缘起

近一年来做信号处理，感觉数学公式的推导非常重要。公式只要理解深刻，后续做matlab，跑数据，写c程序，上产品都有了提纲挈领之感。而信号处理中用的最多的自适应滤波器，它的公式推导过程中很多都需要用到复数或复矩阵的求导，我用的信号处理的书默认大家都有了很好的数学基础，这方面写得语焉不详，尤其喜欢在证明中间忽略一两步，看的人非常郁闷。于是最近花了一些业余时间搜集了一些资料好好学习了一下，这篇文章就是学习过程中的一些总结，希望能对信号处理的初学者有所帮助。

复数的求导

解析函数

连续

如果极限 $\lim\limits_{\Delta z \to 0} f(z_0 + \Delta z) = f(z_0)$ 存在，则说$f(z)$在z_0连续。

可导

设函数$w = f(z)$定义于区域D。z_0是D中的一点，点$z_0 + \Delta z$不出D的范围。如果极限

$$\lim_{\Delta z \to 0} \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z}$$

存在，则认为该函数在z_0可导。这个极限值就是$f(z)$在z_0处的导数。
注意，这个极限的下标$\Delta z \to 0$的方式有很多种。 因为z是一个复数，我们可以想象它是复平面上的一点，那么它逼近原点的方式是多种多样的。

可导的充要条件

设函数$f(z) = u(x, y) + iv(x, y)$定义在区域$D$内，则$f(z)$在$D$内一点$z = x + iy$可导的充要条件是：$u(x, y)$与$v(x, y)$在点$(x, y)$可微，并且满足**柯西-黎曼**方程:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

解析

如果函数$f(z)$在z_0及z_0的邻域内处处可导，那么称$f(z)$在z_0解析，如果$f(z)$在区域D内每一点解析，则称$f(z)$在D内解析，或称$f(z)$是D内的一个解析函数，也叫全纯函数或者正则函数。

复变函数教科书里对导数的定义到此为止了。

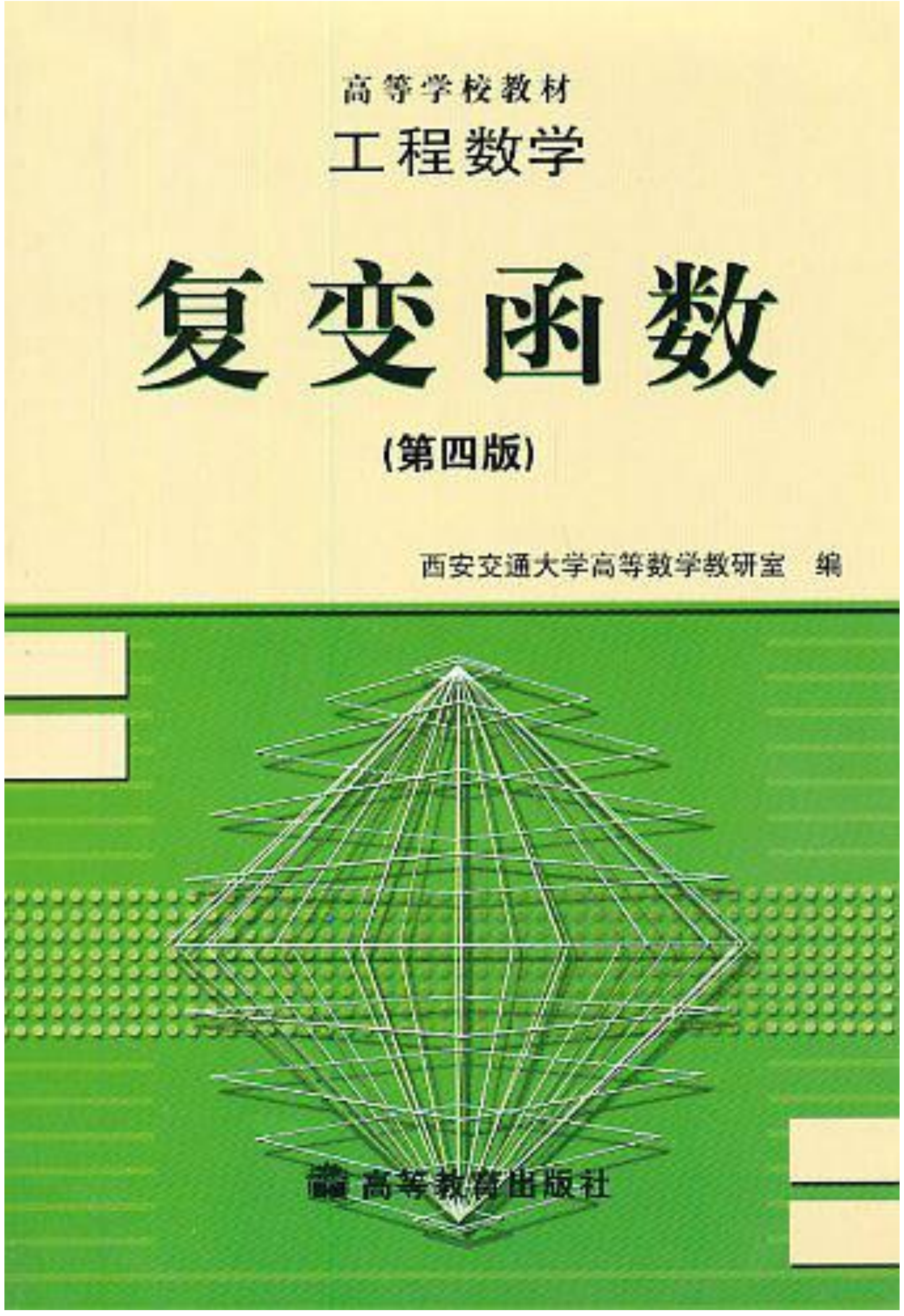

图1 作者的本科教科书

非解析函数

但是，在信号处理领域，需要求导的函数往往是不解析的。

比如自适应滤波器中的损失函数：

均方误差(MSE)

$$Cost = E[|e(k)^2|]$$

最小二乘(LS)

$$Cost = \frac{1}{k+1}\sum_{i=0}^{k}|e(k-i)|^2$$

加权最小二乘(WLS)

$$Cost = \sum_{i=0}^{k}\lambda^i|e(k-i)|^2, \qquad \lambda < 1$$

瞬时值(LSV)

$$Cost = |e(k)|^2$$

我们发现，若用复数表示$e(k)$，所有的损失函数都可以表示为$|e(k)|^2 = zz^*$的和的形式。
对于最优化问题，需要对损失函数求导，但从严格的数学角度来看，这个函数是不可导的。证明如下：

$$zz^* = x^2 + y^2$$

则：

$$\frac{\partial u}{\partial x} = 2x$$

$$\frac{\partial v}{\partial x} = 0$$

显然不满足柯西-黎曼方程，因此不可导。

那最优化问题如何求解?

非解析函数的求导

假设$f(z)$解析，则$f(z)$可以展开为z的泰勒级数的形式，在这个展开式中没有z^*的身影，说明f(z)和z^*没有任何关系。换句话说，一个解析的复变函数只跟z有关，和z^*无关。

实际上，$f(z)$解析可以**等价**为$\frac{\partial f}{\partial z^*}=0$。

那么，不解析的函数就同时依赖于z和z^*。

可以轻松推出一个事实：**实函数$f(z)$都是不解析的。**

因为它必须同时依赖于z和z^*，否则它的虚部无法被消掉。因此$f(z)=zz^*$作为一个实函数，也印证了上面用柯西-黎曼方程推导出不解析的结果。

所以，一个不解析的复变函数由于同时跟z, z^*以及x, y有关，可以写成：

$f(z)=f(z,z^*), \qquad f(z)=f(x,y)$

这两种形式。根据全微分公式：

$$df=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy$$

因此：

$$df=\frac{\partial f}{\partial x}\frac{dz+dz^*}{2}+\frac{\partial f}{\partial y}\frac{dz-dz^*}{2i}$$

$$=\frac{1}{2}(\frac{\partial f}{\partial x}-i\frac{\partial f}{\partial y})dz+\frac{1}{2}(\frac{\partial f}{\partial x}+i\frac{\partial f}{\partial y})dz^*$$

而根据全微分公式又同样可得：

$$df=\frac{\partial f}{\partial z}dz+\frac{\partial f}{\partial z^*}dz^*$$

因此：

$$\frac{\partial f}{\partial z}=\frac{1}{2}(\frac{\partial f}{\partial x}-i\frac{\partial f}{\partial y})$$

$$\frac{\partial f}{\partial z^*}=\frac{1}{2}(\frac{\partial f}{\partial x}+i\frac{\partial f}{\partial y})$$

同样的方式又易得：

$$\frac{\partial f}{\partial x}=\frac{\partial f}{\partial z}+\frac{\partial f}{\partial z^*}$$

$$\frac{\partial f}{\partial y}=i(\frac{\partial f}{\partial z}-\frac{\partial f}{\partial z^*})$$

根据上面的公式，得到推论：

$$\frac{\partial z}{\partial z}=1$$

$$\frac{\partial z^*}{\partial z^*} = 1$$

$$\frac{\partial z}{\partial z^*} = 0$$

$$\frac{\partial z^*}{\partial z} = 0$$

$$\frac{\partial (zz^*)}{\partial z} = z^*$$

$$\frac{\partial (zz^*)}{\partial z^*} = z$$

向量和矩阵的求导

当一个函数有很多变量，需要对多变量求导时，为了简化运算，把变量组成向量或者矩阵，作为一个实体来看待，这就是对向量和矩阵的求导。
可能存在的求导有标量，向量，矩阵分别对各自求导，根据排列组合一共有9种情况，比如“标量对标量的求导，向量对标量的求导”等等。要想详细了解可参考:
https://en.wikipedia.org/wiki/Matrix_calculus

这里我们只看最常用的**标量对矢量的求导**。
假设$f(x)$是一个标量，x是一个矢量，x可表示为

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

则:

$$\frac{\partial f(x)}{\partial \mathbf{x}} = \begin{bmatrix} \frac{f(x)}{x_1} \\ \frac{f(x)}{x_2} \\ \vdots \\ \frac{f(x)}{x_N} \end{bmatrix}$$

为什么是这样呢？答案是全微分公式：

$$df = \frac{f(x)}{x_1}dx_1 + \frac{f(x)}{x_2}dx_2 + \cdots + \frac{f(x)}{x_N}dx_N$$

写成向量的乘法就是：

$$df = (\begin{bmatrix} \frac{f(x)}{x_1} \\ \frac{f(x)}{x_2} \\ \vdots \\ \frac{f(x)}{x_N} \end{bmatrix})^T d\mathbf{x}$$

根据这个定义，易得：

$$\frac{\partial \mathbf{x}^T \mathbf{b}}{\partial \mathbf{x}} = \mathbf{b}$$

$$\frac{\partial \mathbf{b}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{b}$$

复数向量和矩阵求导

标量和向量都可以认为是一种简化的矩阵，所以一个函数对标量，向量和矩阵的求导都可以统一写为复数对矩阵求导的形式。

假设函数F的参数是一个Q行，N列的复数矩阵（对于标量，它就是一个1行1列的矩阵，对于向量，它就是一个Q行1列的矩阵），则可以认为F是Q*N个复数参数的函数，而每一个复数又包含x，y或者z，z^*两个变量，因此根据全微分公式：

$$df = \sum_{k=0}^{Q-1}\sum_{l=0}^{N-1}\frac{\partial f}{\partial x_{k,l}}dx_{k,l} + \sum_{k=0}^{Q-1}\sum_{l=0}^{N-1}\frac{\partial f}{\partial y_{k,l}}dy_{k,l}$$

$$= \sum_{k=0}^{Q-1} \sum_{l=0}^{N-1} \frac{\partial f}{\partial z_{k,l}} dz_{k,l} + \sum_{k=0}^{Q-1} \sum_{l=0}^{N-1} \frac{\partial f}{\partial z_{k,l}^*} dz_{k,l}^*$$

所以，复数导数F可以看成是矩阵Z和矩阵Z^*的函数，即$F(\mathbf{Z}, \mathbf{Z}^*)$。

我们回忆一下微分的基本原理，对一个函数$f(x)$微分：

$$f(x + dx) - f(x) = FirstOrder(dx) + HighOrder(dx)$$

FirstOrder指的是一阶的dx，HighOrder指的是高阶的dx，微分的计算通常是扔掉高阶的dx，保留一阶的dx，即：

$$df = FirstOrder(dx)$$

对$F(\mathbf{Z}, \mathbf{Z}^*)$求导是一样的道理：

$$F(Z + dZ, Z^* + dZ^*) - F(Z, Z^*)$$

$$= FirstOrder(dZ, dZ^*) + HighOrder(dZ, dZ^*)$$

$$dF(Z, Z^*) = FirstOrder(dZ, dZ^*)$$

由此推导出一些结论：

$$d(\mathbf{AZB}) = \mathbf{A}(d\mathbf{Z})\mathbf{B}$$

$$d(a\mathbf{Z}) = ad\mathbf{Z}$$

$$d\mathbf{Z}_0\mathbf{Z}_1 = (d\mathbf{Z}_0)\mathbf{Z}_1 + \mathbf{Z}_0 d\mathbf{Z}_1$$

$$d\mathbf{Z}^{-1} = -\mathbf{Z}^{-1}(d\mathbf{Z})\mathbf{Z}^{-1}$$

$$d\mathbf{Z}^* = (d\mathbf{Z})^*$$

$$d\mathbf{Z}^H = (d\mathbf{Z})^H$$

还有对信号处理非常重要的一些结论，均是标量函数对复向量的求导，为了看起来比较清楚，就不写成矩阵的形式了。

$$dF(\mathbf{a}^T\mathbf{z}) = \mathbf{a}d\mathbf{z}$$

$$dF(\mathbf{z}^T\mathbf{a}) = \mathbf{a}d\mathbf{z}$$

注意$\mathbf{a}^T\mathbf{z} = \mathbf{z}^T\mathbf{a}$

$$dF(\mathbf{a}^T\mathbf{z}^*) = \mathbf{a}d\mathbf{z}^*$$

$$dF(\mathbf{z}^H\mathbf{a}) = \mathbf{a}d\mathbf{z}^*$$

注意$\mathbf{a}^T\mathbf{z}^* = \mathbf{z}^H\mathbf{a}$

$$dF(\mathbf{z}^T\mathbf{A}\mathbf{z}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{z}d\mathbf{z}$$

$$dF(\mathbf{z}^H\mathbf{A}\mathbf{z}) = \mathbf{A}^T\mathbf{z}^*d\mathbf{z} + \mathbf{A}\mathbf{z}d\mathbf{z}^*$$

$$dF(\mathbf{z}^H\mathbf{A}\mathbf{z}^*) = (\mathbf{A} + \mathbf{A}^T)\mathbf{z}^*d\mathbf{z}^*$$

如果A是广义平稳过程的相关矩阵，那么它是一个埃尔米特矩阵，即$A^H = A$，因此上面三个式子可以写成：

$$dF(\mathbf{z}^T\mathbf{A}\mathbf{z}) = 2\mathbf{A}\mathbf{z}d\mathbf{z}$$

$$dF(\mathbf{z}^H\mathbf{A}\mathbf{z}) = \mathbf{A}\mathbf{z}^*d\mathbf{z} + \mathbf{A}\mathbf{z}d\mathbf{z}^*$$

$$dF(\mathbf{z}^H\mathbf{A}\mathbf{z}^*) = 2\mathbf{A}\mathbf{z}^*d\mathbf{z}^*$$

我们把$\mathbf{z}^H\mathbf{A}\mathbf{z}$称为**二次型**（quadratic form），它的求导在DSP的公式推导中非常重要。

应用

1. 维纳滤波（Wiener Filter）

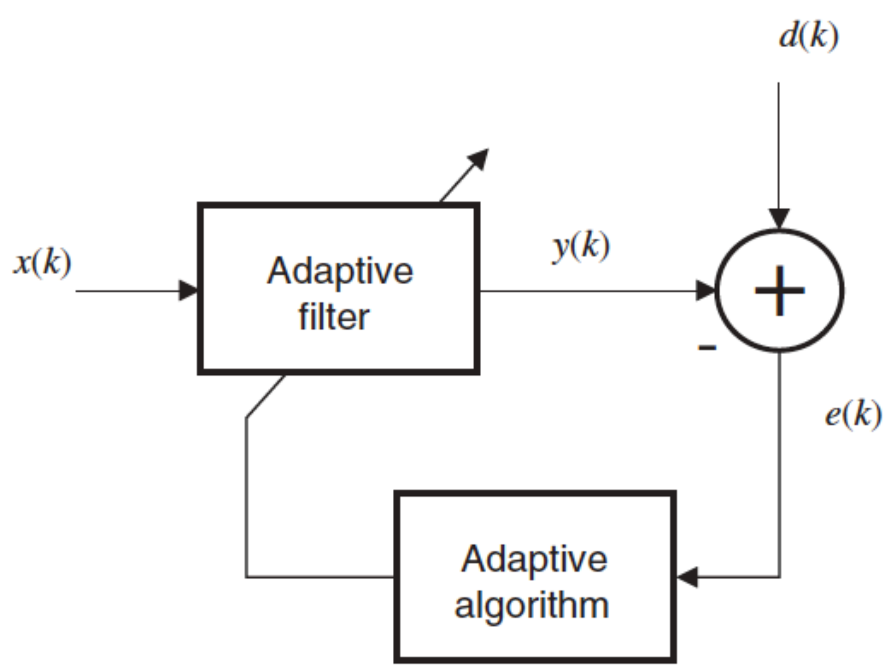

图2 自适应滤波问题

维纳滤波以 $MSE = E(|e(k)^2|)$ 为损失函数。

$e(k) = d(k) - \mathbf{w}^H\mathbf{x}$

$|e(k)|^2 = (d(k) - \mathbf{w}^H\mathbf{x})^H(d(k) - \mathbf{w}^H\mathbf{x})$

$= (d^H(k) - \mathbf{x}^H\mathbf{w})(d(k) - \mathbf{w}^H\mathbf{x})$

$= |d(k)|^2 - d^H(k)\mathbf{w}^H\mathbf{x} - d(k)\mathbf{x}^H\mathbf{w} + \mathbf{x}^H\mathbf{w}\mathbf{w}^H\mathbf{x}$

因为$\mathbf{x}^H\mathbf{w}$和$\mathbf{w}^H\mathbf{x}$都是标量，所以：

$= |d(k)|^2 - d^H(k)\mathbf{w}^H\mathbf{x} - d(k)\mathbf{x}^H\mathbf{w} + \mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}$

根据二次型的求导公式：

$$\frac{\partial |e(k)|^2}{\partial w^*} = -d^H\mathbf{x} + \mathbf{x}\mathbf{x}^H\mathbf{w}$$

加上期望符号E

$$\frac{\partial E(|e(k)|^2)}{\partial w^*} = -E(d^H\mathbf{x}) + E(\mathbf{x}\mathbf{x}^H)\mathbf{w}$$

而$E[xx^H] = R_{xx}$

设$-E(d^H\mathbf{x}) = P$，

则$E(|e(k)|^2)$最小时，$\frac{\partial E(|e(k)|^2)}{\partial w^*} = 0$，

可得

$-\mathbf{P} + \mathbf{R}\mathbf{w} = 0$

$\mathbf{w} = \mathbf{R}^{-1}\mathbf{P}$

2. LMS（Least Mean Square）

仍然使用图2中的符号。

MSE这种损失函数实际上是无法求得的，因为求了期望，就要求对信号进行无限次测量。为了能获得维纳解，人们设计了一些方式来进行迭代求解，也就是在维纳的解空间内进行搜索，逐渐逼近维纳解。比如最陡梯度下降法。

意思就是：

$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\hat{\mathbf{g}}(\mathbf{k})$

$\hat{g}(k)$就是第k次迭代的梯度。

我们利用维纳对$|e(k)|^2$求导的结果，可得梯度为$-d^H\mathbf{x} + \mathbf{x}\mathbf{x}^H\mathbf{w}$

利用最近一次的$\mathbf{x}(k)\mathbf{x}^H(k)$替换掉$E[xx^H]$，注意，$\mathbf{x}$也是向量。

可得：

$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu(-d^H(k)\mathbf{x}(k) + \mathbf{x}(k)\mathbf{x}(k)^H\mathbf{w}(k))$

$= \mathbf{w}(k) + \mu \mathbf{x}(k)(d^H(k) - \mathbf{x}(k)^H \mathbf{w}(k))$

注意 $e(k) = d(k) - \mathbf{w}(k)^H \mathbf{x}(k)$，两边取共轭转置，可得$e(k)^H = d^H(k) - \mathbf{x}^H(k)\mathbf{w}(k)$

代入上式，可得$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \mathbf{x}(k) e(k)^*$，

注意$e(k)$是个标量，共轭转置等于共轭。

3. APA（Affine Projection Algorithm）

仍然使用图2中的符号。

APA是LMS的一种变形。

首先定义

$\mathbf{e}_{ap}(k) = \mathbf{d}_{ap}(k) - \mathbf{X}_{ap}^T \mathbf{w}^*(k)$

迭代用的是滤波器系数的共轭，这就意味着我们需要对滤波器系数的共轭求导。其实这也是一个trick，因为对一个复数的共轭求导会得到关于这个复数的方程。

注意，上面的式子全是向量和矩阵，没有标量。

APA依然是一个带约束的最优化问题，

约束条件是后验误差是0，即$\mathbf{d}_{ap}(k) - \mathbf{X}_{ap}^T(k)\mathbf{w}^*(k+1) = \mathbf{0}$，$\mathbf{w}^*(k+1)$是更新后的滤波器系数。

最优化目标是使得这个值$\frac{1}{2}||\mathbf{w}(k+1) - w(k)||^2$最小化，这被称为最小距离原理。

采用拉格朗日乘子法：

$\frac{1}{2}(\mathbf{w}(k+1) - \mathbf{w}(k))^H(\mathbf{w}(k+1) - \mathbf{w}(k)) + \lambda^T(\mathbf{d}_{ap}(k) - \mathbf{X}_{ap}^T(k)\mathbf{w}^*(k+1))$

注意，λ是一个实数的列向量，转置是为了相乘后得到一个标量。

上面的式子是一个标量，为了得到关于$\mathbf{w}(k)$的式子，对向量$\mathbf{w}^*(k+1)$求导，求导可得：

$\frac{1}{2}(\mathbf{w}(k+1) - \mathbf{w}(k)) - (\lambda^T \mathbf{X}_{ap}^T(k))^T$

最优解是上式为0，则$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mathbf{X}_{ap}(k)\lambda$　(3.1)

对约束条件两边取共轭，可得

$\mathbf{d}_{ap}^*(k) - \mathbf{X}_{ap}^H(k)\mathbf{w}(k+1) = 0$

把$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mathbf{X}_{ap}(k)\lambda$ 代入上式，可得

$\mathbf{d}_{ap}^*(k) - \mathbf{X}_{ap}^H(k)\mathbf{w}(k) - 2\mathbf{X}_{ap}^H(k)\mathbf{X}_{ap}(k)\lambda = \mathbf{0}$

$\mathbf{d}_{ap}^*(k) - \mathbf{X}_{ap}^H(k)\mathbf{w}(k) = 2\mathbf{X}_{ap}^H(k)\mathbf{X}_{ap}(k)\lambda$ (3.2)

由于

$\mathbf{e}_{ap}(k) = \mathbf{d}_{ap}(k) - \mathbf{X}_{ap}^T\mathbf{w}^*(k)$

两边取共轭

$\mathbf{e}_{ap}^*(k) = \mathbf{d}_{ap}^*(k) - \mathbf{X}_{ap}^H\mathbf{w}(k)$

带入（3.2）

则$\mathbf{e}_{ap}^*(k) = 2\mathbf{X}_{ap}^H(k)\mathbf{X}_{ap}(k)\lambda$

$\lambda = \frac{1}{2}(\mathbf{X}_{ap}^H(k)\mathbf{X}_{ap}(k))^{-1}\mathbf{e}_{ap}^*(k)$

带入(3.1)

$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\mathbf{X}_{ap}(k)(\mathbf{X}_{ap}^H(k)\mathbf{X}_{ap}(k))^{-1}\mathbf{e}_{ap}^*(k)$

4. LCMV（Linearly Constrained Minimum Variance）

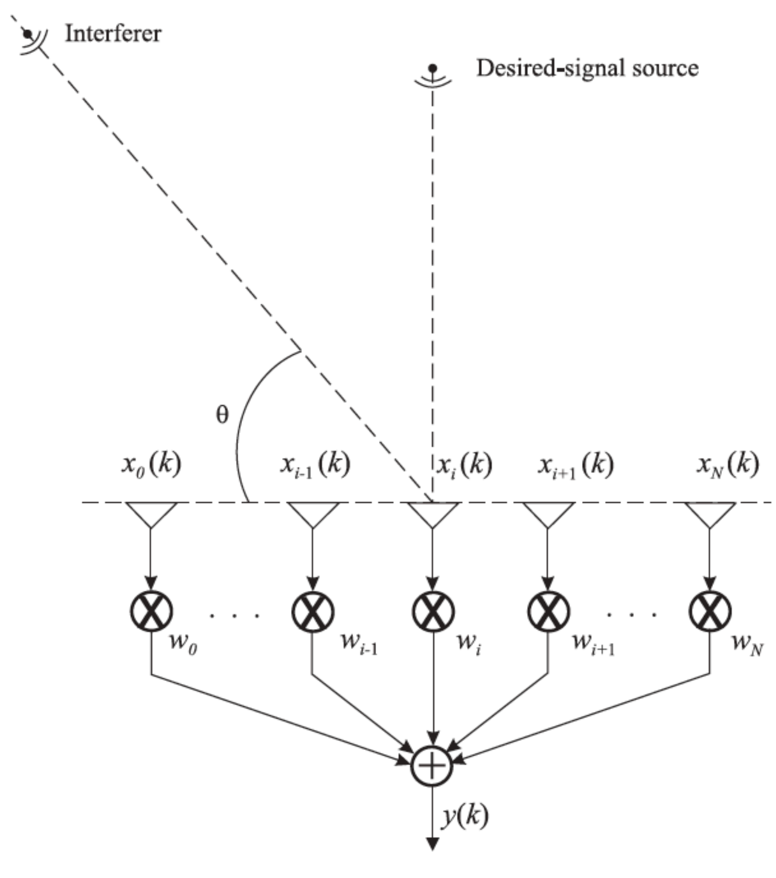

图3 波束形成问题

线性约束最小方差的意思是：在满足约束条件的情况下，让输出信号的功率最小。这个约束条件就是某一方向的信号不衰减。

某方向信号不衰减可以表示为$\mathbf{w}^H\mathbf{c}=f$，推导如下：

整个推导都是在频域，输出是输入信号和滤波器的点乘，假设$\mathbf{y}=\mathbf{w}^*\cdot\mathbf{x}$是滤波器的输出，

则$\sum_{i=0}^{N} y_i=\mathbf{w}^H\mathbf{x}$。

$x_i(k)=x_0(t_0+i\frac{dcos\theta}{c})$

当$\mathbf{w}^H\mathbf{x}=\mathbf{x_0}$

$\mathbf{x}_0=[x_0x_0\cdots x_0]^T$

而$\mathbf{x}=x_0\cdot[1,e^{-jw\frac{dcons\theta}{c}},e^{-j2w\frac{dcons\theta}{c}},\cdots,e^{-j(N-1)w\frac{dcons\theta}{c}}]^T$

这个可以当做约束条件。

从而又变成了带约束条件的最优化问题，可得式子为：

$E[\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}]+\lambda(\mathbf{w}^H\mathbf{c}-f)$

注意，$\mathbf{c}$就是steering vector，而λ是一个实数标量。

在约束情况下求最优化问题，还是用拉尔朗日乘子法。

$E[\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}]+\lambda(\mathbf{w}^H\mathbf{c}-f)$

上式对$\mathbf{w}^*$求导，可得：

$\mathbf{R}_{xx}\mathbf{w}+\lambda\mathbf{c}$

令上式为0，则$\mathbf{w}=-\lambda\mathbf{R}^{-1}\mathbf{c}$ （4.1）

代入$\mathbf{w}^H\mathbf{c}=f$，因为是标量，两边取转置不变，这个式子可以写成$\mathbf{c}^H\mathbf{w}=f$， 代入 （4.1）

$-\lambda\mathbf{c}^H\mathbf{R}^{-1}\mathbf{c}=f$， 则

$\lambda=-(\mathbf{c}^H\mathbf{R}^{-1}\mathbf{c})^{-1}f$

最后代入式子 (4.1)

$\mathbf{w}=\mathbf{R}^{-1}\mathbf{c}(\mathbf{c}^H\mathbf{R}^{-1}\mathbf{c})^{-1}f$

5. MVDR（Minimum Variance Distortionless Response）

仍然使用图3中的符号。

输入信号是：

$\mathbf{x}=\mathbf{s}+\mathbf{i}+\mathbf{n}$

其中$\mathbf{s}$是纯净语音信号，$\mathbf{i}$是干扰信号，$\mathbf{n}$是噪音信号。

目标是要求得一个滤波器系数$\mathbf{w}$，在约束条件$\mathbf{w}^H\mathbf{a}=1$下，使得SINR（信干燥比）最大。

$$SINR = E[\frac{|\mathbf{w}^H\mathbf{s}|^2}{|\mathbf{w}^H(\mathbf{i}+\mathbf{n})|^2}]$$

E是求期望。

我们希望分母最小，当然还是选择拉格朗日乘子法：

$E[\mathbf{w}^H\mathbf{R}_{i+n}\mathbf{w}]+\lambda(\mathbf{w}^H\mathbf{a}-1)$，$\lambda$是一个实数标量。

对$\mathbf{w}^*$求导，则$\mathrm{g}_w=\mathbf{R}_{i+n}\mathbf{w}+\lambda\mathbf{a}$

类似LCMV，代入$\mathbf{a}^H\mathbf{w}=1$

$-\lambda\mathbf{a}^H\mathbf{R}_{i+n}^{-1}\mathbf{a}=1$

$$\mathbf{w}=\frac{\mathbf{R}_{i+n}^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{R}_{i+n}^{-1}\mathbf{a}}$$