# Stochastic Adaptive Control (02421)

Lecture 8

Tobias K. S. Ritschel

Assistant Professor

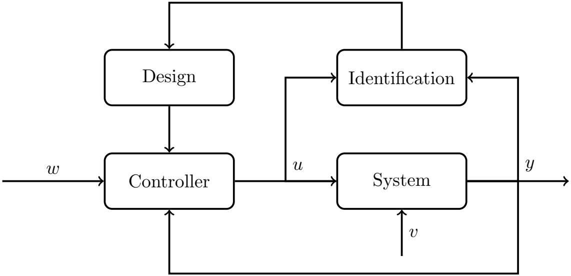
Section for Dynamical Systems, DTU Compute

**DTU Compute**

Department of Applied Mathematics and Computer Science

# Lecture Plan

1. ARX models
2. ARX prediction + control
3. ARX estimation
4. ARX model validation + adaptive control
5. ARMAX control
6. ARMAX estimation + adaptive control
7. Systems and control theory
8. Stochastic systems + Kalman filtering
9. SS estimation (recursive) + control
10. SS control
11. SS estimation (batch)
12. SS estimation (recursive)
13. SS nonlinear control



# Today's Agenda

- Multivariate probability theory
- Stochastic state space models
- Discretization
- State estimation
- Projection theorem
- Kalman filter

# Multivariate probability theory

## Stochastic Vectors

Vector-valued random variables

$$\mathbf{X} = \left[X_1, \ldots, X_n\right]^T \tag{1}$$

$$\text{cdf: } F_{\mathbf{X}}(\mathbf{x}) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n, ) \tag{2}$$

$$\text{marginal cdf: } F_{X_1}(x_1) = \Pr(X_1 \leq x_1) \tag{3}$$

1st and 2nd order moments

$$\mathbf{m}_x = \mathbb{E}[\mathbf{X}] = \left[\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]\right]^T \tag{4}$$

$$P_x = P_x^T = \operatorname{Cov}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbf{m}_x)(\mathbf{X} - \mathbf{m}_x)^T] \succeq 0 \tag{5}$$

Positive semi-definiteness ($\succeq$) means that $x^T P_x x \geq 0$.

Covariance matrices are diagonalizable, e.g., for $n = 2$

$$P_x = \begin{bmatrix} \operatorname{Var}(X_1) & \operatorname{Cov}(X_1, X_2) \\ \operatorname{Cov}(X_2, X_1) & \operatorname{Var}(X_2) \end{bmatrix} \tag{6}$$

# Mathematical properties of Moments: 1st and 2nd

Let $A$ and $m$ be a constant matrix and vector

Expectations

$$\mathbb{E}[\mathbf{X}+m] = \mathbb{E}[\mathbf{X}] + m \tag{7}$$
$$\mathbb{E}[A\mathbf{X}] = A\mathbb{E}[\mathbf{X}] \tag{8}$$
$$\mathbb{E}[\mathbf{X}+\mathbf{Y}] = \mathbb{E}[\mathbf{X}] + \mathbb{E}[\mathbf{Y}] \tag{9}$$
$$\mathbb{E}[\mathbf{X}^T A\mathbf{X}] = \mathrm{Tr}(A\,\mathrm{Cov}(\mathbf{X})) + \mathbb{E}[\mathbf{X}]^T A\mathbb{E}[\mathbf{X}] \tag{10}$$

Covariances

$$\mathrm{Cov}(\mathbf{X}) = \mathbb{E}[\mathbf{X}\mathbf{X}^T] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{X}]^T \tag{11}$$
$$\mathrm{Cov}(\mathbf{X}+m) = \mathrm{Cov}(\mathbf{X}) \tag{12}$$
$$\mathrm{Cov}(A\mathbf{X}) = A\,\mathrm{Cov}(\mathbf{X})A^T \tag{13}$$
$$\mathrm{Cov}(\mathbf{X}+\mathbf{Y}) = \mathrm{Cov}(\mathbf{X}) + \mathrm{Cov}(\mathbf{Y}) + \mathrm{Cov}(\mathbf{X},\mathbf{Y}) + \mathrm{Cov}(\mathbf{X},\mathbf{Y})^T \tag{14}$$

# Vector Covariance and Variance

Further covariances

$$\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \mathbb{E}[(\mathbf{X}-m_x)(\mathbf{Y}-m_y)^T] \tag{15}$$

$$\mathrm{Cov}(\mathbf{X},\mathbf{X}) = \mathrm{Cov}(\mathbf{X}) = P_x \tag{16}$$

$$\mathrm{Cov}(\mathbf{Y},\mathbf{X}) = \mathrm{Cov}(\mathbf{X},\mathbf{Y})^T \tag{17}$$

$$\mathrm{Cov}(A\mathbf{X},\mathbf{Y}) = A\,\mathrm{Cov}(\mathbf{X},\mathbf{Y}) \tag{18}$$

$$\mathrm{Cov}(\mathbf{X},A\mathbf{Y}) = \mathrm{Cov}(\mathbf{X},\mathbf{Y})A^T \tag{19}$$

$$\mathrm{Cov}(\mathbf{X}+\mathbf{V},\mathbf{Y}) = \mathrm{Cov}(\mathbf{X},\mathbf{Y}) + \mathrm{Cov}(\mathbf{V},\mathbf{Y}) \tag{20}$$

Principal directions of the variance (PCA)

$$[\Lambda,\mathbf{V}] = \mathrm{eig}(P_x) \tag{21}$$

$$P_x\mathbf{V}_i = \lambda_i\mathbf{V}_i \tag{22}$$

The columns in **V** indicate the main directions of the variation and the elements of $\Lambda$ indicate the associated variance

# Stochastic state space models

# Stochastic State-Space Models

Discrete-time system

$$x_{k+1} = Ax_k + Bu_k + Gv_k, \quad v_k \sim N(\mu_v, R_1) \tag{23a}$$

$$y_k = Cx_k + Du_k + Fe_k, \quad e_k \sim N(\mu_e, R_2) \tag{23b}$$

Mean and covariance

$$\mu_{k+1} = A\mu_k + Bu_k + G\mu_v, \quad \mu_0 = \mathbb{E}[x_0], \tag{24a}$$

$$P_{k+1} = AP_kA^T + GR_1G^T, \quad P_0 = \mathrm{Cov}(x_0) \tag{24b}$$

Note that $u_k$ is deterministic.

How do the different terms on the right-hand side of (23a) affect the distribution of the states over time?

# Continuous-Time Stochastic Processes (SDE)

Continuous-time system

$$\dot{x}(t) = f(t, x(t), u(t)) \tag{27}$$

First attempt at stochastic differential equation

$$\dot{x}(t) = f(t, x(t), u(t)) + g(t, x(t), u(t))v(t) \tag{28}$$

Process noise

- $v(t) \perp v(s)$ for any $t \neq s$ (independence)
- $v(t)$ is continuous and has bounded variance
- $\mathbb{E}[v(t)] = 0$ (zero-mean)

Theorem 4.1 in Chapter 3 of the book "Stochastic Control Theory" by Åström (1970): $\mathbb{E}[v^2(t)] = 0$

## Continuous-Time Stochastic Processes (SDE)

Stochastic difference equation

$$x(t+\Delta t) - x(t) = f(t, x(t))\Delta t + g(t, x(t))v(t)\Delta t + o(\Delta t). \tag{29}$$

Replace $v(t)\Delta t$ with $\Delta w(t) = w(t+\Delta t) - w(t)$, which has stationary independent zero-mean increments (Wiener process)

$$\Delta x(t) = f(t, x(t))\Delta t + g(t, x(t))\Delta w(t) + o(\Delta t). \tag{30}$$

Take the limit $\Delta t \to 0$

$$\mathrm{d}x(t) = f(t, x(t))\,\mathrm{d}t + g(t, x(t))\,\mathrm{d}w(t) \tag{31}$$

$$x(t) = x(t_0) + \int_{t_0}^{t} f(\tau, x(\tau))\,\mathrm{d}\tau + \int_{t_0}^{t} g(\tau, x(\tau))\,\mathrm{d}w(\tau) \tag{32}$$

# Continuous-Time Stochastic Processes (SDE)

The two first conditional moments of the difference process

$$\mathbb{E}\Big[\Delta x(t) \mid x(t)\Big] = f(t, x(t))\Delta t + o(\Delta t) \tag{33a}$$

$$\mathrm{Var}\Big(\Delta x(t) \mid x(t)\Big) = g^2(t, x(t))\Delta t + o(\Delta t) \tag{33b}$$

Variance of process noise increment

$$\mathbb{E}\Big[\Delta w^2(t)\Big] = \Delta t \tag{34}$$

Note that the variance is proportional to $\Delta t$ and not $\Delta t^2$

Distribution of process noise increment (increment of Wiener process)

$$\Delta w(t) = w(t + \Delta t) - w(t) \sim N(0, \Delta t) \tag{35}$$

# Linear Stochastic Differential Equations

Linear stochastic differential equations

$$\mathrm{d}x(t) = (A(t)x(t) + B(t)u(t))\,\mathrm{d}t + G(t)\,\mathrm{d}w(t), \quad x(t_0) \sim N(m_0, P_0) \tag{36}$$

$A(t)$ and $B(t)$ are continuous functions of time

State expectation

$$\mathbb{E}[x(t)] = \mathbb{E}[x_0] + \mathbb{E}\left[\int_{t_0}^{t} A(\tau)x(\tau) + B(\tau)u(\tau)\,\mathrm{d}\tau\right] + \mathbb{E}\left[\int_{t_0}^{t} G(\tau)\,\mathrm{d}w(\tau)\right] \tag{37}$$

$$= \mathbb{E}[x_0] + \int_{t_0}^{t} A(\tau)\mathbb{E}[x(\tau)] + B(\tau)u(\tau)\,\mathrm{d}\tau = m_x(t) \tag{38}$$

Expected value

$$\dot{m}_x(t) = Am_x(t) + B(t)u(t), \quad m_x(t_0) = m_0. \tag{39}$$

# Linear Stochastic Differential Equations

State-transition matrix

$$\frac{\partial \Phi(t;t_0)}{\partial t} = A(t)\Phi(t;t_0), \qquad \Phi(t_0;t_0) = I. \tag{40}$$

Auto-covariance of $x$ $(s \geq t)$

$$R(s,t) = \mathrm{Cov}(x(s), x(t)) = \Phi(s,t)P(t) \tag{41}$$

Covariance

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) + G(t)G^T(t), \qquad P(t_0) = P_0, \tag{42}$$

# Discretization

# Linear stochastic differential equation

Linear continuous-time state space model

$$\mathrm{d}x(t) = (Ax(t) + Bu(t))\,\mathrm{d}t + G\,\mathrm{d}w(t), \qquad \mathrm{d}w(t) \sim N(0, I\,\mathrm{d}t), \tag{43a}$$

$$y(t) = Cx(t) + Du(t) + Fe(t), \qquad e(t) \sim N(m_e, R_e) \tag{43b}$$

Zero-order-hold parametrization of manipulated inputs

$$u(t) = u_k, \qquad t \in [t_k, t_{k+1}[ \tag{44}$$

Approximation of process noise (not rigorous)

$$\mathrm{d}w(t) = \tilde{w}(t)\,\mathrm{d}t, \qquad \tilde{w}(t) \sim N(0, I) \tag{45}$$

Analytical solution

$$x(t_{k+1}) = e^{A(t_{k+1}-t_k)}x(t_k) + \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)}Bu(t_k)\,\mathrm{d}\tau + v(t_k), \tag{46a}$$

$$y(t_k) = Cx(t_k) + Du(t_k) + Fe(t_k), \quad e(t_k) \sim N(m_e, R_e) \tag{46b}$$

# Zero-order-hold parametrization and approximation

Discrete-time process noise

$$v(t_k) = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)} G\tilde{w}(\tau)\,\mathrm{d}\tau \tag{47a}$$

Mean

$$\mathbb{E}[v(t_k)] = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)} G\mathbb{E}[\tilde{w}(\tau)]\,\mathrm{d}\tau = 0 \tag{48}$$

# Zero-order-hold parametrization and approximation

## Covariance

$$\mathrm{Cov}(v(t_k)) = \mathbb{E}[v(t_k)v^T(t_k)] \tag{49}$$

$$= \mathbb{E}\left[\int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)}G\tilde{w}(\tau)\,\mathrm{d}\tau \int_{t_k}^{t_{k+1}} \tilde{w}^T(s)G^T e^{A^T(t_{k+1}-s)}\,\mathrm{d}s\right] \tag{50}$$

$$= \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)}G\mathbb{E}\left[\tilde{w}(\tau)\tilde{w}^T(s)\right]G^T e^{A^T(t_{k+1}-s)}\,\mathrm{d}\tau\,\mathrm{d}s \tag{51}$$

$$= \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)}G\mathbb{E}\left[\tilde{w}(\tau)\tilde{w}^T(\tau)\right]G^T e^{A^T(t_{k+1}-\tau)}\,\mathrm{d}\tau \tag{52}$$

$$= \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)}GG^T e^{A^T(t_{k+1}-\tau)}\,\mathrm{d}\tau \tag{53}$$

# Stochastic discrete-time state space models

Linear discrete-time stochastic state space model

$$x_{k+1} = A_d x_k + B_d u_k + v_k, \qquad v_k \sim N(0, R_1), \tag{54}$$

$$y_k = C_d x_k + D_d u_k + e_k, \qquad e_k \sim N(0, R_2) \tag{55}$$

System matrices in the state equation

$$\begin{bmatrix} A_d & B_d \\ 0 & I \end{bmatrix} = \exp\left( \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} T_s \right), \tag{56}$$

$$\begin{bmatrix} A_d & \tilde{R}_1 \\ 0 & A_d^{-T} \end{bmatrix} = \exp\left( \begin{bmatrix} A & GG^T \\ 0 & -A^T \end{bmatrix} T_s \right), \qquad R_1 = \tilde{R}_1 A_d^T \tag{57}$$

$$\begin{bmatrix} A_d^{-1} & \tilde{R}_1^T \\ 0 & A_d^T \end{bmatrix} = \exp\left( \begin{bmatrix} -A & GG^T \\ 0 & A^T \end{bmatrix} T_s \right), \qquad R_1 = A_d \tilde{R}_1^T \tag{58}$$

System matrices in the measurement equation

$$C_d = C, \qquad D_d = D, \qquad R_2 = F R_e F^T \tag{59}$$

# Proof of discretization

Exponential form

$$M = \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix} = \exp\left(\begin{bmatrix} F & G \\ 0 & H \end{bmatrix} t\right) = \exp(Kt) \tag{60}$$

Differential form

$$\dot{M} = KM, \quad M(t_0) = I \tag{61}$$

Individual differential equations

$$\dot{X} = FX, \qquad X(t_0) = I, \tag{62}$$

$$\dot{Y} = FY + GZ, \qquad Y(t_0) = 0, \tag{63}$$

$$\dot{Z} = HZ, \qquad Z(t_0) = I \tag{64}$$

Solutions

$$X = e^{F(t-t_0)}X(t_0) = e^{F(t-t_0)}, \tag{65}$$

$$Z = e^{H(t-t_0)}Z(t_0) = e^{H(t-t_0)}, \tag{66}$$

$$Y = e^{F(t-t_0)}Y(t_0) + \int_{t_0}^{t} e^{F(t-\tau)}Ge^{H(\tau-t_0)}Z(t_0)\,\mathrm{d}\tau = \int_{t_0}^{t} e^{F(t-\tau)}Ge^{H(\tau-t_0)}\,\mathrm{d}\tau$$

# Proof of discretization

Let $F = A$, $H = -A^T$, $G = GG^T$, $t_0 = t_k$, and $t = t_{k+1}$ $(t_{k+1} - t_k = T_s)$

$$X = e^{AT_s} = A_d, \tag{67}$$

$$Z = e^{-A^T T_s} = A_d^{-T}, \tag{68}$$

$$Y = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)} GG^T e^{A^T(t_{k+1}-\tau)} \, \mathrm{d}\tau \, e^{-A^T T_s} = \tilde{R}_1 = R_1 A_d^T \tag{69}$$

The proof of the other approach is similar, but has one more step:
A change of variables in the integral

# State estimation

# Filters and Estimation

Objective: Obtain estimate, $\hat{x}_t$, of the signal $x_t$ based on measurements

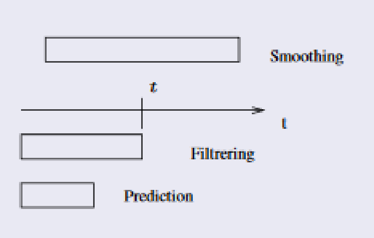
$$Y_{0:N} = \begin{bmatrix} y_0 & y_1 & \cdots & y_N \end{bmatrix} \tag{70}$$

and a state-output relation, e.g.,

$$y_t = Cx_t + e_t \tag{71}$$

Different types of estimation

1. Smoothing ($t < t_N$): Use both past and future data to estimate the states
2. Filtering: ($t = t_N$): Estimate the current states based on current and past data
3. Prediction: ($t > t_N$): Predict future states based on past data



# Filter Theory

Stochastic discrete-time system

$$x_{t+1} = Ax_t + Bu_t + v_t, \qquad x_0 \sim N(m_0, P_0), \qquad v_t \sim N(0, R_1), \tag{72}$$

$$y_t = Cx_t + e_t, \qquad e_t \sim N(0, R_2) \tag{73}$$

We will only consider filtering in this lecture

# Filter Theory - The Good Estimate

**Core concepts** of filter design

1. Characteristics of the signal and noise
2. Observation model (relation between $y$, $x$, $e$)
3. Criterion (what is a good estimate)
4. Restrictions (what information is available)

**Characteristics:** Nature of the states, dynamics, and noises

**Observation:** Relation between the output, $y$, the state $x$, and the noise

**The criterion:** A good estimate minimizes the expected squared deviation

$$\mathbb{E}[||x - \hat{x}||^2] \tag{74}$$

**Restrictions:** What data, $Y$, is available (filter, predict, or smoothe?)

# The filter problem: A good estimator

The law of total expectation

$$\mathbb{E}[g(x)] = \mathbb{E}_Y[\mathbb{E}[g(x)|Y]] \tag{75}$$

Introduce "inner" objective function

$$\begin{aligned} J &= \mathbb{E}[\|x - \hat{x}\|^2] && (76) \\ &= \mathbb{E}[(x - \hat{x})^T (x - \hat{x})] && (77) \\ &= \mathbb{E}_Y[\mathbb{E}[(x - \hat{x})^T (x - \hat{x})|Y]] = \mathbb{E}_Y[J_{in}] && (78) \end{aligned}$$

Inner objective function

$$\begin{aligned} J_{in} &= \mathbb{E}[x^T x - \hat{x}^T x - x^T \hat{x} + \hat{x}^T \hat{x}|Y], && (79) \\ &= \mathbb{E}[x^T x|Y] - \hat{x}^T \mathbb{E}[x|Y] - \mathbb{E}[x|Y]^T \hat{x} + \hat{x}^T \hat{x} && (80) \end{aligned}$$

Optimal estimate

$$\begin{aligned} \nabla_{\hat{x}} J_{in} &= 2\hat{x} - 2\mathbb{E}[x|Y] = 0, && (81) \\ \hat{x} &= \mathbb{E}[x|Y] && (82) \end{aligned}$$

# Projection theorem

# Filter Theory - Projection Theorem

Normally distributed vector

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left( \begin{bmatrix} m_x \\ m_y \end{bmatrix}, \begin{bmatrix} P_x & P_{xy} \\ P_{xy}^T & P_y \end{bmatrix} \right) \tag{83}$$

**Projection theorem**: The conditional distribution $X|Y \sim N(m_{x|y}, P_{x|y})$ is

$$m_{x|y} = m_x + P_{xy} P_y^{-1} (y - m_y), \tag{84}$$

$$P_{x|y} = P_x - P_{xy} P_y^{-1} P_{xy}^T, \tag{85}$$

$$X - \hat{x} \perp Y \tag{86}$$

# Filter Theory - Projection Theorem

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left( \begin{bmatrix} m_x \\ m_y \end{bmatrix}, \begin{bmatrix} P_x & P_{xy} \\ P_{xy}^T & P_y \end{bmatrix} \right), \tag{87}$$

$$\hat{x} = m_x + P_{xy} P_y^{-1} (y - m_y) \tag{88}$$

$$P_{x|y} = P_x - P_{xy} P_y^{-1} P_{xy}^T \tag{89}$$

Assume that $X$ and $Y$ are scalar.

1. What happens if we measure exactly the value we expected?
2. What happens if the measurement is an outlier?
3. What if $X$ and $Y$ are uncorrelated?
4. Can $P_{x|y}$ become negative?
5. What happens as $P_x$ or $P_y$ approach zero?

Think about it for yourself for one minute and
then discuss with the person next to you for five minutes.

# Filter Theory - Proof of Projection Theorem

Probability density functions

$$f_{X,Y}(x,y) = \frac{1}{(2\pi)^{n_x+n_y}\sqrt{\det(P_z)}} e^{-\frac{1}{2}(z-m_z)^T P_z^{-1}(z-m_z)}, \tag{90}$$

$$f_Y(y) = \frac{1}{(2\pi)^{n_y}\sqrt{\det(P_y)}} e^{-\frac{1}{2}(y-m_y)^T P_y^{-1}(y-m_y)} \tag{91}$$

Probability density function of conditional normal distribution

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{92}$$

$$= \sqrt{\frac{\det(P_y)}{(2\pi)^{n_x}\det(P_z)}} e^{-\frac{1}{2}(z-m_z)^T P_z^{-1}(z-m_z)+\frac{1}{2}(y-m_y)^T P_y^{-1}(y-m_y)}$$

$$= \kappa e^{-\frac{1}{2}\alpha} \tag{93}$$

# Filter Theory - Proof of Projection Theorem

Schur complement

$$D = P_x - P_{xy} P_y^{-1} P_{xy}^T \tag{94}$$

Use Woodbury matrix identity on $P_z^{-1}$

$$P_z^{-1} = \begin{bmatrix} D^{-1} & -D^{-1} P_{xy} P_y^{-1} \\ -P_y^{-1} P_{xy}^T D^{-1} & P_y^{-1} + P_y^{-1} P_{xy}^T D^{-1} P_{xy} P_y^{-1} \end{bmatrix} \tag{95}$$

Determinant

$$\det(P_z) = \det(P_y)\det(D) \qquad \Leftrightarrow \qquad \frac{\det(P_y)}{\det(P_z)} = \frac{1}{\det(D)} \tag{96}$$

Factor

$$\kappa = \sqrt{\frac{\det(P_y)}{(2\pi)^{n_x}\det(P_z)}} = \frac{1}{\sqrt{(2\pi)^{n_x}\det(D)}} \tag{97}$$

Exponent

$$\begin{aligned} \alpha &= (z - m_z)^T P_z^{-1} (z - m_z) - (y - m_y)^T P_y^{-1} (y - m_y) & (98) \\ &= [x - (m_x + P_{xy} P_y^{-1}(y - m_y))]^T D^{-1} [x - (m_x + P_{xy} P_y^{-1}(y - m_y)] & (99) \end{aligned}$$

# Filter Theory - Proof of Projection Theorem

Mean and covariance of conditional distribution

$$\mathbb{E}[X|Y] = m_{x|y} = m_x + P_{xy}P_y^{-1}(y - m_y) \tag{100}$$

$$\mathrm{Cov}(X|Y) = P_{x|y} = D = P_x - P_{xy}P_y^{-1}P_{xy}^T \tag{101}$$

Covariance (are the variables independent?)

$$\mathrm{Cov}(X - m_{x|y}, Y) = \mathrm{Cov}(X, Y) - P_{xy}P_y^{-1}\,\mathrm{Cov}(Y, Y) \tag{102}$$

$$= P_{xy} - P_{xy}P_y^{-1}P_y = 0 \tag{103}$$

As $X$ and $Y$ are Gaussian, they are independent

# Kalman filter

# State estimation

Stochastic discrete-time system

$$x_{t+1} = Ax_t + Bu_t + v_t, \qquad x_0 \sim N(\hat{x}_0, P_0), \qquad v_t \sim N(0, R_1), \tag{104}$$

$$y_t = Cx_t + e_t, \qquad e_t \sim N(0, R_2) \tag{105}$$

$v_t \perp x_s$ for all $s \leq t$ and $e_t \perp x_s$ for all $s$

Mean and covariance of joint distribution

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} | Y_{t-1} \sim N\left( \begin{bmatrix} \times \\ \times \end{bmatrix}, \begin{bmatrix} \times & \times \\ \times & \times \end{bmatrix} \right), \quad Y_t = \begin{bmatrix} Y_{t-1} \\ y_t \end{bmatrix} \tag{106}$$

Conditional state distributions

$$x_t | Y_{t-1} \sim N(\hat{x}_{t|t-1}, P_{t|t-1}) \tag{107}$$

$$x_t | Y_t = x_t | y_t, Y_{t-1} \sim N(\hat{x}_{t|t}, P_{t|t}) \tag{108}$$

# State estimation: A recursion

Measurement equation

$$y_t = Cx_t + e_t, \qquad e_t \sim N(0, R_2), \qquad e_t \perp x_s \tag{109}$$

Mean

$$\mathbb{E}[y_t|Y_{t-1}] = C\mathbb{E}[x_t|Y_{t-1}] + \mathbb{E}[e_t|Y_{t-1}] = C\hat{x}_{t|t-1} \tag{110}$$

Covariance

$$\begin{aligned}\operatorname{Cov}(y_t|Y_{t-1}) = C\operatorname{Cov}(x_t|Y_{t-1})C^T + C\operatorname{Cov}(x_t, e_t|Y_{t-1}) \\ + \operatorname{Cov}(e_t, x_t|Y_{t-1})C^T + \operatorname{Cov}(e_t|Y_{t-1}) = CP_{t|t-1}C^T + R_2\end{aligned} \tag{111}$$

Cross-covariance

$$\operatorname{Cov}(y_t, x_t|Y_{t-1}) = C\operatorname{Cov}(x_t|Y_{t-1}) + C\operatorname{Cov}(x_t, e_t|Y_{t-1}) \tag{112}$$

$$= CP_{t|t-1} \tag{113}$$

Conditional distribution

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} |Y_{t-1} \sim N\left( \begin{bmatrix} \hat{x}_{t|t-1} \\ C\hat{x}_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}C^T \\ CP_{t|t-1} & CP_{t|t-1}C^T + R_2 \end{bmatrix} \right) \tag{114}$$

# State Estimation 4: The Current Estimate

Conditional distribution

$$x_t | y_t, Y_{t-1} = x_t | Y_t \sim N(\hat{x}_{t|t}, P_{t|t}) \quad (115)$$

Use projection theorem

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + P_{t|t-1} C^T (C P_{t|t-1} C^T + R_2)^{-1} (y_t - C \hat{x}_{t|t-1}), \quad (116)$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} C^T (C P_{t|t-1} C^T + R_2)^{-1} C P_{t|t-1} \quad (117)$$

# State Estimation 5: Prediction Estimate

State equation

$$x_{t+1} = Ax_t + Bu_t + v_t, \quad v_t \sim N(0, R_1), \quad v_t \perp x_s \text{ for all } s \leq t \tag{118}$$

Mean

$$\mathbb{E}[x_{t+1}|Y_t] = A\mathbb{E}[x_t|Y_t] + B\mathbb{E}[u_t|Y_t] + \mathbb{E}[v_t|Y_t] \tag{119}$$
$$= A\hat{x}_{t|t} + Bu_t \tag{120}$$

Covariance

$$\mathrm{Cov}(x_{t+1}|Y_t) = A\,\mathrm{Cov}(x_t|Y_t)A^T + \mathrm{Cov}(v_t|Y_t) + A\,\mathrm{Cov}(x_t, v_t|Y_t) \tag{121}$$
$$+ \mathrm{Cov}(v_t, x_t|Y_t)A^T = AP_{t|t}A^T + R_1 \tag{122}$$

Prediction estimate

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \tag{123}$$
$$P_{t+1|t} = AP_{t|t}A^T + R_1 \tag{124}$$

# Filter Theory - Kalman Filter

Data/measurement-update (inference)

$$
\boxed{\begin{aligned}
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + \kappa_t(y_t - C\hat{x}_{t|t-1}), \\
\kappa_t &= P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}, \\
P_{t|t} &= P_{t|t-1} - \kappa_t C P_{t|t-1}
\end{aligned}}
\begin{aligned}
&(125)\\
&(126)\\
&(127)
\end{aligned}
$$

Time-update (prediction)

$$
\boxed{\begin{aligned}
\hat{x}_{t+1|t} &= A\hat{x}_{t|t} + Bu_t, & \hat{x}_{0|0} &= \hat{x}_0, \\
P_{t+1|t} &= AP_{t|t}A^T + R_1, & P_{0|0} &= P_0
\end{aligned}}
\begin{aligned}
&(128)\\
&(129)
\end{aligned}
$$

How do the measurements, $y_t$, affect the covariances and the Kalman gain?
Is it intuitive that it is that way, and can we use it to our advantage?

Think about it for yourself for one minute and
then discuss with the person next to you for one minute.

# Example: Pseudocode - Kalman Filter/Simulation Implementation

Initial values: $x_{0|-1}$, $P_{0|-1}$, $x_0$

---

for $t = 0, \ldots, N$

Measurement from true system:

$y_t = \texttt{Measurement}(x_t, e_t)$

---

Data update:

$[\hat{x}_{t|t}, P_{t|t}, \kappa_t] = \texttt{DataUpdate}(y_t, \hat{x}_{t|t-1}, P_{t|t-1}; C, R_2)$

---

Compute control:

$u_t = \texttt{Actuator}(\hat{x}_{t|t})$

Apply control:

$x_{t+1} = \texttt{Simulator}(x_t, u_t, v_t)$

---

Time update:

$[\hat{x}_{t+1|t}, P_{t+1|t}] = \texttt{TimeUpdate}(\hat{x}_{t|t}, P_{t|t}, u_t; A, B, R_1)$

end

# Example: Estimation of constant

Estimate scalar constant

$$x_{t+1} = x_t, \tag{130}$$

$$y_t = x_t + e_t, \quad e_t \in N(0, r_2) \tag{131}$$

Define $q_t = p_t^{-1}$

$$\kappa_t = \frac{p_t}{p_t + r_2} = \frac{1}{1 + r_2 q_t}, \tag{132}$$

$$p_{t+1} = (1 - \kappa_t)p_t = \left(1 - \frac{1}{1 + r_2 q_t}\right)p_t = \frac{r_2 q_t}{1 + r_2 q_t}p_t = \frac{r_2}{1 + r_2 q_t}, \tag{133}$$

$$q_{t+1} = \frac{1}{p_{t+1}} = \frac{1 + r_2 q_t}{r_2} = q_t + \frac{1}{r_2} = q_0 + \frac{t+1}{r_2}, \tag{134}$$

$$\hat{x}_{t+1} = \hat{x}_t + \kappa_t(y_t - \hat{x}_t) \tag{135}$$

If $q_0 = 0$ $(p_0 = \infty)$,

$$\hat{x}_{t+1} = \hat{x}_t + \frac{1}{1+t}(y_t - \hat{x}_t) \quad \text{or} \quad \hat{x}_t = \frac{1}{t}\sum_{i=0}^{t-1} y_i \tag{136}$$

# Two different forms of the Kalman filter

**Ordinary** Kalman filter

$$\underbrace{\begin{bmatrix} \hat{x}_{t|t} \\ P_{t|t} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t+1|t} \\ P_{t+1|t} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t+1|t+1} \\ P_{t+1|t+1} \end{bmatrix}}_{\text{Ordinary Kalman Filter}} \tag{137}$$

**Predictive** Kalman filter

$$\underbrace{\begin{bmatrix} \hat{x}_{t|t-1} \\ P_{t|t-1} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t|t} \\ P_{t|t} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t+1|t} \\ P_{t+1|t} \end{bmatrix}}_{\text{Predictive Kalman Filter}} \tag{138}$$

# Ordinary Kalman filter

Time-update (prediction)

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \qquad \hat{x}_{0|0} = \hat{x}_0, \tag{139}$$

$$P_{t+1|t} = AP_{t|t}A^T + R_1, \qquad P_{0|0} = P_0 \tag{140}$$

Data-update (inference)

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \kappa_t(y_t - C\hat{x}_{t|t-1}), \tag{141}$$

$$\kappa_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}, \tag{142}$$

$$P_{t|t} = P_{t|t-1} - \kappa_t CP_{t|t-1} \tag{143}$$

**Ordinary** Kalman filter

$$\boxed{\begin{aligned} \hat{x}_{t|t} &= (I - \kappa_t C)(A\hat{x}_{t-1|t-1} + Bu_{t-1}) + \kappa_t y_t, \\ P_{t|t} &= AP_{t-1|t-1}A^T + R_1 - \kappa_t C(AP_{t-1|t-1}A^T + R_1), \\ \kappa_t &= (AP_{t-1|t-1}A^T + R_1)C^T(C(AP_{t-1|t-1}A^T + R_1)C^T + R_2)^{-1} \end{aligned}}$$

(144)

(145)

(146)

# Predictive Kalman filter

Time-update (prediction)

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \qquad \hat{x}_{0|0} = \hat{x}_0, \tag{147}$$

$$P_{t+1|t} = AP_{t|t}A^T + R_1, \qquad P_{0|0} = P_0 \tag{148}$$

Data-update (inference)

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \kappa_t(y_t - C\hat{x}_{t|t-1}), \tag{149}$$

$$\kappa_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}, \tag{150}$$

$$P_{t|t} = P_{t|t-1} - \kappa_t CP_{t|t-1} \tag{151}$$

**Predictive** Kalman filter

$$\hat{x}_{t+1|t} = (A - K_tC)\hat{x}_{t|t-1} + Bu_t + K_ty_t, \tag{152}$$

$$P_{t+1|t} = AP_{t|t-1}A^T + R_1 - K_tCP_{t|t-1}A^T, \tag{153}$$

$$K_t = A\kappa_t = AP_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1} \tag{154}$$

# Questions?