Stochastic Adaptive Control (02421)

Lecture 6

Tobias K. S. Ritschel

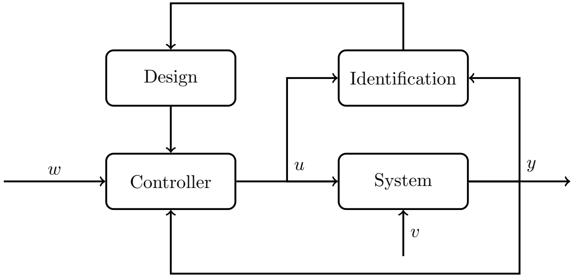
Section for Dynamical Systems

Department of Applied Mathematics and Computer Science

Technical University of Denmark

Lecture Plan

1. System theory
2. Stochastics
3. State estimation 1
4. State estimation 2
5. Optimal control 1
6. **System identification 1 + adaptive control 1**
7. External models + prediction
8. Optimal control 2
9. Optimal control 3
10. System identification 2
11. System identification 3 + model validation
12. System identification 4 + adaptive control 2
13. Adaptive control 3



Today's Agenda

- Follow-up from last time
- Least-squares parameter estimation
- Maximum likelihood parameter estimation
- Fisher's information matrix and the Cramér-Rao lower bound (CRLB)
- Recursive parameter estimation (extended Kalman filter)
- Adaptive control

Follow-up from last time: General pole placement

System

$$x_{k+1} = \begin{bmatrix} 3/10 & 0 \\ 7/10 & 3/2 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_k + v_k, \quad v_k \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/100 & 0 \\ 0 & 1/50 \end{bmatrix} \right) \tag{1}$$

Task: Find a control law that places the closed-loop poles in $1/5$

Control law

$$u_k = -Lx_k \tag{2}$$

Closed-loop system

$$x_{k+1} = (A - BL)x_k + v_k \tag{3}$$

Follow-up from last time: General pole placement

Eigenvalues

$$\text{eig}\left(\begin{bmatrix} 3/10 & 0 \\ 7/10 & 3/2 \end{bmatrix}\right) = \{1.5, 0.3\} \tag{4}$$

Polynomials

$$A_d(q) = (q-0.2)(q-0.2) = q^2 - (0.2+0.2)q + 0.2^2 \tag{5}$$
$$= q^2 - 0.4q + 0.04, \quad \Rightarrow \quad \alpha_1 = -0.4 \quad \alpha_2 = 0.04, \tag{6}$$
$$A(q) = (q-1.5)(q-0.3) = q^2 - (1.5+0.3)q + 1.5 \cdot 0.3 \tag{7}$$
$$= q^2 - 1.8q + 0.45, \quad \Rightarrow \quad a_1 = -1.8 \quad a_2 = 0.45 \tag{8}$$

Controller gain (for controller canonical form)

$$L_{cc} = \begin{bmatrix} \alpha_1 - a_1 & \alpha_2 - a_2 \end{bmatrix} = \begin{bmatrix} -0.4 - (-1.8) & 0.04 - 0.45 \end{bmatrix} \tag{9}$$
$$= \begin{bmatrix} 1.4 & -0.41 \end{bmatrix} \tag{10}$$

Follow-up from last time: General pole placement

System matrices in controller canonical form

$$A_{cc} = \begin{bmatrix} -a_1 & -a_2 \\ 1 & 0 \end{bmatrix} \qquad B_{cc} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{11}$$

$$= \begin{bmatrix} 1.8 & -0.45 \\ 1 & 0 \end{bmatrix} \tag{12}$$

Controllability matrices

$$W_{c,cc} = \begin{bmatrix} B_{cc} & A_{cc}B_{cc} \end{bmatrix}, \qquad W_c = \begin{bmatrix} B & AB \end{bmatrix} \tag{13}$$

$$= \begin{bmatrix} 1 & 1.8 \\ 0 & 1 \end{bmatrix} \qquad = \begin{bmatrix} 1 & 0.3 \\ 0 & 0.7 \end{bmatrix} \tag{14}$$

Similarity transformation matrix

$$T = W_{c,cc} W_c^{-1} = \begin{bmatrix} 1 & -1.5 \\ 0 & 0.7 \end{bmatrix} \tag{15}$$

Follow-up from last time: General pole placement

Controller gain

$$L = L_{cc}T = \begin{bmatrix} 1.4 & 2.4143 \end{bmatrix} \tag{16}$$

Closed-loop system matrix

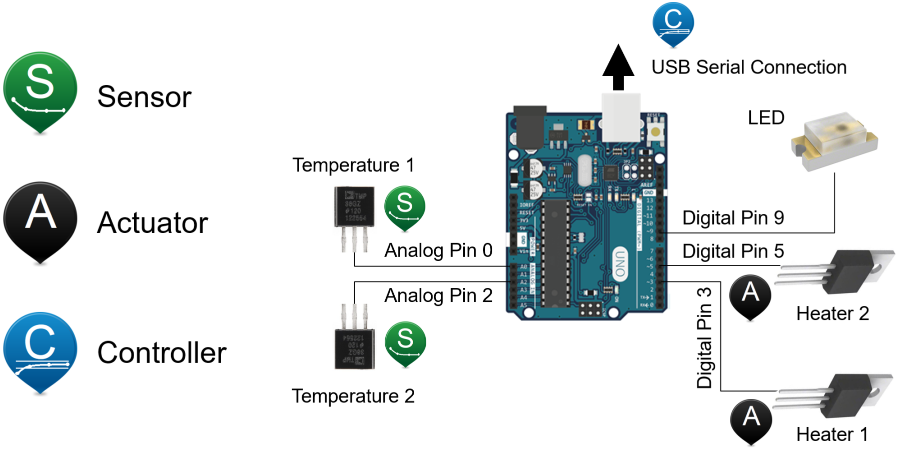
$$A_{cl} = A - BL = \begin{bmatrix} -1.1 & 2.4143 \\ 0.7 & 1.5 \end{bmatrix} \tag{17}$$

Closed-loop eigenvalues (poles)

$$\text{eig}(A_{cl}) = \{0.2, 0.2\} \tag{18}$$

Demonstration

Temperature control laboratory (TCLab)



Link: `https://apmonitor.com/pdc/index.php/Main/ArduinoTemperatureControl`

TCLab model

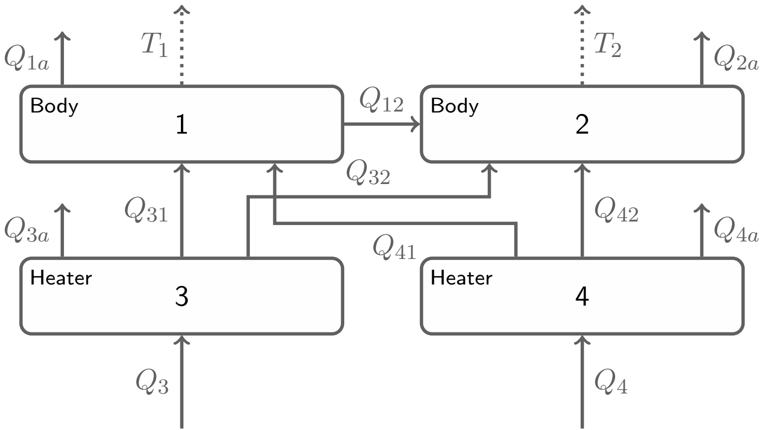


Figure: Four-compartment model of TCLab device.

Least-squares parameter estimation

System

System

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t + G(\theta)v_t, \tag{19}$$

$$y_t = C(\theta)x_t + D(\theta)u_t + F(\theta)e_t, \tag{20}$$

Stochastic vectors

$$x_0 \sim N(m_0(\theta), P_0(\theta)), \qquad v_t \sim N(0, I), \qquad e_t \sim N(0, I) \tag{21}$$

Note: The process and measurement noise are *standard* normal such that we can linearize wrt. θ in the recursive formulation

Least-squares parameter estimation problem

State prediction

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t, \qquad \hat{x}_0 = m_0 \tag{22}$$

Output prediction

$$\hat{y}_t = C\hat{x}_t + Du_t \tag{23}$$

Residuals

$$\epsilon_t = y_t - \hat{y}_t \tag{24}$$

Least-squares parameter estimation problem

$$\hat{\theta} = \arg\min_{\theta} J_N(\theta; Y_N), \qquad J_N(\theta; Y_N) = \frac{1}{2}\sum_{t=0}^{N} \epsilon_t^T \epsilon_t \tag{25}$$

This problem is nonlinear in the parameters, θ, and the solution must be approximated numerically, e.g., using Matlab's `fmincon`.

Matlab's `fmincon`

Syntax

```
x = fmincon(fun,x0,A,b)
x = fmincon(fun,x0,A,b,Aeq,beq)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon,options)
x = fmincon(problem)
[x,fval] = fmincon(___)
[x,fval,exitflag,output] = fmincon(___)
[x,fval,exitflag,output,lambda,grad,hessian] = fmincon(___)
```

Description

Nonlinear programming solver.
Finds the minimum of a problem specified by

$$\min_x f(x) \text{ such that } \begin{cases} c(x) \leq 0 \\ ceq(x) = 0 \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub, \end{cases} \tag{26}$$

b and beq are vectors, A and Aeq are matrices, c(x) and ceq(x) are functions that return vectors, and f(x) is a function that returns a scalar. f(x), c(x), and ceq(x) can be nonlinear functions. x, lb, and ub can be passed as vectors or matrices; see Matrix Arguments.

Link: `https://de.mathworks.com/help/optim/ug/fmincon.html`

Matlab's `fmincon`: Least-squares objective function

```
function JN = least_squares_objective_function(theta, Y, U, m0, p)
    % Create system matrices (provided by the user)
    [A, B, C, D] = p.create_system_matrices(theta, p);

    % Initial state
    xhatt = m0;

    % Objective function
    JN = 0;

    for t = 1:N+1 % (the real t is actually 0:N)
        % Manipulated input and output
        ut = U(:, t);
        yt = Y(:, t);

        % Predict output and residual
        yhatt    = C*xhatt + D*ut;
        epsilont = yt - yhatt;

        % Predict state
        xhattp1 = A*xhatt + B*ut;

        % Add to objective function
        JN = JN + 0.5*(epsilont'*epsilont);

        % Update states
        xhatt = xhattp1;
    end
end
```

Maximum likelihood parameter estimation

Likelihood

Multiplication rule

$$P(A,B) = P(A|B)P(B), \qquad P(A,B|C) = P(A|B,C)P(B|C) \tag{27}$$

Probability density function

$$p(y_N, y_{N-1}, \ldots, y_0|\theta) = p(y_N|y_{N-1}, \ldots, y_0, \theta)p(y_{N-1}, \ldots, y_0|\theta) \tag{28}$$

Likelihood

$$\mathcal{L}(\theta) = p(y_N, y_{N-1}, \ldots, y_0|\theta) = p(y_0|\theta)\prod_{t=1}^{N} p(y_t|y_{t-1}, \ldots, y_0, \theta) \tag{29}$$

Log-likelihood

$$\ln \mathcal{L}(\theta) = \ln p(y_0|\theta) + \sum_{t=1}^{N} \ln p(y_t|y_{t-1}, \ldots, y_0, \theta) \tag{30}$$

Likelihood for normally distributed variables

Probability density of normal distribution

$$p(y_t|y_{t-1},\dots,y_0,\theta) = \frac{1}{\sqrt{(2\pi)^{n_y}\det P_{y,t|t-1}}}\exp\left(-\frac{1}{2}\epsilon_t^T P_{y,t|t-1}^{-1}\epsilon_t\right) \quad (31)$$

Residuals

$$\epsilon_t = y_t - \hat{y}_{t|t-1} \quad (32)$$

Logarithm of probability density of normal distribution

$$\ln p(y_t|y_{t-1},\dots,y_0,\theta) = -\frac{n_y}{2}\ln 2\pi - \frac{1}{2}\ln\det P_{y,t|t-1} - \frac{1}{2}\epsilon_t^T P_{y,t|t-1}^{-1}\epsilon_t \quad (33)$$

Log-likelihood

$$\mathcal{L}(\theta) = -\frac{(N+1)n_y}{2}\ln 2\pi - \frac{1}{2}\sum_{t=0}^{N}\left(\ln\det P_{y,t|t-1} + \epsilon_t^T P_{y,t|t-1}^{-1}\epsilon_t\right) \quad (34)$$

Maximum likelihood estimation

Maximum likelihood estimation

$$\hat{\theta} = \arg\max_{\theta} \quad \mathcal{L}(\theta) \tag{35}$$

Equivalent formulation

$$\hat{\theta} = \arg\min_{\theta} \quad J_N(\theta), \qquad\qquad J_N(\theta) = -\ln \mathcal{L}(\theta) \tag{36}$$

Negative log-likelihood function

$$\boxed{J_N(\theta) = \frac{(N+1)n_y}{2} \ln 2\pi + \frac{1}{2} \sum_{t=0}^{N} \left(\ln \det P_{y,t|t-1} + \epsilon_t^T P_{y,t|t-1}^{-1} \epsilon_t \right)} \tag{37}$$

Key differences to least-squares objective function

1. Determinant of **covariance** is penalized
2. Residuals are weighted by the inverse of the **covariance**

Kalman filter equations

Measurement update (vectors)

$$\hat{y}_{t|t-1} = C\hat{x}_{t|t-1} + Du_t, \quad (38)$$

$$\epsilon_t = y_t - \hat{y}_{t|t-1}, \quad (39)$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \kappa_t \epsilon_t \quad (40)$$

Measurement update (matrices)

$$P_{y,t|t-1} = CP_{t|t-1}C^T + R_2, \quad (41)$$

$$P_{xy,t|t-1} = P_{t|t-1}C^T, \quad (42)$$

$$\kappa_t = P_{xy,t|t-1}P_{y,t|t-1}^{-1}, \quad (43)$$

$$P_{t|t} = P_{t|t-1} - \kappa_t P_{xy,t|t-1}^T \quad (44)$$

Time update

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \quad (45)$$

$$P_{t+1|t} = AP_{t|t}A^T + R_1 \quad (46)$$

Matlab's `fmincon`: Negative log-likelihood objective function

```matlab
function JN = maximum_likelihood_objective_function(theta, Y, U, m0, p)
    % Create system matrices (provided by the user)
    [A, B, G, C, D, F] = p.create_system_matrices(theta, p);

    % Initial state and covariance
    xhatttm1 = m0;
    Pttm1    = P0;

    % Objective function
    JN = 0.5*numel(Ybar)*log(2*pi);

    for t = 1:N+1 % (the real t is actually 0:N)
        % Manipulated input and output
        ut = U(:, t);
        yt = Y(:, t);

        % Predicted output and covariance, residual, and Kalman gain
        yhatttm1 = C*xhatttm1 + D*ut;
        Pyttm1   = C*Pttm1*C' + R2;
        epsilont = yt - yhatttm1;
        kappat   = Pttm1*C'/Pyttm1;

        % Measurement update
        xhattt = xhatttm1 + kappat*epsilont;
        Ptt    = Pttm1 - kappat*C*Pttm1;

        % Time update
        xhatttp1 = A*xhattt + B*ut;
        Pttp1    = A*Ptt*A' + R1;

        % Add to objective function
        JN = JN + 0.5*(log(det(Pyttm1)) + epsilont'*(Pyttm1\epsilont));

        % Update states
        xhatttm1 = xhattp1;
    end
end
```

Fisher's information matrix and Cramér-Rao's lower bound

Cramér-Rao lower bound

Lower bound on individual parameter variances[1]

$$\mathrm{Cov}(\hat{\theta}) \succeq F^{-1} \tag{47}$$

where $A \succeq B$ means that $A - B$ is positive semidefinite.

Fisher's information matrix

$$F_{ij} = \mathbb{E}\left[\frac{\partial J_N}{\partial \theta_i}(\theta; Y_N)\frac{\partial J_N}{\partial \theta_j}(\theta; Y_N)\right] \tag{48}$$

Typically, maximum likelihood estimators are *efficient*, which means that equality holds in the bound

[1] Theorem 4.4 in the book by A. van den Bos, 2007. Parameter estimation for scientists and engineers. Wiley.

The prediction problem: Compact notation (recap)

Compact notation

$$X_N = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad Y_N = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad U_N = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \end{bmatrix}, \quad V_N = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_N \end{bmatrix}, \quad E_N = \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_N \end{bmatrix}, \tag{49}$$

$$\Phi_{xx} = \begin{bmatrix} I \\ A \\ \vdots \\ A^N \end{bmatrix}, \quad \Gamma_{xu} = \begin{bmatrix} 0 & & & \\ B & 0 & & \\ \vdots & \ddots & \ddots & \\ A^{N-1}B & \cdots & B & 0 \end{bmatrix}, \quad \Gamma_{xv} = \begin{bmatrix} 0 & & & \\ G & 0 & & \\ \vdots & \ddots & \ddots & \\ A^{N-1}G & \cdots & G & 0 \end{bmatrix}, \tag{50}$$

$$\Phi_{yx} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^N \end{bmatrix}, \quad \Gamma_{yu} = \begin{bmatrix} D & & & \\ CB & D & & \\ \vdots & \ddots & \ddots & \\ CA^{N-1}B & \cdots & CB & D \end{bmatrix}, \quad \Gamma_{yv} = \begin{bmatrix} 0 & & & \\ CG & 0 & & \\ \vdots & \ddots & \ddots & \\ CA^{N-1}G & \cdots & CG & 0 \end{bmatrix}, \tag{51}$$

$$\Gamma_{ye} = \begin{bmatrix} F & & & \\ & F & & \\ & & \ddots & \\ & & & F \end{bmatrix}, \quad R_V = I, \tag{52}$$

$$X_N = \Phi_{xx} x_0 + \Gamma_{xu} U_N + \Gamma_{xv} V_N, \qquad V_N \sim N(0, R_V), \tag{53}$$

$$Y_N = \Phi_{yx} x_0 + \Gamma_{yu} U_N + \Gamma_{yv} V_N + \Gamma_{ye} E_N, \qquad E_N \sim N(0, R_E) \tag{54}$$

*The superscript N on the matrices and the dependencies on the parameters, θ, have been omitted for brevity of notation.

Output expectation and covariance (recap)

Expectation

$$\mathbb{E}[Y_N] = \Phi_{yx}\mathbb{E}[x_0] + \Gamma_{yu}U_N + \Gamma_{yv}\mathbb{E}[V_N] + \Gamma_{ye}\mathbb{E}[E_N] \tag{55}$$

$$= \Phi_{yx}m_0 + \Gamma_{yu}U_N \tag{56}$$

Deviation from expectation

$$Y_N - \mathbb{E}[Y_N] = \Phi_{yx}(x_0 - m_0) + \Gamma_{yv}V_N + \Gamma_{ye}E_N \tag{57}$$

Covariance

$$\mathrm{Cov}(Y_N) = \mathbb{E}\left[(Y_N - \mathbb{E}[Y_N])(Y_N - \mathbb{E}[Y_N])^T\right] \tag{58}$$

$$= \Phi_{yx}\mathbb{E}[(x_0 - m_0)(x_0 - m_0)^T]\Phi_{yx}^T \tag{59}$$

$$+ \Gamma_{yv}\mathbb{E}[V_N V_N^T]\Gamma_{yv}^T + \Gamma_{ye}\mathbb{E}[E_N E_N^T]\Gamma_{ye}^T + \cdots \tag{60}$$

$$= \Phi_{yx}P_0\Phi_{yx}^T + \Gamma_{yv}R_V\Gamma_{yv}^T + \Gamma_{ye}R_E\Gamma_{ye}^T \tag{61}$$

Distribution of prediction

Distribution of output prediction

$$Y_N \sim N(m_Y, P_Y) \tag{62}$$

Mean

$$m_Y = \mathbb{E}[Y_N] = \Phi_{yx} m_0 + \Gamma_{yu} U_N \tag{63}$$

Covariance

$$P_Y = \mathrm{Cov}(Y_N) = \Phi_{yx} P_0 \Phi_{yx}^T + \Gamma_{yv} R_V \Gamma_{yv}^T + \Gamma_{ye} R_E \Gamma_{ye}^T \tag{64}$$

Negative log-likelihood function (multiplication rule not used)

$$J_N(\theta) = -\ln \mathcal{L}(\theta; Y_N) = \frac{(N+1)n_y}{2} \ln(2\pi) + \frac{1}{2} \ln \det P_Y \tag{65}$$

$$+ \frac{1}{2} (Y_N - m_Y)^T P_Y^{-1} (Y_N - m_Y) \tag{66}$$

Linear algebra hints

Problem 1: `det(PY)` might round to zero even though $\ln\det P_Y > -\infty$

Compute logarithm of determinant of covariance matrix using eigenvalues

$$\ln\det P_Y = \sum_{i=1}^{N_Y} \ln\lambda_i, \qquad N_Y = (N+1)n_y \tag{67}$$

λ_i for $i = 1, \ldots, N_Y$ are eigenvalues of P_Y

Problem 2: `PY\epsilon` might give imprecise results (and a warning)

Use an LDL factorization ($P_Y = LDL^T$)

$$LZ = \epsilon, \qquad Z = DL^T X, \tag{68}$$

$$DY = Z, \qquad Y = L^T X, \tag{69}$$

$$L^T X = Y \tag{70}$$

Derivative of negative log-likelihood

First-order derivatives

$$\frac{\partial J_N}{\partial \theta_i}(\theta; Y_N) = \frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right) - \frac{1}{2}\left(\frac{\partial m_Y}{\partial \theta_i}\right)^T P_Y^{-1}(Y_N - m_Y) \tag{71}$$

$$- \frac{1}{2}(Y_N - m_Y)^T P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}(Y_N - m_Y) \tag{72}$$

$$- \frac{1}{2}(Y_N - m_Y)^T P_Y^{-1}\frac{\partial m_Y}{\partial \theta_i} \tag{73}$$

Simplify

$$\frac{\partial J_N}{\partial \theta_i}(\theta; Y_N) = \frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right) - (Y_N - m_Y)^T P_Y^{-1}\frac{\partial m_Y}{\partial \theta_i} \tag{74}$$

$$- \frac{1}{2}(Y_N - m_Y)^T P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}(Y_N - m_Y) \tag{75}$$

Derivatives of mean and covariance

Derivative of mean

$$\frac{\partial m_Y}{\partial \theta_i} = \frac{\partial \Phi_{yx}}{\partial \theta_i} m_0 + \Phi_{yx} \frac{\partial m_0}{\partial \theta_i} + \frac{\partial \Gamma_{yu}}{\partial \theta_i} U_N \tag{76}$$

Derivative of covariance

$$\frac{\partial P_Y}{\partial \theta_i} = \frac{\partial \Phi_{yx}}{\partial \theta_i} P_0 \Phi_{yx}^T + \Phi_{yx} \frac{\partial P_0}{\partial \theta_i} \Phi_{yx}^T + \Phi_{yx} P_0 \left(\frac{\partial \Phi_{yx}}{\partial \theta_i} \right)^T \tag{77}$$

$$+ \frac{\partial \Gamma_{yv}}{\partial \theta_i} R_V \Gamma_{yv}^T + \Gamma_{yv} \frac{\partial R_V}{\partial \theta_i} \Gamma_{yv}^T + \Gamma_{yv} R_V \left(\frac{\partial \Gamma_{yv}}{\partial \theta_i} \right)^T \tag{78}$$

$$+ \frac{\partial \Gamma_{ye}}{\partial \theta_i} R_E \Gamma_{ye}^T + \Gamma_{ye} \frac{\partial R_E}{\partial \theta_i} \Gamma_{ye}^T + \Gamma_{ye} R_E \left(\frac{\partial \Gamma_{ye}}{\partial \theta_i} \right)^T \tag{79}$$

Derivatives for prediction matrices

Derivative of matrix power

$$\frac{\partial A^N}{\partial \theta_i} = \sum_{k=1}^{N} A^{k-1} \frac{\partial A}{\partial \theta_i} A^{N-k} \tag{80}$$

$$= \frac{\partial A^{N-1}}{\partial \theta_i} A + A^{N-1} \frac{\partial A}{\partial \theta_i} \tag{81}$$

Derivative of matrix product

$$\frac{\partial}{\partial \theta_i} \left(C A^N B \right) = \frac{\partial C}{\partial \theta_i} A^N B + C \frac{\partial A^N}{\partial \theta_i} B + C A^N \frac{\partial B}{\partial \theta_i} \tag{82}$$

Product in Fisher's information matrix

Product

$$\frac{\partial J_N}{\partial \theta_i}(\theta;Y_N)\frac{\partial J_N}{\partial \theta_j}(\theta;Y_N) = \frac{1}{4}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}\right) \tag{83}$$

$$+\left(\frac{\partial m_Y}{\partial \theta_i}\right)^T P_Y^{-1}(Y_N - m_Y)(Y_N - m_Y)^T P_Y^{-1}\frac{\partial m_Y}{\partial \theta_j} \tag{84}$$

$$+\frac{1}{4}(Y_N - m_Y)^T P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}(Y_N - m_Y)(Y_N - m_Y)^T P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}P_Y^{-1}(Y_N - m_Y) \tag{85}$$

$$-\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)(Y_N - m_Y)^T P_Y^{-1}\frac{\partial m_Y}{\partial \theta_j} \tag{86}$$

$$-\frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)(Y_N - m_Y)^T P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}P_Y^{-1}(Y_N - m_Y) \tag{87}$$

$$+(Y_N - m_Y)^T P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}(Y_N - m_Y)(Y_N - m_Y)^T P_Y^{-1}\frac{\partial m_Y}{\partial \theta_j} \tag{88}$$

Expectations of quadratic, cubic, and quartic forms

Let $e \in \mathbb{R}^n$ be a zero-mean normally distributed variable, i.e., $e \sim N(0, R)$, let $b \in \mathbb{R}^n$, and let $A, B \in \mathbb{R}^{n \times n}$ be symmetric matrices. Then,

1. $\mathbb{E}[e^T A e] = \mathrm{Tr}(AR)$,
2. $\mathbb{E}[e^T A e e^T b] = 0$, and
3. $\mathbb{E}[e^T A e e^T B e] = \mathrm{Tr}(AR)\,\mathrm{Tr}(BR) + 2\,\mathrm{Tr}(ARBR)$.

From Isserlis' theorem[2]

$$\mathbb{E}[e_i e_j] = R_{ij}, \tag{89}$$

$$\mathbb{E}[e_i e_j e_k] = 0, \tag{90}$$

$$\mathbb{E}[e_i e_j e_k e_l] = \mathbb{E}[e_i e_j]\mathbb{E}[e_k e_l] + \mathbb{E}[e_i e_k]\mathbb{E}[e_j e_l] + \mathbb{E}[e_i e_l]\mathbb{E}[e_j e_k] \tag{91}$$

$$= R_{ij}R_{kl} + R_{ik}R_{jl} + R_{il}R_{jk} \tag{92}$$

[2] J. V. Michalowicz, J. M. Nicols, F. Bucholtz, C. C. Olson, 2009. An Isserlis' theorem for mixed Gaussian variables: Application to the auto-bispectral density. Journal of Statistical Physics 136, pp. 89–102. DOI: 10.1007/s10955-009-9768-3

Proof of expectation of quadratic and form

Expand the quadratic form

$$\mathbb{E}[e^T Ae] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} e_i A_{ij} e_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[e_i e_j] A_{ij} \tag{93}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij} A_{ij} = \mathrm{Tr}(AR^T) = \mathrm{Tr}(AR) \tag{94}$$

The last equality follows from the symmetry of the covariance matrix, R.

Expand the cubic form

$$\mathbb{E}[e^T Aee^T b] = \mathbb{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n} e_i A_{ij} e_j\right)\left(\sum_{k=1}^{n} e_k b_k\right)\right] \tag{95}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \mathbb{E}[e_i e_j e_k] A_{ij} b_k = 0 \tag{96}$$

Proof of expectation of quartic form

Expand the quartic form

$$\mathbb{E}[e^T A e e^T B e] = \mathbb{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n} e_i A_{ij} e_j\right)\left(\sum_{k=1}^{n}\sum_{l=1}^{n} e_k B_{kl} e_l\right)\right] \tag{97}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \mathbb{E}[e_i e_j e_k e_l] A_{ij} B_{kl} \tag{98}$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij} A_{ij}\right)\left(\sum_{k=1}^{n}\sum_{l=1}^{n} R_{kl} B_{kl}\right) \tag{99}$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \left(A_{ij} R_{ik} B_{kl} R_{jl} + A_{ij} R_{jk} B_{kl} R_{il}\right) \tag{100}$$

$$= \mathrm{Tr}(AR^T)\,\mathrm{Tr}(BR^T) + \mathrm{Tr}(A(RBR^T)^T) + \mathrm{Tr}(ARBR^T) \tag{101}$$

$$= \mathrm{Tr}(AR)\,\mathrm{Tr}(BR) + 2\,\mathrm{Tr}(ARBR) \tag{102}$$

In the last equality, we have used the symmetry of A, B, and R.

Fisher's information matrix

Fisher's information matrix

$$F_{ij} = \mathbb{E}\left[\frac{\partial J_N}{\partial \theta_i}(\theta; Y_N)\frac{\partial J_N}{\partial \theta_j}(\theta; Y_N)\right] \tag{103}$$

$$= \frac{1}{4}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}\right) + \left(\frac{\partial m_Y}{\partial \theta_i}\right)^T P_Y^{-1} P_Y P_Y^{-1}\frac{\partial m_Y}{\partial \theta_j} \tag{104}$$

$$+ \frac{1}{4}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}P_Y\right)\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}P_Y^{-1}P_Y\right) \tag{105}$$

$$+ \frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}P_Y P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}P_Y^{-1}P_Y\right) \tag{106}$$

$$- \frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}P_Y^{-1}P_Y\right) \tag{107}$$

Fisher's information matrix

Simplify expression

$$F_{ij} = \frac{1}{4}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}\right) + \left(\frac{\partial m_Y}{\partial \theta_i}\right)^T P_Y^{-1}\frac{\partial m_Y}{\partial \theta_j} \tag{108}$$

$$+ \frac{1}{4}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}\right) + \frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}\right) \tag{109}$$

$$- \frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}\right)\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}\right) \tag{110}$$

Simplify further

$$\boxed{F_{ij} = \left(\frac{\partial m_Y}{\partial \theta_i}\right)^T P_Y^{-1}\frac{\partial m_Y}{\partial \theta_j} + \frac{1}{2}\operatorname{Tr}\left(P_Y^{-1}\frac{\partial P_Y}{\partial \theta_i}P_Y^{-1}\frac{\partial P_Y}{\partial \theta_j}\right)} \tag{111}$$

Recursive parameter estimation (extended Kalman filter)

Augmented system: Parameter model

Parameter model

$$\theta_{t+1} = \theta_t + H\eta_t, \qquad \eta_t \sim N(0, I) \tag{112}$$

Augmented system

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t + G(\theta)v_t, \qquad v_t \sim N(0, I), \tag{113}$$

$$\theta_{t+1} = \theta_t + H\eta_t, \qquad \eta_t \sim N(0, I), \tag{114}$$

$$y_t = C(\theta)x_t + D(\theta)u_t + F(\theta)e_t, \qquad e_t \sim N(0, I) \tag{115}$$

Compact notation

$$\bar{x}_t = \begin{bmatrix} x_t \\ \theta_t \end{bmatrix}, \qquad \bar{v}_t = \begin{bmatrix} v_t \\ \eta_t \end{bmatrix}, \tag{116}$$

$$\bar{x}_{t+1} = f(\bar{x}_t, u_t, \bar{v}_t), \quad f(\bar{x}_t, u_t, \bar{v}_t) = \begin{bmatrix} A(\theta)x_t + B(\theta)u_t + G(\theta)v_t \\ \theta_t + H\eta_t \end{bmatrix}, \tag{117}$$

$$y_t = g(\bar{x}_t, u_t, e_t), \quad g(\bar{x}_t, u_t, e_t) = C(\theta)x_t + D(\theta)u_t + F(\theta)e_t \tag{118}$$

Linearized system

Linearize around $\bar{x}^*$, u^*, $\bar{v}^* = 0$, and $e^* = 0$

$$f(\bar{x}_t, u_t, \bar{v}_t) \approx f(\bar{x}^*, u^*, \bar{v}^*) + \frac{\partial f}{\partial \bar{x}}(\bar{x}_t - \bar{x}^*) + \frac{\partial f}{\partial u}(u_t - u^*) + \frac{\partial f}{\partial \bar{v}}(\bar{v}_t - \bar{v}^*), \tag{119}$$

$$g(\bar{x}_t, u_t, e_t) \approx g(\bar{x}^*, u^*, e^*) + \frac{\partial g}{\partial \bar{x}}(\bar{x}_t - \bar{x}^*) + \frac{\partial g}{\partial u}(u_t - u^*) + \frac{\partial g}{\partial e}(e_t - e^*) \tag{120}$$

The Jacobian matrices are evaluated in the linearization point, e.g., $\frac{\partial f}{\partial \bar{x}} = \frac{\partial f}{\partial \bar{x}}(\bar{x}^*, u^*, \bar{v}^*)$

Offsets

$$d^* = f(\bar{x}^*, u^*, \bar{v}^*) - \frac{\partial f}{\partial \bar{x}}\bar{x}^* - \frac{\partial f}{\partial u}u^* - \frac{\partial f}{\partial \bar{v}}\bar{v}^*, \tag{121}$$

$$r^* = g(\bar{x}^*, u^*, e^*) - \frac{\partial g}{\partial \bar{x}}\bar{x}^* - \frac{\partial g}{\partial u}u^* - \frac{\partial g}{\partial e}e^* \tag{122}$$

Jacobians

Jacobians

$$\bar{A} = \frac{\partial f}{\partial \bar{x}} = \begin{bmatrix} A & S \\ & I \end{bmatrix}, \quad \bar{B} = \frac{\partial f}{\partial u} = \begin{bmatrix} B \\ \end{bmatrix}, \quad \bar{G} = \frac{\partial f}{\partial \bar{v}} = \begin{bmatrix} G & \\ & H \end{bmatrix}, \tag{123}$$

$$\bar{C} = \frac{\partial g}{\partial \bar{x}} = \begin{bmatrix} C & M \end{bmatrix}, \quad \bar{D} = \frac{\partial g}{\partial u} = D, \quad \bar{F} = \frac{\partial g}{\partial e} = F \tag{124}$$

Jacobians wrt. parameters

$$S_{\cdot i} = \frac{\partial A}{\partial \theta_i} x^* + \frac{\partial B}{\partial \theta_i} u^* + \frac{\partial G}{\partial \theta_i} v^*, \tag{125}$$

$$M_{\cdot i} = \frac{\partial C}{\partial \theta_i} x^* + \frac{\partial D}{\partial \theta_i} u^* + \frac{\partial F}{\partial \theta_i} e^* \tag{126}$$

All matrices are evaluated in θ^*

Linearized system

Linearized system

$$\bar{x}_{t+1} = \bar{A}\bar{x}_t + \bar{B}u_t + \bar{G}\bar{v}_t + d_t, \tag{127}$$

$$y_t = \bar{C}\bar{x}_t + \bar{D}u_t + \bar{F}e_t + r_t \tag{128}$$

Offset models

$$d_{t+1} = d_t, \qquad d_0 \sim N(d^*, 0), \tag{129}$$

$$r_{t+1} = r_t, \qquad r_0 \sim N(r^*, 0) \tag{130}$$

Further augmented system

$$\begin{bmatrix} \bar{x}_{t+1} \\ d_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} \bar{A} & I & \\ & I & \\ & & I \end{bmatrix} \begin{bmatrix} \bar{x}_t \\ d_t \\ r_t \end{bmatrix} + \begin{bmatrix} \bar{B} \\ \\ \end{bmatrix} u_t + \begin{bmatrix} \bar{G} \\ \\ \end{bmatrix} \bar{v}_t, \tag{131}$$

$$y_t = \begin{bmatrix} \bar{C} & 0 & I \end{bmatrix} \begin{bmatrix} \bar{x}_t \\ d_t \\ r_t \end{bmatrix} + \bar{D}u_t + \bar{F}e_t \tag{132}$$

Linearized system: Compact notation

Compact notation

$$\tilde{A} = \begin{bmatrix} \bar{A} & I & \\ & I & \\ & & I \end{bmatrix}, \qquad \tilde{B} = \begin{bmatrix} \bar{B} \\ \\ \\ \end{bmatrix}, \qquad \tilde{G} = \begin{bmatrix} \bar{G} \\ \\ \\ \end{bmatrix}, \tag{133}$$

$$\tilde{C} = \begin{bmatrix} \bar{C} & 0 & I \end{bmatrix}, \qquad \tilde{D} = \bar{D}, \qquad \tilde{F} = \bar{F} \tag{134}$$

$$X_{t+1} = \tilde{A} X_t + \tilde{B} u_t + \tilde{G} \bar{v}_t, \tag{135}$$

$$y_t = \tilde{C} X_t + \tilde{D} u_t + \tilde{F} e_t \tag{136}$$

Linearized Kalman filter

Measurement update

$$\hat{X}_{t|t} = \hat{x}_{t|t-1} + \tilde{\kappa}_t(y_t - (\tilde{C}\hat{X}_{t|t-1} + \tilde{D}u_t)), \quad (137)$$

$$\tilde{\kappa}_t = \tilde{P}_{t|t-1}\tilde{C}^T(\tilde{C}\tilde{P}_{t|t-1}\tilde{C}^T + \tilde{F}\tilde{F}^T)^{-1}, \quad (138)$$

$$\tilde{P}_{t|t} = \tilde{P}_{t|t-1} - \tilde{\kappa}_t\tilde{C}\tilde{P}_{t|t-1} \quad (139)$$

Time update

$$\hat{X}_{t+1|t} = \tilde{A}\hat{X}_{t|t} + \tilde{B}u_t, \quad (140)$$

$$\tilde{P}_{t+1|t} = \tilde{A}\tilde{P}_{t|t}\tilde{A}^T + \tilde{G}\tilde{G}^T \quad (141)$$

Initial distribution

$$\hat{X}_{0|-1} = \begin{bmatrix} \hat{\bar{x}}_{0|-1} \\ d^* \\ r^* \end{bmatrix}, \qquad \tilde{P}_{0|-1} = \begin{bmatrix} P_{\bar{x}\bar{x},0|-1} & & \\ & & \\ & & \end{bmatrix} \quad (142)$$

Linearized Kalman filter: Time update

Mean update

$$\hat{\bar{x}}_{t+1|t} = \bar{A}\hat{\bar{x}}_{t|t} + \bar{B}u_t + \hat{d}_{t|t}, \quad (143)$$

$$\hat{d}_{t+1|t} = \hat{d}_{t|t}, \quad (144)$$

$$\hat{r}_{t+1|t} = \hat{r}_{t|t} \quad (145)$$

Covariance update

$$P_{\bar{x}\bar{x},t+1|t} = \bar{A}P_{\bar{x}\bar{x},t|t}\bar{A}^T + \bar{A}P_{\bar{x}d,t|t} + P_{d\bar{x},t|t}\bar{A}^T + P_{dd,t|t} + \bar{G}\bar{G}^T, \quad (146)$$

$$P_{\bar{x}d,t+1|t} = \bar{A}P_{\bar{x}d,t|t} + P_{dd,t|t}, \quad (147)$$

$$P_{\bar{x}r,t+1|t} = \bar{A}P_{\bar{x}r,t|t} + P_{dr,t|t}, \quad (148)$$

$$P_{dd,t+1|t} = P_{dd,t+1|t}, \quad (149)$$

$$P_{dr,t+1|t} = P_{dr,t|t}, \quad (150)$$

$$P_{rr,t|t} = P_{rr,t|t} \quad (151)$$

Linearized Kalman filter: Measurement update

Measurement update

$$\begin{bmatrix} \hat{\bar{x}}_{t|t} \\ \hat{d}_{t|t} \\ \hat{r}_{t|t} \end{bmatrix} = \begin{bmatrix} \hat{\bar{x}}_{t|t-1} \\ \hat{d}_{t|t-1} \\ \hat{r}_{t|t-1} \end{bmatrix} + \begin{bmatrix} \kappa_{\bar{x},t} \\ \kappa_{d,t} \\ \kappa_{r,t} \end{bmatrix} \left(y_t - \left(\begin{bmatrix} \bar{C} & 0 & I \end{bmatrix} \begin{bmatrix} \hat{\bar{x}}_{t|t-1} \\ \hat{d}_{t|t-1} \\ \hat{r}_{t|t-1} \end{bmatrix} + \bar{D} u_t \right) \right), \quad (152)$$

$$\begin{bmatrix} \kappa_{\bar{x},t} \\ \kappa_{d,t} \\ \kappa_{r,t} \end{bmatrix} = \begin{bmatrix} P_{\bar{x}\bar{x},t|t-1} & P_{\bar{x}d,t|t-1} & P_{\bar{x}r,t|t-1} \\ P_{d\bar{x},t|t-1} & P_{dd,t|t-1} & P_{dr,t|t-1} \\ P_{r\bar{x},t|t-1} & P_{rd,t|t-1} & P_{rr,t|t-1} \end{bmatrix} \begin{bmatrix} \bar{C}^T \\ 0 \\ I \end{bmatrix} \quad (153)$$

$$\left(\begin{bmatrix} \bar{C} & 0 & I \end{bmatrix} \begin{bmatrix} P_{\bar{x}\bar{x},t|t-1} & P_{\bar{x}d,t|t-1} & P_{\bar{x}r,t|t-1} \\ P_{d\bar{x},t|t-1} & P_{dd,t|t-1} & P_{dr,t|t-1} \\ P_{r\bar{x},t|t-1} & P_{rd,t|t-1} & P_{rr,t|t-1} \end{bmatrix} \begin{bmatrix} \bar{C}^T \\ 0 \\ I \end{bmatrix} + \bar{F}\bar{F}^T \right)^{-1}, \quad (154)$$

$$\begin{bmatrix} P_{\bar{x}\bar{x},t|t} & P_{\bar{x}d,t|t} & P_{\bar{x}r,t|t} \\ P_{d\bar{x},t|t} & P_{dd,t|t} & P_{dr,t|t} \\ P_{r\bar{x},t|t} & P_{rd,t|t} & P_{rr,t|t} \end{bmatrix} = \begin{bmatrix} P_{\bar{x}\bar{x},t|t-1} & P_{\bar{x}d,t|t-1} & P_{\bar{x}r,t|t-1} \\ P_{d\bar{x},t|t-1} & P_{dd,t|t-1} & P_{dr,t|t-1} \\ P_{r\bar{x},t|t-1} & P_{rd,t|t-1} & P_{rr,t|t-1} \end{bmatrix} \quad (155)$$

$$- \begin{bmatrix} \kappa_{\bar{x},t} \\ \kappa_{d,t} \\ \kappa_{r,t} \end{bmatrix} \begin{bmatrix} \bar{C} & 0 & I \end{bmatrix} \begin{bmatrix} P_{\bar{x}\bar{x},t|t-1} & P_{\bar{x}d,t|t-1} & P_{\bar{x}r,t|t-1} \\ P_{d\bar{x},t|t-1} & P_{dd,t|t-1} & P_{dr,t|t-1} \\ P_{r\bar{x},t|t-1} & P_{rd,t|t-1} & P_{rr,t|t-1} \end{bmatrix} \quad (156)$$

Linearized Kalman filter: Measurement update

Output prediction mean and covariance

$$\hat{y}_{t|t-1} = \bar{C}\hat{\bar{x}}_{t|t-1} + \bar{D}u_t + \hat{r}_{t|t-1}, \tag{157}$$

$$P_{y,t|t-1} = \bar{C}P_{\bar{x}\bar{x},t|t-1}\bar{C}^T + \bar{C}P_{\bar{x}r,t|t-1} + P_{r\bar{x},t|t-1}\bar{C}^T + P_{rr,t|t-1} + \bar{F}\bar{F}^T \tag{158}$$

Measurement update

$$\hat{\bar{x}}_{t|t} = \hat{\bar{x}}_{t|t-1} + \kappa_{\bar{x},t}(y_t - \hat{y}_{t|t-1}), \tag{159}$$

$$\hat{d}_{t|t} = \hat{d}_{t|t-1} + \kappa_{d,t}(y_t - \hat{y}_{t|t-1}), \tag{160}$$

$$\hat{r}_{t|t} = \hat{r}_{t|t-1} + \kappa_{r,t}(y_t - \hat{y}_{t|t-1}), \tag{161}$$

$$\kappa_{\bar{x},t} = \left(P_{\bar{x}\bar{x},t|t-1}\bar{C}^T + P_{\bar{x}r,t|t-1}\right)P_{y,t|t-1}^{-1}, \tag{162}$$

$$\kappa_{\bar{d},t} = \left(P_{d\bar{x},t|t-1}\bar{C}^T + P_{dr,t|t-1}\right)P_{y,t|t-1}^{-1}, \tag{163}$$

$$\kappa_{\bar{r},t} = \left(P_{r\bar{x},t|t-1}\bar{C}^T + P_{rr,t|t-1}\right)P_{y,t|t-1}^{-1}, \tag{164}$$

$$P_{\bar{x}\bar{x},t|t} = P_{\bar{x}\bar{x},t|t-1} - \kappa_{\bar{x},t}(\bar{C}P_{\bar{x}\bar{x},t|t-1} + P_{r\bar{x},t|t-1}), \quad P_{\bar{x}d,t|t} = -\kappa_{\bar{x},t}(\bar{C}P_{\bar{x}d,t|t-1} + P_{rd,t|t-1}), \tag{165}$$

$$P_{dd,t|t} = P_{dd,t|t-1} - \kappa_{d,t}(\bar{C}P_{\bar{x}d,t|t-1} + P_{rd,t|t-1}), \quad P_{\bar{x}r,t|t} = -\kappa_{\bar{x},t}(\bar{C}P_{\bar{x}r,t|t-1} + P_{rr,t|t-1}), \tag{166}$$

$$P_{rr,t|t} = P_{rr,t|t-1} - \kappa_{r,t}(\bar{C}P_{\bar{x}r,t|t-1} + P_{rr,t|t-1}), \quad P_{dr,t|t} = -\kappa_{d,t}(\bar{C}P_{\bar{x}r,t|t-1} + P_{rr,t|t-1}) \tag{167}$$

Linearized Kalman filter

Trivial estimates

$$\hat{d}_{t|t} = \hat{d}_{t|t-1} = d^*, \qquad \hat{r}_{t|t} = \hat{r}_{t|t-1} = r^*, \tag{168}$$

$$P_{\bar{x}d,t|t} = P_{\bar{x}d,t|t-1} = 0, \qquad P_{\bar{x}r,t|t} = P_{\bar{x}r,t|t-1} = 0, \tag{169}$$

$$P_{dd,t|t} = P_{dd,t|t-1} = 0, \qquad P_{dr,t|t} = P_{dr,t|t-1} = 0, \tag{170}$$

$$P_{rr,t|t} = P_{rr,t|t-1} = 0, \tag{171}$$

Linearized Kalman filter

Measurement update

$$\hat{y}_{t|t-1} = \bar{C}\hat{\bar{x}}_{t|t-1} + \bar{D}u_t + r^*, \tag{172}$$

$$\hat{\bar{x}}_{t|t} = \hat{\bar{x}}_{t|t-1} + \kappa_{\bar{x},t}(y_t - \hat{y}_{t|t-1}), \tag{173}$$

$$\kappa_{\bar{x},t} = P_{\bar{x}\bar{x},t|t-1}\bar{C}^T(\bar{C}P_{\bar{x}\bar{x},t|t-1}\bar{C}^T + \bar{F}\bar{F}^T)^{-1}, \tag{174}$$

$$P_{\bar{x}\bar{x},t|t} = P_{\bar{x}\bar{x},t|t-1} - \kappa_{\bar{x},t}\bar{C}P_{\bar{x}\bar{x},t|t-1} \tag{175}$$

Time update

$$\hat{\bar{x}}_{t+1|t} = \bar{A}\hat{\bar{x}}_{t|t} + \bar{B}u_t + d^*, \tag{176}$$

$$P_{\bar{x}\bar{x},t+1|t} = \bar{A}P_{\bar{x}\bar{x},t|t}\bar{A}^T + \bar{G}\bar{G}^T \tag{177}$$

Linearized nonlinear equations

$$\hat{\bar{x}}_{t+1|t} = f(\bar{x}^*, u^*, \bar{v}^*) + \frac{\partial f}{\partial \bar{x}}(\hat{\bar{x}}_{t|t} - \bar{x}^*) + \frac{\partial f}{\partial u}(u_t - u^*), \tag{178}$$

$$\hat{y}_{t|t-1} = g(\bar{x}^*, u^*, e^*) + \frac{\partial g}{\partial \bar{x}}(\hat{\bar{x}}_{t|t-1} - \bar{x}^*) + \frac{\partial g}{\partial u}(u_t - u^*) \tag{179}$$

Extended Kalman filter

Linearization point (time update)

$$\bar{x}^* = \hat{\bar{x}}_{t|t} = \begin{bmatrix} \hat{x}_{t|t} \\ \hat{\theta}_{t|t} \end{bmatrix}, \qquad u^* = u_t, \qquad \bar{v}^* = 0 \tag{180}$$

Transform back to nonlinear equations

$$\begin{bmatrix} \hat{x}_{t+1|t} \\ \hat{\theta}_{t+1|t} \end{bmatrix} = \hat{\bar{x}}_{t+1|t} = f(\hat{\bar{x}}_{t|t}, u_t, 0) = \begin{bmatrix} A(\hat{\theta}_{t|t})\hat{x}_{t|t} + B(\hat{\theta}_{t|t})u_t, \\ \hat{\theta}_{t|t} \end{bmatrix} \tag{181}$$

Linearization point (measurement update)

$$\bar{x}^* = \hat{\bar{x}}_{t|t-1} = \begin{bmatrix} \hat{x}_{t|t-1} \\ \hat{\theta}_{t|t-1} \end{bmatrix}, \qquad u^* = u_t, \qquad e^* = 0 \tag{182}$$

Transform back to nonlinear equations

$$\hat{y}_{t|t-1} = g(\hat{\bar{x}}_{t|t-1}, u_t, 0) = C(\hat{\theta}_{t|t-1})\hat{x}_{t|t-1} + D(\hat{\theta}_{t|t-1})u_t \tag{183}$$

Extended Kalman filter: Measurement update

Predicted output

$$\boxed{\hat{y}_{t|t-1} = C\hat{x}_{t|t-1} + Du_t} \tag{184}$$

Filtered estimates

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \kappa_{x,t}(y_t - \hat{y}_{t|t-1}), \tag{185}$$

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + \kappa_{\theta,t}(y_t - \hat{y}_{t|t-1}) \tag{186}$$

Kalman gains

$$\kappa_{x,t} = \left(P_{xx,t|t-1}C^T + P_{x\theta,t|t-1}M^T\right)\left(CP_{xx,t|t-1}C^T + CP_{x\theta,t|t-1}M^T\right. \tag{187}$$

$$\left. + MP_{\theta x,t|t-1}C^T + MP_{\theta\theta,t|t-1}M^T + FF^T\right)^{-1}, \tag{188}$$

$$\kappa_{\theta,t} = \left(P_{\theta x,t|t-1}C^T + P_{\theta\theta,t|t-1}M^T\right)\left(CP_{xx,t|t-1}C^T + CP_{x\theta,t|t-1}M^T\right. \tag{189}$$

$$\left. + MP_{\theta x,t|t-1}C^T + MP_{\theta\theta,t|t-1}M^T + FF^T\right)^{-1}, \tag{190}$$

Covariances

$$P_{xx,t|t} = P_{xx,t|t-1} - \kappa_{x,t}\left(CP_{xx,t|t-1} + MP_{\theta x,t|t-1}\right), \tag{191}$$

$$P_{x\theta,t|t} = P_{x\theta,t|t-1} - \kappa_{x,t}\left(CP_{x\theta,t|t-1} + MP_{\theta\theta,t|t-1}\right), \tag{192}$$

$$P_{\theta\theta,t|t} = P_{\theta\theta,t|t-1} - \kappa_{\theta,t}\left(CP_{x\theta,t|t-1} + MP_{\theta\theta,t|t-1}\right) \tag{193}$$

C, D, and F are evaluated in $\hat{\theta}_{t|t-1}$ and M is evaluated in $\hat{x}_{t|t-1}$ and $\hat{\theta}_{t|t-1}$

Extended Kalman filter: Time update

Predicted estimates

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \tag{194}$$

$$\hat{\theta}_{t+1|t} = \hat{\theta}_{t|t}, \tag{195}$$

Covariances

$$P_{xx,t+1|t} = AP_{xx,t|t}A^T + AP_{x\theta,t|t}S^T \tag{196}$$

$$+ SP_{\theta x,t|t}A^T + SP_{\theta\theta,t|t}S^T + GG^T, \tag{197}$$

$$P_{x\theta,t+1|t} = AP_{x\theta,t|t} + SP_{\theta\theta,t|t}, \tag{198}$$

$$P_{\theta\theta,t+1|t} = P_{\theta\theta,t|t} + HH^T \tag{199}$$

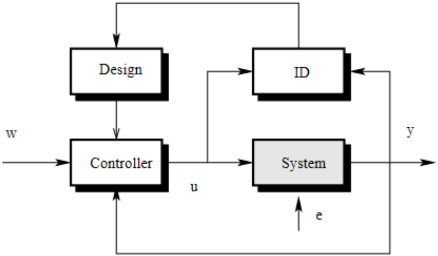
A, B, and G are evaluated in $\hat{\theta}_{t|t}$ and S is evaluated in $\hat{x}_{t|t}$ and $\hat{\theta}_{t|t}$

Adaptive control

Self-tuning adaptive control

Self-tuning methods: Combine identification, design, and control

Furthermore, assume that the certainty equivalence principle holds



Certainty equivalence principle

Replace the true value by an estimate

$$\theta \to \hat{\theta} \tag{200}$$

Control law

$$u_t = -Lx_t \quad \to \quad u_t = -L\hat{x}_t \tag{201}$$

Basis for separation theorem between the design of Kalman filters and LQR

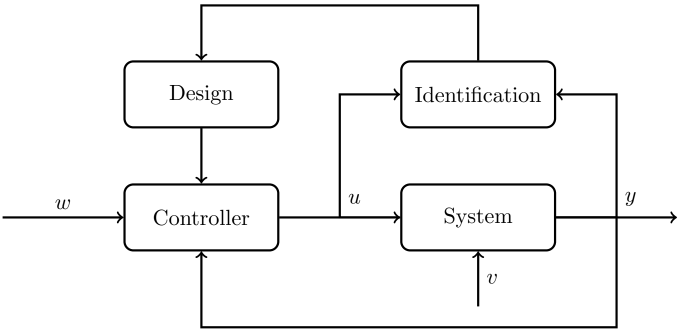
Control law in adaptive control

$$u_t = -Lx_t \quad \to \quad u_t = -\hat{L}\hat{x}_t \tag{202}$$

The principle does not guarantee optimality, but is assumed for convenience

Explicit self-tuning controllers: Pseudoalgorithm

1. Use y_t to compute filtered estimate of states, $\hat{x}_{t|t}$
2. Use y_t or $\{y_{t-j}\}_{j=0}^{N}$ to estimate parameters, $\hat{\theta}_{t|t}$
3. Update controller feedback gain
4. Use $\hat{x}_{t|t}$ to compute the manipulated input, u_t



Questions

Questions?

Appendix

Derivatives

Negative log-likelihood function: Derivatives

First-order derivative

$$\frac{\partial J_N}{\partial \theta_i} = \frac{1}{2}\sum_{t=0}^{N}\left(\mathrm{Tr}\left(P_{y,t|t-1}\frac{\partial P_{y,t|t-1}}{\partial \theta_i}\right) + \left(\frac{\partial \epsilon_t}{\partial \theta_i}\right)^T P_{y,t|t-1}^{-1}\epsilon_t \right. \tag{203}$$

$$\left. - \epsilon_t^T P_{y,t|t-1}^{-1}\frac{\partial P_{y,t|t-1}}{\partial \theta_i}P_{y,t|t-1}^{-1}\epsilon_t + \epsilon_t^T P_{y,t|t-1}^{-1}\frac{\partial \epsilon_t}{\partial \theta_i}\right) \tag{204}$$

$$= \sum_{t=0}^{N}\left(\frac{1}{2}\mathrm{Tr}\left(P_{y,t|t-1}\frac{\partial P_{y,t|t-1}}{\partial \theta_i}\right) + \left(\frac{\partial \epsilon_t}{\partial \theta_i}\right)^T P_{y,t|t-1}^{-1}\epsilon_t \right. \tag{205}$$

$$\left. - \frac{1}{2}\epsilon_t^T P_{y,t|t-1}^{-1}\frac{\partial P_{y,t|t-1}}{\partial \theta_i}P_{y,t|t-1}^{-1}\epsilon_t\right) \tag{206}$$