RECURSIVE ESTIMATION

https://gitlab.ethz.ch/norrisg/re_summary

NORRISG Rev: 20230720

1 PROBABILITY

DRV – DISCRETE RANDOM VARIABLES

DRV – Discrete Random Variable defined by $p_x, \mathcal{X}$
- **Set** $\mathcal{X} \subset \mathbb{Z}$ of all possible outcomes
- **PDF** p_x satisfies $p_x(\bar{x}) \geq 0$ 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	$p(\bar{x}\mid\bar{z}) = \sum_{\bar{y}\in\mathcal{Y}} p_{xy\mid z}(\bar{x},\bar{y}\mid\bar{z})$
Cond. / Product Rule	$p_{x\mid yz}(\bar{x}\mid\bar{y},\bar{z}) := \frac{p_{xy\mid z}(\bar{x},\bar{y}\mid\bar{z})}{p_{y\mid z}(\bar{y}\mid\bar{z})}$
Total Prob. Theorem	$p_x(\bar{x}) := \sum_{\bar{y}\in\mathcal{Y}} p_{x\mid y}(\bar{x}\mid\bar{y})p_y(\bar{y})$
Bayes Theorem	$p(x\mid y) = \frac{p(y\mid x)p(x)}{p(y)}$

Independence

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

Conditional Independence

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

CRV – 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 $\int_{\mathcal{X}} p_x(\bar{x})d\bar{x} = 1$

ACHTUNG $p_x(\bar{x})$ no longer probability $\to$ prob. defined on interval, p_x is the **PDF**

EXPECTATION

Expectation	$\mathbb{E}_x\{g(x)\} = \sum_{\bar{x}\in\mathcal{X}} g(\bar{x})p_x(\bar{x})$
Linearity	$\mathbb{E}_{xy}\{a + bx + cy\} = a + b\,\mathbb{E}_x\{x\} + c\,\mathbb{E}_y\{y\}$
Multi Variable	$\mathbb{E}_{xy}\{g(x,y)\} = \sum_{\bar{y},\bar{x}} g(\bar{x},\bar{y})p_{xy}(\bar{x},\bar{y})$
Independence	$\mathbb{E}_{xy}\{g(x,y)\} = \mathbb{E}_y\Big\{\mathbb{E}_x\{g(x,y)\}\Big\}$
Law of Total Exp.	$\mathbb{E}_y\Big\{\mathbb{E}_{x\mid y}\{x\}\Big\} = \mathbb{E}_x\{x\}$

Variance (generally a matrix)

$$\underset{x}{\mathrm{Var}}\{x\} = \underset{x}{\mathbb{E}}\left\{(x - \mathbb{E}_x\{x\})(x - \mathbb{E}_x\{x\})^\top\right\} = \underset{x}{\mathbb{E}}\left\{x^2\right\} - \underset{x}{\mathbb{E}}\left\{x\right\}^2$$

Linearity $\mathrm{Var}_x\{a + bx\} = b\,\mathrm{Var}_x\{x\}b^\top$

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

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

DISTRIBUTION SAMPLING

CDF – Cum. Dist. Function $\widehat{F}_x(\bar{x}) := \sum_{i=-\infty}^{\bar{x}} \widehat{p}_x(i) = Pr(x \leq \bar{x})$

A1 – 1 DRV $\bar{u}$ sample of $u \sim \mathcal{U}$, solve for $\bar{x}$ such that: $\widehat{F}_x(\bar{x}-1) < \bar{u} \leq \widehat{F}_x(\bar{x})$

A2 – Multiple DRV

Option 1 Want p_{xy}, x, y scalar-valued, finite DRVS. N_x, N_y number of elements in x, y. Define $\mathcal{Z} := \{1, \dots, N_x \cdot N_y\}$

$$\widehat{p}_{\mathcal{Z}}(1) = \widehat{p}_{xy}(1,1), \dots, \widehat{p}_{\mathcal{Z}}(N_x \cdot N_y) = \widehat{p}_{xy}(N_x, N_y)$$

Then apply **A1** to this

Option 2 This works for finite & infinite # of elements
- Decompose $\widehat{p}(\bar{x},\bar{y}) = \widehat{p}_{x|y}(\bar{x}|\bar{y})\widehat{p}_y(\bar{y})$
- Apply **A1** to get sample $\bar{y}$ via $\widehat{p}_y(\bar{y})$
- With $\bar{y}$ fixed, apply **A1** to get sample $\bar{x}$ via $\widehat{p}_{x|y}(\bar{x}|\bar{y})$

A3 – 1 CRV Desired PDF $\widehat{p}_x$ given, assume pw-continuous, bounded. Let $u \sim \mathcal{U}(0,1)$. Let $\bar{x}$ be **any** sol'n to $\bar{u} = F_x(\bar{x})$. Then x has PDF $p_x = \widehat{p}_x$

CHANGE OF VARIABLES

DRV $x = g(y) \rightsquigarrow$ calc. p_x $\quad p_x(\bar{x}) = \sum_{\bar{y}\in\mathcal{Y}: g(\bar{y}) = \bar{x}} p_y(\bar{y})$

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

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

EXAMPLES

Uniform Distribution

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

2 BAYESIAN TRACKING

PROBLEM STATEMENT

Model

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

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

BAYESIAN TRACKING

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 $\mathbf{a}^i_{k|k-1} := p_{x(k)|z(1:k-1)}(i|\bar{z}(1:k-1))$
4. **Algorithm**

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

Recursion $(k > 0)$

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

3 ESTIMATE EXTRACTION

GENERAL MEASUREMMENT

General Measurement
- m measure.
- n states

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

Noise $w_i \sim \mathcal{N}(0,1)$ – Zero-mean additive gaussian

Independence x, w_i $\forall i$ assumed mutually independent

ML – MAXIMUM LIKELIHOOD

Interpretation Choice of x to make $\bar{z}$ most likely

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

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

MAP – MAXIMUM A POSTERIORI

Interpretation ML given observation & prior belief about x

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

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

LS – LEAST SQUARES

Interpretation ML estimate when errors are indep. 0-mean, same variance, $\mathcal{N}$-distributed

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

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

WLS – WEIGHTED LEAST SQUARES

Interpretation Given $1:k$, e-terms with var$\downarrow$ weighed $\uparrow$

$$\begin{aligned} \widehat{x}^{WLS} &:= \mathrm{argmin}_x\left[(\bar{z} - H\widehat{x})^\top R^{-1}(\bar{z} - H\widehat{x})\right] \\ &= (H^\top RH)^{-1}H^\top R^{-1}\bar{z} \end{aligned}$$

Incorporate Prior Knowledge on x

Define: $r \sim \mathcal{N}(0, P_x) \Rightarrow x := \widehat{x} + r$, $\mathbb{E}\{x\} = \widehat{x}_0$, $\mathrm{Var}\{x\} = P_x$

Introduce Extended System:

$$\tilde{\bar{z}} = \begin{bmatrix} \widehat{x}_0 \\ z(1:k) \end{bmatrix},\ \tilde{H} = \begin{bmatrix} \mathbb{I} \\ H(1:k) \end{bmatrix},\ \tilde{w} = \begin{bmatrix} -r \\ w(1:k) \end{bmatrix},\ \tilde{R} = \begin{bmatrix} P_x & 0 \\ 0 & R(1:k) \end{bmatrix}$$

RLS – RECURSIVE LEAST SQUARES

Interpretation Update $\widehat{x}$ recursively when new $\bar{z}(k)$ arrives

Estimation Error (Matrices at k) $e(k) = (\mathbb{I} - KH)\,e(k-1) - Kw(k)$

Objective Minimize MSE with $P(k) = \mathrm{Var}\{e(k)\}$

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

Algorithm

Initialization: $\widehat{x}(0) = \widehat{x}_0, \quad P(0) = P_x = \mathrm{Var}\{x\}$

Recursion: Observe $\bar{z}(k)$, then update:

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

4 KALMAN FILTER

MODEL

Linear Time-Varying System

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

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

GRV – GAUSSIAN RANDOM VARIABLE

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

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

$\rightsquigarrow$ fully characterized by mean $\mu \in \mathbb{R}^D$ and var $\Sigma \in \mathbb{R}^{D\times D}$

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

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

JGRV – JOINTLY GAUSSIAN RANDOM VARIABLES

Jointly GRV if joint vector RV (x,y) a GRV

Property 1 – Affine Transformation of GRV is GRV
y a GRV $\rightsquigarrow x = My + b$ (M,b constant) $\rightsquigarrow x$ also GRV

Property 2 – Linear Comb. of 2 Jointly GRV is GRV
x, y JGRV, $\rightsquigarrow z = M_x x + M_y y$ (M_x, M_y constant matrices) $\rightsquigarrow z$ is a GRV $z \sim \mathcal{N}(\mu_z, \Sigma_z)$

ACHTUNG $x \sim \mathcal{N}(\mu_x, \Sigma_x)$, $y \sim \mathcal{N}(\mu_y, \Sigma_y)$ does not imply (x,y) jointly GRV $\rightsquigarrow$ **Must be additionally independent**

AUXILIARY 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

$$\begin{aligned} &\textbf{Init:} & x_m(0) &= x(0) \\ &\textbf{S1:} & x_p(k) &= A(k{-}1)x_m(k{-}1) + u(k{-}1) + v(k{-}1) \\ &\textbf{S2:} & z_m(k) &= H(k)x_p(k) + w(k) \\ & & p_{x_m(k)}(\xi) &:= p_{x_p(k)|z_m(k)}(\xi|\bar{z}(k)) \quad \forall\xi \end{aligned}$$

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 $k = 1, 2, \dots$

$$\widehat{x}_p(k) := \mathbb{E}\{x_p(k)\} \qquad P_p(k) := \mathrm{Var}\{x_p(k)\}$$
$$\widehat{x}_m(k) := \mathbb{E}\{x_m(k)\} \qquad P_m(k) := \mathrm{Var}\{x_m(k)\}$$

KF – KALMAN FILTER

Recursive Update Equations

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

S1 – Prior Update / Prediction Step

$$\widehat{x}_p(k) = A(k-1)\widehat{x}_m(k-1) + u(k-1)$$
$$P_p(k) = A(k-1)P_m(k-1)A^\top(k-1) + Q(k-1)$$

S2 – A Posteriori Update / Measurement Update

$$P_m(k) = \left(P_p^{-1}(k) + H^\top(k)R^{-1}(k)H(k)\right)^{-1}$$
$$\widehat{x}_m(k) = \widehat{x}_p(k) + \underbrace{P_m(k)H^\top(k)R^{-1}(k)}_{\text{correctionfactor}}\underbrace{(\bar{z}(k) - H(k)\widehat{x}_p(k))}_{\text{predictionerror}}$$

Alternative Equations (all matrices/vectors at time k)

S2 – Kalman Filter Gain

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

ACHTUNG 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

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

ACHTUNG Filter is linear, time-varying

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

ASYMPTOTIC PROPERTIES OF KF

Model assume **time-invariant 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** $q > 0$

DETECTABILITY / STABILIZABILITY

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. $\lim_{k\to\infty} x(k) = 0 \quad \forall x_0 \in \mathbb{R}^n$
- $\Leftrightarrow$ $\begin{bmatrix} A-\lambda\mathbb{I} & B \end{bmatrix}$ 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 if $(A^\top, B^\top)$ detectable

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

Reachability Matrix $\begin{bmatrix} A & A^{n-1}B \end{bmatrix}$ implies stabilizablity

STEADY-STATE KF

DARE – Discrete Algebraic Ricatti Equation

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

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

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 $e(k) = x(k) - x_f(k) \equiv \text{true} - \hat{x}$ with z as RV

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

- $(\mathbb{I} - K_\infty H)A$ 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

5 EKF – EXTENDED KALMAN FILTER

EKF – MODELING

System nonlinear, DT system

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

- $x(0), \{v\}, \{w\}$ mutually indep.
- q_{k-1}, h_k assumed cont. diff'bar

	$\mathbb{E}\{\cdot\}$	$\mathrm{Var}\{\cdot\}$
$x(0)$	x_0	P_0
$v(k-1)$	0	$Q(k-1)$
$w(k)$	0	$R(k)$

Distribution not assumed GRV, but at least unimodal

EKF DT EQUATIONS

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

S1 – Prior Update / Prediction Step **(Vec/Mat. at k-1 unless noted)**

$$\hat{x}_p(k) = q_{k-1}(\hat{x}_m(k-1), 0) \qquad A(k-1) := \frac{\partial q_{k-1}}{\partial x}(\hat{x}_m(k-1), 0)$$
$$P_p(k) = AP_mA^\top + LQL^\top \qquad L(k-1) := \frac{\partial q_{k-1}}{\partial v}(\hat{x}_m(k-1), 0)$$

S2 – A Posteriori / Measurement Update **(at k unless noted)**

$$K = P_pH^\top\left(HP_pH^\top + MRM^\top\right)^{-1} \qquad H := \frac{\partial h_k}{\partial x}(\hat{x}_p(k, 0))$$
$$\hat{x}_m = \hat{x}_p + K\left(\bar{z} - h_k(\hat{x}_p(k), 0)\right) \qquad M := \frac{\partial h_k}{\partial w}(\hat{x}_p(k, 0))$$
$$P_m = (\mathbb{I} - K(k)H(k))\, P_p(k)$$

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

ACHTUNG

- 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.**

HYBRID EKF – MODELING

System **Process** – nlin, CT, **Measurement** – nlin, DT

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

CT White Noise

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

ACHTUNG **True CT White noise does not exist** $\rightsquigarrow$ 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

ACHTUNG

- Process & measurement noise **assumed** 0-mean
- ODEs typically solved num. Accuracy $\uparrow \rightsquigarrow$ Comp. Cost $\uparrow$

6 PF – PARTICLE FILTER

MCS – MONTE CARLO SAMPLING – DRV

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

Helper Variables

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

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

MCS – 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

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

PF – MODELING

System **NL, DT**

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

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

Objective **Approximate** Bayesian State Estimator with MCS

Auxiliary Variables Same as KF

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.

PF – 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 $\{\bar{x}_m^n(k-1)\}$
$$\bar{x}_p^n(k) := q_{k-1}(\bar{x}_m^n(k-1), \bar{v}^n(k-1)), \qquad n = 1, \dots, N$$
Requires N noise samples from $p_{v(k-1)}$

S2 – A Posteriori Update / Measurement Update

Scale each particle by measurement likelihood:
$$\beta_n = \alpha p_{z(k)\mid x(k)}(\bar{z}(k)\mid\bar{x}_p^n(k)), \qquad n = 1, \dots, N$$
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.

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 $\Delta x_i^n(k)$
$$\sigma_i = KE_iN^{-\frac{1}{d}}$$

$K:$	tuningparam.$K \ll 1$	$d:$	dim.ofstatespace
$E_i:$	maxinter − samplevariablility $\max_{n_1,n_2}\lvert\bar{x}_{m,i}^{n_1} - \bar{x}_{m,i}^{n_2}\rvert$	$N^{-\frac{1}{d}}:$	nodespacingongrid

7 OBSERVER-BASED CONTROL

LTI OBSERVER

LTI System

- v, w 0-mean noise CRVs

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

- $\mathbb{E}\{\hat{x}(k)\} \xrightarrow{k\to\infty} \mathbb{E}\{x(k)\}$, $\mathrm{Var}\{\hat{x}(k)\}$ 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-dyn. of observ. 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

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

Error dynamics

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

STATIC STATE-FEEDBACK 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

LQR Feedback Controller Assuming (A, B) stabil., (A, G) detect.
$$F = -(B^\top PB + \bar{R})^{-1}B^\top PA$$
DARE Same as KF but $A \to A^\top, H \to B^\top, Q \to \bar{Q}, R \to \bar{R}$
$$P = A^\top A + \bar{Q} - A^\top PB(B^\top PB + \bar{R})^{-1}B^\top PA$$

SEPARATION PRINCIPLE

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

Dynamics w/o noise

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

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

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

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. Profit.

8 APPENDIX

$$xy^\top = \begin{bmatrix} x_1y & \dots & x_ny \end{bmatrix} \qquad \langle x, y\rangle = x^\top y = \sum x_iy_i$$

$$\frac{\partial}{\partial x}x^\top A = \frac{\partial}{\partial x}A^\top x = A \qquad \frac{\partial}{\partial x}x^\top Ax = (A + A^\top)x$$

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

$$\frac{\partial f}{\partial x^\top} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \qquad \frac{\partial f}{\partial A} = \begin{bmatrix} \frac{\partial f}{\partial A_{11}} & \cdots & \frac{\partial f}{\partial A_{1m}} \\ \cdots & \cdots & \cdots \\ \frac{\partial f}{\partial A_{n1}} & \cdots & \frac{\partial f}{\partial A_{nm}} \end{bmatrix}$$

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

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