

Optimal State Estimation Kalman, H_∞ , and Nonlinear Approaches Notes

Luke Wicent Sy

October 21, 2019

This is my personal attempt to the "digest" or "solve the end chapter problems in" the book "Simon, Dan. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. John Wiley & Sons, 2006". This document is by no means a perfect solution manual for the said book, but was written simple for my personal enrichment.

1 Linear Systems Theory

1.1 Matrix Algebra and Matrix Calculus

Significance: most if not all optimal state estimation are formulated with matrices.

1.1.1 Matrix Algebra

- Transpose: A^T
- Hermitian: A^H (complex conjugate + transpose)
- $HPH^T = \sum_{j,k} H_j P_{jk} H_k^T$
- Determinant: $|A|$
- $|A| = \sum_{j=1}^n (-1)^{i+j} A_{ij} |A^{(i,j)}|$
- $|A| = \sum_{i=1}^n (-1)^{i+j} A_{ij} |A^{(i,j)}|$
- $|AB| = |A||B|$
- $|A| = \prod_{i=1}^n \lambda_i$
- Trace: $Tr(A)$ = sum of its eigenvalues
- $Tr(xx^T) = \|x\|_2^2$
- Positive definite ($x^T Ax > 0$), Positive semidefinite ($x^T Ax \geq 0$), Negative definite ($x^T Ax < 0$), Negative semidefinite ($x^T Ax \leq 0$), Indefinite. For all (non zero) $n \times 1$ vector x .
- Weighted 2 Norm $\|x\|_Q^2 = \sqrt{x^T Q x}$
- Eigenvalues λ_i : transformation from 1 vector space to itself
- If $A = n \times n$, A has n eigenvalues.
- Singular values σ : transformation from 1 vector space to a different vector space
- If $A = n \times m$, A has $\min(n, m)$ singular values.
- $\sigma^2(A) = \lambda(A^T A) = \lambda(AA^T)$

1.1.2 Matrix Inversion Lemma

Purpose: reduce computational effort of matrix inversion if you already know A and A^{-1} , and the change in A is only in a certain location.

Given

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

, we have the ff. formulas:

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} \quad (1.1)$$

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A||D - CA^{-1}B| = |D||A - BD^{-1}C| \quad (1.2)$$

1.1.3 Matrix Calculus

With respect to scalar t

$$\dot{A}(t) = \begin{bmatrix} \dot{A}_{11}(t) & \dots & \dot{A}_{1n}(t) \\ \vdots & \ddots & \vdots \\ \dot{A}_{m1}(t) & \dots & \dot{A}_{mn}(t) \end{bmatrix} \quad (1.3)$$

$$\frac{d}{dt}(A^{-1}) = -A^{-1}\dot{A}A^{-1} \quad (1.4)$$

where $A = m \times n$ matrix, and t is scalar.

With respect to vector x (or y)

$$\frac{\delta f}{\delta x} = [\delta f / \delta x_1 \quad \dots \quad \delta f / \delta x_n] \quad (1.5)$$

$$\frac{\delta(x^T y)}{\delta x} = y^T \quad (1.6)$$

$$\frac{\delta(x^T y)}{\delta y} = x^T \quad (1.7)$$

$$\frac{\delta(x^T Ax)}{\delta x} = x^T A^T + x^T A \quad (1.8)$$

$$\frac{\delta(Ax)}{\delta x} = \frac{\delta(x^T A)}{\delta x} = A \quad (1.9)$$

$$\frac{\delta g}{\delta x} = \begin{bmatrix} \delta g_1 / \delta x_1 & \dots & \delta g_1 / \delta x_n \\ \vdots & \ddots & \vdots \\ \delta g_m / \delta x_1 & \dots & \delta g_m / \delta x_n \end{bmatrix} \quad (1.10)$$

$$\frac{\delta g^T}{\delta x} = \frac{\delta g}{\delta x^T} = \left(\frac{\delta g}{\delta x} \right)^T \quad (1.11)$$

$$\frac{\delta g^T}{\delta x^T} = \frac{\delta g}{\delta x} \quad (1.12)$$

where $f(x)$ is a function, A is a $n \times n$ matrix, x, y are $n \times 1$ vectors, and $g(x) = [g_1(x) \quad \dots \quad g_m(x)]^T$.

With respect to matrix A

$$\frac{\delta f}{\delta A} = \begin{bmatrix} \delta f / \delta \dot{A}_{11} & \dots & \delta f / \delta \dot{A}_{1n} \\ \vdots & \ddots & \vdots \\ \delta f / \delta \dot{A}_{m1} & \dots & \delta f / \delta \dot{A}_{mn} \end{bmatrix} \quad (1.13)$$

$$\frac{\delta Tr(ABA^T)}{\delta A} = AB^T + AB \quad (1.14)$$

1.2 Linear Systems

Many process in our world can be described by state-space systems. If we can derive a mathematical model for a process, then we can use the tools of mathematics to control the process and obtain information about the process.

General Form:

$$\dot{x} = Ax + Bu, y = Cx$$

where A is called system matrix, B is input matrix, and C is output matrix.

Solution (always works even if A, B, C is scalar, matrix, or vector):

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^t e^{A(t-\tau)}Bu(\tau)d\tau \quad (1.15)$$

$$y(t) = Cx(t)$$

where the integral part goes to zero on "zero input case".

e^{At} properties (this is also referred to as state transition matrix)

$$e^{At} = \sum_{j=0}^{\infty} \frac{(At)^j}{j!} \quad (1.16)$$

$$= \mathcal{L}^{-1} \left[(sI - A)^{-1} \right] \quad (1.17)$$

$$= Qe^{\hat{A}t}Q^{-1} \quad (1.18)$$

$$\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A \quad (1.19)$$

$$e^{\hat{A}t} = \begin{bmatrix} e^{\hat{A}_{11}t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{\hat{A}_{nn}t} \end{bmatrix} \quad (1.20)$$

$$e^{-At} = Qe^{-\hat{A}t}Q^{-1} \quad (1.21)$$

where Q matrix' columns comprise the eigenvectors of A , and \hat{A} is the Jordan form of A .

1.3 Nonlinear Systems

General Form:

$$\dot{x} = f(x, u, w)$$

$$y = h(x, v)$$

Linear Approximation using Taylor Series

$$D_{\bar{x}}^k f = \left(\sum_{i=1}^n \tilde{x}_i \frac{\delta}{\delta x_i} \right)^k f(x) \Big|_{\bar{x}} \quad (1.22)$$

$$f(x) = f(\bar{x}) + D_{\bar{x}}^1 f + \frac{1}{2!} D_{\bar{x}}^2 f + \dots \quad (1.23)$$

$$\approx f(\bar{x}) + A\tilde{x} \quad (1.24)$$

where \bar{x} is the nominal operating point and $\tilde{x} = x - \bar{x}$.

Approximated General Form:

$$\begin{aligned} \dot{\tilde{x}} &= A\tilde{x} + B\tilde{u} + L\tilde{w} \\ \tilde{y} &= C\tilde{x} + D\tilde{v} \end{aligned} \quad (1.25)$$

where $\tilde{w} = w$ and $\tilde{v} = v$ because $\bar{w} = \bar{v} = 0$, $A = \left. \frac{\delta f}{\delta x} \right|_{\bar{x}}$, $B = \left. \frac{\delta f}{\delta u} \right|_{\bar{x}}$, $L = \left. \frac{\delta f}{\delta w} \right|_{\bar{x}}$, $C = \left. \frac{\delta h}{\delta x} \right|_{\bar{x}}$, and $D = \left. \frac{\delta h}{\delta v} \right|_{\bar{x}}$.

1.4 Discretization

base equation (source of general form)

$$x(t_k) = e^{A\Delta t}x(t_{k-1}) + e^{A\Delta t} \int_0^{\Delta t} e^{-A\alpha} d\alpha Bu(t_{k-1}) \quad (1.26)$$

$$\int_0^{\Delta t} e^{-A\alpha} d\alpha = [I - e^{-A\Delta t}]A^{-1} \quad (1.27)$$

where $\Delta t = t_k - t_{k-1}$ and $\alpha = \tau - t_{k-1}$.

Discrete General Form:

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + G_{k-1}u_{k-1} \\ F_{k-1} &= e^{A\Delta t} \\ G_{k-1} &= e^{A\Delta t}[I - e^{-A\Delta t}]A^{-1}B \end{aligned} \quad (1.28)$$

1.5 Simulation

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} f[x(t), u(t), t]dt$$

1.5.1 Rectangular Integration / Euler Integration (error = $O(T)$)

$$x(t_{n+1}) \approx x(t_n) + f[x(t_n), u(t_n), t_n]T \quad (1.29)$$

1.5.2 Trapezoidal Integration (error = $O(T^2)$)

$$\begin{aligned} \Delta x_1 &= f[x(t_n), u(t_n), t_n]T \\ \Delta x_2 &= f[x(t_n) + \Delta x_1, u(t_{n+1}), t_{n+1}]T \end{aligned} \quad (1.30)$$

$$x(t_{n+1}) \approx x(t_n) + \frac{1}{2}(\Delta x_1 + \Delta x_2)$$

1.5.3 Runge Kutta Integration (4th order) (error = $O(T^4)$ or $O(T^n)$)

$$\begin{aligned} \Delta x_1 &= f[x(t_k), u(t_k), t_k]T \\ \Delta x_2 &= f[x(t_k) + \Delta x_1/2, u(t_{k+1/2}), t_{k+1/2}]T \\ \Delta x_3 &= f[x(t_k) + \Delta x_2/2, u(t_{k+1/2}), t_{k+1/2}]T \\ \Delta x_4 &= f[x(t_k) + \Delta x_3, u(t_{k+1}), t_{k+1}]T \\ x(t_{n+1}) &\approx x(t_k) + \frac{1}{6}(\Delta x_1 + 2\Delta x_2 + 2\Delta x_3 + \Delta x_4) \end{aligned} \quad (1.31)$$

	LTI Continuous-time, System	LTI Discrete-time
General Equation	$\dot{x} = Ax + Bu$ $y = Cx$ $x(t) = \exp(At)x(0)$	$x_k = Fx_{k-1} + Gu_{k-1}$ $y_k = Hx_k$ $x_k = A^k x_0$
Marginal / Lyapunov Stability	<p>If the state $x(t)$ is bounded for all t and for all bounded initial states $x(0)$.</p> <p>If and only if $\lim_{t \rightarrow \infty} \exp(At) \leq M < \infty$ for some constant matrix M.</p> <p>If and only if one of the ff. conditions holds.</p> <ol style="list-style-type: none"> All of the eigenvalues of A have negative real parts. All of the eigenvalues of A have negative or zero real parts, and those with real parts equal to zero have a geometric multiplicity equal to their algebraic multiplicity. That is, the Jordan blocks that are associated with the eigenvalues that have real parts equal to zero are first order. 	<p>If the state x_k is bounded for all k and for all bounded initial states x_0.</p> <p>If and only if $\lim_{k \rightarrow \infty} A^k \leq M < \infty$ for some constant matrix M.</p> <p>If and only if one of the ff. conditions holds.</p> <ol style="list-style-type: none"> All of the eigenvalues of A have magnitude less than one. All of the eigenvalues of A have magnitude less than or equal to one, and those with magnitude equal to one have a geometric multiplicity equal to their algebraic multiplicity. That is, the Jordan blocks that are associated with the eigenvalues that have magnitude equal to one are first order.
Asymptotic Stability	<p>If, for all bounded initial states $x(0)$, $\lim_{t \rightarrow \infty} x(t) = 0$.</p> <p>If and only if $\lim_{t \rightarrow \infty} \exp(At) = 0$</p> <p>If and only if all of the eigenvalues of A have negative real parts.</p>	<p>If $\lim_{k \rightarrow \infty} x_k = 0$ for all bounded initial states x_0.</p> <p>If and only if $\lim_{k \rightarrow \infty} A^k = 0$.</p> <p>If and only if all of the eigenvalues of A have magnitude less than one.</p>
Controllable	<p>If for any initial state $x(0)$ and any final time $t > 0$ there exists a control that transfers the state to any desired value at time t.</p> <p>The n-state system has the controllability matrix P defined by $P = [B \ AB \ \dots \ A^{n-1}B]$. The system is controllable if and only if $\text{rank}(P) = n$.</p> <p>If and only if the controllability grammian defined by $\int_0^t e^{A\tau} B B^T e^{A^T \tau} d\tau$ is positive definite for some $t \in (0, \infty)$.</p> <p>If and only if the differential Lyapunov equation $W(0) = 0$ and $\dot{W} = WA^T + AW + BB^T$ has a positive definite solution $W(t)$ for some $t \in (0, \infty)$. This is also called a Sylvester equation.</p>	<p>If for any initial state x_0 and some final time k there exists a control that transfers the state to any desired value at time k.</p> <p>The n-state system has the controllability matrix P defined by $P = [G \ FG \ \dots \ F^{n-1}G]$. The system is controllable if and only if $\text{rank}(P) = n$.</p> <p>If and only if the controllability grammian defined by $\sum_{i=0}^k A^{k-i} B B^T (A^T)^{k-i}$ is positive definite for some $k \in (0, \infty)$.</p> <p>If and only if the difference Lyapunov equation $W_0 = 0$ and $W_{i+1} = FW_i F^T + GG^T$ has a positive definite solution W_k for some $k \in (0, \infty)$. This is also called a Stein equation.</p>
Observable	<p>If for any initial state $x(0)$ and any final time $t > 0$ the initial state $x(0)$ can be uniquely determined by knowledge of the input $u(\tau)$ and output $y(\tau)$ for all $\tau \in [0, t]$.</p> <p>The n-state system has the observability matrix Q defined by $Q = [C \ CA \ \dots \ CA^{n-1}]^T$. The system is observable if and only if $\text{rank}(Q) = n$.</p> <p>If and only if the observability grammian defined by $\int_0^t e^{A^T \tau} C^T C e^{A\tau} d\tau$ is positive definite for some $t \in (0, \infty)$.</p> <p>If and only if the differential Lyapunov equation $W(t) = 0$ and $-\dot{W} = WA + A^T W + C^T C$ has a positive definite solution $W(\tau)$ for some $\tau \in (0, t)$. This is also called a Sylvester equation.</p>	<p>If for any initial state x_0 and some final time k the initial state x_0 can be uniquely determined by knowledge of the input u_i and output y_i for all $i \in [0, k]$.</p> <p>The n-state system has the observability matrix Q defined by $Q = [H \ HF \ \dots \ HF^{n-1}]^T$. The system is observable if and only if $\text{rank}(Q) = n$.</p> <p>If and only if the observability grammian defined by $\sum_{i=0}^k (F^T)^i H^T H F^i$ is positive definite for some $k \in (0, \infty)$.</p> <p>If and only if the difference Lyapunov equation $W_k = 0$ and $W_i = F^T W_{i+1} F + H^T H$ has a positive definite solution W_0 for some $k \in (0, \infty)$. This is also called a Stein equation.</p>

Table 1: Stability, Controlability, and Observability

1.6 Stability, Controllability, and Observability

Linear Time Invariant (LTI). Refer to table 1.

Stabilizability and Detectability

The modes of a system are all of the decoupled states after the system is transformed into Jordan form. Given

$$\dot{x} = Ax + Bu = Cx + Du$$

and

$$M = [v_1 \ \dots \ v_n]$$

(note that M is guaranteed to be invertible),

$$\begin{aligned} \dot{\tilde{x}} &= M^{-1}AM\tilde{x} + M^{-1}B \\ &= \tilde{A}\tilde{x} + \tilde{B}u \\ y &= CM\tilde{x} + Du \\ &= \tilde{C}\tilde{x} + Du \end{aligned} \quad (1.32)$$

- If a system is controllable or stable, then it is also stabilizable. If a system is uncontrollable or unstable, then it is stabilizable if its uncontrollable modes are stable.
- If a system is observable or stable, then it is also detectable. If a system is unobservable or unstable, then it is detectable if its unobservable modes are stable.

2 Probability Theory

2.1 Probability

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (2.1)$$

$$P(A) = \frac{\# \text{ of times A occurs}}{\text{total } \# \text{ of outcomes}} \text{ apriori} \quad (2.2)$$

$$P(A|B) = \frac{P(A, B)}{P(B)} \text{ aposteriori} \quad (2.3)$$

$$= \frac{P(B|A)P(A)}{P(B)} \quad (2.4)$$

$$(2.5)$$

If A and B are independent:

$$P(A, B) = P(A) * P(B) \quad (2.6)$$

$$P(A|B) = P(A) \quad (2.7)$$

$$P(B|A) = P(B) \quad (2.8)$$

2.2 Random Variables (RV)

2.2.1 Probability Distribution Function

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ F_X(x) &\in [0, 1] \\ F_X(-\infty) &= 0 \\ F_X(\infty) &= 1 \\ F_X(a) &\leq F_X(b) \text{ if } a \leq b \\ P(a < x \leq b) &= F_X(b) - F_X(a) \end{aligned} \quad (2.9)$$

2.2.2 Probability Density Function

$$\begin{aligned} f_X(x) &= \frac{dF_X(x)}{dx} \\ F_X(x) &= \int_{-\infty}^x f_X(x)dx \\ f_X(x) &\geq 0 \\ \int_{-\infty}^{\infty} f_X(x)dx &= 1 \\ P(a < x \leq b) &= \int_a^b f_X(x)dx \\ Q(x) &= 1 - F_X(x) \\ &= P(X > x) \end{aligned} \quad (2.10)$$

$$\begin{aligned} \text{Chapman Kolmogorov:} \\ f[x_1|x_2, x_3, x_4]f[x_2, x_3|x_4] &= f[x_1, x_2, x_3|x_4] \end{aligned} \quad (2.11)$$

2.2.3 i th Moment and Central Moment

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \quad (2.13)$$

$$E[(x - \bar{x})^2] = \int_{-\infty}^{\infty} (x - \bar{x})^2 f_X(x)dx \quad (2.14)$$

$$= E(x^2) - \bar{x}^2 \quad (2.15)$$

$$= \sigma_x^2 \quad (2.16)$$

$$x \sim (\bar{x}, \sigma^2) \quad (2.17)$$

$$i\text{th moment of } x = E(x^i) \quad (2.18)$$

$$i\text{th central moment of } x = E[(x - \bar{x})^i] \quad (2.19)$$

$$\text{skew} = 3\text{rd central moment} \quad (2.20)$$

$$\text{skewness} = \text{skew}/\sigma^3 \quad (2.21)$$

Uniform RV:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, x \in [a, b] \\ 0, \text{ otherwise} \end{cases} \quad (2.22)$$

2.2.4 Gaussian or Normal RV

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \bar{x})^2}{2\sigma^2}\right] \quad (2.23)$$

$$x \sim N(\bar{x}, \sigma^2) \quad (2.24)$$

$$F_{X0}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{(z)^2}{2}\right] dz \quad (2.25)$$

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{(z - \bar{x})^2}{2\sigma^2}\right] dz \quad (2.26)$$

$$= F_{X0}\left(\frac{x - \bar{x}}{\sigma}\right) \quad (2.27)$$

$$\approx 1 - \left[\frac{1}{(1-a)x + a\sqrt{x^2 + b}}\right] \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$$

$$\text{for } x \geq 0, a = 0.339, b = 5.510 \quad (2.28)$$

If $f_X(x)$ is odd, odd i th moment = 0

2.3 Transformations of RV

$$\begin{aligned} Y &= g(X) \\ X &= g^{-1}(Y) = h(Y) \end{aligned} \quad (2.29)$$

$$\begin{aligned} P(X \in [x, x + dx]) &= P(Y \in [y, y + dy]) \\ f_Y(y) &= |h'(y)|f_X[h(y)] \end{aligned} \quad (2.30)$$

$$= \sum_i f_X(x_i)/|g'(x_i)| \quad (2.31)$$

where $g(\cdot)$ is non-monotonic

2.4 Multiple RV

$$\begin{aligned} F_{XY}(x, y) &= P(X \leq x, Y \leq y) \\ F(x, -\infty) &= F(-\infty, y) = 0 \\ F(\infty, \infty) &= 1 \\ F(x, \infty) &= F(x) \\ F(\infty, y) &= F(y) \end{aligned} \quad (2.32)$$

$$\begin{aligned} f_{XY}(x, y) &= \frac{d^2 F_{XY}(x, y)}{dxdy} \\ F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(z_1, z_2) dz_1 dz_2 \\ P(a < x \leq b, c < y \leq d) &= \int_c^d \int_a^b f(x, y) dxdy \end{aligned} \quad (2.33)$$

Marginal Density Function:

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ f(y) &= \int_{-\infty}^{\infty} f(x, y) dx \end{aligned} \quad (2.34)$$

$$E[g(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dxdy \quad (2.35)$$

$$C_{XY} = E[(x - \bar{x})(y - \bar{y})] \text{ Covariance} \quad (2.36)$$

$$= E(XY) - \bar{X}\bar{Y} \quad (2.37)$$

$$\rho = \frac{C_{XY}}{\sigma_X \sigma_Y} \text{ Correlation coefficient} \quad (2.38)$$

$$R_{XY} = E(XY) \text{ Correlation} \quad (2.39)$$

2.4.1 Statistical Independence

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \text{ for all } x, y$$

$$\begin{aligned} F_{XY}(x, y) &= F_X(x)F_Y(y) \\ f_{xy}(x, y) &= f_X(x)f_Y(y) \end{aligned} \quad (2.40)$$

$$R_{XY} = E(X)E(Y) \quad (2.41)$$

$$E(X + Y) = E(X) + E(Y) \quad (2.42)$$

- Central Limit Theorem - sum of independent RVs tend toward a gaussian RV.
- Correlation coefficient $\rho = 0$ if X and Y are independent.
- $R_{XY} = 0$ means 2 RVs are orthogonal.

2.4.2 Multivariate Statistics

$$R_{XY} = E(XY^T) \quad (2.43)$$

$$= \begin{bmatrix} E(X_1Y_1) & \dots & E(X_1Y_m) \\ \vdots & \ddots & \vdots \\ E(X_nY_1) & \dots & E(X_nY_m) \end{bmatrix} \quad (2.44)$$

$$C_{XY} = E[(X - \bar{X})(Y - \bar{Y})] \quad (2.45)$$

$$= E(XY^T) - \bar{X}\bar{Y}^T \quad (2.46)$$

$$R_X = E[XX^T] \quad (2.47)$$

$$R_X = R_X^T \quad (2.48)$$

$$z^T R_X z = E[(z^T X)^2] \geq 0 \text{ for all } z \quad (2.49)$$

$$C_X = E[(X - \bar{X})(X - \bar{X})^T] \quad (2.50)$$

$$= \begin{bmatrix} \sigma_{11}^2 & \dots & \sigma_{1n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \dots & \sigma_{nn}^2 \end{bmatrix} \quad (2.51)$$

$$C_X = C_X^T \quad (2.52)$$

$$z^T C_X z \geq 0 \text{ for all } z \text{ (positive semidefinite)} \quad (2.53)$$

Gaussian:

$$f(X) = \frac{1}{(2\pi)^{\frac{n}{2}} |C_X|^{\frac{1}{2}}} \exp \left[\frac{-(x - \bar{x})^T C_X^{-1} (x - \bar{x})}{2} \right] \quad (2.54)$$

$$\text{Given } Y = Ax + b \quad (2.55)$$

$$y \sim N(A\bar{x} + b, AC_X A^T) \quad (2.56)$$

2.5 Stochastic Process

Stochastic process - RV X that changes with time. It can be a combination of 1) continuous OR discrete signal and 2) continuous (process) OR discrete (sequence) time.

2.5.1 Definition

1st Order:

$$\begin{aligned} F_X(x, t) &= P(X(t) \leq x) \\ f(x, t) &= \frac{dF_X(x, t)}{dx_1 \dots dx_n} \\ \bar{x} &= \int_{-\infty}^{\infty} x f(x, t) dx \\ C_X(t) &= \int_{-\infty}^{\infty} [x - \bar{x}(t)][x - \bar{x}(t)]^T f(x, t) dx \end{aligned} \quad (2.57)$$

2nd Order:

$$\begin{aligned} F_X(x_1, x_2, t_1, t_2) &= P(X(t_1) \leq x_1, X(t_2) \leq x_2) \\ f(x_1, x_2, t_1, t_2) &= \frac{dF_X(x_1, x_2, t_1, t_2)}{dx_1 dx_2} \end{aligned} \quad (2.58)$$

Autocorrelation:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)^T] \quad (2.59)$$

Autocovariance:

$$C_X(t_1, t_2) = E\{[X(t_1) - \bar{X}(t_1)][X(t_2) - \bar{X}(t_2)]^T\} \quad (2.60)$$

2.5.2 Stationary

$$E[X(t)] = \bar{X} \quad (2.61)$$

$$E[X(t_1)X(t_2)^T] = R_X(t_2 - t_1) \quad (2.62)$$

- Strict Sense Stationary - pdf does not change w/ time.
- Wide Sense Stationary - satisfies the 2 eq. above, but is not Strict Sense Stationary.

Wide Sense Stationary Eq:

$$R_X(0) = E[X(t)X(t)^T] \quad (2.63)$$

$$R_X(-\tau) = R_X(\tau) \quad (2.64)$$

$$|R_X(\tau)| \leq R_X(0) \text{ for scalar} \quad (2.65)$$

2.5.3 Another Representation

Time Average of $x(t)$:

$$A[X(t)] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt \quad (2.66)$$

Time Autocorrelation:

$$R[X(t), \tau] = A[X(t)X(t + \tau)^T] \quad (2.67)$$

Ergodic Process:

$$A[X(t)] = E(X) \quad (2.68)$$

$$R[X(t), \tau] = R_X(\tau) \quad (2.69)$$

Multiple Process:

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^T] \quad (2.70)$$

$$C_{XY}(t_1, t_2) = E\{[X(t_1) - \bar{X}(t_1)][Y(t_2) - \bar{Y}(t_2)]^T\} \quad (2.71)$$

$$R_{XY}(t_1, t_2) = E[X(t_1)]E[Y(t_2)^T] \text{ if uncorrelated} \quad (2.72)$$

2.6 White Noise and Colored Noise

2.6.1 Definition

- White noise - $x(t_1)$ is uncorrelated w/ $x(t_2)$ for all $t_1 \neq t_2$.
- Colored noise - otherwise

$$\mathcal{F}[f(t)] = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt \quad (2.73)$$

$$\mathcal{F}^{-1}[F(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega \quad (2.74)$$

2.6.2 Wiener Khintchine Relation

Power Density Spectrum / Power Spectral Density:

$$S_X(\omega) = \mathcal{F}[R_X(\tau)] \quad (2.75)$$

$$R_X(\tau) = \mathcal{F}^{-1}[S_X(\omega)] \quad (2.76)$$

$$\text{Power } P_X = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega \quad (2.77)$$

$$S_{XY}(\omega) = \mathcal{F}[R_{XY}(\tau)] \quad (2.78)$$

2.6.3 White Noise Properties

$$\text{Kronecker } \delta_k = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.79)$$

$$R_X(k) = \sigma^2 \delta_k \quad (2.80)$$

$$S_X(\omega) = R_X(0) \quad (2.81)$$

$$R_X(\tau) = R_X(0)\delta(\tau) \quad (2.82)$$

2.7 Simulating Correlated Noise

1. Find the eigenvalues u_1^2, \dots, u_n^2 and eigenvectors $D = [d_1 \dots d_n]$ from covariance matrix Q .
2. Compute v where $v_i = u_i r_i$ and r_i is an independent RV with $\sigma^2 = 1$.
3. Compute $w = Dv$ and return w .

3 Least Squares Estimation

3.1 Estimation of a constant

$$\begin{aligned} y &= Hx + v \\ \hat{y} &= H\hat{x} \\ \epsilon_y &= y - H\hat{x} \end{aligned} \quad (3.1)$$

Minimize $\epsilon_y^T \epsilon_y$ by deriving wrt to \hat{x} and equating to 0.

$$\epsilon_y^T \epsilon_y = (y - H\hat{x})^T (y - H\hat{x}) \quad (3.2)$$

$$= y^T y - \hat{x}^T H^T y - y^T H \hat{x} + \hat{x}^T H^T H \hat{x} \quad (3.3)$$

$$0 = -2y^T H + 2\hat{x}^T H^T H \quad (3.4)$$

$$\hat{x} = (H^T H)^{-1} H^T y \quad (3.5)$$

3.2 Weighted Least Squares Estimation

$$R = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_k^2 \end{bmatrix} \quad (3.6)$$

$$J = \epsilon_y^T R^{-1} \epsilon_y \quad (3.7)$$

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} y \quad (3.8)$$

\hat{x} was derived by minimizing J .

3.3 Recursive Least Squares Estimation

$$\begin{aligned} y_k &= H_k x + v_k \\ \hat{x}_k &= \hat{x}_{k-1} + K_k (y_k - H_k \hat{x}_{k-1}) \end{aligned} \quad (3.9)$$

K_k is called "Estimator gain matrix". $y_k - H_k \hat{x}_{k-1}$ is called "correction term". y_k and \hat{x}_k are called "unbiased estimator".

1. Initialize $\hat{x}_0 = E[X]$ and $P_0 = E[(x - \hat{x}_0)(x - \hat{x}_0)^T]$. If no knowledge, $P_0 = \infty I$. If perfect knowledge, $P_0 = 0$.
2. For $k = 1, 2, \dots$

- Obtain y_k where v_k is zero mean vector w/ covariance R_k . Note $E[v_i v_k] = R_X \delta_{k-i}$ (white noise).
- Update the ff. equations

$$K_k = P_{k-1} H_k^T (H_k P_{k-1} H_k^T + R_k)^{-1} \quad (3.10)$$

$$= P_k H_k^T R_k^{-1} \quad (3.11)$$

$$\hat{x}_k = \hat{x}_{k-1} + K_k (y_k - H_k \hat{x}_{k-1}) \quad (3.12)$$

$$P_k = (I - K_k H_k) P_{k-1} (I - K_k H_k)^T + K_k R_k K_k^T \quad (3.13)$$

$$= (I - K_k H_k) P_{k-1} \text{ sub } K_k \text{ opt} \quad (3.14)$$

$$= [P_{k-1}^{-1} + H_k^T R_k^{-1} H_k]^{-1} \text{ matrix inv lemma} \quad (3.15)$$

3.4 Wiener Filtering

LTI filter to extract a signal from noise, approaching the problem from the frequency domain perspective.

$$R_y(\alpha) = \int \int g(\tau) g(\gamma) R_x(\alpha + \tau - \gamma) d\tau d\gamma \quad (3.16)$$

$$S_y(\omega) = G(-\omega) G(\omega) S_x(\omega) \quad (3.17)$$

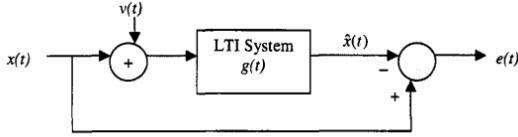


Figure 3.2 Wiener filter representation.

$$E(\omega) = [1 - G(\omega)] X(\omega) - G(\omega) V(\omega) \quad (3.18)$$

$$S_e(\omega) = [1 - G(\omega)][1 - G(-\omega)] S_x(\omega) - G(\omega) G(-\omega) S_v(\omega) \quad (3.19)$$

$$E[e^2(t)] = \frac{1}{2\pi} \int S_e(\omega) d\omega \quad (3.20)$$

3.4.1 Parametric Filter Optimization

Assuming $G(\omega)$ is a first order low pass filter with BW $\frac{1}{T}$, we have $G(\omega) = \frac{1}{1 + Tj\omega}$. Suppose $S_x(\omega)$ and $S_v(\omega)$ are as follows, we have T_{opt} :

$$S_x(\omega) = \frac{2\sigma^2\beta}{\omega^2 + \beta^2} \quad (3.21)$$

$$S_v(\omega) = A \quad (3.22)$$

$$T_{opt} = \frac{\text{sqr}t A}{\sigma\sqrt{2\beta} - \beta\sqrt{A}} \quad (3.23)$$

3.4.2 General Filter Optimization

Use calculus of variation to differentiate and minimize $E[e^2(t)]$:

$$E[e^2(t)] = E[x^2(t)] - 2 \int g(u) R_x(u) du + \quad (3.24)$$

$$\int \int g(u) g(\gamma) [R_x(u-v) + R_v(u-v)] du d\gamma \quad (3.25)$$

$$g(t) \rightarrow g(t) + \epsilon \nu(t) \quad (3.26)$$

$$\frac{\delta E(e^2(t))}{\delta \epsilon} \Big|_{\epsilon=0} = 0 \quad (3.27)$$

$$(3.28)$$

Then solve $g(t)$ from $\int \nu(\tau) [-R_x(\tau) + \int g(u) [R_x(u-\tau) + R_v(u-\tau)] du] d\tau = 0$.

3.4.3 Noncausal Filter Optimization

From the equation to be solved in "Parametric Filter Optimization", if we do not have any restriction on causality of our filter, then $g(t)$ and $\nu(t)$ can be nonzero for $t < 0$ which will give us the ff:

$$R_x(\tau) = g(\tau) * [R_x(\tau) + R_v(\tau)] \quad (3.29)$$

$$S_x(\omega) = G(\omega) [S_x(\omega) + S_v(\omega)] \quad (3.30)$$

$$G(\omega) = \frac{S_x(\omega)}{S_x(\omega) + S_v(\omega)} \quad (3.31)$$

3.4.4 Causal Filter Optimization

From the equation to be solved in "Parametric Filter Optimization", if we require a causal filter for signal estimation, then $g(t) = 0$ and $\nu(t) = 0$ for $t < 0$. Using Wiener-Hopf equation we can solve this as follows:

$$a(\tau) = \begin{cases} \text{some number, } t > 0 \\ 0, t \geq 0 \end{cases} \quad (3.32)$$

$$= R_x(\tau) - \int g(u) [R_x(u-\tau) + R_v(u-\tau)] du \quad (3.33)$$

$$A(\omega) = S_x(\omega) - G(\omega) [S_x(\omega) + S_v(\omega)] \quad (3.34)$$

$$S_{xv}(\omega) = S_x(\omega) + S_v(\omega) \quad (3.35)$$

$$S_{xv}^+(\omega) = \text{poles \& zeros at LHP of } S_{xv}(\omega) \quad (3.36)$$

$$S_{xv}^-(\omega) = \text{poles \& zeros at RHP of } S_{xv}(\omega) \quad (3.37)$$

$$G(\omega) = \frac{1}{S_{xv}^+(\omega)} [\text{causal part of } \frac{S_x(\omega)}{S_{xv}^-(\omega)}] \quad (3.38)$$

4 Propagation of states and covariances

4.1 Discrete Time Systems

$$x_k = F_{k-1} x_{k-1} + G_{k-1} u_{k-1} + w_{k-1} \quad (4.1)$$

$$\bar{x}_k = E(x_k) = F_{k-1} \bar{x}_{k-1} + G_{k-1} u_{k-1} \quad (4.2)$$

$$P_k = E[(x_k - \bar{x}_k)(\dots)^T] = F_{k-1} P_{k-1} F_{k-1}^T + Q_{k-1} \quad (4.3)$$

Note: x_k is a linear combination of x_0 , $\{w_i\}$, $\{u_i\}$. If we assume $\{u_i\}$ is known, x_0 and $\{w_i\}$ as Gaussian, we can fully characterize x_k as $x_k \sim N(\bar{x}_k, P_k)$.

Theorem 21 Discrete-time Lyapunov Eq. Consider the equation $P = FPF^T + Q$ where F and Q are real matrices. Denote by $\lambda_i(F)$ the eigenvalues of the F matrix.

1. A unique solution P exists if and only if $\lambda_i(F)\lambda_j(F) \neq 1$ for all i, j . This unique solution is symmetric.
2. If F is stable then the discrete-time Lyapunov equation has a solution P that is unique and symmetric. $P = \sum_{i=0}^{\infty} F^i Q (F^T)^i$
3. If F is stable and Q is positive (semi)definite, then the unique solution P is symmetric and positive (semi)definite.
4. If F is stable, Q is positive semidefinite, and $(F, Q^{1/2})$ is controllable, then P is unique, symmetric, and positive definite. Note that $Q^{1/2}(Q^{1/2})^T = Q$.

Many times, process noise is first multiplied by some matrix before it enters the system dynamics. We can represent this as:

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + L_{k-1}w_{k-1} \quad (4.4)$$

$$w_{k-1} \sim (0, Q_k)$$

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} \quad (4.5)$$

$$w_{k-1} \sim (0, L_k Q_k L_k^T)$$

$$y_k = H_k x_k + L_k v_k, v_k \sim (0, R_k) \quad (4.6)$$

$$y_k = H_k x_k + v_k, v_k \sim (0, L_k R_k L_k^T) \quad (4.7)$$

4.2 Sampled Data Systems

Definition: system whose dynamics are described by a continuous time differential equation, but the input only changes at discrete time instants because the input is generated by a digital computer.

$$\dot{x} = Ax + Bu + w \quad (4.8)$$

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + \int_{t_{k-1}}^{t_k} e^{A(t_k-\tau)} w(\tau) d\tau \quad (4.9)$$

$$F_k = e^{A\Delta t} \quad (4.10)$$

$$G_k = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)} B(\tau) d\tau$$

$$\bar{x}_k = E(x_k) = F_{k-1}\bar{x}_{k-1} + G_{k-1}u_{k-1} \quad (4.11)$$

$$P_k = F_{k-1}P_{k-1}F_{k-1}^T + Q_{k-1} \quad (4.12)$$

$$Q_{k-1} = \int_{t_{k-1}}^{t_k} e^{A(t_k-\tau)} Q_c(\tau) e^{A^T(t_k-\tau)} d\tau \quad (4.13)$$

$$e^{A(t_k-\tau)} \approx I \text{ for } \tau \in [t_{k-1}, t_k] \quad (4.14)$$

$$Q_{k-1} \approx Q_c(t_k)\Delta t$$

4.3 Continuous Time Systems

$$\bar{x}_k = e^{A\Delta t}\bar{x}_{k-1} + \int_{t_{k-1}}^{t_k} e^{A(t_k-\tau)} B(\tau) u(\tau) d\tau \quad (4.15)$$

$$\dot{x} = Ax + Bu + w \quad (4.16)$$

$$\dot{\bar{x}} = A\bar{x} + Bu \quad (4.17)$$

$$P_k = F_{k-1}P_{k-1}F_{k-1}^T + Q_{k-1} \quad (4.18)$$

The mean and covariance for the continuous time system was obtained equating the left hand side to $\frac{\bar{x}_k - \bar{x}_{k-1}}{\Delta t}$ and using $F \approx I + A\Delta t$.

Theorem 22 Continuous-time Lyapunov/Sylvester Eq. Consider the equation $\dot{P} = AP + PA^T + Q_c$ where A and Q_c are real matrices. Denote by $\lambda_i(A)$ the eigenvalues of the A matrix.

1. A unique solution P exists if and only if $\lambda_i(A)\lambda_j(A) \neq 1$ for all i, j . This unique solution is symmetric.
2. If A is stable then the continuous-time Lyapunov equation has a solution P that is unique and symmetric. $P = \sum_{i=0}^{\infty} e^{A^T \tau} Q_c e^{A\tau} d\tau$
3. If A is stable and Q is positive (semi)definite, then the unique solution P is symmetric and positive (semi)definite.
4. If A is stable, Q is positive semidefinite, and $[A, (Q_c^{1/2})^T]$ is controllable, then P is unique, symmetric, and positive definite. Note that $Q_c^{1/2}(Q_c^{1/2})^T = Q_c$.

5 The Discrete-time Kalman Filter

5.1 Derivation

- Kalman filter operates by propagating the mean and covariance of the state through time.
- Our goal is to estimate the state x_k based on our knowledge of the system dynamics and the availability of the noisy measurements y_k .
- P_k = covariance of the estimation error.
- Estimates:

$$\hat{x}_k^+ = E[x_k | y_1 \dots y_k] = \text{a posteriori} \quad (5.1)$$

$$\hat{x}_k^- = E[x_k | y_1 \dots y_{k-1}] = \text{a priori} \quad (5.2)$$

$$\hat{x}_{k|k+N} = E[x_k | y_1 \dots y_{k+N}] = \text{smoothed} \quad (5.3)$$

$$\hat{x}_{k|k-M} = E[x_k | y_1 \dots y_{k-M}] = \text{predicted} \quad (5.4)$$

1. Dynamic system definition

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

$$y_k = H_k x_k + v_k$$

$$E(w_k w_j^T) = Q_k \delta_{k-j}$$

$$E(v_k v_j^T) = R_k \delta_{k-j}$$

$$E(w_k v_j^T) = 0$$

$$(5.5)$$

2. Initialization: $P_0^+ = 0$ if perfect knowledge, $P_0^+ = \infty I$ if no knowledge.

$$\begin{aligned}\hat{x}_0^+ &= E(x_0) \\ P_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]\end{aligned}\quad (5.6)$$

3. KF is computed for each time step $k = 1, 2, \dots$

- (a) Time Update:

$$P_k^- = F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1} \quad (5.7)$$

$$\hat{x}_k^- = F_{k-1} \hat{x}_{k-1}^+ + G_{k-1} u_{k-1} \quad (5.8)$$

- (b) Kalman Gain:

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (5.9)$$

$$= P_k^+ H_k^T R_k^{-1} \quad (5.10)$$

- (c) Measurement Update:

$$\hat{x}_k^+ = \hat{x}_k^- + K_k (y_k - H_k \hat{x}_k^-) \quad (5.11)$$

$$P_k^+ = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T \quad (5.12)$$

$$= [(P_k^-)^{-1} + H_k^T R_k^{-1} H_k]^{-1} \quad (5.13)$$

$$= (I - K_k H_k) P_k^- \quad (5.14)$$

- Eq. 5.12 is also called "Joseph stabilized version". More stable and robust. It is symmetric positive definite given P_k^- is symmetric positive definite.
- Eq. 5.13 is rarely implemented.
- Eq. 5.14 is computationally simpler.
- K_k can be pre-calculated offline except for nonlinear systems.

5.2 KF Properties

Given $\tilde{x}_k = x_k - \hat{x}_k$, \tilde{x} is also a RV. Suppose we want $\min E[\tilde{x}_k^T S_k \tilde{x}_k]$ where S_k is a positive definite user-defined weighting matrix.

- If w_k and v_k are Gaussian, zero-mean, uncorrelated, and white, then the KF is the solution to the above problem.
- If w_k and v_k are zero-mean, uncorrelated, and white, then KF is the optimal linear filter (best filter that is a linear combination of the measurements).
- If w_k and v_k are correlated or colored, or for nonlinear systems, KF can be modified to solve the problem.

The quantity $(y_k - H_k \hat{x}_k^-)$ is called the innovations. It contains new information about the state. It is zero-mean and white with covariance $(H_k P_k^- H_k^T + R_k)$. In fact, KF can be interpreted as a filter that whitens the measurement and extracts the maximum possible info from the measurement. If mean and covariance of innovation is not as expected, then either the system model is incorrect or the assumed noise statistics are incorrect.

5.3 One-Step KF Equations

$$\hat{x}_{k+1}^- = F_k (I - K_k H_k) \hat{x}_k^- + F_k K_k y_k + G_k u_k \quad (5.15)$$

$$P_{k+1}^- = F_k P_k^- F_k^T + Q_k -$$

$$F_k P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} H_k P_k^- F_k^T \quad (5.16)$$

= discrete Riccati eq.

$$\hat{x}_k^+ = (I - K_k H_k) (F_{k-1} \hat{x}_{k-1}^+ + G_{k-1} u_{k-1}) + K_k y_k \quad (5.17)$$

$$P_k^+ = (I - K_k H_k) (F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1}) \quad (5.18)$$

5.4 Alternate Propagation of Covariance

5.4.1 Multiple State Systems

If P_k^- can be factored as $P_k^- = A_k B_k^{-1}$, then $P_{k+1}^- = A_{k+1} B_{k+1}^{-1}$. A and B can propagate as follows:

$$\begin{bmatrix} A_{k+1} \\ B_{k+1} \end{bmatrix} = \begin{bmatrix} (F_k + Q_k F_k^{-T} H_k^T R_k^{-1} H_k) & Q_k F_k^{-T} \\ F_k^{-T} H_k^T R_k^{-1} H_k & F_k^{-T} \end{bmatrix} \begin{bmatrix} A_k \\ B_k \end{bmatrix} \quad (5.19)$$

$$= \Phi \begin{bmatrix} A_k \\ B_k \end{bmatrix} = \Phi^k \begin{bmatrix} P_1^- \\ I \end{bmatrix} \quad (5.20)$$

$$\begin{bmatrix} A_\infty \\ B_\infty \end{bmatrix} \approx \Phi^{2^{\text{large } p}} \begin{bmatrix} P_1^- \\ I \end{bmatrix} \quad (5.21)$$

5.4.2 Scalar Systems

$$\begin{bmatrix} A_k \\ B_k \end{bmatrix} = \Phi^{k-1} \begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = M \begin{bmatrix} \lambda_1^{k-1} & 0 \\ 0 & \lambda_2^{k-1} \end{bmatrix} M^{-1} \begin{bmatrix} P_1^- \\ 1 \end{bmatrix} \quad (5.22)$$

$$P_k^- = \frac{\tau_1 \mu_1^{k-1} (2RH^2 P_1^- - \tau_2) - \tau_2 \mu_2^{k-1} (2H^2 P_1^- - \tau_1)}{2H^2 \mu_1^{k-1} (2RH^2 P_1^- - \tau_2) - 2H^2 \mu_2^{k-1} (2H^2 P_1^- - \tau_1)} \quad (5.23)$$

$$\lambda_{1,2} = \frac{H^2 Q + R(F^2 + 1) \pm \sigma}{2FR} \quad (5.24)$$

$$\sigma = \sqrt{H^2 Q + R(F+1)^2} \sqrt{H^2 Q + R(F-1)^2} \quad (5.25)$$

$$\tau_{1,2} = H^2 Q + R(F^2 - 1) \pm \sigma \quad (5.26)$$

$$\mu_{1,2} = H^2 Q + R(F^2 + 1) \pm \sigma \quad (5.27)$$

$$M = \begin{bmatrix} \frac{\tau_1}{2H^2} & \frac{\tau_2}{2H^2} \\ 1 & 1 \end{bmatrix} \quad (5.28)$$

$$M^{-1} = \frac{1}{\tau_1(R-1) + 2\sigma} \begin{bmatrix} 2RH^2 & -\tau_1 \\ -2RH^2 & R\tau_1 \end{bmatrix} \quad (5.29)$$

$$\lim_{k \rightarrow \infty} P_k^- = \frac{\tau_1}{2H^2} \quad (5.30)$$

5.5 Divergence Issues

- (Co)variance increases during time update and decreases during measurement update.
- Primary cause of KF failure are finite precision arithmetic and modeling errors.

- KF assumes model is precisely known; noise sequences w_k and v_k are pure white, zero-mean, and completely uncorrelated.
- To improve filter performance:
 1. Increase arithmetic precision.
 2. Use some form of square root filtering. This effectively increases arithmetic precision at the cost of adding complication.
 3. Symmetrize P at each time step: $P = (P + P^T)/2$, change lower triangle, or force eigenvalues to be positive.
 4. Initialize P appropriately to avoid large changes in P .
 5. Use fading memory filter. Force to forget measurement from distant past and more emphasis on recent measurements. Exchange optimality with stability and convergence.
 6. Use fictitious process noise (especially for estimating "constants"), effectively telling the filter not to trust the model as much.
- If a system model has too much noise, it becomes difficult to estimate. If a system model has too little noise, it is susceptible to modeling errors.

6 Alternate KF formulations

6.1 Sequential KF

KF implementation w/out matrix inversion. Requirement: 1) R_k is diagonal (eq. 6.2) OR 2) R_k is constant (eq. 6.3). Normal KF is sometimes called Sequential/Recursive/Batch KF.

1. System definition

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} \quad (6.1)$$

$$y_k = H_k x_k + v_k$$

$$R_k = \text{diag}(R_{1k}, \dots, R_{rk}) \quad (6.2)$$

$$R = S \hat{R} S^{-1}$$

$$\begin{aligned} \bar{y}_k &= S^{-1}y_k = S^{-1}(H_k x_k + v_k) \\ &= \bar{H}_k x_k + \bar{v}_k \end{aligned} \quad (6.3)$$

2. Initialization:

$$\begin{aligned} \hat{x}_0^+ &= E(x_0) \\ P_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \end{aligned} \quad (6.4)$$

3. At each time step k , time update:

$$\begin{aligned} P_k^- &= F_{k-1}P_{k-1}^+ F_{k-1}^T + Q_{k-1} \\ \hat{x}_k^- &= F_{k-1}\hat{x}_{k-1}^+ + G_{k-1}u_{k-1} \end{aligned} \quad (6.5)$$

4. At each time step k , measurement update:

(a) Initialize:

$$\hat{x}_{0k}^+ = \hat{x}_{0k}^- \quad (6.6)$$

$$\hat{P}_{0k}^+ = \hat{P}_{0k}^- \quad (6.7)$$

(b) For $i = 1, \dots, r$:

$$\begin{aligned} K_{ik} &= \frac{P_{i-1,k}^+ H_{i,k}^T}{H_{i,k} P_{i-1,k}^+ H_{i,k}^T + R_{ik}} \\ &= \frac{P_{i,k}^+ H_{i,k}^T}{R_{ik}} \\ \hat{x}_{ik}^+ &= \hat{x}_{i-1,k}^+ + K_{ik}(y_{ik} - H_{i,k}\hat{x}_{i-1,k}^+) \\ P_{ik}^+ &= (I - K_{ik}H_{i,k})P_{i-1,k}^+(I - K_{ik}H_{i,k})^T \\ &\quad + K_{ik}R_{ik}K_{ik}^T \\ &= [(P_{i-1,k}^+)^{-1} + H_{i,k}^T H_{i,k}/R_{ik}]^{-1} \\ &= (I - K_{ik}H_{i,k})P_{i-1,k}^+ \end{aligned} \quad (6.8)$$

(c) End:

$$\begin{aligned} \hat{x}_k^+ &= \hat{x}_{rk}^+ \\ \hat{P}_k^+ &= \hat{P}_{rk}^+ \end{aligned} \quad (6.9)$$

6.2 Information Filtering

KF that propagates information matrix $\mathcal{I} = P^{-1}$. Computationally efficient if $r \gg n$ (much more measurements than states). More mathematically precise for the zero initial certain case, while KF is more precise for the zero initial uncertainty case.

1. Dynamic system definition

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} \\ y_k &= H_k x_k + v_k \end{aligned} \quad (6.10)$$

2. Initialization: $P_0^+ = 0$ if perfect knowledge, $P_0^+ = \infty I$ if no knowledge.

$$\begin{aligned} \hat{x}_0^+ &= E(x_0) \\ \mathcal{I}_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]^{-1} \end{aligned} \quad (6.11)$$

3. For each $k = 1, 2, \dots$

$$\mathcal{I}_k^- = Q_{k-1}^{-1} \quad (6.12)$$

$$- Q_{k-1}^{-1} F_{k-1} (\mathcal{I}_{k-1}^+ + F_{k-1}^T Q_{k-1}^{-1} F_{k-1})^{-1} F_{k-1}^T Q_{k-1}^{-1} \quad (6.13)$$

$$\mathcal{I}_k^+ = \mathcal{I}_k^- + H_k^T R_k^{-1} H_k \quad (6.14)$$

$$K_k = (\mathcal{I}_k^+)^{-1} H_k^T R_k^{-1} \quad (6.15)$$

$$\hat{x}_k^- = F_{k-1}\hat{x}_{k-1}^+ + G_{k-1}u_{k-1} \quad (6.16)$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - H_k\hat{x}_k^-) \quad (6.17)$$

6.3 Square Root Filtering

To solve numerical precision problem that arises from cases in which some elements of the state-vector x are estimated to much greater precision than other elements of x .

6.3.1 Condition Number

$$\sigma^2(P) = \lambda(P^T P) = \lambda(PP^T) \quad (6.18)$$

$$\kappa(P) = \frac{\sigma_{max}(P)}{\sigma_{min}(P)} \geq 1 \quad (6.19)$$

Our goal is to find S such that $P = SS^T$. If P is symmetric positive definite then it always has a square root (can be more than one).

The Cholesky Matrix Square Root Algorithm {

$$\begin{aligned} &\text{For } i = 1, \dots, n \\ &\{ \\ &\quad S_{ii} = \sqrt{P_{ii} - \sum_{j=1}^{i-1} S_{ij}^2} \\ &\quad \text{For } j = 1, \dots, n \\ &\quad \{ \\ &\quad \quad S_{ji} = 0 \quad j < i \\ &\quad \quad S_{ji} = \frac{1}{S_{ii}} \left(P_{ji} - \sum_{k=1}^{i-1} S_{jk} S_{ik} \right) \quad j > i \\ &\quad \} \\ &\} \\ &\} \end{aligned}$$

$$\sigma^2(P) = [\sigma^2(S)]^2 \quad (6.20)$$

$$\frac{\sigma_{max}(P)}{\sigma_{min}(P)} = \frac{\sigma_{max}^2(S)}{\sigma_{min}^2(S)} \quad (6.21)$$

$$\kappa(P) = \kappa^2(S) \quad (6.22)$$

6.3.2 Square Root Time Update Eq.

$$\begin{bmatrix} (S_k^-)^T \\ 0 \end{bmatrix} = T \begin{bmatrix} (S_{k-1}^+)^T F_{k-1}^T \\ Q_{k-1}^{T/2} \end{bmatrix} \quad (6.23)$$

where T is a $2n \times 2n$ orthogonal matrix computed using numerical linear algebra methods.

6.3.3 Potter's Square Root Measurement Update Eq.

Potter's square root measurement-update algorithm

1. After the *a priori* covariance square root S_k^- and the *a priori* state estimate \hat{x}_k^- have been computed, initialize

$$\begin{aligned} \hat{x}_{0k}^+ &= \hat{x}_k^- \\ S_{0k}^+ &= S_k^- \end{aligned} \quad (6.73)$$

2. For $i = 1, \dots, r$ (where r is the number of measurements), perform the following.

- (a) Define H_{ik} as the i th row of H_k , y_{ik} as the i th element of y_k , and R_{ik} as the variance of the i th measurement (assuming that R_k is diagonal).
- (b) Perform the following to find the square root of the covariance after the i th measurement has been processed:

$$\begin{aligned} \phi_i &= S_{i-1,k}^{+T} H_{ik}^T \\ a_i &= \frac{1}{\phi_i^T \phi_i + R_{ik}} \\ \gamma_i &= \frac{1}{1 \pm \sqrt{a_i R_{ik}}} \\ S_{ik}^+ &= S_{i-1,k}^+ (I - a_i \gamma_i \phi_i \phi_i^T) \end{aligned} \quad (6.74)$$

- (c) Compute the Kalman gain for the i th measurement as

$$K_{ik} = a_i S_{ik}^+ \phi_i \quad (6.75)$$

- (d) Compute the state estimate update due to the i th measurement as

$$\hat{x}_{ik}^+ = \hat{x}_{i-1,k}^+ + K_{ik} (y_{ik} - H_{ik} \hat{x}_{i-1,k}^+) \quad (6.76)$$

3. Set the *a posteriori* covariance square root and the *a posteriori* state estimate as

$$\begin{aligned} S_k^+ &= S_{rk}^+ \\ \hat{x}_k^+ &= \hat{x}_{rk}^+ \end{aligned} \quad (6.77)$$

6.3.4 Square Root Time Update Eq

$$\begin{bmatrix} (R_k + H_k P_k^- H_k^T)^{T/2} & \tilde{K}_k^T \\ 0 & (S_k^+)^T \end{bmatrix} = \tilde{T} \begin{bmatrix} R_k^{T/2} & 0 \\ (S_k^-)^T H_k^T & (S_k^-)^T \end{bmatrix} \quad (6.24)$$

$$\tilde{K}_k = K_k (R_k + H_k P_k^- H_k^T)^{T/2} \quad (6.25)$$

where \tilde{T} is a $(n+r) \times (n+r)$ matrix computed using numerical linear algebra methods.

6.3.5 Algos for Orthogonal Transformations

6.3.5.1 The Householder algorithm The algorithm presented here was developed by Alston Householder [Hou64, Chapter 5], applied to least squares estimation by Gene Golub [Gol65], and summarized for Kalman filtering by Paul Kaminski [Kam71].

1. Suppose that we have a $2n \times n$ matrix $A^{(1)}$, and we want to find an $n \times n$ matrix W such that

$$T A^{(1)} = \begin{bmatrix} W \\ 0 \end{bmatrix} \quad (6.98)$$

where T is an orthogonal $2n \times 2n$ matrix, and 0 is the $n \times n$ matrix consisting of all zeros. Note that this problem statement is in the same form as Equation (6.58). Also note that we do not necessarily need to find T ; our goal is to find W .

2. For $k = 1, \dots, n$ perform the following:

- (a) Compute the scalar σ_k as

$$\sigma_k = \text{sgn}(A_{kk}^{(k)}) \sqrt{\sum_{i=k}^{2n} (A_{ik}^{(k)})^2} \quad (6.99)$$

where $A_{ik}^{(k)}$ is the element in the i th row and k th column of $A^{(k)}$. The $\text{sgn}(\cdot)$ function is defined to be equal to $+1$ if its argument is greater than or equal to zero, and -1 if its argument is less than zero.

- (b) Compute the scalar β_k as

$$\beta_k = \frac{1}{\sigma_k (\sigma_k + A_{kk}^{(k)})} \quad (6.100)$$

- (c) For $i = 1, \dots, 2n$ perform the following:

$$u_i^{(k)} = \begin{cases} 0 & i < k \\ \sigma_k + A_{kk}^{(k)} & i = k \\ A_{ik}^{(k)} & i > k \end{cases} \quad (6.101)$$

This gives a $2n$ -element column vector $u^{(k)}$.

- (d) For $i = 1, \dots, n$ perform the following:

$$y_i^{(k)} = \begin{cases} 0 & i < k \\ 1 & i = k \\ \beta_k u^{(k)T} A_i^{(k)} & i > k \end{cases} \quad (6.102)$$

where $A_i^{(k)}$ is the i th column of $A^{(k)}$. This gives an n -element column vector $y^{(k)}$.

- (e) Compute the $2n \times n$ matrix $A^{(k+1)}$ as

$$A^{(k+1)} = A^{(k)} - u^{(k)} y^{(k)T} \quad (6.103)$$

3. After the above steps have been executed, $A^{(n+1)}$ has the form

$$A^{(n+1)} = \begin{bmatrix} W \\ 0 \end{bmatrix} \quad (6.104)$$

where W is the $n \times n$ matrix that we are trying to solve for. Note that if $\sigma_k = 0$ at any stage of the algorithm, that means $A^{(1)}$ is rank deficient and the algorithm will fail. Also note that the above algorithm does not compute the T matrix. However, we can find the T matrix as

$$\begin{aligned} T &= T^{(n)} T^{(n-1)} \dots T^{(1)} \\ T^{(k)} &= I - \beta_k u^{(k)} u^{(k)T} \quad i = 1, \dots, n \end{aligned} \quad (6.105)$$

6.3.5.2 The modified Gram–Schmidt algorithm The modified Gram–Schmidt algorithm for orthonormalization that is presented here is discussed in most linear systems books [Kai80, Bay99, Che99]. It was first given in [Bjo67] and was summarized for Kalman filtering in [Kam71].

1. Suppose that we have a $2n \times n$ matrix $A^{(1)}$, and we want to find an $n \times n$ matrix W such that

$$TA^{(1)} = \begin{bmatrix} W \\ 0 \end{bmatrix} \quad (6.106)$$

where T is an orthogonal $2n \times 2n$ matrix, and 0 is the $n \times n$ matrix consisting of all zeros. Note that this problem statement is in the same form as Equation (6.58).

2. For $k = 1, \dots, n$ perform the following.

- (a) Compute the scalar σ_k as

$$\sigma_k = \sqrt{A_k^{(k)T} A_k^{(k)}} \quad (6.107)$$

where $A_i^{(k)}$ is the i th column of $A^{(k)}$.

- (b) Compute the k th row of W as

$$W_{kj} = \begin{cases} 0 & j = 1, \dots, k-1 \\ \sigma_k & j = k \\ A_k^{(k)T} A_j^{(k)} / \sigma_k & j = k+1, \dots, n \end{cases} \quad (6.108)$$

- (c) Compute the k th row of T as

$$T_k = A_k^{(k)T} / \sigma_k \quad (6.109)$$

- (d) If $(k < n)$, compute the last $(n - k)$ columns of $A^{(k+1)}$ as

$$A_j^{(k+1)} = A_j^{(k)} - W_{kj} A_k^{(k)} / \sigma_k \quad j = k+1, \dots, n \quad (6.110)$$

Note that the first k columns of $A^{(k+1)}$ are not computed in this algorithm.

As with the Householder algorithm, if $\sigma_k = 0$ at any stage of the algorithm, that means $A^{(1)}$ is rank deficient and the algorithm fails. After this algorithm completes, we have the first n rows of T , and T is an $n \times 2n$ matrix. If we want to know the last n rows of T , we can compute them using a regular Gram–Schmidt algorithm as follows [Hor85, Gol89, Moo00].

1. Fill out the T matrix that was begun above by appending a $2n \times 2n$ identity matrix to the bottom of it. This ensures that the rows of T span the entire $2n$ -dimensional vector space:

$$T = \begin{bmatrix} T \\ I \end{bmatrix} \quad (6.111)$$

Note that this T is a $3n \times 2n$ matrix.

2. Now we perform a standard Gram–Schmidt orthonormalization procedure on the last $2n$ rows of T (with respect to the already obtained first n rows of T). For $k = n+1, \dots, 3n$, compute the k th row of T as

$$\begin{aligned} T_k &= T_k - \sum_{i=1}^{k-1} (T_k T_i^T) T_i \\ T_k &= \frac{T_k}{\|T_k\|_2} \end{aligned} \quad (6.112)$$

If T_k is zero then that means that it is a linear combination of the previous rows of T . In that case, the division in the above equation will be a divide by zero, so instead T_k should be discarded. This discard will actually occur exactly n times so that this procedure will compute n additional rows of T and we will end up with an orthogonal $2n \times 2n$ matrix T .

$$U = \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (6.26)$$

$$D = \text{diag}(d_{11}, d_{22}, d_{33}) \quad (6.27)$$

$$\begin{aligned} p_{11} &= d_{11} + d_{22}u_{12}^2 + d_{33}u_{13}^2 \\ p_{12} &= d_{22}u_{12} + d_{33}u_{13}u_{23} \\ p_{13} &= d_{33}u_{13} \\ p_{22} &= d_{22} + d_{33}u_{23}^2 \\ p_{23} &= d_{33}u_{23} \\ p_{33} &= d_{33} \end{aligned} \quad (6.28)$$

$$\begin{aligned} \bar{U} \bar{D} \bar{U}^T &= [D_{i-1} - \frac{1}{\alpha_i} (D_{i-1} U_{i-1}^T H_i^T) (D_{i-1} U_{i-1}^T H_i^T)^T] \\ U_i &= U_{i-1} \bar{U} \\ D_i &= \bar{D} \end{aligned} \quad (6.29)$$

$$u(k, j) = \frac{w_k \hat{D} v_j^T}{v_j \hat{D} v_j^T} \quad \text{for } j, k = 1, \dots, n \quad (6.30)$$

$$W = U^{-1} V$$

The U-D measurement update

1. We start with the *a priori* estimation covariance P^- at time k . Define $P_0 = P^-$.
2. For $i = 1, \dots, r$ (where r is the number of measurements), perform the following:
 - (a) Define H_i as the i th row of H , R_i as the i th diagonal entry of R , and $\alpha_i = H_i P_{i-1} H_i^T + R_i$.
 - (b) Perform a U-D factorization of P_{i-1} to obtain U_{i-1} and D_{i-1} , and then form the matrix on the right side of Equation (6.120).
 - (c) Find the U-D factorization of the matrix on the right side of Equation (6.120) and call the factors \bar{U} and \bar{D} .
 - (d) Compute U_i and D_i from Equation (6.122).
3. The *a posteriori* estimation covariance is given as $P^+ = U_r D_r U_r^T$.

The U-D time update

1. Begin with $P^+ = U^+ D^+ U^{+T}$ (from the measurement update equation).
2. Define the following matrices.

$$\begin{aligned} W &= \begin{bmatrix} F U^+ & I \end{bmatrix} \\ \hat{D} &= \begin{bmatrix} D^+ & 0 \\ 0 & Q \end{bmatrix} \end{aligned} \quad (6.135)$$

3. Use the rows of W along with the Gram–Schmidt orthogonalization procedure to generate v_i vectors that are orthogonal with respect to the \hat{D} inner product. The algorithm for generating the v_i vectors is given in Equation (6.128).
4. Form the V matrix using the v_i vectors as rows; see Equation (6.132).
5. Use \hat{D} inner products to form the unit upper triangular matrix U^- ; see Equations (6.129) and (6.132).
6. Define D^- as $D^- = V \hat{D} V^T$.

6.4 U-D Filtering

KF that has twice as much precision but requires less computation than square root filter. Base on factorization $P = U D U^T$ where U is an upper triangle matrix and D is a diagonal matrix. It also has the same requirements as sequential KF.

7 KF Generalizations

7.1 Correlated Process and Measurement Noise

General Discrete KF

1. System and measurement equations

$$\begin{aligned}
x_k &= F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} \\
y_k &= H_k x_k + v_k \\
w_k &\approx (0, Q_k) \\
v_k &\approx (0, R_k) \\
E[w_k w_j^T] &= Q_k \delta_{k-j} \\
E[v_k v_j^T] &= R_k \delta_{k-j} \\
E[w_k, v_j^T] &= M_k \delta_{k-j+1}
\end{aligned} \tag{7.1}$$

2. Initialization

$$\begin{aligned}
\hat{x}_0^+ &= E(x_0) \\
\hat{P}_0^+ &= E[(x_0 - \hat{x}_0^+)(\dots)^T]
\end{aligned} \tag{7.2}$$

3. For $k = 1, 2, \dots$

$$\begin{aligned}
P_k^- &= F_{k-1}P_{k-1}^+ F_{k-1}^T + Q_{k-1} \\
K_k &= (P_k^- H_k^T + M_k)(H_k P_k^- H_k^T + H_k M_k + \\
&\quad M_k^T H_k^T + R_k)^{-1} \\
&= P_k^+ (H_k^T + (P_k^-)^{-1} M_k)(R_k - M_k^T (P_k^-)^{-1} M_k)^{-1} \\
\hat{x}_k^- &= F_{k-1} \hat{x}_{k-1}^+ + G_{k-1} u_{k-1} \\
\hat{x}_k^+ &= \hat{x}_k^- + K_k (y_k - H_k \hat{x}_k^-) \\
P_k^+ &= (I - K_k H_k) P_k^- (I - K_k H_k)^T + \\
&\quad K_k (H_k M_k + M_k^T H_k^T + R_k) K_k^T - M_k K_k^T - K_k M_k^T \\
&= [(P_k^-)^{-1} + (H_k^T + (P_k^-)^{-1} M_k) \times \\
&\quad (R_k - M_k^T (P_k^-)^{-1} M_k)^{-1} (H_k + M_k^T (P_k^-)^{-1})]^{-1} \\
&= P_k^- - K_k (H_k P_k^- + M_k^T)
\end{aligned} \tag{7.3}$$

7.2 Colored Process and Measurement Noise

7.2.1 Colored Process Noise

$$w_k = \psi w_{k-1} + \zeta_{k-1} \tag{7.4}$$

$$\begin{bmatrix} x_k \\ w_k \end{bmatrix} = \begin{bmatrix} F & I \\ 0 & \psi \end{bmatrix} \begin{bmatrix} x_{k-1} \\ w_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \zeta_{k-1} \end{bmatrix} \tag{7.5}$$

$$x'_k = F'_{k-1} x'_{k-1} + w'_{k-1} \tag{7.6}$$

$$E(w'_k w_k'^T) = \begin{bmatrix} 0 & 0 \\ 0 & E(\zeta_k \zeta_k^T) \end{bmatrix} = Q'_k \tag{7.7}$$

7.3 Colored measurement noise: State augmentation

$$v_k = \psi_{k-1} v_{k-1} + \zeta_{k-1} \tag{7.8}$$

$$\begin{bmatrix} x_k \\ v_k \end{bmatrix} = \begin{bmatrix} F & 0 \\ 0 & \psi \end{bmatrix} \begin{bmatrix} x_{k-1} \\ v_{k-1} \end{bmatrix} + \begin{bmatrix} w_{k-1} \\ \zeta_{k-1} \end{bmatrix} \tag{7.9}$$

$$y_k = [H_k \quad I] \begin{bmatrix} x_k \\ v_k \end{bmatrix} + 0 \tag{7.10}$$

$$E(w'_k w_k'^T) = \begin{bmatrix} Q_k & 0 \\ 0 & Q_{\zeta k} \end{bmatrix} \tag{7.11}$$

$$E[v'_k v_k'^T] = 0 \tag{7.12}$$

However, a singular measurement-noise covariance often results in numerical problems.

7.3.1 Colored measurement noise: Measurement differencing

1. Definition

$$v_k = \psi_{k-1} v_{k-1} + \zeta_{k-1} \tag{7.13}$$

$$y'_{k-1} = y_k - \psi_{k-1} y_{k-1} \tag{7.14}$$

$$= (H_k F_{k-1} - \psi_{k-1} H_{k-1}) x_{k-1} + (H_k w_{k-1} + \zeta_{k-1}) \tag{7.15}$$

$$= H'_{k-1} x_{k-1} + v'_{k-1} \tag{7.16}$$

2. At each time step

$$\begin{aligned}
\hat{x}_k^+ &= \hat{x}_k^- + K_k (y'_k - H'_k \hat{x}_k^-) \\
\hat{x}_{k+1}^- &= F_k \hat{x}_k^+ + C_k (y'_k - H'_k \hat{x}_k^+) \\
K_k &= P_k^- H_k'^T (H_k' P_k^- H_k'^T + R_k)^{-1} \\
M_k &= Q_k H_{k+1}^T \\
C_k &= M_k (H_k' P_k^- H_k'^T + R_k)^{-1} \\
P_k^+ &= (I - K_k H_k') P_k^- (I - K_k H_k')^T + K_k R_k K_k^T \\
P_{k+1}^- &= F_k P_k^+ F_k^T + Q_k - C_k M_k^T - \\
&\quad F_k K_k M_k - M_k^T K_k^T F_k^T
\end{aligned} \tag{7.17}$$

7.4 Steady-State Filtering

Ways of calculating Kalman gain:

1. numerical simulation
2. discrete algebraic Riccati equation (DARE): assume $P_k^- = P_{k+1}^-$ in P update eq. then P_∞ to compute for K_∞
 - (a) may not converge to steady state value
 - (b) may converge to different steady state value depending on P_0
 - (c) may converge to steady state value but result in unstable KF

$$\hat{x}_k^- = F \hat{x}_{k-1}^+ \tag{7.18}$$

$$\hat{x}_k^+ = (I - K_\infty H) F \hat{x}_{k-1}^+ + K_\infty y_k \tag{7.19}$$

DARE Theorems Def: matrix pair (F, G) is controllable on the unit circle if there exists some matrix K such that $(F - GK)$ does not have eigenvalues with magnitude 1.

		Results	Conditions (iff.)	
Thm	P_∞	# of sol'n \rightarrow ss KF	(F, H)	$(F - MR^{-1}H, G)$
23	unique pos. def.	1 stable	detectable	stabilizable
24	≥ 1 pos. semidef.	1 stable	detectable	controllable on unit circle
25	≥ 1 pos. def.	1 stable	detectable	controllable on and inside unit circle
26	≥ 1 pos. semidef.	geq1 marginally stable	detectable	N/A

7.4.1 $\alpha - \beta$ filtering

Newton dynamic system with state position and velocity.

$$x_k = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x_{k+1} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} w'_{k-1} \quad (7.20)$$

$$y_k = [1 \ 0] x_k + v_k$$

$$w'_k = (0, \sigma_w^2)$$

$$K = [K_1 \ K_2] = [\alpha \ \beta/T]^T \quad (7.21)$$

$$K_1 = -\frac{1}{8}(\lambda^2 + 8\lambda - (\lambda + 4)\sqrt{\lambda^2 + 8\lambda})$$

$$K_2 = \frac{1}{4T}(\lambda^2 + 4\lambda - \lambda\sqrt{\lambda^2 + 8\lambda})$$

$$P_{11}^- = \frac{K_1 \sigma_w^2}{1 - K_1} \quad (7.22)$$

$$P_{12}^- = \frac{K_2 \sigma_w^2}{1 - K_1}$$

$$P_{22}^- = \left(\frac{K_1}{T} + \frac{K_2}{2}\right) P_{12}^-$$

$$\lambda = \frac{\sigma_w^2 T^2}{R}$$

$$P_{11}^+ = K_1 R$$

$$P_{12}^+ = K_2 R \quad (7.23)$$

$$P_{22}^+ = \left(\frac{K_1}{T} - \frac{K_2}{2}\right) P_{12}^-$$

7.4.2 $\alpha - \beta - \gamma$ filtering

Newton dynamic system with state position, velocity, and acceleration.

$$x_k = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} x_{k+1} + \begin{bmatrix} T^2/2 \\ T \\ 1 \end{bmatrix} w'_{k-1} \quad (7.24)$$

$$y_k = [1 \ 0 \ 0] x_k + v_k$$

$$w'_k = (0, \sigma_w^2)$$

$$K = [K_1 \ K_2 \ K_3] = [\alpha \ \beta/T \ \phi/2T^2]^T \quad (7.25)$$

$$\alpha = 1 - s^2$$

$$\beta = 2(1 - s)^2 \quad (7.26)$$

$$\phi = 2\lambda s$$

$$b = \lambda/2 - 3$$

$$c = \lambda/2 + 3$$

$$p = c - b^2/3$$

$$q = \frac{2b^3}{27} - \frac{bc}{3} - 1 \quad (7.27)$$

$$z = \left[\frac{-q + \sqrt{q^2 + 4p^3/27}}{2} \right]^{1/3}$$

$$s = z - p/(3z) - b/3$$

$$P_{11}^+ = \alpha R$$

$$P_{12}^+ = \beta R/T$$

$$P_{13}^+ = \phi R/2T^2$$

$$P_{22}^+ = \frac{8\alpha\beta + \phi(\beta - 2\alpha - 4)}{8T^2(1 - \alpha)} R \quad (7.28)$$

$$P_{23}^+ = \frac{\beta(2\beta - \phi)R}{4T^3(1 - \alpha)}$$

$$P_{33}^+ = \frac{\phi(2\beta - \phi)R}{4T^4(1 - \alpha)}$$

7.4.3 Hamiltonian Approach

1. Form the $2nx2n$ Hamiltonian matrix for an n state KF.

$$\mathcal{H} = \begin{bmatrix} F^{-T} & F^{-T} H^T R^{-1} H \\ QF^{-T} & F + QF^{-T} H^T R^{-1} H \end{bmatrix} \quad (7.29)$$

2. Compute the eigenvalues of \mathcal{H} . If any of them are on the unit circle, then we cannot go any further with this procedure; the Riccati eq. does not have a steady-state sol'n.
3. Collect the n eigenvectors that corresponds to the n eigenvalues that are outside the unit circle. Column i is the i^{th} eigenvector.

$$\begin{bmatrix} \Phi_{12} \\ \Phi_{22} \end{bmatrix} \quad (7.30)$$

4. Compute the steady-state Riccati eq. sol'n. Φ_{12} must be invertible.

$$P_\infty^- = \Phi_{22} \Phi_{12}^{-1} \quad (7.31)$$

\mathcal{H} is a symplectic matrix. It satisfies the ff:

1. $J^{-1} H^T J = H^{-1}$ where $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$.
2. None of the eigenvalues are 0.
3. If λ is an eigenvalue, then so is $1/\lambda$.
4. The determinant is ± 1 .

7.5 KF with Fading Memory

Fading-memory filter

1. System equation

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

$$y_k = H_k x_k + v_k$$

$$E(w_k w_j^T) = Q_k \delta_{k-j}$$

$$E(v_k v_j^T) = R_k \delta_{k-j}$$

$$E(w_k v_j^T) = 0$$

$$(7.32)$$

2. Initialization

$$\hat{x}_0^+ = E(x_0)$$

$$\tilde{P}_0^+ = E[(x_0 - \hat{x}_0^+)(\dots)^T] \quad (7.33)$$

3. Choose $\alpha \geq 1$ based on how much you want the filter to forget past measurements. $\alpha = 1$ is like standard KF. $\alpha = \infty$ is like taking most recent measurement only.

4. For each time step $k = 1, 2, \dots$

$$\begin{aligned}
\tilde{P}_k^- &= \alpha^2 F_{k-1} \tilde{P}_{k-1}^+ F_{k-1}^T + Q_{k-1} \\
K_k &= \tilde{P}_k^- H_k^T (H_k \tilde{P}_k^- H_k^T + R_k)^{-1} \\
&= \tilde{P}_k^+ H_k^T R_k^{-1} \\
\hat{x}_k^- &= F_{k-1} \hat{x}_{k-1}^+ + G_{k-1} u_{k-1} \\
\hat{x}_k^+ &= \hat{x}_k^- + K_k (y_k - H_k \hat{x}_k^-) \\
\tilde{P}_k^+ &= (I - K_k H_k) \tilde{P}_k^- (I - K_k H_k)^T + K_k R_k K_k^T \\
&= \left[(\tilde{P}_k^-)^{-1} + H_k^T R_k^{-1} H_k \right]^{-1} \\
&= \tilde{P}_k^- K_k H_k \tilde{P}_k^-
\end{aligned} \tag{7.34}$$

7.6 Constrained KF

7.6.1 Model reduction

1. (-) makes interpretation less natural and more difficult (loses physical meaning)
2. (-) cannot extend inequality constraints
3. (+) straightforward and (usually) easily implemented

7.6.2 Perfect measurements

Add constraints to rows of measurement.

1. (-) singular covariance increase the possibility of numerical problems
2. (-) inequality constraints are implemented as soft constraints. difficult to control how close the state estimate gets to the constraint boundary.

7.6.3 Projection approaches

* to be added *

7.6.4 Pdf truncation approach

* to be added *

8 Nonlinear KF

All systems are ultimate nonlinear.

x_0, u_0, y_0, w_0 and v_0 are the nominal state, control, output, system noise and measurement noise.

CT system equation:

$$\begin{aligned}
\dot{x} &= f(x, u, w, t) \\
y &= h(x, v, t) \\
\tilde{w}(0, Q) \\
\tilde{v}(0, R) \\
\dot{x}_0 &= f(x_0, u_0, 0, t) \\
y_0 &= h(x_0, 0, t)
\end{aligned} \tag{8.1}$$

DT system equation:

$$\begin{aligned}
x_k &= f_{k-1}(x_{k-1}, u_{k-1}, w_{k-1}) \\
y_k &= h(x_k, v_k) \\
\tilde{w}_k(0, Q_k) \\
\tilde{v}_k(0, R_k)
\end{aligned} \tag{8.2}$$

8.1 Linearized KF

1. CT system equation (eq. 8.1)
2. Compute

$$\begin{aligned}
A &= \left. \frac{\delta f}{\delta x} \right|_0 \\
L &= \left. \frac{\delta f}{\delta w} \right|_0 \\
C &= \left. \frac{\delta h}{\delta x} \right|_0 \\
M &= \left. \frac{\delta h}{\delta v} \right|_0 \\
\tilde{Q} &= LQL^T \\
\tilde{R} &= MRM^T
\end{aligned} \tag{8.3}$$

3. $\Delta y = y - y_0$
- 4.

$$\begin{aligned}
\Delta \hat{x}(0) &= 0 \\
P(0) &= E[(\Delta x(0) - \Delta \hat{x}(0))(\dots)^T] \\
\Delta \dot{\hat{x}} &= A \Delta \hat{x} + K(\Delta y - C \Delta \hat{x}) \\
K &= PC^T \tilde{R}^{-1} \\
\dot{P} &= AP + PA^T + \tilde{Q} - PC^T \tilde{R}^{-1} CP
\end{aligned} \tag{8.5}$$

5. $\hat{x} = x_0 + \Delta \hat{x}$

Derivation Notes:

1. Taylor series linearization gives us a nominal trajectory which is \approx actual.
2. $\left. \right|_0$ means evaluated at nominal control, state, output, and noise values.
3. Taylor approximation gives us:

$$\dot{x} \approx f(x_0, u_0, w_0, t) + A \Delta x + B \Delta u + L \Delta w \tag{8.6}$$

$$y \approx h(x_0, v_0, t) + C \Delta x + M \Delta v \tag{8.7}$$

4. Assume $w_0 = v_0 = 0$ and $u(t) = u_0$ (perfectly known).
5. P = covariance of est. error + linearization error

8.2 Extended KF (EKF)

8.2.1 Continuous-time

1. CT system equation (eq. 8.1)

2. Compute

$$\begin{aligned}
A &= \left. \frac{\delta f}{\delta x} \right|_{\hat{x}} \\
L &= \left. \frac{\delta f}{\delta w} \right|_{\hat{x}} \\
C &= \left. \frac{\delta h}{\delta x} \right|_{\hat{x}} \\
M &= \left. \frac{\delta h}{\delta v} \right|_{\hat{x}} \\
\tilde{Q} &= LQL^T \\
\tilde{R} &= MRMT^T
\end{aligned} \tag{8.8}$$

3.

$$\begin{aligned}
\hat{x}(0) &= E[x(0)] \\
P(0) &= E[(x(0) - \hat{x}(0))(\dots)^T] \\
\dot{\hat{x}} &= f(\hat{x}, u, w_0, t) + K[y - h(\hat{x}, v_0, t)] \tag{8.10} \\
K &= PC^T \tilde{R}^{-1} \\
\dot{P} &= AP + PA^T + \tilde{Q} - PC^T \tilde{R}^{-1} CP
\end{aligned}$$

Derivation: EKF came from equation $\dot{x}_0 + \Delta \dot{\hat{x}} = f(x_0, u_0, w_0, t) + A\Delta \hat{x} + K[y - y_0 - C(\hat{x} - x_0)]$, and then choosing $x_0(t) = \hat{x}(t)$.

8.2.2 Hybrid

Continuous time dynamics + discrete time measurements.

1. CT system equation (eq. 8.1)
2. Initialization:

$$\begin{aligned}
\hat{x}_0^+ &= E(x_0) \\
P_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]
\end{aligned} \tag{8.11}$$

3. For $k = 1, 2, \dots$

(a) Integrate with the ff. eq to get \hat{x}_k^- and P_k^-

$$\dot{\hat{x}} = f(\hat{x}, u, 0, t) \tag{8.12}$$

$$\dot{P} = AP + PA^T + LQL^T \tag{8.13}$$

(b) Substitute R_k with $M_k R_k M_k^T$ and then use eq. 5.9-5.10 and 5.12-5.14 to update K_k and P_k^+ .

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - h_k(\hat{x}_k^-, 0, t_k)) \tag{8.14}$$

Note that between (discrete) measurements times, $R = \infty$ because we don't have new measurements.

8.2.3 Discrete-time

1. DT system equation (eq. 8.2)
2. Initialization:

$$\begin{aligned}
\hat{x}_0^+ &= E(x_0) \\
P_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]
\end{aligned} \tag{8.15}$$

3. For $k = 1, 2, \dots$

(a) Time update

$$\begin{aligned}
F_{k-1} &= \left. \frac{\delta f_{k-1}}{\delta x} \right|_{\hat{x}_{k-1}^+} \\
L_{k-1} &= \left. \frac{\delta f_{k-1}}{\delta w} \right|_{\hat{x}_{k-1}^+}
\end{aligned} \tag{8.16}$$

$$\begin{aligned}
P_k^- &= F_{k-1} P_{k-1}^+ F_{k-1}^T + L_{k-1} Q_{k-1} L_{k-1}^T \\
\hat{x}_k^- &= f_{k-1}(\hat{x}_{k-1}^+, u_{k-1}, 0)
\end{aligned} \tag{8.17}$$

(b) Measurement update: substitute R_k with $M_k R_k M_k^T$ and then use eq. 5.9-5.10 and 5.12-5.14 to update K_k and P_k^+ .

$$H_k = \left. \frac{\delta h_k}{\delta x} \right|_{\hat{x}_k^-} \tag{8.18}$$

$$\begin{aligned}
M_k &= \left. \frac{\delta h_k}{\delta v} \right|_{\hat{x}_k^-} \\
\hat{x}_k^+ &= \hat{x}_k^- + K_k(y_k - h_k(\hat{x}_k^-, 0, t_k))
\end{aligned} \tag{8.19}$$

8.3 Higher Order Approaches

8.3.1 Iterated EKF

Idea: reformulate f and h from previous step since we have an even better estimate of x_k

Similar to sec. 8.2.3 except at measurement update (3b), you do for $i = 0, \dots, N$. You start with $\hat{x}_{k,0}^+ = \hat{x}_k^-$ and $P_{k,0}^+ = P_k^-$; and end with $\hat{x}_k^+ = \hat{x}_{k,N+1}^+$ and $P_k^+ = P_{k,N+1}^+$.

8.3.2 2nd Order EKF

2nd order hybrid EKF

1. CT system equation (eq. 8.1)
2. Initialization:

$$\begin{aligned}
\hat{x}_0^+ &= E(x_0) \\
P_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]
\end{aligned} \tag{8.20}$$

3. Time update

$$\begin{aligned}
\dot{\hat{x}} &= f(\hat{x}, u, 0, t) + \frac{1}{2} \sum_{i=1}^n \phi_i \text{Tr} \left[\left. \frac{\delta^2 f_i}{\delta x^2} \right|_{\hat{x}} P \right] \\
\dot{P} &= FP + PF^T + LQL^T \\
\phi_i &= [0 \dots 1 \dots 0]^T \text{1 on } i\text{th element} \\
F &= \left. \frac{\delta f}{\delta x} \right|_{\hat{x}} \\
L &= \left. \frac{\delta f}{\delta w} \right|_{\hat{x}}
\end{aligned} \tag{8.21}$$

4. Measurement update

$$\begin{aligned}
\hat{x}_k^+ &= \hat{x}_k^- + K_k[y_k - h(\hat{x}_k^-)] - \pi_k \\
\pi_k &= \frac{1}{2}K_k \sum_{i=1}^m \phi_i Tr[D_{k,i}P_k^-] \\
D_{k,i} &= \left. \frac{\delta^2 h_i(x_k, t_k)}{\delta x^2} \right|_{\hat{x}_k^-} \\
K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k + \Lambda_k)^{-1} \\
H_k &= \left. \frac{\delta h(x_k, t_k)}{\delta x} \right|_{\hat{x}_k^-} \\
\Lambda_k(i, j) &= \frac{1}{2}Tr(D_{k,i}P_k^- D_{k,j}P_k^-) \\
P_k^+ &= P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + R_k + \Lambda_k)^{-1} H_k P_k^-
\end{aligned} \tag{8.22}$$

2nd order discrete EKF

1. DT system equation (eq. 8.2)
2. Initialization:

$$\begin{aligned}
\hat{x}_0^+ &= E(x_0) \\
P_0^+ &= E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T]
\end{aligned} \tag{8.23}$$

3. Time update

$$\begin{aligned}
\hat{x}_{k+1}^- &= f(\hat{x}, u, 0, t) + \frac{1}{2} \sum_{i=1}^n \phi_i Tr \left[\left. \frac{\delta^2 f_i}{\delta x^2} \right|_{\hat{x}_k^+} P_k^+ \right] \\
\dot{P}_{k+1}^- &= F P_k^+ F^T + Q_k \\
\phi_i &= [0 \dots 1 \dots 0]^T \text{ 1 on } i\text{th element} \\
F &= \left. \frac{\delta f}{\delta x} \right|_{\hat{x}_k^+}
\end{aligned} \tag{8.24}$$

4. Measurement update

$$\begin{aligned}
\hat{x}_k^+ &= \hat{x}_k^- + K_k[y_k - h(\hat{x}_k^-)] - \pi_k \\
\pi_k &= \frac{1}{2}K_k \sum_{i=1}^m \phi_i Tr[D_{k,i}P_k^-] \\
D_{k,i} &= \left. \frac{\delta^2 h_i(x_k, t_k)}{\delta x^2} \right|_{\hat{x}_k^-} \\
K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \\
H_k &= \left. \frac{\delta h(x_k, t_k)}{\delta x} \right|_{\hat{x}_k^-} \\
\Lambda_k(i, j) &= \frac{1}{2}Tr(D_{k,i}P_k^- D_{k,j}P_k^-) \\
P_k^+ &= (I - K_k H_k) P_k^-
\end{aligned} \tag{8.25}$$

8.3.3 Other Approaches

Gaussian sum filter

1. DT system equation (eq. 8.2)
2. Initialization: (a_{0i} must sum to 1)

$$pdf(\hat{x}_0^+) = \sum_{i=1}^M a_{0i} N(\hat{x}_{0i}^+, P_{0i}^+) \tag{8.26}$$

3. For $k = 1, 2, \dots$

- (a) Time update. For $i = 1, \dots, M$:

$$\begin{aligned}
\hat{x}_{ki}^- &= f_{k-1}(\hat{x}_{k-1,i}^+, u_{k-1}, 0) \\
F_{k-1,i} &= \left. \frac{\delta f_{k-1}}{\delta x_{k-1}} \right|_{\hat{x}_{k-1,i}^+} \\
P_{ki}^- &= F_{k-1,i} P_{k-1,i}^+ F_{k-1,i}^T + Q_{k-1} \\
a_{ki} &= a_{k-1,i}
\end{aligned} \tag{8.27}$$

$$pdf(\hat{x}_k^-) = \sum_{i=1}^M a_{ki} N(\hat{x}_{ki}^+, P_{ki}^+) \tag{8.28}$$

- (b) Measurement update. For $i = 1, \dots, M$:

$$\begin{aligned}
H_{ki} &= \left. \frac{\delta h_k}{\delta x_k} \right|_{\hat{x}_{ki}^-} \\
K_{ki} &= P_{ki}^- H_{ki}^T (H_{ki} P_{ki}^- H_{ki}^T + R_k)^{-1} \\
P_{ki}^+ &= P_{ki}^- - K_{ki} H_{ki} P_{ki}^- \\
\hat{x}_{ki}^+ &= \hat{x}_{ki}^- + K_{ki}[y_k - h_k(\hat{x}_{ki}^-, 0)]
\end{aligned} \tag{8.29}$$

$$\begin{aligned}
r_{ki} &= y_k - h_k(\hat{x}_{ki}^-, 0) \\
S_{ki} &= H_{ki} P_{ki}^- H_{ki}^T + R_k \\
\beta_{ki} &= \frac{exp[-r_{ki}^T S_{ki}^{-1} r_{ki}/2]}{(2\pi)^{n/2} |S_{ki}|^{1/2}} \\
a_{ki} &= \frac{a_{k-1,i} \beta_{ki}}{\sum_{j=1}^M a_{k-1,j} \beta_{kj}} \\
pdf(\hat{x}_k^+) &= \sum_{i=1}^M a_{ki} N(\hat{x}_{ki}^+, P_{ki}^+)
\end{aligned} \tag{8.30}$$

$$pdf(\hat{x}_k^+) = \sum_{i=1}^M a_{ki} N(\hat{x}_{ki}^+, P_{ki}^+) \tag{8.31}$$

- grid based filtering - valud of the pdf of the state is approximated, stored, propogated, and updated at discrete points in space.
- compute the theoretical optimal nonlinear filter and then linearize the nonlinear filter. Theoretical optimal is very difficult to compute.

8.4 Parameter Estimation

Estimate not only the state of the system, but also the parameters of the system.

9 Others

<http://www.doc88.com/p-714869662960.html>

$$A, b \tag{9.1}$$

$$A, b \tag{9.2}$$

$$\mathbf{A}, \mathbf{b} \tag{9.3}$$

$$\mathbf{A}, \mathbf{b} \tag{9.4}$$