统计信号处理

第五章

最大似然估计

清华大学电子工程系
李洪 副教授

2023.3

内容概要

- 一、引言
- 二、最大似然估计（MLE）
- 三、MLE的性质
- 四、数值解法
- 五、矢量参数的MLE
- 六、MLE的应用
- 七、小结

一、引言

例：白噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\ldots,N-1$$

待估计参数为信号幅度 $A(A>0)$，$w[n]$ 为高斯白噪声，且其方差为 A，即 $w[n]\sim N(0,A)$。求 A 的MVU估计量？

1. 方法一：CRLB

$$p(\mathbf{x};A)=\frac{1}{(2\pi A)^{N/2}}\exp\left\{-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2\right\}$$

$$\ln p(\mathbf{x};A)=-\frac{N}{2}\ln(2\pi A)-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2$$

$$\frac{\partial \ln p(\mathbf{x};A)}{\partial A}=-\frac{N}{2A}+\frac{1}{A}\sum_{n=0}^{N-1}(x[n]-A)+\frac{1}{2A^2}\sum_{n=0}^{N-1}(x[n]-A)^2$$

$$\frac{\partial^2 \ln p(\mathbf{x};A)}{\partial A^2}=\frac{N}{2A^2}-\frac{1}{A^2}\sum_{n=0}^{N-1}(x[n]-A)-\frac{N}{A}-\frac{1}{A^3}\sum_{n=0}^{N-1}(x[n]-A)^2-\frac{1}{2A^2}\sum_{n=0}^{N-1}2(x[n]-A)$$

$$\frac{\partial^2 \ln p(\mathbf{x};A)}{\partial A^2} = \frac{N}{2A^2} - \frac{1}{A^2}\sum_{n=0}^{N-1}(x[n]-A) - \frac{N}{A} - \frac{1}{A^3}\sum_{n=0}^{N-1}(x[n]-A)^2 - \frac{1}{2A^2}\sum_{n=0}^{N-1}2(x[n]-A)$$

$$\Rightarrow E\left(\frac{\partial^2 \ln p(\mathbf{x};A)}{\partial A^2}\right) = \frac{N}{2A^2} - \frac{N}{A} - \frac{1}{A^3}NA = -\frac{N+2NA}{2A^2}$$

$$\Rightarrow \operatorname{var}(\hat{A}) \geq \frac{1}{-E\left(\frac{\partial^2 \ln p(\mathbf{x};A)}{\partial A^2}\right)} = \frac{A^2}{N\left(A+\frac{1}{2}\right)}$$

CRLB!

$$\frac{\partial \ln p(\mathbf{x};A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A}\sum_{n=0}^{N-1}(x[n]-A) + \frac{1}{2A^2}\sum_{n=0}^{N-1}(x[n]-A)^2$$

$$\boxed{\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)\left(g(\mathbf{x})-\theta\right)}$$

$\Rightarrow g(\mathbf{x})$?

2. 方法二：充分统计量方法

$$p(\mathbf{x};\theta)=g\left(T(\mathbf{x}),\theta\right)h(\mathbf{x})$$

$$p(\mathbf{x};A)=\frac{1}{(2\pi A)^{N/2}}\exp\left\{-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2\right\}$$

$$=\frac{1}{(2\pi A)^{N/2}}\exp\left\{-\frac{1}{2A}\sum_{n=0}^{N-1}\left(x^2[n]+A^2-2x[n]A\right)\right\}$$

$$=\underbrace{\frac{1}{(2\pi A)^{N/2}}\exp\left\{-\frac{1}{2A}\sum_{n=0}^{N-1}x^2[n]-\frac{NA}{2}\right\}}_{g\left(\sum_{n=0}^{N-1}x^2[n],A\right)}\underbrace{\exp\left\{\sum_{n=0}^{N-1}x[n]\right\}}_{h(\mathbf{x})}$$

（1）构造无偏估计：

即构造 f，使得 $E\left(f\left(\sum_{n=0}^{N-1}x^2[n]\right)\right)=A$

$$E\left(\sum_{n=0}^{N-1}x^2[n]\right)=\sum_{n=0}^{N-1}E\left(x^2[n]\right)$$

$$=\sum_{n=0}^{N-1}\left(E^2(x[n])+\operatorname{var}(x[n])\right)=N\left(A^2+A\right)$$

$$f\left(N\left(A+A^2\right)\right)=A \qquad f\ ?$$

（2）条件期望：

$$\breve{A}=x[0]$$

$$E\left(x[0]\left|\sum_{n=0}^{N-1}x^2[n]\right.\right) \qquad ?$$

- 求解MVU方法有时比较难！
- 需另觅它途！

二、最大似然估计

- Maximum likelihood estimation (MLE)

对观测数据 $\mathbf{x}=\left[x[0],x[1],x[2],\ldots,x[N-1]\right]^T$，使 $p(\mathbf{x};\theta)$ 最大的 θ 值，即使得似然函数最大的 θ 值，称为 θ 的最大似然估计

$$\hat{\theta}=\arg\max_{\theta}\left\{p(\mathbf{x};\theta)\right\}$$

若可导

$$\begin{cases}\left.\dfrac{\partial p(\mathbf{x};\theta)}{\partial\theta}\right|_{\hat{\theta}}=0\\[2ex]\left.\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\hat{\theta}}=0\end{cases}$$

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}=I(\theta)\left(g(\mathbf{x})-\theta\right)$$

当有效估计量存在时，使用最大似然估计方法就可以求得

- **MLE Vs MVU**

MVU: $$\begin{cases}\min\left\{E\left(\left(\hat{\theta}-\theta\right)^2\right)\right\}\\ s.t. \quad E\left(\hat{\theta}\right)=\theta\end{cases}$$

Vs

MLE: $\hat{\theta}=\arg\max_{\theta}\left\{p(\mathbf{x};\theta)\right\}$

例：白噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\ldots,N-1$$

待估计参数为信号幅度 $A(A>0)$，$w[n]$ 为高斯白噪声，且其方差为 A，即 $w[n]\sim N(0,A)$。求 A 的MVU估计量？

$$p(\mathbf{x};A)=\frac{1}{(2\pi A)^{N/2}}\exp\left\{-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2\right\}$$

$$\ln p(\mathbf{x};A)=-\frac{N}{2}\ln(2\pi A)-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2$$

$$\left.\begin{aligned}&\frac{\partial\ln p(\mathbf{x};A)}{\partial A}=-\frac{N}{2A}+\frac{1}{A}\sum_{n=0}^{N-1}(x[n]-A)+\frac{1}{2A^2}\sum_{n=0}^{N-1}(x[n]-A)^2\\&\left.\frac{\partial\ln p(\mathbf{x};A)}{\partial A}\right|_{\hat{A}}=0\end{aligned}\right\}$$

$$\Rightarrow\ \hat{A}^2+\hat{A}-\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]=0$$

$$\Rightarrow\ \hat{A}=-\frac{1}{2}\pm\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]+\frac{1}{4}}\quad\Rightarrow\ \hat{A}=-\frac{1}{2}+\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]+\frac{1}{4}}$$

MLE!

$$\hat{A}=-\frac{1}{2}+\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]+\frac{1}{4}}$$

- **数学期望：**

$$E\left(\hat{A}\right)=-\frac{1}{2}+E\left(\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]+\frac{1}{4}}\right)\neq-\frac{1}{2}+\left(\sqrt{E\left(\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]\right)+\frac{1}{4}}\right)$$

$$=-\frac{1}{2}+\left(\sqrt{E\left(x^2[n]\right)+\frac{1}{4}}\right)$$

$$=-\frac{1}{2}+\left(\sqrt{A+A^2+\frac{1}{4}}\right)$$

$$=A$$ 有偏!

但，根据大数定理 $\frac{1}{N}\sum_{n=0}^{N-1}x^2[n] \longrightarrow E\left(x^2[n]\right)=A+A^2$

$$\Rightarrow \quad \hat{A}=-\frac{1}{2}+\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]+\frac{1}{4}} \longrightarrow -\frac{1}{2}+\sqrt{E\left(x^2[n]\right)+\frac{1}{4}}=A$$

$\hat{A}$为一致估计量!

当 $N\to\infty$ 时，$\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]$ 将集中在期望值 $A+A^2$ 附近

可对函数（A 的估计量）进行线性化

$$\hat{A}=-\frac{1}{2}+\sqrt{\underbrace{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]}_{u}+\frac{1}{4}}$$

$$g(u)=-\frac{1}{2}+\sqrt{u+\frac{1}{4}}$$

线性化处理：$g(u)\approx g(u_0)+\left.\frac{dg(u)}{du}\right|_{u_0}(u-u_0)$

$$u_0=A+A^2$$

$$\hat{A}\approx-\frac{1}{2}+\sqrt{A+A^2+\frac{1}{4}}+\frac{1}{2\left(A+\frac{1}{2}\right)}\left(\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]-A-A^2\right)$$

$$=A+\frac{1}{2\left(A+\frac{1}{2}\right)}\left(\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]-A-A^2\right)\Rightarrow E\left(\hat{A}\right)\to A$$

渐近无偏！

$$\hat{A} = -\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}$$

- **方差：**

$$\mathrm{var}\left(\hat{A}\right) = \mathrm{var}\left(-\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}\right)$$

线性化处理结果：$\hat{A} \approx A + \dfrac{1}{2\left(A+\frac{1}{2}\right)}\left(\dfrac{1}{N}\sum_{n=0}^{N-1} x^2[n] - A - A^2\right)$

$$\mathrm{var}\left(\hat{A}\right) \approx \left(\frac{1}{2\left(A+\frac{1}{2}\right)}\right)^2 \mathrm{var}\left(\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - A - A^2\right)$$

$$= \left(\frac{1}{2\left(A+\frac{1}{2}\right)}\right)^2 \frac{1}{N}\underline{\mathrm{var}\left(x^2[n]\right)}$$

$$= \left(\frac{1}{2\left(A+\frac{1}{2}\right)}\right)^2 \frac{1}{N}\left(2A^2 + 4A^3\right) = \frac{A^2}{N\left(A+\frac{1}{2}\right)}$$

若$x[n] \sim N\left(\mu, \sigma^2\right)$，则

$$E\left(x^2[n]\right) = \mu^2 + \sigma^2$$

$$E\left(x^4[n]\right) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

$$\mathrm{var}\left(x^2[n]\right) = E\left(x^4[n]\right) - \left(E\left(x^2[n]\right)\right)^2 = 2\sigma^4 + 4\mu^2\sigma^2$$

- 渐近达到了CRLB！
- 渐近有效!

三、MLE的性质

更一般地：

如果数据 $\mathbf{x}$ 的PDF $p(\mathbf{x};\theta)$ 满足“正则”条件，那么对于足够多的数据记录，未知参数 θ 的MLE渐近服从

$$\hat{\theta} \overset{a}{\sim} N\left(\theta, I^{-1}(\theta)\right)$$

其中 $I(\theta)$ 是在未知参数真值处计算的Fisher信息

MLE的渐近有效性

- MLE是渐近无偏的
- MLE渐近达到CRLB

- MLE的方差可大于、等于、小于CRLB！（不同于MVU估计）
- 但数据量足够多时，将与CRLB接近
- 因此，可利用CRLB评估MLE的性能

- MLE是渐近有效的
- MLE是渐近最佳的

“足够多”数据：大量能带来新信息的数据

若参数$\alpha = g(\theta)$，则 α 的MLE由下式给出

$$\hat{\alpha} = g\left(\hat{\theta}\right)$$

其中 $\hat{\theta}$ 是θ 的MLE。若g非一对一函数，那么$\hat{\alpha}$是使修正后的似然函数 $p_T(\mathbf{x};\alpha)$ 最大者

$$\hat{\alpha} = \arg\max_{\alpha} p_T(\mathbf{x};\alpha)$$

$$p_T(\mathbf{x};\alpha) = \max_{\{\theta:\alpha=g(\theta)\}} p(\mathbf{x};\theta)$$

MLE的不变性

- 该性质对函数 g 无线性变换要求，对任意函数均成立
- 对比MVU
  - ✓ 无偏性、有效性仅对线性变换成立
  - ✓ 对非线性变换不能保持，但渐近无偏、渐近有效

例：噪声功率估计：

$$x[n]=w[n],\ n=0,1,\dots,N-1$$

$w[n]$ 为方差为σ^2(未知)的高斯白噪声。求功率P(dB)的MLE?

$$P=10\log_{10}\sigma^2$$

$$p\left(\mathbf{x};\sigma^2\right)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right\}$$

$$\Rightarrow \ln p\left(\mathbf{x};\sigma^2\right)=-\frac{N}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]$$

$$\Rightarrow \frac{\partial\ln p\left(\mathbf{x};\sigma^2\right)}{\partial\sigma^2}=-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{n=0}^{N-1}x^2[n]$$

σ^2的MLE估计为： $\hat{\sigma}^2=\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]$

根据MLE不变性： $\hat{P}=10\log_{10}\hat{\sigma}^2=10\log_{10}\left(\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]\right)$

四、数值解法

最大似然估计：$\left.\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right|_{\hat{\theta}} = 0$ 是否好求解?

例：指数信号估计：

$$x[n] = r^n + w[n],\ n = 0,1,\ldots,N-1$$

$w[n]$ 为方差为 σ^2 的高斯白噪声。估计信号指数因子 $r\,(r>0)$ ？

$$p(\mathbf{x};r) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)^2\right\}$$ 使之最大化

$\Updownarrow$

最小化：$J(r) = \sum_{n=0}^{N-1}\left(x[n]-r^n\right)^2$

$\Downarrow$

$$\frac{\partial J(r)}{\partial r} = 0 \Rightarrow \sum_{n=0}^{N-1}\left(x[n]-r^n\right) n r^{n-1} = 0 \quad ?$$

1. 网格搜索法

目标：$\left.\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right|_{\hat{\theta}} = 0$

令 $g(\theta) = \dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}$

$$\underbrace{g(\theta) \approx g(\theta_0) + \left.\frac{dg(\theta)}{d\theta}\right|_{\theta_0} (\theta - \theta_0)}_{=0}$$

$$\theta_1 = \theta_0 - \frac{g(\theta_0)}{\left.\dfrac{dg(\theta)}{d\theta}\right|_{\theta=\theta_0}}$$

$$\theta_{k+1} = \theta_k - \frac{g(\theta_k)}{\left.\dfrac{dg(\theta)}{d\theta}\right|_{\theta=\theta_k}}$$

$$\theta_{k+1} = \theta_k - \left.\frac{\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}}{\dfrac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}}\right|_{\theta=\theta_k}$$

在 $\theta_{k+1} = \theta_k$ 处收敛，此时 $g(\theta_k) = 0$

2. Newton-Raphson迭代法

1. 迭代有可能不收敛。当对数似然函数二阶导数较小时尤为明显。每次迭代间修正项起伏可能较大

2. 即使收敛，求得的可能不是全局最大值而是局部最大值、甚至是局部最小值。

3. 使用时，建议多采用几个起始点，在收敛点中选择使对数似然函数最大的那个点

例：指数信号估计：

$$x[n]=r^n+w[n],\ n=0,1,\ldots,N-1$$

$w[n]$ 为方差为 σ^2 的高斯白噪声。估计信号指数因子 $r\,(r>0)$？

$$p(\mathbf{x};r)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)^2\right\}$$

$$\ln p(\mathbf{x};r)=-\frac{N}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)^2$$

$$\Rightarrow\quad \frac{\partial \ln p(\mathbf{x};r)}{\partial r}=\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)nr^{n-1}$$

$$\Rightarrow\quad \frac{\partial^2 \ln p(\mathbf{x};r)}{\partial r^2}=\frac{1}{\sigma^2}\sum_{n=0}^{N-1}nr^{n-2}\left[(n-1)x[n]-(2n-1)r^n\right]$$

$$\Rightarrow\quad r_{k+1}=r_k-\frac{\sum_{n=0}^{N-1}\left(x[n]-r_k^n\right)nr_k^{n-1}}{\sum_{n=0}^{N-1}nr_k^{n-2}\left[(n-1)x[n]-(2n-1)r_k^n\right]}$$

Newton-Raphson迭代法:

$$\theta_{k+1}=\theta_k-\frac{\left.\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_k}}{\left[\dfrac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]^{-1}}$$

$$\left.\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right|_{\theta=\theta_k}\approx -I(\theta_k)$$

$$\theta_{k+1}=\theta_k+I^{-1}(\theta)\left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right|_{\theta=\theta_k}$$

3. 得分法

所面临的问题与Newton-Raphson迭代法基本相同

$$p(\mathbf{x};\theta)=p\left(\left[x(0),x(1),\ldots,x(N-1)\right];\theta\right)$$

若数据独立同分布，则有

$$p(\mathbf{x};\theta)=\prod_{n=0}^{N-1}p(x(n);\theta)$$

$$\ln p(\mathbf{x};\theta)=\sum_{n=0}^{N-1}\ln p(x(n);\theta)$$

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}&=\sum_{n=0}^{N-1}\frac{\partial^2 \ln p(x(n);\theta)}{\partial\theta^2}\\&=N\frac{1}{N}\sum_{n=0}^{N-1}\frac{\partial^2 \ln p(x(n);\theta)}{\partial\theta^2}\\&\approx NE\left(\frac{\partial^2 \ln p(x(n);\theta)}{\partial\theta^2}\right)\\&=-Ni(\theta)\\&=-I(\theta)\end{aligned}$$

例：指数信号估计：

$$x[n]=r^n+w[n],\ n=0,1,\dots,N-1$$

$w[n]$ 为方差为 σ^2 的高斯白噪声。估计信号指数因子 $r\,(r>0)$ ？

$$p(\mathbf{x};r)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)^2\right\}$$

$$\ln p(\mathbf{x};r)=-\frac{N}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)^2$$

$$\frac{\partial\ln p(\mathbf{x};r)}{\partial r}=\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)nr^{n-1}$$

$$\frac{\partial^2\ln p(\mathbf{x};r)}{\partial r^2}=\frac{1}{\sigma^2}\sum_{n=0}^{N-1}nr^{n-2}\left[(n-1)x[n]-(2n-1)r^n\right]$$

$$I(\theta)=\frac{1}{\sigma^2}\sum_{n=0}^{N-1}n^2r^{2n-2}$$

Newton-Raphson迭代法:

$$r_{k+1}=r_k-\frac{\sum_{n=0}^{N-1}\left(x[n]-r_k^n\right)nr_k^{n-1}}{\sum_{n=0}^{N-1}nr_k^{n-2}\left[(n-1)x[n]-(2n-1)r_k^n\right]}$$

Vs

得分法:

$$r_{k+1}=r_k+\frac{\sum_{n=0}^{N-1}\left(x[n]-r_k^n\right)nr_k^{n-1}}{\sum_{n=0}^{N-1}n^2r_k^{2n-2}}$$

4. EM算法

$$\hat{\theta} = \arg\max_{\theta} \left\{ \ln p\left(\boldsymbol{x};\theta \right) \right\} \xrightarrow{\boldsymbol{x} = g\left(\boldsymbol{y} \right)} \hat{\theta} = \arg\max_{\theta} \left\{ \ln p\left(\boldsymbol{y};\theta \right) \right\}$$

1. **数学期望 (E)**：确定完备数据的平均对数似然函数

$$U\left(\theta,\theta_k\right) = \int \ln p_y\left(\boldsymbol{y};\theta\right) p\left(\boldsymbol{y} \mid \boldsymbol{x};\theta_k\right) d\boldsymbol{y}$$

2. **最大化 (M)**：求使完备数据的平均对数似然函数最大的 θ 值

$$\theta_{k+1} = \arg\max_{\theta} \; U\left(\theta,\theta_k\right)$$

- *Feder, M., E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", IEEE Trans. Acoust., Speech, Signal Process., Vol.36, pp.477-489, April, 1988*
- *Jeffrey A. Fessler, Alfred O. Hero, "Space-alternating generalized expectation-maximization algorithm", IEEE Trans. on Signal Processing, Vol.42, pp.2664-2677, Oct., 1994*
- *Todd. K. Moon, " The expectation-maximization algorithm", IEEE Signal Processing Magazine, pp. 47-60, Nov., 1996*

五、矢量参数的MLE

标量参数MLE：$\hat{\theta}=\arg\max_{\theta}\{p(\mathbf{x};\theta)\}$ $\xrightarrow{\text{若可导}}$ $\left.\dfrac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right|_{\hat{\theta}}=0$

$\Downarrow$

矢量参数MLE：$\hat{\boldsymbol{\theta}}=\arg\max_{\boldsymbol{\theta}}\{p(\mathbf{x};\boldsymbol{\theta})\}$ $\xrightarrow{\text{若可导}}$ $\left.\dfrac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}}=\mathbf{0}$

$$\boxed{\begin{pmatrix} \dfrac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_1} \\ \dfrac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \dfrac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_p} \end{pmatrix}}$$

例：电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\ldots,N-1$$

$w[n]$为高斯白噪声，且 $w[n]\sim N(0,\sigma^2)$，待估计参数 $\boldsymbol{\theta}=\left[A,\sigma^2\right]^T$

$$p(\mathbf{x};\boldsymbol{\theta})=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right\}$$

$$\Rightarrow \ln p(\mathbf{x};\boldsymbol{\theta})=-\frac{N}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2$$

$$\Rightarrow \left\{\begin{aligned}\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A}&=\frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)\\ \frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\sigma^2}&=-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{n=0}^{N-1}(x[n]-A)^2\end{aligned}\right.$$

$$\left.\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}}=\mathbf{0}$$

$$\Rightarrow \hat{\boldsymbol{\theta}}=\begin{bmatrix}\bar{x}\\ \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2\end{bmatrix}$$

例：电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\dots,N-1$$

$w[n]$为高斯白噪声，且 $w[n]\sim N(0,\sigma^2)$，待估计参数 $\boldsymbol{\theta}=\left[A,\sigma^2\right]^T$

MLE估计量：

$$\hat{\boldsymbol{\theta}}=\begin{bmatrix}\bar{x}\\ \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2\end{bmatrix}\qquad \mathbf{C}_{\hat{\theta}}=\begin{bmatrix}\frac{\sigma^2}{N} & 0\\ 0 & \frac{N-1}{N}\frac{2\sigma^4}{N}\end{bmatrix}$$

MVU估计量：

$$\hat{\boldsymbol{\theta}}=\begin{bmatrix}\bar{x}\\ \frac{1}{N-1}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2\end{bmatrix}\qquad \mathbf{C}_{\hat{\theta}}=\begin{bmatrix}\frac{\sigma^2}{N} & 0\\ 0 & \frac{2\sigma^4}{N-1}\end{bmatrix}$$

（见第三章课件例题）

Fisher信息矩阵逆：

$$\mathbf{I}^{-1}(\boldsymbol{\theta})=\begin{bmatrix}\frac{\sigma^2}{N} & 0\\ 0 & \frac{2\sigma^4}{N}\end{bmatrix}$$

矢量参数MLE的性质

1. 渐近有效性

如果数据 $\mathbf{x}$ 的PDF $p(\mathbf{x};\boldsymbol{\theta})$ 满足“正则”条件，那么对于足够多的数据记录，未知参数 $\boldsymbol{\theta}$ 的MLE渐近服从 $\hat{\boldsymbol{\theta}} \overset{a}{\sim} N\left(\boldsymbol{\theta}, \mathbf{I}^{-1}(\boldsymbol{\theta})\right)$。其中 $\mathbf{I}(\boldsymbol{\theta})$ 是在未知参数真值处计算的Fisher信息矩阵

- MLE是渐近无偏的
- MLE渐近达到CRLB
- MLE是渐近有效的
- MLE是渐近最佳的

2.不变性

若参数 $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$，则 $\boldsymbol{\alpha}$ 的MLE由下式给出：$\hat{\boldsymbol{\alpha}} = \mathbf{g}(\hat{\boldsymbol{\theta}})$。其中 $\hat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta}$ 的MLE。若 $\mathbf{g}$ 非一对一函数，那么 $\hat{\boldsymbol{\alpha}}$ 是使修正后的似然函数 $p_T(\mathbf{x};\boldsymbol{\alpha})$ 最大者：$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} p_T(\mathbf{x};\boldsymbol{\alpha})$，其中 $p_T(\mathbf{x};\boldsymbol{\alpha}) = \max_{\{\boldsymbol{\theta}:\boldsymbol{\alpha}=\mathbf{g}(\boldsymbol{\theta})\}} p(\mathbf{x};\boldsymbol{\theta})$。

一般线性模型：

$$\boldsymbol{x} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{w}$$

其中 $\mathbf{H}$ 是已知的 $N \times p$ 矩阵，$\boldsymbol{\theta}$ 是$p \times 1$的待估计量，$\boldsymbol{w}$ 是均值为零、协方差为 $\mathbf{C}$ 的高斯噪声矢量。

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}\{\det(\mathbf{C})\}^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{x}-\mathbf{H}\boldsymbol{\theta})\right\}$$

$$\Rightarrow \quad \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{H}^T\mathbf{C}^{-1}\boldsymbol{x} - \mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta}$$

MLE估计量：$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\boldsymbol{x}$

——MVU估计量！

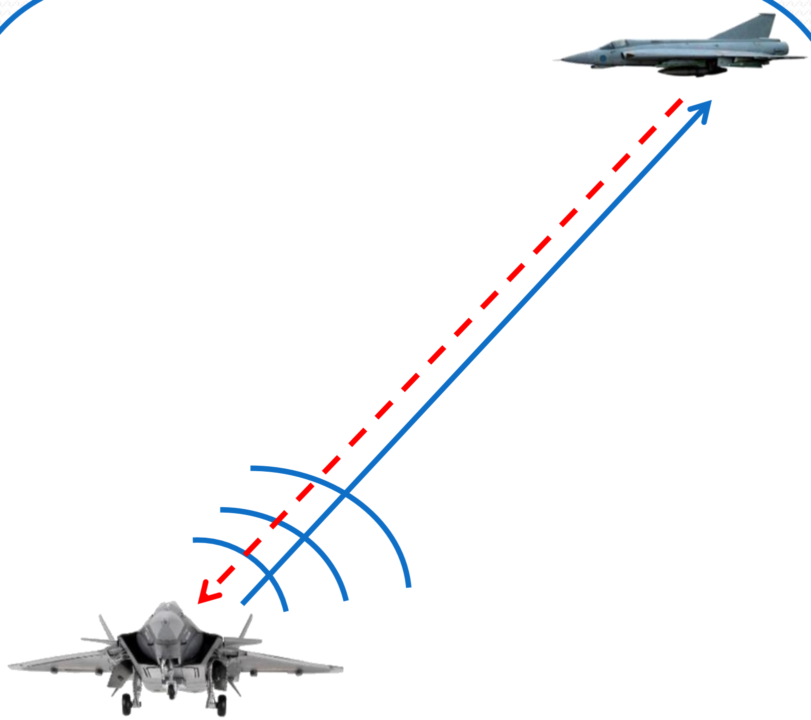
$$\mathbf{C}_{\hat{\theta}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}$$

$$\hat{\boldsymbol{\theta}} \sim N\left(\boldsymbol{\theta}, \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\right)$$

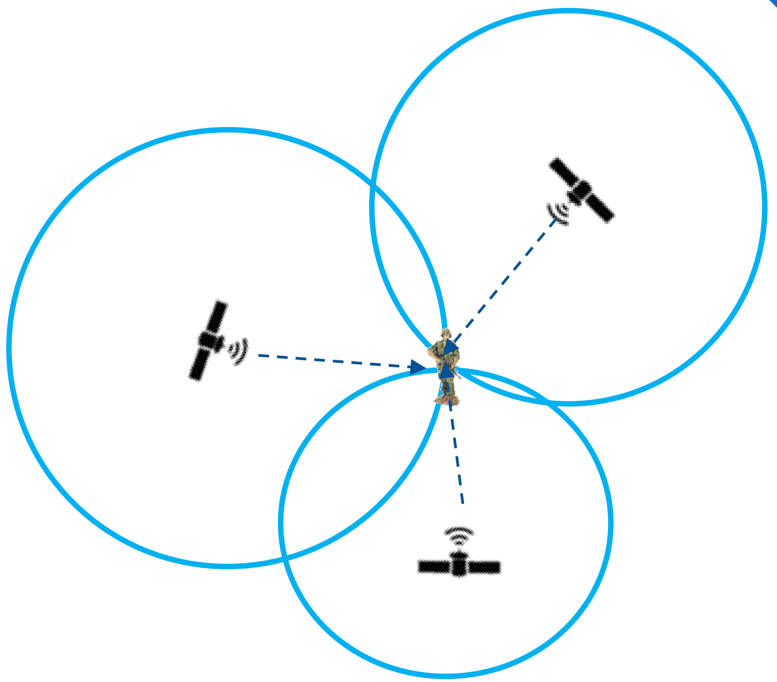
对一般线性模型，MLE是MVU估计量，达到CRLB，是有效的、最佳的！

六、MLE的应用

- 雷达与导航测距问题



雷达测距：信号从发射到遇到“目标”反射回来的时延 × 光速



导航测距：信号从卫星发出至用户接收的时延 × 光速

信号模型：

$$x(t)=s(t-\tau_0)+w(t),\quad 0\le t\le T$$

接收到的波形　发射信号　时延　WGN噪声

对雷达测距 $R=c\tau_0/2$

对导航测距 $R=c\tau_0$

假定信号带宽为 B，以奈奎斯特采样率 $\Delta=1/(2B)$ 进行采样则所得离散信号数据模型可表示为：

$$x[n]=s(n\Delta-\tau_0)+w[n]$$

假定信号只在 $[0,T_s]$ 内非零，因此信号模拟可进一步简化为

$$x[n]=\begin{cases} w[n], & 0\le n\le n_0-1 \\ s(n\Delta-\tau_0)+w[n], & n_0\le n\le n_0+M-1 \\ w[n], & n_0+M\le n\le N-1 \end{cases}$$

其中 $n_0\approx\tau_0/\Delta$，N 表示观察数据样本数，M 表示采样信号长度

MLE方法:

$$x[n]=\begin{cases} w[n], & 0\le n\le n_0-1 \\ s\left(n\Delta-\tau_0\right)+w[n], & n_0\le n\le n_0+M-1 \\ w[n], & n_0+M\le n\le N-1 \end{cases}$$

似然函数:

$$p\left(\boldsymbol{x};n_0\right)=\prod_{n=0}^{n_0-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}x^2[n]\right\}$$

$$\times\prod_{n=n_0}^{n_0+M-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left(x[n]-\underbrace{s[n-n_0]}_{s\left(n\Delta-\tau_0\right)}\right)^2\right\}$$

$$\times\prod_{n=n_0+M}^{N-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}x^2[n]\right\}$$

可化简为: $p\left(\boldsymbol{x};n_0\right)=\dfrac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right\}$

$$\times\exp\left\{-\frac{1}{2\sigma^2}\underline{\sum_{n=n_0}^{n_0+M-1}\left(-2x[n]s[n-n_0]+s^2[n-n_0]\right)}\right\}$$

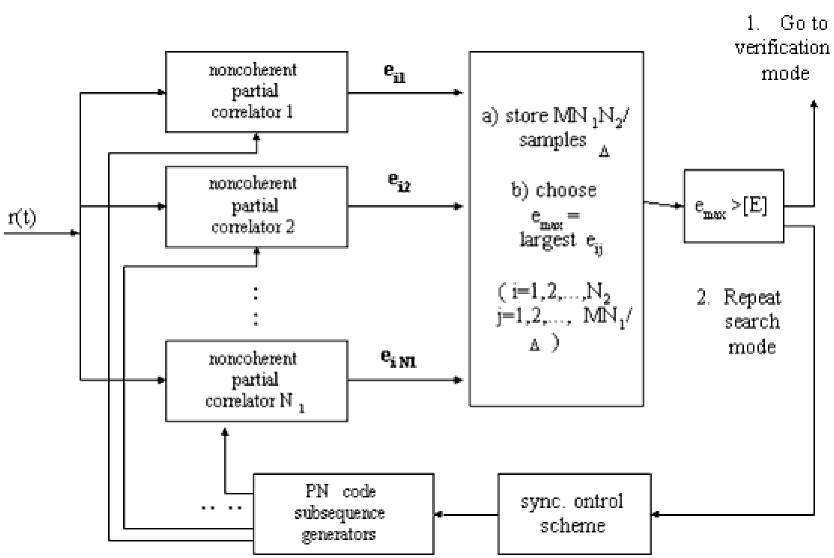
$$\min\left\{\sum_{n=n_0}^{n_0+M-1}\left(-2x[n]s[n-n_0]+\underline{s^2[n-n_0]}\right)\right\}$$

$$\sum_{n=n_0}^{n_0+M-1} s^2[n-n_0]=\sum_{n=0}^{M-1} s^2[n]$$ ——与延时无关

$$\max\left\{\underline{\sum_{n=n_0}^{n_0+M-1} x[n]s[n-n_0]}\right\}$$

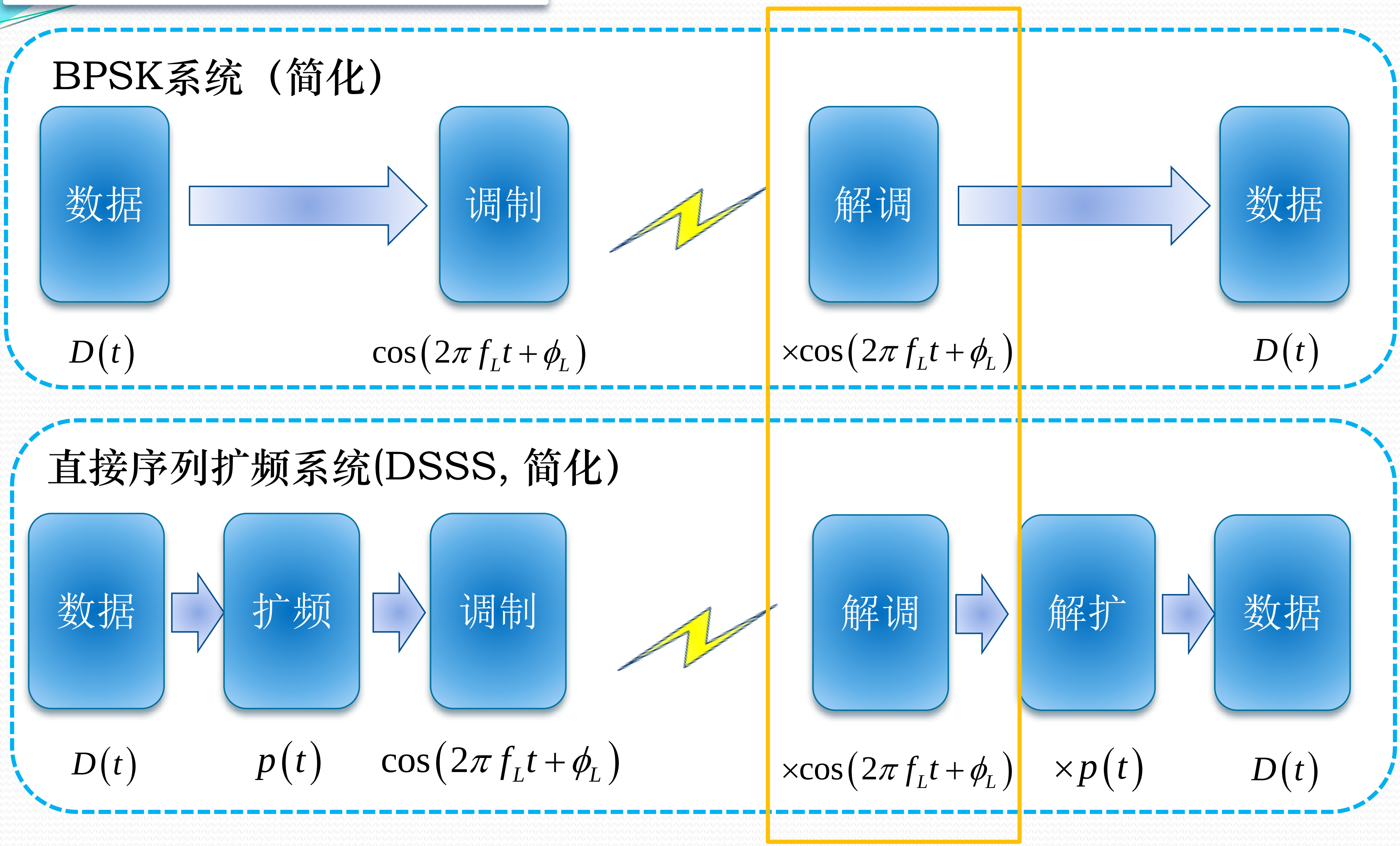
物理含义：将各起始时刻的数据与信号进行相关，选择最大值，其对应的 n_0 即为延时的MLE，反之亦然

导航系统中的伪码捕获问题（雷达、通信等其它直接序列扩频系统（DSSS）均适用）



- 最大值捕获法

● 正弦信号参数估计



$\cos(2\pi f_0 t+\phi)$

关键问题，复现载波：$\cos(2\pi f_0 t+\phi)$

此外还受信号**衰减、噪声**等的影响

信号模型：

$$x[n]=A\cos\left(2\pi f_0 n+\phi\right)+w[n],\ n=0,1,\ldots,N-1$$

$w[n]$为方差为 σ^2 的高斯白噪声。待估计参数包括：A $(A>0)$，$f_0\,(0<f_0<1/2)$，ϕ 。求其最大似然估计

$$p(\mathbf{x};\boldsymbol{\theta})=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\cos\left(2\pi f_0 n+\phi\right)\right)^2\right\}$$

$$\min\left\{\sum_{n=0}^{N-1}\left(x[n]-A\cos\left(2\pi f_0 n+\phi\right)\right)^2\right\}$$

$$J\left(A,f_0,\phi\right)=\sum_{n=0}^{N-1}\left(x[n]-A\cos\left(2\pi f_0 n+\phi\right)\right)^2$$

$$=\sum_{n=0}^{N-1}\left(x[n]-\underline{A\cos\phi}\cos\left(2\pi f_0 n\right)+\underline{A\sin\phi}\sin\left(2\pi f_0 n\right)\right)^2$$

$$\begin{cases}\alpha_1=A\cos\phi\\ \alpha_2=-A\sin\phi\end{cases}\quad\Longleftrightarrow\quad\begin{cases}A=\sqrt{\alpha_1^2+\alpha_2^2}\\ \phi=\arctan\left(\dfrac{-\alpha_2}{\alpha_1}\right)\end{cases}$$

$$J\left(\alpha_1,\alpha_2,f_0\right)=\sum_{n=0}^{N-1}\left(x[n]-\alpha_1\cos\left(2\pi f_0 n\right)-\alpha_2\sin\left(2\pi f_0 n\right)\right)^2$$

令 $\boldsymbol{c}=\left[1,\cos\left(2\pi f_0\right),\cos\left(4\pi f_0\right),\cos\left(6\pi f_0\right),\ldots,\cos\left(2\pi f_0\left(N-1\right)\right)\right]^T$

$$\boldsymbol{s}=\left[0,\sin\left(2\pi f_0\right),\sin\left(4\pi f_0\right),\sin\left(6\pi f_0\right),\ldots,\sin\left(2\pi f_0\left(N-1\right)\right)\right]^T$$

$$J\left(\alpha_1,\alpha_2,f_0\right)=\left(\boldsymbol{x}-\alpha_1\boldsymbol{c}-\alpha_2\boldsymbol{s}\right)^T\left(\boldsymbol{x}-\alpha_1\boldsymbol{c}-\alpha_2\boldsymbol{s}\right)$$

$$=\left(\boldsymbol{x}-\mathbf{H}\boldsymbol{\alpha}\right)^T\left(\boldsymbol{x}-\mathbf{H}\boldsymbol{\alpha}\right)$$ ，其中 $\mathbf{H}=[\boldsymbol{c},\boldsymbol{s}],\boldsymbol{\alpha}=\left[\alpha_1,\alpha_2\right]^T$

$$\frac{\partial J\left(\alpha_1,\alpha_2,f_0\right)}{\partial\boldsymbol{\alpha}}=0 \Rightarrow \hat{\boldsymbol{\alpha}}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}$$

$$\Rightarrow J\left(\alpha_1,\alpha_2,f_0\right)=\left(\boldsymbol{x}-\mathbf{H}\hat{\boldsymbol{\alpha}}\right)^T\left(\boldsymbol{x}-\mathbf{H}\hat{\boldsymbol{\alpha}}\right)$$

$$=\boldsymbol{x}^T\left(\mathbf{I}-\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\left(\mathbf{I}-\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\boldsymbol{x}$$

$$=\boldsymbol{x}^T\left(\mathbf{I}-\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\boldsymbol{x}$$

目标： $\min\left\{J\left(A,f_0,\phi\right)\right\} \Leftrightarrow \max\left\{\boldsymbol{x}^T\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}\right\}$

$$\boldsymbol{x}^T\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}=\left[\boldsymbol{x}^T\boldsymbol{c},\boldsymbol{x}^T\boldsymbol{s}\right]\begin{bmatrix}\boldsymbol{c}^T\boldsymbol{c} & \boldsymbol{c}^T\boldsymbol{s}\\ \boldsymbol{s}^T\boldsymbol{c} & \boldsymbol{s}^T\boldsymbol{s}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{c}^T\boldsymbol{x}\\ \boldsymbol{s}^T\boldsymbol{x}\end{bmatrix}=\left[\boldsymbol{x}^T\boldsymbol{c},\boldsymbol{x}^T\boldsymbol{s}\right]\begin{bmatrix}N/2 & 0\\ 0 & N/2\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{c}^T\boldsymbol{x}\\ \boldsymbol{s}^T\boldsymbol{x}\end{bmatrix}$$

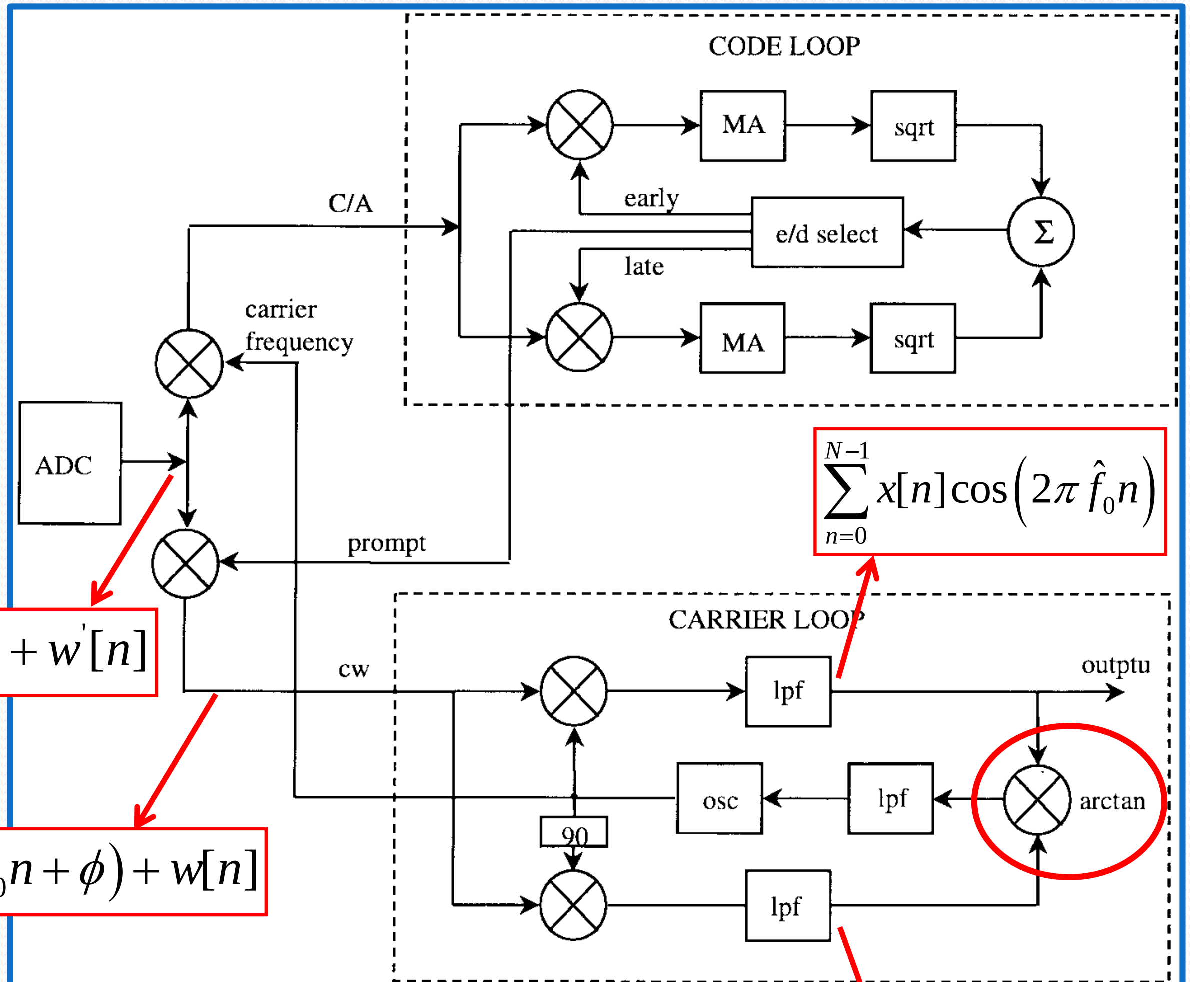
$$=\frac{2}{N}\left|\sum_{n=0}^{N-1}x[n]\exp\left(-j2\pi f_0 n\right)\right|^2$$

$$\max\left\{\boldsymbol{x}^T\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}\right\}$$

$$\Rightarrow\quad \hat{\boldsymbol{\alpha}}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\boldsymbol{x}=\frac{2}{N}\begin{bmatrix}\hat{\boldsymbol{c}}^T\boldsymbol{x}\\ \hat{\boldsymbol{s}}^T\boldsymbol{x}\end{bmatrix}=\begin{bmatrix}\frac{2}{N}\sum_{n=0}^{N-1}x[n]\cos\left(2\pi\hat{f}_0 n\right)\\ \frac{2}{N}\sum_{n=0}^{N-1}x[n]\sin\left(2\pi\hat{f}_0 n\right)\end{bmatrix}$$

$$\Rightarrow\quad \begin{cases}\hat{A}=\sqrt{\hat{\alpha}_1^2+\hat{\alpha}_2^2}=\frac{2}{N}\left|\sum_{n=0}^{N-1}x[n]\exp\left(-j2\pi\hat{f}_0 n\right)\right|\\ \hat{\phi}=\arctan\left(\frac{-\hat{\alpha}_2}{\hat{\alpha}_1}\right)=\arctan\left\{\frac{-\sum_{n=0}^{N-1}x[n]\sin\left(2\pi\hat{f}_0 n\right)}{\sum_{n=0}^{N-1}x[n]\cos\left(2\pi\hat{f}_0 n\right)}\right\}\end{cases}$$

● 导航信号/DSSS信号载波跟踪



$$\sum_{n=0}^{N-1} x[n]\cos\left(2\pi \hat{f}_0 n\right)$$

$$AD(t)\,p(t)\cos\left(2\pi f_0 n+\phi\right)+w^{'}[n]$$

$$x[n]=A\cos\left(2\pi f_0 n+\phi\right)+w[n]$$

$$-\sum_{n=0}^{N-1} x[n]\sin\left(2\pi \hat{f}_0 n\right)$$

- Figure8.3, “*Fundamentals of Global Positioning System Receivers: A Software Approach*” , James Bao-Yen Tsui, p174

六、总结

- 最大似然估计（MLE）：$\hat{\theta} = \arg\max_{\theta}\left\{p\left(\mathbf{x};\theta\right)\right\}$
  - 找概率最大的那个 ——核心思想：大概率事件优先发生 Vs MVU
  - 若存在有效估计量，那么MLE可以求得
  - MLE渐近有效性：渐近无偏、渐近达到CRLB，因此渐近有效、渐近最佳
  - MLE的不变性：无论是否线性变换

  MLE应用广泛
- 数值解法：
  - 网格搜索法
  - 迭代法（Newton-Raphson、得分法、EM算法等）
- 矢量最大似然估计：$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left\{p\left(\mathbf{x};\boldsymbol{\theta}\right)\right\}$
- 最大似然估计的应用