm j\pi/4}$,激励白噪声方差为 1。用此模型产生 200 个数据,作为数据集,设计一个 LS 前向预测器,设预测器阶数为 8,使用协方差方法构成数据矩阵,计算预测系数和平均估计误差。另设阶数为 4 和 12,重做此题,比较平均估计误差的大小。