S.→symmetric；P.S.D.→positive semi-definite；var.-variance; autocorrel.-autocorrelation

Lecture 2 Stochastic Theory

event A given event B：$P(A|B)=\frac{P(A,B)}{P(B)}$

Bayers' Rule $P(A,B)=P(A|B)P(B)=P(B|A)P(A)$

独立P(A, B) = P(A)P(B)、$P(B|A)=P(B)$、$P(A|B)=P(A)$

Random variables（RVs）Gaussian $Z\sim N(\eta,\sigma^2)$，$f_Z(z)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(z-\eta)^2}{2\sigma^2}}$

Expected value/ Variance $E[g(X)]=\int_{-\infty}^{\infty}g(x)f_X(x)dx$，$\sigma_X^2=E(X_2)-(EX)^2$

$\sigma_X^2=E[(X-EX)^2]=\int_{-\infty}^{\infty}(x-EX)^2f_X(x)dx$，Standard deviation σ

$Y=g(X),X=g^{-1}(Y)=h(Y)$ known the pdf of x, compute the pdf of Y as follows

$\&P(X\in[x,x+dx])=P(Y\in[y,y+dy])(dx>0)$

$\&\int_x^{x+dx}f_X(z)\,dz=\begin{cases}\int_y^{y+dy}-f_Y(z)dz & \text{if dy>0}\\ -\int_y^{y+dy}f_Y(z)dz & \text{if dy<0}\end{cases}$

$\&f_X(x)dx\&=f_Y(y)|dy|,f_Y(y)=\left|\frac{dx}{dy}\right|f_X[h(y)]=|h'(y)|f_X[h(y)]$

Multiple RVs $F(x,\infty)=F(x),f(x)=\int_{-\infty}^{\infty}f(x,y)dy$, marginal distribution/density

Expectation $E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f(x,y)dxdy$

Covariance 协方差 $C_{XY}=E[(X-E(X))(Y-E(Y))]=E(XY)-E(X)E(Y)$

独立性$P(X\le x,Y\le y)=P(X\le x)P(Y\le y),\forall x,y$，$F_{XY}(x,y)=F_X(x)F_Y(y)$, F or f

相关系数 $\rho=\frac{C_{XY}}{\sigma_X\sigma_Y}$，相关性 $R_{XY}=E(XY)$；uncorrelated if $\rho=0$ or $R_{XY}=E(X)E(Y)$.

独立 ⊊ 不相关，正交 if $R_{XY}=0$||条件密度$f_Y(y|X=x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\frac{f_X(x|Y=y)f_Y(y)}{f_X(x)}$

Multivariate statistics $R_{XY}=E(XY^T)=\begin{bmatrix}E(X_1Y_1)&\cdots&E(X_1Y_m)\\ \vdots&&\vdots\\ E(X_nY_1)&\cdots&E(X_nY_m)\end{bmatrix}$

协方差$C_{XY}=E[(X-E(X))(Y-E(Y))^T]=E(XY^T)-E(X)E(Y)^T$

自相关$R_X=E[XX^T]=\begin{bmatrix}E(X_1^2)&\cdots&E(X_1X_n)\\ \vdots&&\vdots\\ E(X_nX_1)&\cdots&E(X_n^2)\end{bmatrix}$，symmetric &P.S.D.

Autocovariance 自协方差$C_X=E[(X-E(X))(X-E(X)^T)]=\begin{bmatrix}\sigma_1^2&\cdots&\sigma_{1n}\\ \vdots&&\vdots\\ \sigma_{n1}&\cdots&\sigma_n^2\end{bmatrix}$，S.&P.S.D.

Linear transformation of Gaussian RV: An n-element RV X is Gaussian (normal) if

$$\mathrm{pdf}(X)=\frac{1}{(2\pi)^{n/2}|det(C_X)|^{1/2}}\exp\left[-\frac{1}{2}(x-E(X))^TC_X^{-1}(x-E(X))\right]$$

考虑 Y = g(X) = AX + b,A is invertible, $f_Y(y)=|h'(y)|f_X[h(y)]=\frac{1}{(2\pi)^{n/2}|det(C_XA^T)|^{1/2}}$.

$\exp\left[-\frac{1}{2}(y-E(Y))^T(AC_XA^T)^{-1}(y-E(Y))\right]$；i.e.，$Y\sim\mathcal{N}(AE(X)+b,AC_XA^T)$

Matrix derivative:$\mathbb{R}^n\to\mathbb{R}$、$\mathbb{R}^n\to\mathbb{R}^m$、$\mathbb{R}^{n\times m}\to\mathbb{R}$

$$\nabla_xf=\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{bmatrix},\nabla_x^2f=\frac{\partial^2f}{\partial x\partial x^T}=\begin{bmatrix}\frac{\partial^2f}{\partial x_1\partial x_1}&\cdots&\frac{\partial^2f}{\partial x_1\partial x_n}\\ \vdots&&\vdots\\ \frac{\partial^2f}{\partial x_n\partial x_1}&\cdots&\frac{\partial^2f}{\partial x_n\partial x_n}\end{bmatrix}$$

$$\nabla_xf=\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{\partial f_1}{\partial x_1}&\cdots&\frac{\partial f_m}{\partial x_1}\\ \vdots&&\vdots\\ \frac{\partial f_1}{\partial x_n}&\cdots&\frac{\partial f_m}{\partial x_n}\end{bmatrix},\ \nabla_xf=\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{\partial f}{\partial x_{11}}&\cdots&\frac{\partial f}{\partial x_{1m}}\\ \vdots&\ddots&\vdots\\ \frac{\partial f}{\partial x_{n1}}&\cdots&\frac{\partial f}{\partial x_{nm}}\end{bmatrix}$$

Ellipsoid：A is real S.&P.D.‖ $(x-v)^TA(x-v)=1$，centered at v. A 的特征向量是椭球体的主轴，A 的特征值是半轴平方的倒数：a^{-2}、b^{-2}、c^{-2} 。

对于协方差矩阵 Σ, SVD 分解为$\Sigma=U\Lambda U^{-1}$, U 和 Λ 分别表示 Σ 的特征向量和特征值，特征向量是表示数据最大方差方向的单位向量，而特征值表示相应方向上方差的大小。

PCA：给定数据 {x1, …, xm}，计算协方差矩阵 $\Sigma=\frac{1}{m}\sum_{i=1}^m(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T$，主成分即协方差矩阵的特征向量，其对应特征值越大，越重要。

Stochastic processes theory(S.P.)

Covar.$C_X(t)=E\{[X(t)-\bar{x}(t)][X(t)-\bar{x}(t)]^T\}$；correlation BTW X(t1) and X(t2)

$R_X(t_1,t_2)=E[X(t_1)X^T(t_2)]$；Autocovar.$C_X(t_1,t_2)=E\{[X(t_1)-\bar{x}(t_1)][X(t_2)-\bar{x}(t_2)]^T\}$

Stationary S.P.:1.Strict-sense stationary: $F_X(x(t_1+\tau),\ldots,x(t_n+\tau))=F_X(x(t_1),\ldots,x(t_n))$;

Wide-sense stationary：$E[X(t)]=\bar{x},E[X(t_1)X^T(t_2)]=R_X(t_2-t_1)$ strict →wide

$R_X(0)=E[X(t)X^T(t)],R_X(-\tau)=R_X^T(\tau)$and for scalar $|R_X(\tau)|\le R_X(0)$

Time average/autocorrel. $A[X(t)]=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)dt$，$R[X(t),\tau]=A[X(t)X^T(t+\tau)]$

Ergodic process：各态历经过程是平稳随机过程 if $A[X(t)]=E(X)$，$R[X(t),\tau]=R_X(\tau)$

Two S.P. ：cross correlation of X &Y: $R_{XY}(t_1,t_2)=E[X(t_1)Y^T(t_2)]$//互协方差

If X&Y are uncorrelated $R_{XY}(t_1,t_2)=E[X(t_1)]E[Y(t_2)]^T$ for all t

cross covariance：$C_{XY}(t_1,t_2)=E\{[X(t_1)-\bar{X}(t_1)][Y(t_2)-\bar{Y}(t_2)]^T\}$

Several Kinds of Stochastic Processes：Markov chain 未来只依赖当前状态，不依赖之前发生的事$P\{X_{n+1}=i_{n+1}|X_0=i_0,X_1=i_1,\ldots,X_n=i_n\}=P\{X_{n+1}=i_{n+1}|X_n=i_n\}$

Ergodic M.C./遍历链：可以从每个状态转移到每个状态；任何没有 0 的转移矩阵决定一个正则 regular M.C.，但正则 M.C.可能有一个含零转移矩阵，正则链都是遍历的

Hidden M.M. :$P(Y_n\in A|X_1=x_1,\ldots,X_n=x_n)=\mathbf{P}(Y_n\in A|X_n=x_n)$

1-dimensional Random Walk、Wiener Process、Poisson Processes

White Noise：power spectral density $P=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x^2(t)dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_X(\omega)d\omega$

power spectrum：$S_X(\omega)=\int_{-\infty}^{\infty}R_X(\tau)e^{-j\omega\tau}d\tau$；autocorrel.$R_X(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_X(\omega)e^{j\omega\tau}d\omega$

X(t) and Y (t): $S_{XY}(\omega)=\int_{-\infty}^{\infty}R_{XY}(\tau)e^{-j\omega\tau}d\tau$, $R_{XY}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_{XY}(\omega)e^{j\omega\tau}d\omega$

Lecture 3 Estimation Theory

Unbiased/无偏: $E(\hat{s}(n))=E(s(n)),E(\tilde{s}(n))=0$;Asymptotically unbiased:$\lim_{n\to\infty}E(\hat{s}(n))=\lim_{n\to\infty}E(s(n)),\lim_{n\to\infty}E(\tilde{s}(n))=0$，Consistent estimator/一致估计$\lim_{n\to\infty}E(\tilde{s}^2(n))=0$

Maximum Likelihood Estimation(MLE)：Likelihood f:$f_z(z|\mathbf{s}=s)$；$\hat{\mathbf{s}}_{ML}maxf_z(z|\mathbf{s}=s)$

$$\hat{\mathbf{s}}_{ML}=value\ of\ s\ for\ which\ \frac{\partial f_z(z|\mathbf{s}=s)}{\partial s}=0\ or\ \frac{\partial\ln f_z(z|\mathbf{s}=s)}{\partial s}=0$$

Maximum a posteriori Estimation(MAPE)：$\hat{\mathbf{s}}_{MAP}$ maximizes $f_z(z|\mathbf{s}=s)f_s(s)$

$$\hat{\mathbf{s}}_{MAP}=\text{value of s that maximizes }f_s(s|\mathbf{z}=z)=\frac{f_z(z|\mathbf{s}=s)f_s(s)}{f_z(z)}$$

Naive Bayes: $X=[X_1,\ldots,X_n]$，X_i,Y are boolean ; $2(2^n-1)$

Conditionally independence assumption: $P(X_1\cdots X_n|Y)=\prod_iP(X_i|Y)$

Under CIA: $2n\ for\ P(X=x_i|Y=y_j)=\prod_{k=1}^nP(X_k=x_{ik}|Y=y_j);P(Y=y_j\mid X_1,\ldots,X_n)$

$$=\frac{P(Y=y_j)P(X_1,\ldots,X_n\mid Y=y_j)}{\sum_mP(Y=y_m)P(X_1,\ldots,X_n\mid Y=y_m)}\overset{conditional}{\underset{independence}{=}}\frac{P(Y=y_j)\prod_iP(X_i\mid Y=y_j)}{\sum_mP(Y=y_m)\prod_iP(X_i\mid Y=y_m)}$$

Naive Bayes 算法: For y_j: $\pi_j=P(Y=y_j)$;for x_{ik}: $\theta_{ijk}=P(X_i=x_{ik}\mid Y=y_j)$

$$Y^{new}=\mathrm{argmax}_{y_j}P(Y=y_j)\prod_iP(X_i|Y=y_j)=\mathrm{argmax}_{y_j}\pi_j\prod_i\theta_{ijk}$$

discrete-valued: $\pi_j=p(Y=y_j)=\frac{\#D\{Y=y_j\}}{|D|}$, $\theta_{ijk}=P(X_i=x_{ik}\mid Y=y_j)=\frac{\#D\{X_i=x_{ik}\wedge Y=y_j\}}{\#D\{Y=y_j\}}$

MAP estimates (Laplace smoothing for the case $l=1$)

离散情况$\pi_j=p(Y=y_j)=\frac{\#D\{Y=y_j\}+l}{|D|+lR}$，$\theta_{ijk}=P(X_i=x_{ik}|Y=y_j)=\frac{\#D\{X_i=x_{ik}\wedge Y=y_j\}+l}{\#D\{Y=y_j\}+lM}$

Logistic regression：X 实值向量，Y boolean，推导

条件独立：$P(X_i|Y=y_k)\sim N(\mu_{ik},\sigma_i)$，$\omega_0=\ln\frac{1-\pi}{\pi}+\sum_i\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2}$，$\omega_i=\frac{\mu_{i0}-\mu_{i1}}{\sigma_i^2}$

Logistics 分布：$\ln\frac{P(Y=0|X)}{P(Y=1|X)}=\omega_0+\sum_{i=1}^n\omega_ix_i$

Minimum Mean-Square Error Estimation(MMSE)

$$MSE=E[E[\tilde{s}^2|z]]=E[E[(s-\hat{s})^2|z]]=E[(s-\hat{s})^2]$$

$\hat{s}_{MMSE}=E[\mathbf{s}|\mathbf{z}]$；unbiased：$E(\hat{\mathbf{s}})=E(\mathbf{s}),E(\tilde{\mathbf{s}})=0$；is one type of Baysian estimation；

MMSE estimate with Gaussian noise: $\mathbf{z}=\mathbf{s}+\mathbf{v}$ with $v\sim N(0,\sigma_v^2)$and $s\sim N(\bar{s},\sigma_s^2)$，

s v 独立：s v,不相关，MMSE 与 MAP 相同：$E[\mathbf{z}]=E[\mathbf{s}]=\bar{s}$,

$Var[\mathbf{z}]=\mathrm{Var}[\mathbf{s}]+\mathrm{Var}[\mathbf{v}]=\sigma_s^2+\sigma_v^2$; pdf of z $f_z(z)=\frac{1}{\sqrt{2\pi}\sqrt{\sigma_s^2+\sigma_v^2}}\exp\left[-\frac{(z-\bar{s})^2}{2(\sigma_s^2+\sigma_v^2)}\right]$

$$f_s(s|z=z)=\frac{f_s(s)f_v(z-s)}{f_z(z)}=\frac{1}{2\pi\sigma_s\sigma_vf_z(z)}\exp\left\{-\left[\frac{(z-s)^2}{2\sigma_v^2}+\frac{(s-\bar{s})^2}{2\sigma_s^2}\right]\right\}$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_s^2\sigma_v^2}{\sigma_s^2+\sigma_v^2}}}\exp\left[-\frac{(s-\hat{s}_{MAP})^2}{2\frac{\sigma_s^2\sigma_v^2}{\sigma_s^2+\sigma_v^2}}\right];\ \hat{s}_{MMSE}=E[\mathbf{s}|\mathbf{z}=z]=\hat{s}_{MAP}=\bar{s}+\frac{\sigma_s^2}{\sigma_v^2+\sigma_s^2}(z-\bar{s})$$

Orthogonality Principle：$E[(s-E[s|z])\gamma(z)]=0$

$\hat{s}=\alpha(z)$is the MMSE estimate ⟺ $(s-\alpha(z))\perp\gamma(z)$，$i.e.E[(s-\alpha(z))\gamma(z)]=0$

Linear MMSE(LMMSE)

$$\hat{s}=\lambda z,\min_\lambda MSE=E[(s-\lambda z)^2],\hat{s}_{LMMSE}=\alpha(z)=\frac{E(sz)}{E(z^2)}z$$

Orthogonality $(s-\alpha(z))\perp\gamma(z)$，$i.e.E[(s-\alpha(z))\gamma(z)]=0$

LMMSE for vector RVs: $\hat{s}=Mz,P=E[(s-\hat{s})(s-\hat{s})^T]$

$$MSE=tr(P)=E[(s-\hat{s})^T(s-\hat{s})],\hat{s}_{LMMSE}=E(sz^T)[E(zz^T)]^{-1}z$$

Orthogonality for vector RVs: $E[(s-\hat{s})z^T]=0$

$\hat{s}=\alpha(z)$is the MMSE estimate ⟺ $E[(s-\hat{s})z^T]=0$

Lecture 4 Least Squares Estimation

$y(n)=\mathbf{x}^T\mathbf{h}(n)+v(n)$；期望响应：$y(n)\in\mathbb{R}(n=1,\ldots,N)$；输入：$h_k(n),k=1,\ldots,M$；

Aim to estimate $\mathbf{x}=[x_1,\ldots,x_M]^T$ say $\hat{\mathbf{x}}$，and $\hat{\mathbf{y}}(n)$ is predicted output.

Estimation error $e(n)=y(n)-\hat{y}(n)=y(n)-\hat{\mathbf{x}}^T\mathbf{h}(n)$，Matrix Formulation $\mathbf{e}=\mathbf{y}-H\hat{\mathbf{x}}$

$$\begin{bmatrix}e(1)\\ \vdots\\ e(N)\end{bmatrix}=\begin{bmatrix}y(1)\\ \vdots\\ y(N)\end{bmatrix}-\begin{bmatrix}h_1(1)&\cdots&h_M(1)\\ \vdots&\ddots&\vdots\\ h_1(N)&\cdots&h_M(N)\end{bmatrix}\begin{bmatrix}\hat{x}_1\\ \vdots\\ \hat{x}_M\end{bmatrix}$$

$E_e=\mathbf{e}^T\mathbf{e}=\mathbf{y}^T\mathbf{y}-\hat{\mathbf{x}}^TH^T\mathbf{y}-\mathbf{y}^TH\hat{\mathbf{x}}+\hat{\mathbf{x}}^TH^TH\hat{\mathbf{x}}$ 推出$\hat{\mathbf{x}}=(H^TH)^{-1}H^T\mathbf{y}$

Ps: sufficient condition $\frac{\partial^2E_e}{\partial\mathbf{x}\partial\mathbf{x}^T}=H^TH$ is P.D. And LS estimator is unbiased.

uniqueness: $\hat{\mathbf{x}}_{ls}$ is unique if H has full column rank or H^TH is P.D.

v is zero-mean white noise $E(\mathbf{v}\mathbf{v}^T)=\sigma^2I$,LS has the minimum MSE among all the linear unbiased estimate of x. if $\check{\mathbf{x}}=L\mathbf{y}$ & $E(\check{\mathbf{x}})=E(\mathbf{x})$, $E\{[\mathbf{x}-\hat{\mathbf{x}}][\mathbf{x}-\hat{\mathbf{x}}]^T\}\le E\{[\mathbf{x}-\check{\mathbf{x}}][\mathbf{x}-\check{\mathbf{x}}]^T\}$

$$E[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^T]=(H^TH)^{-1}H^TE(\mathbf{v}\mathbf{v}^T)H(H^TH)^{-1}$$
$$=(H^TH)^{-1}H^T\sigma^2H(H^TH)^{-1}=\sigma^2(H^TH)^{-1}$$

Consider Gaussian additive noise, $\mathbf{z}=\mathbf{s}+\mathbf{v}$, assume z(1), . . . , z(N) and v is zero mean,

LS estimation: $\hat{s}_{LS}=(H^TH)^{-1}H^T[z(1),\ldots,z(N)]^T=\frac{1}{N}(z(1)+\cdots+z(N))$

Likelihood function $f(z(1),\ldots,z(N)|\mathbf{s}=s)=\prod_if(z(i)|\mathbf{s}=s)=\prod_i\frac{1}{\sqrt{2\pi}\sigma_v}e^{-(z(i)-s)^2/2\sigma_v^2}$

Log-likelihood f $\log f(z(1),\ldots,z(N)|\mathbf{s}=s)=C-\frac{1}{2\sigma_v^2}\sum_{i=1}^N(z(i)-s)^2$，$\hat{s}_{ML}=\hat{s}_{LS}$

Weighted Least Squares

$E_e=\sum_{n=1}^Nw_n^2[y_n-h(n)^Tx]^2=(y-Hx)^TW(*)$，$W=\mathrm{diag}(w_1^2,\ldots,w_N^2)\ y=[y_1,\ldots,y_N]^T$

WLSE $\hat{\mathbf{x}}=(H^TWH)^{-1}H^TW\mathbf{y}$ and uniqueness of WLSE requires H^TWH to be P.D..

Recursive Least Squares

H is M×n matrix，and A linearly recursive estimator can be written in the form

$$\mathbf{y}_k=H_k\mathbf{x}+\mathbf{v}_k,\hat{\mathbf{x}}_k=\hat{\mathbf{x}}_{k-1}+K_k(\mathbf{y}_k-H_k\hat{\mathbf{x}}_{k-1})$$

$$\begin{aligned}E(\epsilon_{x,k})&=E(\mathbf{x}-\hat{\mathbf{x}}_k)=E[\mathbf{x}-\hat{\mathbf{x}}_{k-1}-K_k(\mathbf{y}_k-H_k\hat{\mathbf{x}}_{k-1})]\\&=E[\epsilon_{x,k-1}-K_k(H_k\mathbf{x}+\mathbf{v}_k-H_k\hat{\mathbf{x}}_{k-1})]=E[\epsilon_{x,k-1}-K_kH_k(\mathbf{x}-\hat{\mathbf{x}}_{k-1})-K_k\mathbf{v}_k]\\&=(I-K_kH_k)E(\epsilon_{x,k-1})-K_kE(\mathbf{v}_k)\text{ where }\epsilon_{x,k}=\mathbf{x}-\hat{\mathbf{x}}_k\end{aligned}$$

And if $E(v_k)=0$ and $E(\epsilon_{x,k-1})=0$ then $E(\epsilon_{x,k})=0$ and if v_k is zero-mean for all k, initial estimate of x is set as $E(\mathbf{x})$, i.e., $\hat{\mathbf{x}}_0=E(\mathbf{x})$ then $\hat{\mathbf{x}}_k=E(\mathbf{x})$ for all k

Recursive least squares estimation STEP

Measurement equation: $\mathbf{y}_k=H_k\mathbf{x}+v_k$，$E(\mathbf{v}_k)=0,E(v_kv_i^T)=R_k\delta_{k-i}$

1. Initialization：$\hat{\mathbf{x}}_0=E(\mathbf{x}),P_0=E[(\mathbf{x}-\hat{\mathbf{x}}_0)(\mathbf{x}-\hat{\mathbf{x}}_0)^T]$，

2.Iteration：$K_k=P_{k-1}H_k^T(H_kP_{k-1}H_k^T+R_k)^{-1}=P_kH_k^TR_k^{-1}$，$\hat{\mathbf{x}}_k=\hat{\mathbf{x}}_{k-1}+K_k(\mathbf{y}_k-H_k\hat{\mathbf{x}}_{k-1})$

$P_k=(I-K_kH_k)P_{k-1}(I-K_kH_k)^T+K_kR_kK_k^T=(P_{k-1}^{-1}+H_k^TR_k^{-1}H_k)^{-1}=(I-K_kH_k)P_{k-1}$

Assumption:1.测量之前没有关于 x 的知识，P0=∞I；测量之前有很好信息，那么 P0=0。

2.测量噪声每次是独立的，且测量噪声是白噪声

RLS2：assume $C(k)=H(k)^TH(k)$

1. $\hat{x}(1)=(H_1^TH_1)^{-1}H_1y(1)$ Statistical properties of the noise is known or unknown

2. Iteration：$\bar{K}(k)=C(k-1)^{-1}H_k^T[I+H_kC(k-1)^{-1}H_k^T]^{-1}=C(k)^{-1}H_k^T$

$$C(k)^{-1}=[I-\bar{K}(k)H_k]C(k-1)^{-1}=(C(k-1)+H_k^TH_k)^{-1}$$
$$\hat{x}(k)=\hat{x}(k)+\bar{K}(k)[y_k-H_k\hat{x}(k-1)]$$

Example use RLS1：$\hat{\mathbf{x}}_0=E(\mathbf{x}),P_0=E[(\mathbf{x}-\hat{\mathbf{x}}_0)(\mathbf{x}-\hat{\mathbf{x}}_0)^T]$，$R_k=R$

$$P_{k-1}=\frac{P_0R}{(k-1)P_0+R},K_k=\frac{P_0}{kP_0+R}$$

$$\hat{x}_k=\hat{x}_{k-1}+K_k(y_k-\hat{x}_{k-1})=(1-K_k)\hat{x}_{k-1}+K_ky_k=\frac{(k-1)P_0+R}{kP_0+R}\hat{x}_{k-1}+\frac{P_0}{kP_0+R}y_k$$

If x is known perfectly a priori $P_0=0$, and then $K=0$ and $\hat{x}_k=\hat{x}_0$

If x is completely unknown a priori, $P_0\to\infty$ $\hat{x}_k=\frac{1}{k}[(k-1)\hat{x}_{k-1}+y_k]$,即测量均值

Using RLS2：$\hat{x}(k+1)=\hat{x}(k)+C(k+1)^{-1}H_{k+1}^T(y_{k+1}-H_{k+1}\hat{x}(k))$

$H_1=1,\hat{x}(1)=y_1,C(1)=1,C(2)=2,\hat{x}(2)=\frac{1}{2}[\hat{x}(1)+y_2]$，$\hat{x}(k)=\frac{1}{k}\left[\sum_{j=1}^{k-1}y_j+y_k\right]$

Lecture 5 Propagation of states and covariances

Consider the following linear discrete-time system $x_k=F_{k-1}x_{k-1}+G_{k-1}u_{k-1}+w_{k-1}$

其中u_k是已知输入，w_k是由**协方差Q_k**的零均值多元正态分布得到的过程噪声。此外，假设初始状态和每一步的噪声向量$\{x_0,\ w_1,\ \ldots\ldots,\ w_k\}$都是相互独立的。

均值$\bar{x}_k=E(x_k)=F_{k-1}\bar{x}_{k-1}+G_{k-1}u_{k-1}$，$P_k=E[x_k-\bar{x}_k][x_k-\bar{x}_k]^T$

方差$(x_k-\bar{x}_k)(*)^T=(F_{k-1}x_{k-1}+G_{k-1}u_{k-1}+w_{k-1}-\bar{x}_k)(*)^T$

$=[F_{k-1}(x_{k-1}-\bar{x}_{k-1})+w_{k-1}][*]^T=F_{k-1}(x_{k-1}-\bar{x}_{k-1})(x_{k-1}-\bar{x}_{k-1})^TF_{k-1}^T+w_{k-1}w_{k-1}^T+F_{k-1}(x_{k-1}-\bar{x}_{k-1})w_{k-1}^T+w_{k-1}(x_{k-1}-\bar{x}_{k-1})^TF_{k-1}^T$

covariance matrix: $P_k=E[(x_k-\bar{x}_k)(\cdots)^T]=F_{k-1}P_{k-1}F_{k-1}^T+Q_{k-1}$

Sampled-data systems

采样数据系统是一个动力学由连续时间微分方程描述的系统，但输入只在离散时间瞬间发生变化，我们感兴趣的是仅在离散时间瞬间获得状态的均值和协方差，连续时间动力学描述为$\dot{x}=Ax+Bu+w$

解为$x(t_k)=e^{A(t_k-t_{k-1})}x(t_{k-1})+\int_{t_{k-1}}^{t_k}e^{A(t_k-\tau)}[Bu(\tau)+w(\tau)]d\tau$,

Define $F_{k-1}=e^{A\Delta t},G_{k-1}=\int_{t_{k-1}}^{t_k}e^{A(t_k-\tau)}Bd\tau$，$\Delta t=t_k-t_{k-1}$

前提:$w(t)\sim N(0,Q_c(t))$，$E[w(t)w^T(\tau)]=Q_c(t)\delta(t-\tau)$

$:x_k=F_{k-1}x_{k-1}+G_{k-1}u_{k-1}+\int_{t_{k-1}}^{t_k}e^{A(t_k-\tau)}w(\tau)\,d\tau,\bar{x}_k=E(x_k)=F_{k-1}\bar{x}_{k-1}+G_{k-1}u_{k-1}$

$:P_k=E[(x_k-\bar{x}_k)(x_k-\bar{x}_k)^T]=F_{k-1}P_{k-1}F_{k-1}^T+\int_{t_{k-1}}^{t_k}e^{A(t_k-\tau)}Q_c(\tau)e^{A^T(t_k-\tau)}d\tau$

Define $Q_{k-1}=\int_{t_{k-1}}^{t_k}e^{A(t_k-\tau)}Q_c(\tau)e^{A^T(t_k-\tau)}d\tau$，$P_k=F_{k-1}P_{k-1}F_{k-1}^T+Q_{k-1}$

And for small values of Δt，$e^{A(t_k-\tau)}\approx I\ for\ \tau\in[t_{k-1},t_k],Q_{k-1}\approx Q_c(t_k)\Delta t$

Continuous-time systems

Lecture6 Kalman filter

Suppose we have a linear discrete-time system given as follows:

$$x_k=F_{k-1}x_{k-1}+G_{k-1}u_{k-1}+w_{k-1}$$
$$y_k=H_kx_k+v_k$$

$w_k\sim\mathcal{N}(0,Q_k),v_k\sim\mathcal{N}(0,R_k)$，不相关

$E[w_kw_j^T]=Q_k\delta_{k-j},E[v_kv_j^T]=R_k\delta_{k-j},E[v_kw_j^T]=0$

a posteriori estimate $\hat{x}_k^+=E[x_k|y_1,y_2,\cdots,y_k]$ 后验估计

a priori estimate $\hat{x}_k^-=E[x_k|y_1,y_2,\cdots,y_{k-1}]$ 先验

$\hat{x}_{k|k+N}=E[x_k|y_1,y_2,\cdots,y_k,\cdots,y_{k+N}]$ =smoothed estimate

$\hat{x}_{k|k-M}=E[x_k|y_1,y_2,\cdots,y_{k-M}]=$ predicted estimate

$\hat{x}_0^+$ our initial estimate of x0 before any measurements are available in general $\hat{x}_0^+=E(x_0)$

the covariance of the estimation error of $\hat{x}_k^-$, $P_k^-=E[(x_k-\hat{x}_k^-)(x_k-\hat{x}_k^-)^T]$

the covariance of the estimation error of $\hat{x}_k^+$, $P_k^+=E[(x_k-\hat{x}_k^+)(x_k-\hat{x}_k^+)^T]$

The discrete-time Kalman filter

dynamic system is given by the following equations:

$x_k=F_{k-1}x_{k-1}+G_{k-1}u_{k-1}+w_{k-1}$，$w_k\sim\mathcal{N}(0,Q_k),v_k\sim\mathcal{N}(0,R_k)$ 不相关
$y_k=H_kx_k+v_k$

$E[w_kw_j^T]=Q_k\delta_{k-j},E[v_kv_j^T]=R_k\delta_{k-j}E[v_kw_j^T]=0$

The Kalman filter is initialized as follows

$$\hat{x}_0^+=E(x_0),\ P_0^+=E[(x_0-\hat{x}_0^+)(x_0-\hat{x}_0^+)^T]$$

The Kalman filter is given by the following equations:

$P_k^-=F_{k-1}P_{k-1}^+F_{k-1}^T+Q_{k-1}$；$K_k=P_k^-H_k^T(H_kP_k^-H_k^T+R_k)^{-1}=P_k^+H_k^TR_k^{-1}$

$\hat{x}_k^-=F_{k-1}\hat{x}_{k-1}^++G_{k-1}u_{k-1}$，a priori state estimate

$\hat{x}_k^+=\hat{x}_k^-+K_k(y_k-H_k\hat{x}_k^-)$，a posteriori state estimate

$P_k^+=(I-K_kH_k)P_k^-(I-K_kH_k)^T+K_kR_kK_k^T=((P_k^-)^{-1}+H_k^TR_k^{-1}H_k)^{-1}=(I-K_kH_k)P_k^-$

P_k^+的第一个表达式称为协方差测量更新方程的 Joseph 稳定版本，它比P_k^+的第三个表达式更稳定和稳健；P_k^+的第一个表达式保证了P_k^+总是对称的半正定的，只要P_k^-是对称的半正定的；P_k^+的第三个表达式在计算上比第一个表达式简单，但不能保证P_k^+的对称性或半正定性；P_k^+的第二种形式很少实现，但在信息过滤器的推导中会很有用

AND 使用K_k第二表达式必须搭配P_k^+的第二表达式

AND if x_k常值，$F_k=I,Q_k=0,u_k=0$，退化为常值向量的 RLS 估计

P_k^+、Kk 和P_k^-的计算不依赖于测量值 yk，而只依赖于系统参数 Fk、Hk、Qk 和 Rk 通过预先计算，可以在实时运行期间节省计算 Kk 的计算量。可以在滤波器实际运行之前研究和评估滤波器的性能（P_k^+ 表示精度）

Example: $\mathbf{x}_k=\begin{bmatrix}1&\Delta t\\0&1\end{bmatrix}\mathbf{x}_{k-1}+\begin{bmatrix}0\\\Delta t\end{bmatrix}\mathbf{u}_{k-1}+\mathbf{w}_{k-1}$，$y_k=\begin{bmatrix}1&0\end{bmatrix}\mathbf{x}_k+v_k$，$v_k\sim\mathcal{N}(0,0.05)$,$\mathbf{w}_k\sim\mathcal{N}(\mathbf{0},(0.1)\mathbf{1}_{2\times2})$ $\hat{x}_0\sim\mathcal{N}\left(\begin{bmatrix}0\\5\end{bmatrix},\begin{bmatrix}0.01&0\\0&1\end{bmatrix}\right)$,$\Delta t=0.5s,u_0=-2[m/s^2],y_1=2.2$

$[m]$;**prediction**:$\hat{\mathbf{x}}_1^-=\mathbf{F}_{k-1}\mathbf{x}_{k-1}+\mathbf{G}_{k-1}\mathbf{u}_{k-1}$: $\begin{bmatrix}\hat{p}_1^-\\\hat{\dot{p}}_1^-\end{bmatrix}=\begin{bmatrix}1&0.5\\0&1\end{bmatrix}\begin{bmatrix}0\\5\end{bmatrix}+\begin{bmatrix}0\\0.5\end{bmatrix}(-2)=\begin{bmatrix}2.5\\4\end{bmatrix}$

$\mathbf{P}_k^-=\mathbf{F}_{k-1}\mathbf{P}_{k-1}^+\mathbf{F}_{k-1}^T+\mathbf{Q}_{k-1}=\begin{bmatrix}1&0.5\\0&1\end{bmatrix}\begin{bmatrix}0.01&0\\0&1\end{bmatrix}\begin{bmatrix}1&0.5\\0&1\end{bmatrix}^T+\begin{bmatrix}0.1&0\\0&0.1\end{bmatrix}=\begin{bmatrix}0.36&0.5\\0.5&1.1\end{bmatrix}$

Correction:

$$\mathbf{K}_1=\hat{\mathbf{P}}_1^-\mathbf{H}_1^T(\mathbf{H}_1\hat{\mathbf{P}}_1^-\mathbf{H}_1^T+\mathbf{R}_1)^{-1}:=\begin{bmatrix}0.36&0.5\\0.5&1.1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\left(\begin{bmatrix}1&0\end{bmatrix}\begin{bmatrix}0.36&0.5\\0.5&1.1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}+0.05\right)^{-1}=\begin{bmatrix}0.88\\1.22\end{bmatrix}$$

$$\hat{\mathbf{x}}_1^+=\hat{\mathbf{x}}_1^-+\mathbf{K}_1(\mathbf{y}_1-\mathbf{H}_1\hat{\mathbf{x}}_1^-):\begin{bmatrix}\hat{p}_1^+\\\hat{\dot{p}}_1^+\end{bmatrix}=\begin{bmatrix}2.5\\4\end{bmatrix}+\begin{bmatrix}0.88\\1.22\end{bmatrix}\left(2.2-[10]\begin{bmatrix}2.5\\4\end{bmatrix}\right)=\begin{bmatrix}2.24\\3.63\end{bmatrix}$$

$$\hat{\mathbf{P}}_1^+=(\mathbf{1}-\mathbf{K}_1\mathbf{H}_1)\hat{\mathbf{P}}_1^-=\begin{bmatrix}0.04&0.06\\0.06&0.49\end{bmatrix}$$

Some properties :unbiased , Consistency of state estimators; overall optimality

Causes of inconsistency: Modeling errors; Numerical errors; Programming errors

Orthogonality principle in discrete-time Kalman filter:

Since $\hat{x}_k^-$ should be the optimum LMMSE estimate, $E[(x_k-\hat{x}_k^-)y_i^T]=0,\forall i=1,2,\ldots,k-1$

Steady-state Kalman filter 稳态卡尔曼滤波

Some properties: the dynamic system is time-invariant; a constant gain K can be pre-computed, $K=PH^T(HPH^T+R)^{-1}$,and P reaches a steady-state $P=FPF^T-FPH^T(HPH^T+R)^{-1}HPF^T+Q$, is called discrete time algebraic Riccati equation (DARE).

Conditions for the existence of the steady-state Kalman filter iff both of the following conditions hold. 1. (F, H) is detectable: A system is detectable if all the unobservable states are stable. **2. (F, J) is stabilizable** (J is any matrix such that $JJ^T=Q$)：A system is said to be stabilizable when all uncontrollable state variables can be made to have stable dynamics.

(F, H) is detectable:[H HF …H^(n-1)F]^T 秩为 n

AND **如果信号处理模型是时不变且渐近稳定的**：3.1 对于任意非负对称初始条件 P_0^-，有$\lim_{k\to\infty}P_k^-=P$，AND P **满足 DARE;3.2 卡尔曼滤波器增益 K 达到恒定值**，矩阵(I-KH)F 稳定。如果 DARE 具有唯一的正半定解，则稳态卡尔曼滤波器是稳定的。

Example in 1-dimension: $x_{k+1}=\varphi x_k+w_k$，$y_k=hx_k+v_k$， wk and vk are stationary random process, with $w_k\sim\mathcal{N}(0,q),v_k\sim\mathcal{N}(0,r)$,DARE as follows

$p=\varphi^2p-\varphi ph\frac{1}{h^2p+r}hp\varphi+q,h^2p^2+(r-\varphi^2r-h^2q)p-qr=0$

Solving the second-order equation (2) gives the solution of steady-state p.

consider two special cases.1.r=0,p=q, k=1/h and $\hat{x}_{k+1}=\frac{\varphi}{h}y_k$，只取决于测量（没有误差）

2. the model is accurate: q = 0. p=0 ,and then $\hat{x}_{k+1}=\varphi\hat{x}_k$

Kalman filter generalizations: the general discrete-time Kalman filter

Suppose we have a linear discrete-time system given as follows:

$$x_k=F_{k-1}x_{k-1}+G_{k-1}u_{k-1}+w_{k-1}$$
$$y_k=H_kx_k+v_k$$

$w_k\sim\mathcal{N}(0,Q_k),v_k\sim\mathcal{N}(0,R_k)$，不相关

$E[w_kw_j^T]=Q_k\delta_{k-j},E[v_kv_j^T]=R_k\delta_{k-j},E[v_kw_j^T]=M_j\delta_{k-j+1}$

The Kalman filter is initialized as follows

$$\hat{x}_0 = E(x_0),\ \hat{P}_0 = E[(x_0-\hat{x}_0)(x_0-\hat{x}_0)^T]$$

The Kalman filter is given by the following equations:

$\check{P}_k = F_{k-1}\hat{P}_{k-1}F_{k-1}^T + Q_{k-1}$; $K_k = (\check{P}_k H_k^T + M_k)(H_k\check{P}_k H_k^T + H_k M_k + M_k^T H_k^T + R_k)^{-1}$

$\check{x}_k = F_{k-1}\hat{x}_{k-1} + G_{k-1}u_{k-1}$, $\hat{x}_k = \check{x}_k + K_k(y_k - H_k\check{x}_k)$,

$\hat{P}_k = (I - K_kH_k)\check{P}_k(I-K_kH_k)^T + K_kR_kK_k^T + K_k(H_kM_k + M_k^TH_k^T)K_k^T - M_kK_k^T - K_kM_k^T$

$= \check{P}_k - K_k(H_k\check{P}_k + M_k^T)$

The discrete-time extended Kalman filter

1. The system and measurement equations are given as follows:

$x_k = f_{k-1}(x_{k-1}, u_{k-1}, w_{k-1}), y_k = h_k(x_k, v_k), w_k \sim \mathcal{N}(0, Q_k), v_k \sim \mathcal{N}(0, R_k)$

2. Initialize the filter as follows: $\hat{x}_0 = E(x_0)$, $\hat{P}_0 = E[(x_0-\hat{x}_0)(x_0-\hat{x}_0)^T]$

3. For k = 1, 2, . . . , perform the following.

compute the partial derivative matrices: $F_{k-1} = \frac{\partial f_{k-1}}{\partial x}|_{(\hat{x}_{k-1},0)}$, $L_{k-1} = \frac{\partial f_{k-1}}{\partial w}|_{(\hat{x}_{k-1},0)}$

perform the time update: $\check{P}_k = F_{k-1}\hat{P}_{k-1}F_{k-1}^T + L_{k-1}Q_{k-1}L_{k-1}^T$, $\check{x}_k = f_{k-1}(\hat{x}_{k-1}, u_{k-1}, 0)$

compute the partial derivative matrices: $H_k = \frac{\partial h_k}{\partial x}|_{(\check{x}_k,0)}, M_k = \frac{\partial h_k}{\partial v}|_{(\check{x}_k,0)}$

perform the measurement update: $K_k = \check{P}_kH_k^T(H_k\check{P}_kH_k^T + M_kR_kM_k^T)^{-1}$

$\hat{x}_k = \check{x}_k + K_k[y_k - h_k(\check{x}_k, 0)]$, $\hat{P}_k = (I - K_kH_k)\check{P}_k$

EKF 的不充分性：动态系统的一阶近似可能会在变换（高斯）随机变量的真实后验均值和协方差中引入较大的误差，这可能导致滤波器性能次优，有时甚至导致滤波器发散。

Example for EKF: $\mathbf{x} = [p, \dot{p}]^T$ $\mathbf{u} = \ddot{p}$, Motion /process model &measurement model

$$\mathbf{X}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) = \begin{bmatrix}1 & \Delta t\\0 & 1\end{bmatrix}\mathbf{x}_{k-1} + \begin{bmatrix}0\\\Delta t\end{bmatrix}\mathbf{u}_{k-1} + \mathbf{w}_{k-1}$$

$$y_k = \phi_k = h(p_k, v_k) = \tan^{-1}\left(\frac{S}{D-p_k}\right) + v_k$$

$\boldsymbol{v_k \sim \mathcal{N}(0, 0.01)}$, $\mathbf{w_k} \sim \boldsymbol{\mathcal{N}(0, (0.1)1_{2\times2})}$

$\mathbf{F}_{k-1} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}_{k-1}}|_{\hat{\mathbf{x}}_{k-1},\mathbf{u}_{k-1},\mathbf{0}} = \begin{bmatrix}1 & \Delta t\\0 & 1\end{bmatrix}, \mathbf{L}_{k-1} = \frac{\partial \mathbf{f}}{\partial \mathbf{w}_{k-1}}|_{\hat{\mathbf{x}}_{k-1},\mathbf{u}_{k-1},\mathbf{0}} = \mathbf{1}_{2\times2}$

$\mathbf{H}_k = \frac{\partial h}{\partial \mathbf{x}_k}|_{\mathbf{x}_k,\mathbf{0}} = [\frac{S}{(D-\tilde{p}_k)^2 + S^2}\ 0], M_k = \frac{\partial h}{\partial v_k}|_{\mathbf{x}_k,\mathbf{0}} = 1$

$\hat{\mathbf{x}}_0 \sim \mathcal{N}\left(\begin{bmatrix}0\\5\end{bmatrix}, \begin{bmatrix}0.01 & 0\\0 & 1\end{bmatrix}\right), \Delta t = 0.5s, u_0 = -2[m/s^2], y_1 = 30[\text{deg}], S = 20[m] D = 40[m]$

Prediction: $\check{\mathbf{x}}_1 = \mathbf{f}_0(\hat{\mathbf{x}}_0, \mathbf{u}_0, \mathbf{0})$; $\begin{bmatrix}\check{p}_1\\\check{\dot{p}}_1\end{bmatrix} = \begin{bmatrix}1 & 0.5\\0 & 1\end{bmatrix}\begin{bmatrix}0\\5\end{bmatrix} + \begin{bmatrix}0\\0.5\end{bmatrix}(-2) = \begin{bmatrix}2.5\\4\end{bmatrix}$, $\check{\mathbf{P}}_1 = \mathbf{F}_0\hat{\mathbf{P}}_0\mathbf{F}_0^T + \mathbf{L}_0\mathbf{Q}_0\mathbf{L}_0^T$, $\check{\mathbf{P}}_1 = \begin{bmatrix}1 & 0.5\\0 & 1\end{bmatrix}\begin{bmatrix}0.01 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0.5 & 1\end{bmatrix} + \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}0.1 & 0\\0 & 0.1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}0.36 & 0.5\\0.5 & 1.1\end{bmatrix}$

Correction: $\mathbf{K}_1 = \check{\mathbf{P}}_1\mathbf{H}_1^T(\mathbf{H}_1\check{\mathbf{P}}_1\mathbf{H}_1^T + \mathbf{M}_1\mathbf{R}_1\mathbf{M}_1^T)^{-1}$

$$\mathbf{K1} = \begin{bmatrix}0.36 & 0.5\\0.5 & 1.1\end{bmatrix}\begin{bmatrix}0.011\\0\end{bmatrix}\left([0.011\ \ 0]\begin{bmatrix}0.36 & 0.5\\0.5 & 1.1\end{bmatrix}\begin{bmatrix}0.011\\0\end{bmatrix} + 1(0.01)(1)\right)^{-1} = \begin{bmatrix}0.39\\0.55\end{bmatrix}$$

$$\hat{\mathbf{x}}_1 = \check{\mathbf{x}}_1 + \mathbf{K}_1(\mathbf{y}_1 - \mathbf{h}_1(\check{\mathbf{x}}_1, \mathbf{0})) = \begin{bmatrix}2.5\\4\end{bmatrix} + \begin{bmatrix}0.39\\0.55\end{bmatrix}(0.52 - 0.49) = \begin{bmatrix}2.51\\4.02\end{bmatrix}$$

$$\hat{\mathbf{P}}_1 = (\mathbf{1} - \mathbf{K}_1\mathbf{H}_1)\check{\mathbf{P}}_1 = \begin{bmatrix}0.3585 & 0.4979\\0.4978 & 1.0970\end{bmatrix}$$

Unscented transform

1.For vectors x ~ N (m, P), the generalization of standard deviation σ is the Cholesky factor $L = \sqrt{P}$: $P = LL^T$ 2. The (2n + 1) sigma points can be formed using columns of L:

$X_0 = m$, $X_i = m + \sqrt{n+\lambda}L_i$, $X_{n+i} = m - \sqrt{n+\lambda}L_i$ []i 表示矩阵第 i 列

3.对于变换 y=g(x)，估计如下：

$$E[g(x)] = \sum_{i=0}^{2n} W_i^{(m)}g(X_i), Cov[g(x)] = \sum_{i=0}^{2n} W_i^{(c)}(g(X_i) - \mu_y)(g(X_i) - \mu_y)^T$$

参数设置：$\lambda = \alpha^2(n+\kappa) - n$；$\alpha$ and κ determine the spread of the sigma points；

Weights $W_i^{(m)}$ and $W_i^{(c)}$ are given as follows:

$$W_0^{(m)} = \frac{\lambda}{n+\lambda}, W_0^{(c)} = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta); W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n+\lambda)}, i = 1, \ldots, 2n$$

β can be used for incorporating priori information on the (non-Gaussian) distribution of x.

Unscented Kalman Filter (UKF)

1.Prediction step n=维数？

1.1 Form the matrix of sigma points:

$X_{k-1} = [\hat{x}_{k-1}\ \cdots\ \hat{x}_{k-1}] + \sqrt{n+\lambda}[0\ \ \sqrt{\hat{P}_{k-1}}\ \ -\sqrt{\hat{P}_{k-1}}]$

1.2 Propagate the sigma points through the dynamic model: $\hat{X}_{k,i} = f(X_{k-1,i}), i = 0, 1, \ldots, 2n$

1.3 Compute the predicted mean and covariance

$\check{x}_k = \sum_i W_i^{(m)}\hat{X}_{k,i}$, $\check{P}_k = \sum_i W_i^{(c)}(\hat{X}_{k,i} - \check{x}_k)(\hat{X}_{k,i} - \check{x}_k)^T + Q_{k-1}$

2. Update step

2.1 Form the matrix of sigma points: $\check{X}_k = [\check{x}_k\ \cdots\ \check{x}_k] + \sqrt{n+\lambda}[0\ \ \sqrt{\check{P}_k}\ \ -\sqrt{\check{P}_k}]$

2.2 Propagate the sigma points through the measurement model:

$$\hat{Y}_{k,i} = h(\check{X}_{k,i}), i = 0, 1, \ldots, 2n$$

2.3 Compute: $\mu_k = \sum_i W_i^{(m)}\hat{Y}_{k,i}$, $S_k = \sum_i W_i^{(c)}(\hat{Y}_{k,i} - \mu_k)(\hat{Y}_{k,i} - \mu_k)^T + R_k$

$C_k = \sum_i W_i^{(c)}(\check{X}_{k,i} - \check{x}_k)(\hat{Y}_{k,i} - \mu_k)$

2.4 Compute the filter gain K_k and the filtered state mean m_k and covariance $\hat{P}_k$, conditional to the measurement y_k: $K_k = C_kS_k^{-1}$, $\hat{x}_k = \check{x}_k + K_k(y_k - \mu_k)$, $\hat{P}_k = \check{P}_k - K_kS_kK_k^T$

EXAMPLE for UKF：**Prediction**: $n = 2, \lambda = 1$ $\sqrt{\hat{P}_0} = \begin{bmatrix}0.1 & 0\\0 & 1\end{bmatrix}$, choose 5 sigma points:

$\hat{x}_0^{(0)} = \hat{x}_0, \hat{x}_0^{(i)} = \hat{x}_0 + \sqrt{3}[\sqrt{\hat{P}_0}]_i, i = 1,2, \hat{x}_0^{(i+2)} = \hat{x}_0 - \sqrt{3}[\sqrt{\hat{P}_0}]_i, i = 1,2$

$\hat{x}_0^{(0)} = [0,5]^T, \hat{x}_0^{(1)} = [0.2,5]^T, \hat{x}_0^{(2)} = [0,6.7]^T, \hat{x}_0^{(3)} = [-0.2,5]^T, \hat{x}_0^{(4)} = [0,3.3]^T$

Prediction: $W_0^{(m)} = W_0^{(c)} = 1/3, W_i^{(m)} = W_i^{(c)} = 1/6, i = 1, \ldots, 4$

$\check{x}_1^{(i)} = f_0(\hat{x}_0^{(i)}, u_0, 0), i = 0, 1, \ldots, 4$

$\check{x}_1^{(0)} = [2.5,4]^T, \check{x}_1^{(1)} = [2.7,4]^T, \check{x}_1^{(2)} = [3.4,5.7]^T, \check{x}_1^{(3)} = [2.3,4]^T, \check{x}_1^{(4)} = [1.6,2.3]^T$

$\check{x}_1 = \sum_{i=0}^4 W_i\check{x}_1^{(i)} = [2.5,4]^T$, $\check{P}_k = \sum_{i=0}^4 W_i(\check{x}_k^i - \check{x}_k)(\check{x}_k^i - \check{x}_k)^T + Q_{k-1} = \begin{bmatrix}0.36 & 0.5\\0.5 & 1.1\end{bmatrix}$

Correction: $\sqrt{\check{P}_1} = \begin{bmatrix}0.51 & 0\\0.98 & 0.20\end{bmatrix}$, choose 5 sigma points,

$\check{x}_1^{(0)} = \check{x}_1, \check{x}_1^{(i)} = \check{x}_1 + \sqrt{3}[\sqrt{\check{P}_1}]_i, i = 1,2, \check{x}_1^{(i+2)} = \check{x}_1 - \sqrt{3}[\sqrt{\check{P}_1}]_i, i = 1,2$; $\check{x}_1^{(0)} = [2.5,4]^T$

$, \check{x}_1^{(1)} = [3.54,5.44]^T, \check{x}_1^{(2)} = [2.5,5.10]^T, \check{x}_1^{(3)} = [1.46,2.56]^T, \check{x}_1^{(4)} = [2.5,2.90]^T$

the output $\hat{y}_1^{(i)} = h_1(\check{x}_1^{(i)}, 0), i = 0, \ldots, 2n$,

$\hat{y}_1^{(0)} = 28.1, \hat{y}_1^{(1)} = 28.7, \hat{y}_1^{(2)} = 28.1, \hat{y}_1^{(3)} = 27.4, \hat{y}_1^{(4)} = 28.1$

$u_1 = \sum_{i=0}^{2n} W_i^{(m)}\hat{y}_1^{(i)} = 28.1, S_1 = \sum_{i=0}^{2n} W_i^{(c)}(\hat{y}_k^{(i)} - \mu_k)(\hat{y}_k^{(i)} - \mu_k)^T + R_k = 0.16$

$C_1 = \sum_{i=0}^{2n} W_c^{(i)}(\check{x}_k^{(i)} - \check{x}_k)(\hat{y}_k^{(i)} - \mu_k)^T = [0.23, 0.32]^T, K_1 = C_1S_1^{-1} = [1.47, 2.05]^T$

$\hat{x}_1 = \check{x}_1 + K_1(y_1 - \mu_1) = [5.33, 7.93]^T, \hat{P}_1 = \check{P}_1 - K_1S_1K_1^T = \begin{bmatrix}0.0143 & 0.0178\\0.0178 & 0.4276\end{bmatrix}$

Comparison of EKF and UKF: 局部近似 vs 大面积近似；需要 F 和 h 的可微性 vs 不需要；封闭形式的导数或期望 vs 不需要这些形式；需要非线性动力学的一阶近似 vs 捕获高阶分布矩 **disadvantage of UKF：Not a truly global approximation,** based on a small set of trial points**. Does not work well with nearly singular covariances,** i.e., with nearly deterministic systems. **Requires more computations** than EKF, e.g., Cholesky factorizations on every step**.** Can only be applied to **models driven by Gaussian noises**

HW2 (1)Suppose that z = s+v, where s and v are independent, jointly distributed RVs with s ~ N (η, σ2) and v ~ N (0, V 2). (a) Derive an expression for E[s|z = z].

(b) Derive an expression for E[s²|z = z].

$f_s(s|\mathbf{z} = z) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma^2V^2}{\sigma^2+V^2}}}\exp\left[-\frac{(s-\eta-\frac{\sigma^2}{V^2+\sigma^2}(z-\eta))^2}{2\frac{\sigma^2V^2}{\sigma^2+V^2}}\right]$,

$E[s|z = z] = \eta + \frac{\sigma^2}{V^2+\sigma^2}(z - \eta)$

$E[s^2|z = z] == D[s|z = z] + E[s|z = z]^2 = \frac{\sigma^2V^2}{\sigma^2+V^2} + \left[\eta + \frac{\sigma^2}{V^2+\sigma^2}(z-\eta)\right]^2$

(2) Suppose that z = s+v, where s and v are independent, jointly distributed, RVs with s ~ N (ηs, σs2) and v ~ N (0, σv2). Assume we have measurements z(1), . . . , z(n),

(a) Derive the maximum likelihood estimate for s;

(b) Derive the maximum a posteriori estimate for s;

(c) Derive the minimum mean square estimate for s;

(d) Derive the linear minimum mean square estimate for s;

MLE: $f(z(1), \ldots, z(N)|\mathbf{s} = s) = \prod_i f(z(i)|\mathbf{s} = s) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_v}e^{-(z(i)-s)^2/2\sigma_v^2}$

$\log f(z(1), \ldots, z(N)|\mathbf{s} = s) = C - \frac{1}{2\sigma_v^2}\sum_{i=1}^N (z(i) - s)^2$, $\hat{s}_{\text{ML}} = \frac{1}{N}(z(1) + \cdots + z(N))$

MAPE: $f(\mathbf{s} = s|z(1), \ldots, z(n)) = \frac{f(z(1),\ldots,z(n)|\mathbf{s}=s)\times f_s(s)}{f(z(1),\ldots,z(n))}, lnf = g(s) = C - \frac{1}{2\sigma_v^2}\sum_{i=1}^n (z(i) - s)^2 - \frac{1}{2\sigma_s^2}(s - \eta_s)^2, g'(s) = -\frac{1}{2\sigma_v^2}\sum_{i=1}^n 2(s - z(i)) - \frac{1}{2\sigma_s^2}2(s - \eta_s), \hat{s}_{MAP} = \frac{1}{n\sigma_s^2+\sigma_v^2}\{\sigma_s^2[z(1)+\ldots+z(n)] + \sigma_v^2\eta_s\}$

MMSE: We first demonstrate that s, z(1), . . . , z(n) are jointly Gaussian, which is true as the linear combination of x, z(1), . . . , z(n) are Gaussian, i.e.,

$Y = a_0\boldsymbol{s} + a_1\mathbf{z}(1) + \cdots + a_n\mathbf{z}(n) = (\sum_{i=0}^n a_i)s + \sum_{i=1}^n a_iv(i)$

is Gaussian with mean $\sum_{i=0}^n a_i\eta_s$, and var $(\sum_{i=0}^n a_i)^2\sigma_s^2 + \sum_{i=1}^n a_i^2\sigma_v^2$. Similarly, z(1), . . . , z(n) are also jointly Gaussian. Assume z = [z(1), . . . , z(n)]T , and as s and z are jointly Gaussian, we have $(s, \mathbf{z}) \sim \mathcal{N}\left(\begin{bmatrix}\mu_s\\\mu_z\end{bmatrix}, \begin{bmatrix}\Sigma_{ss} & \Sigma_{sz}\\\Sigma_{zs} & \Sigma_{zz}\end{bmatrix}\right)$, According to Schur complement, we have

$$\begin{bmatrix}\Sigma_{ss} & \Sigma_{sz}\\\Sigma_{zs} & \Sigma_{zz}\end{bmatrix} = \begin{bmatrix}I & \Sigma_{sz}\Sigma_{zz}^{-1}\\0 & I\end{bmatrix}\begin{bmatrix}\Sigma_{ss} - \Sigma_{sz}\Sigma_{zz}^{-1}\Sigma_{zs} & 0\\0 & \Sigma_{zz}\end{bmatrix}\begin{bmatrix}I & 0\\\Sigma_{zz}^{-1}\Sigma_{zs} & I\end{bmatrix},$$

the joint distribution p(s, z) is $p(s, \mathbf{z}) = \frac{1}{\sqrt{(2\pi)^{n+1}det\Sigma}}\exp\left(-\frac{1}{2}(X - \mu_X)^T\Sigma^{-1}(X - \mu_X)\right)$

in which X = [s, zᵀ]ᵀ, $\Sigma = \begin{bmatrix}\Sigma_{ss} & \Sigma_{sz}\\\Sigma_{zs} & \Sigma_{zz}\end{bmatrix}$,and the quadratic part is

$(X - \mu_X)^T\Sigma^{-1}(X - \mu_X)$

$= [(s-\eta_s)^T, (z-\mu_z)^T]\cdot\begin{bmatrix}I & 0\\-\Sigma_{zz}^{-1}\Sigma_{zs} & I\end{bmatrix}\begin{bmatrix}(\Sigma_{ss} - \Sigma_{sz}\Sigma_{zz}^{-1}\Sigma_{zs})^{-1} & 0\\0 & \Sigma_{zz}^{-1}\end{bmatrix}\begin{bmatrix}I & -\Sigma_{sz}\Sigma_{zz}^{-1}\\0 & I\end{bmatrix}$

$\cdot\begin{bmatrix}s - \eta_s\\z - \mu_z\end{bmatrix} = [(s-\eta_s)^T - (z-\mu_z)^T\Sigma_{zz}^{-1}\Sigma_{zs}z - \mu_z]\cdot\begin{bmatrix}(\Sigma_{ss} - \Sigma_{sz}\Sigma_{zz}^{-1}\Sigma_{zs})^{-1} & 0\\0 & \Sigma_{zz}^{-1}\end{bmatrix}$

$\cdot\begin{bmatrix}s - \eta_s - \Sigma_{ss}\Sigma_{zz}^{-1}(\mathbf{z} - \mu_z)\\\mathbf{z} - \mu_z\end{bmatrix}$

$= [(s-\eta_s) - \Sigma_{sz}\Sigma_{zz}^{-1}(z-\mu_z)]^T(\Sigma_{ss} - \Sigma_{sz}\Sigma_{zz}^{-1}\Sigma_{zs})^{-1}[\cdots] + (z-\mu_z)^T\Sigma_{zz}^{-1}(z-\mu_z)$

the determinant $det\left(\begin{bmatrix}\Sigma_{ss} & \Sigma_{sz}\\\Sigma_{zs} & \Sigma_{zz}\end{bmatrix}\right) = det(\Sigma_{zz})\cdot det(\Sigma_{ss} - \Sigma_{zs}\Sigma_{zz}^{-1}\Sigma_{sz})$

as $p(s, \mathbf{z}) = p(s|\mathbf{z})p(\mathbf{z})$ and then $p(s|\mathbf{z}) = \mathcal{N}(\eta_s + \Sigma_{sz}\Sigma_{zz}^{-1}(\mathbf{z} - \mu_z), \Sigma_{ss} - \Sigma_{sz}\Sigma_{zz}^{-1}\Sigma_{zs})$

and $p(\mathbf{z}) = \mathcal{N}(\mu_z, \Sigma_{zz})$. Hence the MMSE estimate is $E(s|\mathbf{z}) = \eta_s + \Sigma_{sz}\Sigma_{zz}^{-1}(\mathbf{z} - \mu_z)$

(d) The linear MMSE: $\hat{s}_{\text{LMMSE}} = E[\boldsymbol{s}\mathbf{z}^T][E(\mathbf{z}\mathbf{z}^T)]^{-1}\mathbf{z}$ **in which**

$E[\boldsymbol{sz}^T] = E[\boldsymbol{sz}(1), \ldots, \boldsymbol{sz}(n)] = [\eta_s^2 + \sigma_s^2\ \ \cdots\ \ \eta_s^2 + \sigma_s^2]$ and then

$$E[\mathbf{z}\mathbf{z}^T] = \begin{bmatrix}E[\mathbf{z}(1)^2] & \cdots & E[\mathbf{z}(1)\mathbf{z}(n)]\\\vdots & \ddots & \vdots\\E[\mathbf{z}(n)\mathbf{z}(1)] & \cdots & E[\mathbf{z}(n)^2]\end{bmatrix} = \begin{bmatrix}\eta_s^2+\sigma_s^2+\sigma_v^2 & \eta_s^2+\sigma_s^2 & \cdots & \eta_s^2+\sigma_s^2\\\eta_s^2+\sigma_s^2 & \eta_s^2+\sigma_s^2+\sigma_v^2 & \cdots & \eta_s^2+\sigma_s^2\\\vdots & \vdots & \ddots & \vdots\\\eta_s^2+\sigma_s^2 & \eta_s^2+\sigma_s^2 & \cdots & \eta_s^2+\sigma_s^2+\sigma_v^2\end{bmatrix}$$

In order to calculate the inversion of E[zzT], we represent it as

$$E[\mathbf{z}\mathbf{z}^T] = \begin{bmatrix}\sigma_v^2 & \cdots & 0\\\vdots & \ddots & \vdots\\0 & \cdots & \sigma_v^2\end{bmatrix} + \begin{bmatrix}1\\\vdots\\1\end{bmatrix}(\sigma_s^2 + \eta_s^2)[1\ \ \cdots\ \ 1]$$

According to the matrix inversion lemma, i.e.,

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}$$

$$\{E[\mathbf{z}\mathbf{z}^T]\}^{-1} = \frac{1}{\sigma_v^2[\sigma_v^2 + n(\sigma_s^2 + \eta_s^2)]}\begin{bmatrix}\sigma_v^2 + (n-1)(\sigma_s^2 + \eta_s^2) & \cdots & -(\sigma_s^2 + \eta_s^2)\\\vdots & \ddots & \vdots\\-(\sigma_s^2 + \eta_s^2) & \cdots & \sigma_v^2 + (n-1)(\sigma_s^2 + \eta_s^2)\end{bmatrix}$$

Therefore, the LMMSE estimate is

$$\hat{s}_{\text{LMMSE}} = \frac{\sigma_s^2 + \eta_s^2}{\sigma_v^2 + n(\sigma_s^2 + \eta_s^2)}\sum_{i=1}^n z(i)$$

证明俩随机变量联合高斯且不相关，则独立。

Notedx ~ (X̄, 6x²), Y ~ (Ȳ, 6y²)

$$C_z = E[(z - E(z))(z - E(z)^T] = \begin{bmatrix}\mathbb{E}[(x-\bar{x})^2] & \mathbb{E}[(x-\bar{x})(y-\bar{y})]\\\mathbb{E}[(y-\bar{x})(x-\bar{x})] & \mathbb{E}[(y-\bar{x})^2]\end{bmatrix}$$

$$= \begin{bmatrix}6x^2 & \mathbb{E}(XY) - 2\bar{X}\bar{Y} + \bar{X}\bar{Y}\\E(xY) - 2\bar{x}\bar{Y} + \bar{x}\bar{Y} & \sigma_Y^2\end{bmatrix} = \begin{bmatrix}6x^2 & 0\\0 & 6y^2\end{bmatrix}$$

$$C_2^{-1} = \begin{bmatrix}\frac{1}{6x^2} & 0\\0 & \frac{1}{6x^2}\end{bmatrix}, |det(C_z)|^{1/2} = 6x6y$$

带入原始式子即可

$$\text{pdf}(X,Y) = f_{X,Y}(x,y) = \frac{1}{2\pi|det(C_z)|^{1/2}}\exp\left[-\frac{1}{2}(Z - E(Z))^TC_Z^{-1}(Z - E(Z))\right],$$

$$= \frac{1}{\sqrt{2\pi}\cdot 6x}\exp\left[-\frac{1}{2}(x-\bar{x})^2/6x^2\right]\cdot\frac{1}{\sqrt{2\pi}\cdot 6y}\exp\left[-\frac{1}{2}(y-\bar{y})^2/6y^2\right] = f_X(x)\cdot f_Y(y)$$

Consider the signal plus noise z(n) = s + v(n), where s is a RV with E[s] =1, E[s2] = 2, and for each value of n, v(n) ~ N (0, 1). It is known that E[sv(i)] = 1， for all i, and v(i) is independent of v(j) for all i = j.

(1) Compute the autocorrelation function Rz(i, j) of z(n) for all integers i, j.

(2) Is z(n) WSS(Wide sense stationary)? If so, derive a mathematical expression for Rz(k)

$$R_z(i,j) = E[\mathbf{z}(i)\mathbf{z}(j)] = E[(s + \boldsymbol{v}(i))(\boldsymbol{s} + \boldsymbol{v}(j))] = \begin{cases}4 & i \neq j\\5 & i = j\end{cases}$$

Yes, z(n) is wide sense stationary, which is because E[z(n)] = 1 AND $R_z(k) = \begin{cases}4 & k \neq 0\\5 & k = 0\end{cases}$

Linear transformation of Gaussian RV

$$\begin{aligned}f_Y(y) &= |h'(y)|f_X[h(y)]\\&= |\det(A^{-1})|f_X[h(y)]\\&= |\det(A^{-1})|\frac{1}{(2\pi)^{n/2}|\det(C_X)|^{1/2}}\cdot\\&\quad\exp\left\{-\frac{1}{2}\left[A^{-1}(y-b) - E(X)\right]^TC_X^{-1}[*]\right\}\\&= |\det(A^{-1})|\frac{1}{(2\pi)^{n/2}|\det(C_X)|^{1/2}}\cdot\\&\quad\exp\left\{-\frac{1}{2}\left[A^{-1}y - A^{-1}b - \bar{x}\right]^TC_X^{-1}[*]\right\}\\&= \frac{1}{(2\pi)^{n/2}|\det(A)||\det(C_X)|^{1/2}}\cdot\\&\quad\exp\left\{-\frac{1}{2}\left[A^{-1}y - A^{-1}b - A^{-1}\bar{y} + A^{-1}b\right]^TC_X^{-1}[*]\right\}\\&= \frac{1}{(2\pi)^{n/2}|\det(A)|^{1/2}|\det(C_X)|^{1/2}|\det(A^T)|^{1/2}}\exp\left[-\frac{1}{2}(y-\bar{y})^T(A^{-1})^TC_X^{-1}A^{-1}(y-\bar{y})\right]\\&= \frac{1}{(2\pi)^{n/2}|\det(AC_XA^T)|^{1/2}}\exp\left[-\frac{1}{2}(y-\bar{y})^T(AC_XA^T)^{-1}(y-\bar{y})\right]\end{aligned}$$

Random Variables and Stochastic Processes 2-37

Example

The power spectrum

$$\begin{aligned}S_X(\omega) &= \textstyle\int_{-\infty}^{\infty}\sigma^2e^{-\beta|\tau|}e^{-j\omega\tau}d\tau\\&= \textstyle\int_{-\infty}^{0}\sigma^2e^{(\beta-j\omega)\tau}d\tau + \int_0^\infty\sigma^2e^{-(\beta+j\omega)\tau}d\tau\\&= \frac{\sigma^2}{\beta-j\omega} + \frac{\sigma^2}{\beta+j\omega}\\&= \frac{2\sigma^2\beta}{\omega^2+\beta^2}\end{aligned}$$

The variance (also power) of the stochastic process is computed as

$$\begin{aligned}E[X^2(t)] &= R_X(0)\\&= P_X = \tfrac{1}{2\pi}\textstyle\int_{-\infty}^{\infty}S_X(\omega)d\omega\\&= \tfrac{1}{2\pi}\textstyle\int_{-\infty}^{\infty}\frac{2\sigma^2\beta}{\omega^2+\beta^2}d\omega\\&= \tfrac{\sigma^2}{\pi}\arctan\tfrac{\omega}{\beta}\Big|_{-\infty}^{\infty}\\&= \sigma^2\end{aligned}$$

Overall optimality

- When s and z are zero-mean, jointly Gaussian, the LMMSE estimate is also the optimal MMSE estimate.
  - Suppose that s and z have a zero-mean bivariate Gaussian distribution with covariance matrix P given by

$$P = \begin{bmatrix}\sigma_s^2 & \text{Cov}(\mathbf{s},\mathbf{z})\\\text{Cov}(\mathbf{z},\mathbf{s}) & \sigma_z^2\end{bmatrix}$$

  - the conditional density function $f_{\mathbf{s}}(s|\mathbf{z} = z)$ is given by

$$f_{\mathbf{s}}(s|\mathbf{z} = z) = \frac{1}{\sqrt{2\pi}\sigma_s\sqrt{(1-\rho^2)}}\exp\left\{-\frac{1}{2\sigma_s^2(1-\rho^2)}\left(s - \frac{E(\mathbf{sz})}{E(\mathbf{z}^2)}z\right)^2\right\}$$

  - $\hat{s}_{\text{MMSE}} = \hat{s}_{\text{LMMSE}} = \frac{E(\mathbf{sz})}{E(\mathbf{z}^2)}\mathbf{z}$

Estimation Theory 3-80

Comparison of different estimators

	Maximum likelihood (ML)	Maximum *a posteriori* (MAP)
Motivation	Given z, what value of s is most likely to have produced z?	Given z, what value of s is most likely to have occured?
Objective	Maximize the likelihood function $f_{\mathbf{z}}(z\|\mathbf{s} = s)$	maximize the conditional density $f_{\mathbf{s}}(s\|\mathbf{z} = z)$ via Bayes rule, equivalently maximize $f_{\mathbf{z}}(z\|\mathbf{s} = s)f_{\mathbf{s}}(s)$.
Esitmate	$\hat{s}_{\text{ML}} = \text{argmax} f_{\mathbf{z}}(z\|\mathbf{s} = s)$	$\hat{s}_{\text{MAP}} = \text{argmax} f_{\mathbf{z}}(z\|\mathbf{s} = s)f_{\mathbf{s}}(s)$
Required knowledge	likelihood function $f_{\mathbf{z}}(z\|\mathbf{s} = s)$	Density function $f_{\mathbf{s}}(s\|z)$ (or $f_{\mathbf{z}}(z\|\mathbf{s} = s)$ and $f_{\mathbf{s}}(s)$)

Comparison of different estimators

	Minimum mean-square error (MMSE)	Linear MMSE (LMMSE)
Motivation	Given z, what estimate of s gives the smallest MSE?	Given z, what linear function $\hat{s} = \lambda\mathbf{z}$ gives the smallest MSE?
Objective	Minimize the MSE $E[(\mathbf{s} - \hat{\mathbf{s}})^2]$	find λ to minimize $E[(\mathbf{s} - \lambda\mathbf{z})^2]$.
Esitmate	$\hat{\mathbf{s}}_{\text{MMSE}} = E(\mathbf{s}\|\mathbf{z}) = \int_{-\infty}^{\infty} sf_{\mathbf{s}}(s\|\mathbf{z})ds$	$\hat{s}_{\text{LMMSE}} = \lambda\mathbf{z}$, where $\lambda = E[\mathbf{sz}]/E[\mathbf{z}^2]$
Required knowledge	Density $f_{\mathbf{s}}(s\|z)$	Cross-correlation of s and z $E[\mathbf{sz}]$; second moment of z, $E(\mathbf{z}^2)$