RE Summary

Jorit Geurts - jgeurts@ethz.ch

Version: 7. August 2023

1 Appendix

1.1 Linear Algebra

$$xy^\top = [x_1 y \quad \dots x_n y], \quad \langle x, y\rangle = x^\top y = \sum x_i y_i$$

(Semi) Posivite Definite Matrix iff all eigenvalues (≥ 0) > 0

Trace $\text{trace}(\cdot)$ sum of diagonal elements

$$\frac{\partial \text{trace}(ABA^\top)}{\partial A} = 2AB \text{ if } B = B^\top \qquad \frac{\partial \text{trace}(AB)}{\partial A} = B^\top$$

1.2 Calculus

$$\frac{\partial}{\partial x} x^\top A = \frac{\partial}{\partial x} A^\top x = A, \quad \frac{\partial}{\partial x} x^\top A x = (A + A^\top)x$$

Del-Operator (Gradient): $\nabla_x f(x) = \left[\frac{\partial}{\partial x_1} f(x) \quad \cdots \quad \frac{\partial}{\partial x_n} f(x)\right]^\top$

Jacobian $\frac{\partial f}{\partial x^\top} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$

2 Probability Theory

2.1 Random Variables

2.1.1 Discrete Random Variables

DRV – Discrete Random Variable defined by $p_x, \mathcal{X}$
- $\mathcal{X} \subset \mathbb{Z}$ of all possible outcomes
- $p_x(\bar x) \geq 0$ and $\sum_{\bar x\in\mathcal X} p_x(\bar x) = 1$ bzw. $\sum_{\bar x \in \mathcal X} p_{x|y}(\bar x|\bar y) = 1$

Margin. / Sum Rule	$p(\bar x) = \sum_{\bar y\in\mathcal Y} p_{xy}(\bar x,\bar y)$
Cond. / Product Rule	$p_{x\|y}(\bar x\|\bar y) := \frac{p_{xy}(\bar x,\bar y)}{p_y(\bar y)}$
Total Prob. Theorem	$p_x(\bar x) := \sum_{\bar y\in\mathcal Y} p_{x\|y}(\bar x\|\bar y)p_y(\bar y)$

2.1.2 Continuous Random Variables

CRV – Continuous Random Variable: DRV $\sum \rightsquigarrow$ **CRV** $\int$
- **Set** $\mathcal X \subseteq \mathbb R$ section of real line
- **PDF** p_x satisfies $p_x(\bar x)\geq 0$ and $\int_{\mathcal X} p_x(\bar x) d\bar x = 1$

Caution: $Pr(x=\bar x) = 0, \quad Pr(x\in(a,b)) = \int_a^b p(\bar x)d\bar x$

2.1.3 Conditional PDF

Conditional PDF

Margin. / Sum Rule	$p(\bar x\|\bar z) = \sum_{\bar y\in\mathcal Y} p_{xy\|z}(\bar x,\bar y\|\bar z)$
Cond. / Product Rule	$p_{x\|yz}(\bar x\|\bar y,\bar z) := \frac{p_{xy\|z}(\bar x,\bar y\|\bar z)}{p_{y\|z}(\bar y\|\bar z)}$

Independence

$$p(x|y) = p(x) \quad \Leftrightarrow \quad p(y|x) = p(y) \quad \Leftrightarrow \quad p(x,y)=p(x)p(y)$$

$p(x,y,z) = p(x,y|z)p(z) \to$ x,y indp. $p(x,y|z) = p(z|x,y)$

Conditional Independence

The knowledge of z makes x and y independent:

$$p(x|y,z) = p(x|z) \quad \Leftrightarrow \quad p(x,y|z) = p(x|z)p(y|z)$$

Caution!!! in general we still have: $p(x,y)\neq p(x)p(y)$
Caution!!! Independence $\not\Rightarrow$ Conditional Independence

2.2 Expectation and Variance

2.2.1 Expectation

Definition: Integral for CRV!

$$\mathbb E_x[x] = \sum_{\bar x\in\mathcal X} \bar x p_x(\bar x)$$

Linearity	$\mathbb E_{xy}[a+bx+cy] = a + b\mathbb E_x[x] + c\mathbb E_y[y]$
Multi Variable	$\mathbb E_{xy}[g(x,y)] = \sum_{\bar y}\sum_{\bar x} g(\bar x,\bar y)p_{xy}(\bar x,\bar y)$
Independence	$\mathbb E_{xy}[xy] = \mathbb E_x[x]\mathbb E_y[y]$

Law of Unconcious Statistician for $y=g(x)$

$$\mathbb E_y[y] = \sum_{\bar y\in\mathcal Y}\bar y p_y(\bar y) = \sum_{\bar x\in\mathcal X} g(\bar x)p_x(\bar x)$$

2.2.2 Variance (generally a matrix)

$$\underset{x}{\text{Var}}[x] = \underset{x}{\mathbb E}\left[(x-\mathbb E_x[x])(x-\mathbb E_x[x])^\top\right] = \underset{x}{\mathbb E}\left[x^2\right] - \underset{x}{\mathbb E}[x]^2$$

Linearity

$$\text{Var}_x[a+bX+cY] = b^2\,\text{Var}_X[x] + c^2\,\text{Var}_Y[y] + 2bc\text{Cov}[X,Y]$$

Covariance: if independent $\text{Cov}(X,Y)=0$

$$\text{Cov}[X,Y] = \mathbb E[(X-\mathbb E[X])(Y-\mathbb E[Y])]$$

2.3 Distribution Sampling

CDF – Cumulative Distribution Function

$$\hat F_x(\bar x) := \sum_{i=-\infty}^{\bar x}\hat p_x(i) = Pr(x\leq\bar x) = \int_{-\infty}^{\bar x} p_x(\hat x)d\hat x$$

A1 – 1 DRV $\bar u$ sample of $u\sim\mathcal U$, solve for $\bar x$ such that:

$$\hat F_x(\bar x-1) < \bar u \leq \hat F_x(\bar x)$$

A2 – Multiple DRV or CRV
Works for finite & infinite # of elements
- Decompose $\hat p(\bar x,\bar y) = \hat p_{x|y}(\bar x|\bar y)\hat p_y(\bar y)$
- Apply **A1** to get sample $\bar y$ via $\hat p_y(\bar y)$
- With $\bar y$ fixed, apply **A1** to get sample $\bar x$ via $\hat p_{x|y}(\bar x|\bar y)$

A3 – 1 CRV $\hat p_x$: pw-continuous, bounded. Sample $u\sim\mathcal U(0,1)$ and solve:

$$\bar u = F_x(\bar x) \to \bar x = F_x^{-1}(\bar u)$$

2.4 Change of Variables

DRV: $x=g(y)$

$$p_x(\bar x) = \sum_{\bar y\in\mathcal Y: g(\bar y)=\bar x} p_y(\bar y)$$

CRV general case $y=g(x), \to x = g^{-1}(y)$:

$$p_y(\bar y) = \frac{p_x(\bar x)}{\left|\frac{dg}{dx}(\bar x)\right|}, \quad \bar x = g^{-1}(y)$$

CRV – Multivar. & Indep. Case: $z = x+y$

$$p_z(\bar z) = \sum_{\bar x\in\mathcal X} p_x(\bar x)p_y(\bar z-\bar x)$$

CRV – Multivar. & Cond. Case: $z=g(w,x), w=h(z,x)$
Assumptions: - $g(w,x), p_w$ cont. diff'bar
- $g(w,x)$ strictly monotonic
- For scalars $det(x)$ is just x

$$p_{z|x}(\bar z|\bar x) = p_w(h(\bar z,\bar x)|\bar x)\cdot \left|\det\left(\frac{\partial}{\partial w} g(h(\bar z,\bar x),\bar x)\right)\right|^{-1}$$

2.5 Miscellaneous Probability Theory

Independent multiplication: $z = xy \to p(z) = \sum_Y p_x(\frac{z}{y})p_Y(y)$

Uniform Distribution

$$p(x) = \begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & \text{else}\end{cases} \qquad \mathbb E\{x\} = \frac{a+b}{2}, \quad \text{Var}\{x\} = \frac{(b-a)^2}{12}$$

Triangular Distribution

$$p(x) = \begin{cases}\frac{2(x-a)}{(b-a)(c-a)} & a\le x\le c\\ \frac{2(b-x)}{(b-a)(b-c)} & c\le x\le b\\ 0 & \text{else}\end{cases}$$

$$F(x) = \begin{cases}0 & x<a\\ \frac{(x-a)^2}{(b-a)(c-a)} & a\le x\le c\\ 1-\frac{(b-x)^2}{(b-a)(b-c)} & c\le x\le b\\ 1 & x>b\end{cases}$$

$$\mathbb E\{x\} = \frac{a+b+c}{3}, \quad \text{Var}\{x\} = \frac{a^2+b^2+c^2-ab-ac-bc}{18}$$

Gaussian Distribution

PDF Gaussian distributed vector CRV $y=(y_1,\dots,y_D)$

$$p(x) = \frac{1}{(2\pi)^{D/2}\det(\Sigma)^{1/2}}\exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right)$$

$\rightsquigarrow$ fully characterized by mean $\mu\in\mathbb R^D$ and var $\Sigma\in\mathbb R^{D\times D}$
(Scalar: $D=1,\ \sigma^2 = \det(\Sigma)\to\det(\Sigma)^{1/2}=\sigma$)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad \mathbb E\{x\}=\mu, \quad \text{Var}\{x\}=\sigma^2$$

Sum of two independent Gaussians:

$$x_1\sim\mathcal N(\mu_1,\Sigma_1), x_2\sim\mathcal N(\mu_2,\sigma_2)\to z = M_1x_1+M_2x_2$$

$$z\sim\mathcal N(M_1\mu_1+M_2\mu_2,\ M_1\Sigma_1M_1^T + M_2\Sigma_2M_2^T)$$

Diagonal Variance with $\Sigma = \text{diag}(\sigma_1^2,\dots,\sigma_D^2)$, $\mu = \begin{bmatrix}\mu_1\\ \mu_D\end{bmatrix}$
$\rightsquigarrow$ y_i mut.indep. iff Σ diag.

$$p(y) = \prod_{i=1}^D \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{(y_i-\mu_i)^2}{2\sigma_i^2}\right)$$

Dependent Gaussians:

$$\Sigma_{x\pm y} = \Sigma_x+\Sigma_y\pm\Sigma_{xy}\pm\Sigma_{xy}^T,\ \mu_{x\pm y} = \mu_x\pm\mu_y$$

3 Bayesian Tracking

3.1 Bayes Theorem

Definition: (Applies to discrete and continuous random variables)

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

x=system state, $p(x)$=prior belief of state, z=observations related to state
$p(z|x)$=observation likelihood, $p(x|z)$=posterior belief of state
$p(z)$=normalization constant $=\sum_x p(z|x)p(x)$

For multiple observations:

$$\underbrace{p(x|z_1,\dots,z_N)}_{\text{posterior}} = \frac{\overbrace{p(x)}^{\text{prior}}\overbrace{\prod_i p(z_i|x)}^{\text{observation likelihood}}}{\underbrace{\sum_{x\in\mathcal X} p(x)\prod_i p(z_i|x)}_{\text{normalization}}}$$

3.2 Bayesian Tracking

Model

Objective: Calculate pdf of state $p(x(k)|z(1:k))$

$$x(k) = q_{k-1}(x(k-1), v(k-1))\in\mathcal X$$
$$z(k) = h_k(x(k),w(k))$$

Prior Update - State Prediction

$$p(x(k)|z(1\!:\!k-1)) = \sum_{x(k-1)}\underbrace{p(x(k)|x(k-1))}_{\text{process model}}\underbrace{p(x(k-1)|z(1\!:\!k-1))}_{\text{previous iteration}}$$

Measurement Update

$$p(\bar x(k)|\bar z(1\!:\!k)) = \frac{\overbrace{p(\bar z(k)|\bar x(k))}^{\text{measurement model}}\overbrace{p(\bar x(k)|\bar z(1\!:\!k-1))}^{\text{prior}}}{\underbrace{\sum_{i\in\mathcal X} p(\bar z(k)|i)p(i|\bar z(1\!:\!k-1))}_{\text{normalization}}}$$

Computer Implementation
1. **Enumerate** state space $\mathcal X = \{0,1,\dots,N-1\}$
2. **Posterior** define $\mathbf a^i_{k|k} := p_{x(k)|z(1:k)}(i|\bar z(1:k))$
3. **Prior** define $\mathbf a^i_{k|k-1} := p_{x(k)|z(1:k-1)}(i|\bar z(1:k-1))$

Initialization $\mathbf a^i_{0|0} = p_{x(0)}(i) \quad i=0,\dots,N-1$

$$\left.\begin{aligned}\mathbf a^i_{k|k-1} &= \sum_{j=0}^{N-1} p_{x(k)|x(k-1)}(i|j)\mathbf a^j_{k-1|k-1}\\ \mathbf a^i_{k|k} &= \frac{p_{z(k)|x(k)}(\bar z(k)|i)\mathbf a^i_{k|k-1}}{\sum_{j=0}^{N-1} p_{z(k)|x(k)}(\bar z(k)|j)\mathbf a^j_{k|k-1}}\end{aligned}\right\}\ \text{Iterate k} > 0$$

4 Estimate Extraction

4.1 General Measurement

$$\underbrace{\begin{bmatrix}z_1\\ z_m\end{bmatrix}}_{z\,\in\,\mathbb R^m} = \underbrace{\begin{bmatrix}h_{11} & h_{1n}\\ h_{m1} & h_{mn}\end{bmatrix}}_{H\,\in\,\mathbb R^{m\times n}}\cdot\underbrace{\begin{bmatrix}x_1\\ x_n\end{bmatrix}}_{x\,\in\,\mathbb R^n} + \underbrace{\begin{bmatrix}w_1\\ w_m\end{bmatrix}}_{w\,\in\,\mathbb R^m}$$

m Measurements, n States, $w_i\sim\mathcal N(0,R)$ – Zero-mean additive gaussian, $x, w_i\ \ \forall i$ assumed mutually independent

4.2 Likelihood Estimators

4.2.1 ML-Maximum Likelihood

Choose x to make $\bar z$ most likely:

$$\left.\begin{matrix} x\in\mathcal X & \text{unknown}\\ p_x(\bar x) & \text{unknown}\end{matrix}\right\}\quad \hat x^{ML} := \underset{\bar x\in\mathcal X}{\text{argmax}}\left[p_{z|x}(\bar z|\bar x)\right]$$

$p_{z|x}(\bar z|\bar x)$ is var. – technically not a PDF

4.2.2 MAP-Maximum A Posteriori

ML given observation & prior belief about x:

$$\left.\begin{matrix} x\in\mathcal X & \text{unknown}\\ p_x(\bar x) & \textbf{known}\end{matrix}\right\}\quad \hat x^{MAP} := \underset{\bar x\in\mathcal X}{\text{argmax}}\left[p_{z|x}(\bar z|\bar x)p_x(\bar x)\right]$$

$\hat x^{MAP} = \hat x^{ML}$ for constant $p_x(\bar x)$

4.3 Least Squares Estimators

4.3.1 Linear Least Squares

ML estimate when errors are indep. 0-mean, smae variance, $\mathcal N$-distributed:

$$\hat x^{LS} := \underset{x}{\text{argmin}}\left[(z-Hx)^\top(z-Hx)\right] = (H^\top H)^{-1}H^\top z$$

$\hat x^{LS} = \hat x^{ML}$ for gaussian noise

4.3.2 Weighted Least Squares

$$\hat x^{WLS} := \text{argmin}_x\left[(\bar z-H\hat x)^\top R^{-1}(\bar z-H\hat x)\right] = (H^\top R^{-1}H)^{-1}H^\top R^{-1}\bar z$$

Incorporate Prior Knowledge on x:
$r\sim\mathcal N(0,P_x)\Rightarrow x := \bar x + r, \mathbb E\{x\} = \hat x_0, \text{Var}\{x\} = P_x$
Introduce Extended System:

$$\tilde{\bar z} = \begin{bmatrix}\hat x_0\\ z(1:k)\end{bmatrix}, \tilde H = \begin{bmatrix}\mathbb I\\ H(1:k)\end{bmatrix}, \tilde w = \begin{bmatrix}-r\\ w(1:k)\end{bmatrix}, \tilde R = \begin{bmatrix}P_x & 0\\ 0 & R(1:k)\end{bmatrix}$$

4.3.3 Recursive Least Squares

Interpretation Update $\hat x$ recursively when new $\bar z(k)$ arrives
Estimation Error: $e(k) = (\mathbb I - KH)\,e(k-1) - Kw(k)$
Objective Minimize MSE with $P(k) = \text{Var}\{e(k)\}$

$$J(k) := \mathbb E\left\{e^\top(k)e(k)\right\} = \mathbb E\{\text{trace}(P(k))\}$$

Algorithm
Initialization: $\hat x(0) = \hat x_0, \quad P(0)=P_x = \text{Var}\{x\}$
Recursion: Observe $\bar z(k)$, then update:

$$K(k) = P(k-1)H(k)^\top\left(H(k)P(k-1)H(k)^\top + R(k)\right)^{-1}$$
$$\hat x(k) = \hat x(k-1)+K(k)\left(\bar z(k) - H(k)\hat x(k-1)\right)$$
$$P(k) = (\mathbb I-K(k)H(k))P(k-1)(\mathbb I-K(k)H(k))^\top + K(k)R(k)K(k)^\top$$

5 Kalman Filters

5.1 Kalman Filter

Exact Solution for Bayesian Tracking of LTI Systems with Gaussian Noise.

Model (Linear Time-Invariant)

$$x(k) = A(k-1)x(k-1)+u(k-1)+v(k-1)$$
$$z(k) = H(k)x(k)+w(k)$$

$x(0), v(\cdot), w(\cdot)$ ind.	$x(0)\sim\mathcal N(x_0,P_0)$
$v\sim\mathcal N(0,Q(k))$	$w\sim\mathcal N(0,R(k))$

Auxilliary Variables

Problem Formulation want $p_{x(k)|z(1:k)}$ $\rightsquigarrow$ bayesian tracking provides this
BUT we have **CRVs** and don't want integrals
$\Rightarrow$ exploit linearity, GRV to convert to matrix manipulations
Auxilliary Variables:
- **Init:** $x_m(0) = x(0)$
- **S1:** $x_p(k) = A(k-1)x_m(k-1)+u(k-1)+v(k-1)$
- **S2:** $z_m(k) = H(k)x_p(k)+w(k)$

$$p_{x_m(k)}(\xi) := p_{x_p(k)|z_m(k)}(\xi|\bar z(k)) \quad \forall\xi$$

Equivalency x conditioned on $z(1:k-1)/z(1:k)$

$$p_{x_p(k)}(\xi) = p_{x(k)|z(1:k-1)}(\xi|\bar z(1:k-1))$$
$$p_{x_m(k)}(\xi) = p_{x(k)|z(1:k)}(\xi|\bar z(1:k))$$

GRV x_p and x_m are GRVs for $k=1,2,\dots$

$$\hat x_p(k) := \mathbb E\{x_p(k)\} \qquad P_p(k) := \text{Var}\{x_p(k)\}$$
$$\hat x_m(k) := \mathbb E\{x_m(k)\} \qquad P_m(k) := \text{Var}\{x_m(k)\}$$

Recursive Update Equations

Initialization $\hat x_m(0) = x_0, \quad P_m(0)=P_0$

S1 – Prior Update / Prediction Step

$$\hat x_p(k) = A(k-1)\hat x_m(k-1)+u(k-1)$$
$$P_p(k) = A(k-1)P_m(k-1)A^\top(k-1)+Q(k-1)$$

S2 – A Posteriori Update / Measurement Update

$$P_m(k) = \left(P_p^{-1}(k)+H^\top(k)R^{-1}(k)H(k)\right)^{-1}$$
$$\hat x_m(k) = \hat x_p(k) + \underbrace{P_m(k)H^\top(k)R^{-1}(k)}_{\text{correctionfactor}}\underbrace{(\bar z(k)-H(k)\hat x_p(k))}_{\text{predictionerror}}$$

Alternative Equations – all matrices/vectors at time k

S2 – Kalman Filter Gain:

$$K = P_pH^\top\left(HP_pH^\top+R\right)^{-1}$$
$$\hat x_m = \hat x_p + K\left(\bar z - H\hat x_p\right)$$
$$P_m = \underbrace{(\mathbb I-KH)P_p}_{\text{cheaper}} = \underbrace{(\mathbb I-KH)P_p(\mathbb I-KH)^\top+KRK^\top}_{\text{fewernumericalerrors}}$$

Equivalent to RLS $\rightsquigarrow$ among linear, unbiased estimators KF minimizes MSE

KF Remarks

Known Data $A(k), H(k), Q(k), R(k)\ \forall k$ as well as $P_0 \rightsquigarrow$ KF matrices $P_p(k), P_m(k), K(k)$ can be computed offline

$$\hat{x}_p(k) = A(k-1)\hat{x}_m(k-1) + u(k-1)$$

$$\hat{x}_m(k) = \underbrace{(\mathbb{I}-KH)A(k-1)}_{\tilde{A}(k)}\hat{x}_m(k-1) + \underbrace{(\mathbb{I}-KH)}_{\tilde{B}(k)}u(k-1) + K\bar{z}(k)$$

KF not guaranteed to be optimal – non-linear estimators may do better

5.1.1 Asymptotic Properties of KF

Model assume **TI System, stationary Distributions** (A, H, Q, R constants)

Change of Variance

$$P_p(k+1) = \left(AP_p(k)A^\top\right) + Q - \left(AP_p(k)H^\top\right)\left(HP_p(k)H^\top + R\right)^{-1}\left(HP_p(k)A^\top\right)$$

Scalar Case $P_p(k+1) = \frac{a^2 r P_p(k)}{h^2 P_p(k) + r} + q =: f(P_p(k))$

Summary *Variance Converges to unique steady-state sol'n provided either $|a| < 1$ or if $|a| \geq 1 \Rightarrow h \neq 0$ and $q > 0$*

If slope for small P_∞ is $> 1 \Rightarrow$ unstable ($|a| > 1$)

5.1.2 Detectability / Stabilizablilty

Detectability Conditions

Pair (A, H) detectable

$\Leftrightarrow$ For deterministic LTI Sys. ($x(k)=Ax(k-1), z(k)=Hx(k)$) $\lim_{k\to\infty} z(k) = 0 \Rightarrow \lim_{k\to\infty} x(k) = 0 \quad \forall x_0 \in \mathbb{R}^n$

$\Leftrightarrow$ $\begin{bmatrix} A-\lambda\mathbb{I} \\ H \end{bmatrix}$ is full rank for all $\lambda \in \mathbb{C}$ with $|\lambda| \geq 1$ (PBH-Test)

$\Leftrightarrow$ Eigenvalues of $A - LH$ (or $(\mathbb{I} - LH)A$) can be placed within unit circle by suitable choice of $L \in \mathbb{R}^{n\times m}$

Stabilizability Conditions

Pair (A, B) stabilizable

$\Leftrightarrow$ For deterministic LTI Sys. ($x(k)=Ax(k-1)+Bu(k-1)$) $\exists\, u(0:k-1)$ s.t. $\lim_{k\to\infty} x(k) = 0 \quad \forall x_0 \in \mathbb{R}^n$

$\Leftrightarrow$ $[\,A-\lambda\mathbb{I}\ B\,]$ full rank $\forall \lambda \in \mathbb{C}$ with $|\lambda| \geq 1$ (PBH-Test)

$\Leftrightarrow$ Eigenvalues of $A - BK$ (or $(\mathbb{I} - BK)A$) can be placed within unit circle by suitable choice of $K \in \mathbb{R}^{m\times n}$

NOTE Stabilizability is dual of detectability, (A, B) stabilizable

Observability / Reachability

if $(A^\top, B^\top)$ detectable

Observability Matrix implies detectability $\quad \mathcal{O} = \begin{bmatrix} H \\ HA^{n-1} \end{bmatrix}$

Reachability Matrix implies stabilizablilty $\quad \mathcal{R} = [\,A\ A^{n-1}B\,]$

5.1.3 Steady-State KF

DARE – Discrete Algebraic Ricatti Equation (P_∞ is P_p)

$$P_\infty = AP_\infty A^\top + Q - AP_\infty H^\top(HP_\infty H^\top + R)^{-1}HP_\infty A^\top$$

Steady-State KF Gain $\quad K_\infty = P_\infty H^\top(HP_\infty H^\top + R)^{-1}$

Filter Equations Linear Time-Invariant System

$$\hat{x}(k) = \underbrace{(\mathbb{I}-K_\infty H)A}_{\hat{A}}\hat{x}(k-1) + \underbrace{(\mathbb{I}-K_\infty H)}_{\hat{B}}u(k-1) + K_\infty\bar{z}(k)$$

Error dynamics (also holds for non ss-KF) $e(k) = x(k) - x_f(k)$

$$e(k) = (\mathbb{I}-K_\infty H)Ae(k-1) + (\mathbb{I}-K_\infty H)v(k-1) - K_\infty w(k)$$

$$\mathbb{E}\{e(k)\} = (\mathbb{I}-K_\infty H)A\,\mathbb{E}\{e(k-1)\}$$

- $(\mathbb{I} - K_\infty H)A$ must be stable (all $|\lambda| < 1$) for e not to diverge
- $\mathbb{E}\{e(k)\} \xrightarrow{k\to\infty} 0$ iff $(\mathbb{I} - K_\infty H)A$ stable

Theorem for $R > 0, Q \geq 0, G$ s.t. $Q = GG^\top$

$\Leftrightarrow$ (A, H) is detectable, (A, G) is stabilizable

$\Leftrightarrow$ DARE has unique pos. semidef. sol. $P_\infty \geq 0$, resulting $(\mathbb{I} - K_\infty H)A$ is stable and

$$\lim_{k\to\infty} P_p(k) = P_\infty \quad \forall P_p(1) \geq 0$$

(A,H) Detectable: can observe all unstable modes

(A,G) Stabilizable: noise excites unstable modes

Note if $Q > 0 \Rightarrow (A, G)$ always stabilizable If $Q = 0$ multiple valid solutions can be possible

5.2 Extended Kalman Filter

Extension of KF to non-linear systems:

$$x(k) = q_{k-1}(x(k-1), v(k-1))$$

$$z(k) = h_k(x(k), w(k))$$

$x(0), v(\cdot), w(\cdot)$ mutually ind.	$x(0) \sim \mathcal{N}(x_0, P_0)$
$v \sim \mathcal{N}(0, Q(k))$	$w \sim \mathcal{N}(0, R(k))$

EKF Equations

Initialization: $\hat{x}_m(0) = x_0, \quad P_m(0) = P_0$

S1 – Prior Update / Prediction Step

$$\hat{x}_p(k) = q_{k-1}(\hat{x}_m(k-1), 0)$$

$$P_p(k) = A(k-1)P_m(k-1)A^\top(k-1) + L(k-1)Q(k-1)L^\top(k-1)$$

$$A(k-1) = \frac{\partial q_{k-1}(\hat{x}_m(k-1), 0)}{\partial x}, \quad L(k-1) = \frac{\partial q_{k-1}(\hat{x}_m(k-1), 0)}{\partial v}$$

S2 – A Posteriori Update / Measurement Update (at k)

$$K = P_pH^\top\left(HP_pH^\top + MRM^\top\right)^{-1}$$

$$\hat{x}_m = \hat{x}_p + K\left(\bar{z} - h_k(\hat{x}_p(k), 0)\right)$$

$$P_m = (\mathbb{I} - K(k)H(k))\,P_p(k)$$

$$H(k) := \frac{\partial h_k(\hat{x}_p(k,0))}{\partial x}, \quad M(k) := \frac{\partial h_k(\hat{x}_p(k,0))}{\partial w}$$

Intuition

Process Update – predict the mean state estimate fwd using NL process model and update the var. according to lin. eqns

Measurement Update – correct for mismatch between $\bar{z}(k)$ and prediction $h_k(\hat{x}_p(k), 0)$, and correct var. according to lin. eqn

Caution

- Process & measurement noise **assumed** 0-mean
- A, L, H, M lin. about current $\hat{x}_m \rightsquigarrow$ **cannot** compute offline
- Prior update only accurate if $\mathbb{E}\{\cdot\}$ & $q_{k-1}(\cdot)$ commute
- **EKF Variables DO NOT capture true conditional mean & var.**

5.2.1 Hybrid EKF

System Process – non-lin, CT, **Measurement** – non-lin, DT

$$\dot{x}(k) = q(x(t), v(t), t) \qquad x[k] := x(kT)$$

$$z[k] = h_k(x[k], w[k]) \qquad \mathbb{E}\{w[k]\} = 0, \mathrm{Var}\{w[k]\} = R$$

CT White Noise $\delta(t)$ = Dirac pulse

$$\mathbb{E}\{v(t)\} = 0 \text{ and } \mathbb{E}\{v(t)v^\top(t+\tau)\} = \mathcal{Q}_c\delta(\tau)$$

True CT White noise does not exist $\rightarrow$ Would require infinite power

Hybrid EKF Equations

Initialization $\hat{x}_m(0) = x_0, \quad P_m(0) = P_0$

S1 – Prior Update / Prediction Step (Solve ODEs)

Solve, for $(k-1)T \leq t \leq kT$ and $\hat{x}((k-1)T) = \hat{x}_m[k-1]$

$$\dot{\hat{x}}(t) = q(\hat{x}(t), 0, t) \quad \Rightarrow \hat{x}_p[k] := \hat{x}(kT)$$

Solve, for $(k-1)T \leq t \leq kT$ and $P((k-1)T) = P_m[k-1]$

$$\dot{P}(t) = AP + PA^\top + LQ_cL^\top \quad \Rightarrow P_p[k] := P(kT)$$

with

$$A(t) := \frac{\partial q}{\partial x}(\hat{x}(t), 0, t), \qquad L(t) := \frac{\partial q}{\partial v}(\hat{x}(t), 0, t)$$

S2 – A Posteriori Update / Measurement Update

Measurement model still DT – S2 identical to standard EKF

Caution

- Process & measurement noise **assumed** 0-mean
- ODEs typically solved numerically: Accuracy $\uparrow$ = Comp. Cost $\uparrow$

6 Particle Filter

6.1 Monte Carlo Sampling

Samples N DRVs y^n of samples of $y \in \mathcal{Y} = \{1, \ldots, \bar{Y}\} \sim p_y$

Helper Variables

$$s_i^n = \delta(i - y^n) = \begin{cases} 1 & \text{if} y^n = i \\ 0 & \text{otherwise} \end{cases} \quad \begin{array}{|l} i = 1, \ldots, \bar{Y} \\ n = 1, \ldots, N \end{array}$$

Average – s_i converges to $p_y(i)$ by Law of Large Numbers

$$\frac{1}{N}\sum_{n=1}^N s_i^n \xrightarrow{N\to\infty} \mathbb{E}\{s_i^n\} = p_y(i) \qquad p_y(i) \approx \frac{1}{N}\sum_{n=1}^N \bar{s}_i^n$$

Approximation

For $y \in \mathcal{Y}$, it holds $\quad p_y(i) \approx \frac{1}{N}\sum_{n=1}^N \delta(i - \bar{y}^n), \quad i \in \mathcal{Y}$

For $x = g(y)$, with $x \in \mathcal{X} := g(\mathcal{Y})$, it holds

$$p_x(j) \approx \frac{1}{N}\sum_{n=1}^N \delta(j - g(\bar{y}^n)), \quad j \in g(\mathcal{Y})$$

Monte Carlo Sampling CRV

Helper Variables

$$s_a^n = \int_a^{a+\Delta y}\delta(i - y^n)d\xi = \begin{cases} 1 & \text{if} a \leq y^n < a + \Delta y \\ 0 & \text{otherwise} \end{cases}$$

Average – s_a converges to $p_y(\xi)$ by Law of Large Numbers

$$\int_a^{a+\Delta y}\frac{1}{N}\sum_{n=1}^N \delta(\xi - y^n)d\xi \xrightarrow{N\to\infty} \int_a^{a+\Delta y} p_y(\xi)d\xi$$

Approximation

$$\text{For } y \in \mathcal{Y}\text{, it holds } p_y(\xi) \approx \frac{1}{N}\sum_{n=1}^N \delta(\xi - \bar{y}^n)\,\forall\xi$$

$$\text{For } x = g(y)\text{, it holds } p_x(\xi) \approx \frac{1}{N}\sum_{n=1}^N \delta(\xi - g(\bar{y}^n))\,\forall\xi$$

6.2 Particle Filter - Modeling

System *Non-linear, discrete-time*:

$$x(k) = q_{k-1}(x(k-1), v(k-1))$$

$$z(k) = h_k(x(k), w(k))$$

$x(0), \{v\}, \{w\}$ mutually indep. can be D- or CRVs, known PDFs

Objective

Bayesian State Estimator with Monte Carlo Sampling

Init:	$x_m(0)$	$:= x(0)$
S1:	$x_p(k)$	$:= q_{k-1}(x_m(k-1), v(k-1))$
S2:	$z_m(k)$	$:= h_k(x_p(k), w(k))$
	$p_{x_m(k)}(\xi)$	$:= p_{x_p(k)\mid z_m(k)}(\xi\mid\bar{z}(k)) \quad \forall\xi$

Similar proof for $p_{x_p(k)}(\xi), p_{x_m(k)}(\xi)\,\forall\xi$ as for KF aux. var.

6.3 Particle Filter - Equations

Initialization Draw N samples $\{\bar{x}_m^n(0)\}$ from $p_{x(0)}$

S1 – Prior Update / Prediction Step

Apply process equation to the particles $\{\bar{x}_m^n(k-1)\}$

$$\bar{x}_p^n(k) := q_{k-1}(\bar{x}_m^n(k-1), \bar{v}^n(k-1)), \quad n = 1, \ldots, N$$

Requires N noise samples from $p_{v(k-1)}$

S2 – A Posteriori Update / Measurement Update

Scale each particle by measurement likelihood:

$$\beta_n = \alpha p_{z(k)\mid x(k)}(\bar{z}(k)\mid\bar{x}_p^n(k)), \qquad n = 1, \ldots, N$$

with α normalization s.t $\sum_{n=1}^N \beta_n = 1$

$$\alpha = \left(\sum_{n=1}^N p_{z_m(k)\mid x_p(k)}(\bar{z}(k)\mid\bar{x}_p^n(k))\right)^{-1}$$

Resample to get N posterior particles $\{\bar{x}_m^n(k)\}$, w/ eq. weights

Repeat N times:
• Select random number r uniformly on $(0, 1)$ • Pick particle $\bar{n}$ s.t $\sum_{n=1}^{\bar{n}-1}\beta_n < r \leq \sum_{n=1}^{\bar{n}}\beta_n$

New particles all have equal weight

$$p_{x_m(k)}(\xi) \approx \frac{1}{N}\sum_{n=1}^N \delta(\xi - \bar{x}_m^n(k)), \quad \forall\xi$$

Intuition

Process Update – Propagate N particles through system dynamics. Provided N large, $\{x_m^n(k-1)\} \approx p_{x_m(k-1)} \rightsquigarrow \{\bar{x}_p^n(k)\} \approx p_{x_p(k)}$

Measurement Update – treat each particle separately. At points of high prior, we have many particles. Posterior has same particles, but scaled by measurement likelihood.

6.4 Roughening

Sample Impoverishment *Lumpiness* – when we resample, only retain subset of particles. For finite N this is problem.

Roughening

Perturb particles **after** resampling

$$\bar{x}_m^n(k) \leftarrow \bar{x}_m^n(k) + \Delta x^n(k)$$

$\Delta x^n(k)$ drawn from 0-mean, finite-var. distribution

Example Method for Δx^n: σ_i stand. dev. of $\Delta x_i^n(k)$

$$\sigma_i = KE_iN^{-\frac{1}{d}}$$

K: tuning parameter, typically $K \ll 1$

d: dimension of state space

E_i: max inter-sample variability $\max_{n_1,n_2}|\bar{x}_{m,i}^{n1} - \bar{x}_{m,i}^{n2}|$

$N^{-\frac{1}{d}}$: node spacing on grid

7 Observer Based Control

7.1 LTI Observer

LTI System

$$x(k) = Ax(k-1) + Bu(k-1) + v(k-1)$$

$$z(k) = Hx(k) + w(k)$$

v, w 0-mean noise CRVs

$\mathbb{E}\{\hat{x}(k)\} \xrightarrow{k\to\infty} \mathbb{E}\{x(k)\}$, $\mathrm{Var}\{\hat{x}(k)\}$ bounded

Luenberger Observer

Same structure as steady-state KF:

$$\hat{x}(k) = A\hat{x}(k-1) + Bu(k-1) + K(\bar{z}(k) - \hat{z}(k))$$

$$= (\mathbb{I} - K_\infty H)A\hat{x}(k-1) + (\mathbb{I} - K_\infty H)Bu(k-1) + K_\infty\bar{z}(k)$$

$$\hat{z}(k) = H(A\hat{x}(k-1) + Bu(k-1))$$

Error $e(k) = x(k) - \hat{x}(k)$

in absence of noise $\bar{z}(k) = z(k)$

$$e(k) = (\mathbb{I} - KH)Ae(k-1)$$

- $(\mathbb{I} - KH)A$ must be stable (all $|\lambda| < 1$) for $e \xrightarrow{k\to\infty} 0$
- Such K exists iff (A, H, A) detectable $\Leftrightarrow (A, H)$ detectable

Observer Design

If (A, H) detectable, construct K

Pole Placement: place λ of e-dynamics of observer modes at desired locations

Steady State KF: Yields optimal K (min steady-state MSE) given statistics of v, w. Also need (A, G) stabilizable, $\mathrm{Var}\{v(k)\} = GG^\top$

Alternative Formulation

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + K(\bar{z}(k) - \hat{z}(k))$$

$$\hat{z}(k) = H\hat{x}(k) \rightarrow \text{ delayed measurement}$$

Error Dynamics

$$e(k+1) = (A - KH)e(k)$$

7.2 Static State Feeback Control

System Model

LTI, no noise, deterministic, perf. state info $z(k) = x(k)$:

Feedback Law	**CL Dynamics**
$u(k) = Fx(k) = Fz(k)$	$x(k) = (A + BF)x(k-1)$

- $(A + BF)$ must be stable (all $|\lambda| < 1$)
- Such F exists $\Leftrightarrow (A, B)$ stabilizable

Controller Design

Pole Placement: place (controllable) poles at des. CL loc.

LQR: find F that min. quad. cost ($\bar{Q} = \bar{Q}^\top \geq 0, \bar{R} = \bar{R}^\top \geq 0$)

$$J_{\mathrm{LQR}} = \sum_{k=0}^\infty x^\top(k)\bar{Q}x(k) + u^\top(k)\bar{R}u(k)$$

LQR Design

Assuming (A, B) stabil., (A, G) detect.

$$F = -(B^\top PB + \bar{R})^{-1}B^\top PA$$

DARE

Same as KF but $A \rightarrow A^\top, H \rightarrow B^\top, Q \rightarrow \bar{Q}, R \rightarrow \bar{R}$

$$P = A^\top PA + \bar{Q} - A^\top PB(B^\top PB + \bar{R})^{-1}B^\top PA$$

7.3 Seperation Principle

Stable Observer + Controller $\rightsquigarrow u(k) = F\hat{x}(k)$

Dynamics

$$\begin{bmatrix} x(k) \\ e(k) \end{bmatrix} = \begin{bmatrix} A + BF & -BF \\ 0 & (\mathbb{I} - KH)A \end{bmatrix}\begin{bmatrix} x(k-1) \\ e(k-1) \end{bmatrix}$$

Separation Principle

- $(\mathbb{I} - KH)A$ and $(A + BF)$ stable $\rightsquigarrow$ **Overall system stable**
- Still stable with noise, Generalizable to time-varying case
- Does **NOT** hold for nonlinear systems

Separation Theorem

Optimal strategy for control problem

$$J_{\mathrm{LQG}} = \lim_{N\to\infty}\mathbb{E}\left\{\frac{1}{N}\sum_{k=0}^{N-1}\left[x^\top(k)\bar{Q}x(k) + u^\top(k)\bar{R}u(k)\right]\right\}$$

1. Design steady state KF to provide $\hat{x}(k)$ **(no dependency on $\bar{Q}, \bar{R}$)**
2. Design optimal state-FB law $u(k) = Fx(k)$ for determ. LQR problem minimizes J_{LQR} w/o noise **(no dependency on noise statistics Q, R)**
3. Put together