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

Lecture 2 Stochastic Theory

event A given event B：

Bayers’ Rule

独立、、

Random variables（RVs）Gaussian ，

Expected value/ Variance 、

，Standard deviation

known the pdf of x, compute the pdf of Y as follows

&

&

&

Multiple RVs marginal distribution/density

Expectation

Covariance 协方差

独立性、

相关系数 、相关性 ；uncorrelated if

独立 ⊊ 不相关，正交if ||条件密度

Multivariate statistics

协方差

自相关，symmetric &P.S.D.

Autocovariance自协方差，S.&P.S.D.

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

考虑 ,A is invertible,

；i.e.,

Matrix derivative:、、

、

Ellipsoid：A is real S.&P.D.|| ，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： 。

对于协方差矩阵 Σ，SVD分解为，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 ，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.；correlation BTW X(t1) and X(t2)

；Autocovar.

Stationary S.P.:1.Strict-sense stationary：；

Wide-sense stationary： strict →wide

and for scalar

Time average/autocorrel. 、

Ergodic process: 各态历经过程是平稳随机过程if

Two S.P. ：cross correlation of X &Y: //互协方差

If X&Y are uncorrelated

cross covariance：

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

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. :

1-dimensional Random Walk、Wiener Process、Poisson Processes

White Noise：power spectral density

power spectrum：；autocorrel.

X(t) and Y (t):

Lecture 3 Estimation Theory

Unbiased/无偏: ;Asymptotically unbiased:，Consistent estimator/一致估计

Maximum Likelihood Estimation(MLE)：Likelihood f:；

Maximum a posteriori Estimation(MAPE)：

Naive Bayes： are boolean ;

Conditionally independence assumption:

Under CIA:

Naive Bayes算法: For :;for :

discrete-valued：

MAP estimates (Laplace smoothing for the case = 1)

离散情况

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

条件独立：，

Logistics分布：

Minimum Mean-Square Error Estimation(MMSE)

；unbiased：；is one type of Baysian estimation；

MMSE estimate with Gaussian noise: ，

s v独立；s v,不相关，MMSE与MAP相同：

; pdf of z

;

Orthogonality Principle：

Linear MMSE(LMMSE)

Orthogonality

LMMSE for vector RVs:

Orthogonality for vector RVs:

Lecture 4 Least Squares Estimation

期望响应：；输入：；

Aim to estimate say ，and is predicted output.

Estimation error , Matrix Formulation

推出

Ps: sufficient condition is P.D. And LS estimator is unbiased.

is unique if H has full column rank or is P.D.

v is zero-mean white noise ,LS has the minimum MSE among all the linear unbiased estimate of x. if ,

Consider Gaussian additive noise,, assume z(1), . . . , z(N) and v is zero mean,

LS estimation:

Likelihood f

Log-likelihood f ，

Weighted Least Squares

，

WLSE and uniqueness of WLSE requires to be P.D..

Recursive Least Squares

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

And if and if is zero-mean for all k, initial estimate of is set as , i.e., then

Recursive least squares estimation **STEP**

Measurement equation：，

1. Initialization: ，

2.Iteration:

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

1. Statistical properties of the noise is known or unknown

2. Iteration：

Example use RLS1：，

If x is known perfectly a priori

If x is completely unknown a priori, ,即测量均值

Using RLS2：

，

Lecture 5 Propagation of states and covariances

Consider the following linear discrete-time system

其中是已知输入，是由**协方差**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都是相互独立的。

均值，

方差

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，连续时间动力学被描述为

解为,

Define ,

前提:，

:

:

Define ,

And for small values of ,

Continuous-time systems

Lecture6 Kalman filter

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

a posteriori estimate 后验估计

a priori estimate 先验

smoothed estimate

predicted estimate

our initial estimate of x0 before any measurements are available in general

the covariance of the estimation error of ,

the covariance of the estimation error of ,

**The discrete-time Kalman filter**

**dynamic system is given by the following equations:**

，不相关

**The Kalman filter is initialized as follows**

**The Kalman filter is given by the following equations：**

；

的第一个表达式称为协方差测量更新方程的Joseph稳定版本，它比的第三个表达式更稳定和稳健：的第一个表达式保证了总是对称的半正定的，只要是对称的半正定的；的第三个表达式在计算上比第一个表达式简单，但不能保证的对称性或半正定性；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 使用第二表达式必须搭配的第二表达式

AND if 常值，，退化为常值向量的RLS估计

、Kk和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（ 表示精度）

Example:，， ;**prediction**:

**Correction**:

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 should be the optimum LMMSE estimate,

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,

,and P reaches a steady-state , 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 ）：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 …Hn-1F]^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 ，有，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:, wk and vk are stationary random process, with ,,DARE as follows

,

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 , 只取决于测量（没有误差）

2. the model is accurate: q = 0. p=0 ,and then

**Kalman filter generalizations: the general discrete-time Kalman filter**

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

**The Kalman filter is initialized as follows**

**The Kalman filter is given by the following equations：**

；

The discrete-time extended Kalman filter1. The system and measurement equations are given as follows:

2. Initialize the filter as follows:

3. For k = 1, 2, . . . , perform the following.compute the partial derivative matrices:

perform the time update:

compute the partial derivative matrices:

perform the measurement update:

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:Motion /process model &measurement model**

**,**

**Prediction**:,

**Correction:**

**=**

**Unscented transform**

**1.For vectors x ∼ N (m, P), the generalization of standard deviation σ is the Cholesky factor** ： 2. The (2n + 1) sigma points can be formed using columns of L:

[]i表示矩阵第i列

3.对于变换y=g(x), 估计如下：

：

参数设置：； determine the spread of the sigma points；

Weights and are given as follows:

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:

:

1.2 Propagate the sigma points through the dynamic model:

1.3 Compute the predicted mean and covariance

:

**2. Update step**

2.1 Form the matrix of sigma points:

2.2 Propagate the sigma points through the measurement model:

2.3 Compute:

:

2.4 Compute the filter gain and the filtered state mean and covariance , conditional to the measurement :

EXAMPLE for UKF： ，choose 5 sigma points:

**Prediction**：

，

**Correction：**，choose 5 sigma points，

:

:

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].

:

:

:

(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**：

： ，

MAPE：

**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.,

:

is Gaussian with mean , and var 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 , According to Schur complement, we have

,

the joint distribution p(s, z) is

in which X = [s, zT]T, ,and the quadratic part is

the determinant

as and then

and Hence the MMSE estimate is

(d) The linear MMSE:  **in which**

and then

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

:

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

Therefore, the LMMSE estimate is

证明俩随机变量联合高斯且不相关，则独立。

=

带入原始式子即可

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)

Yes, z(n) is wide sense stationary, which is because E[z(n)] = 1 AND









