State Estimation of Linear and Nonlinear Dynamic Systems

Part II: Observability and Stability

James B. Rawlings and Fernando V. Lima

Department of Chemical and Biological Engineering
University of Wisconsin–Madison

AICES Regional School
RWTH Aachen
March 10, 2008

Outline

Observability Property

- Consider the linear system (A, C) with n measurements $Y(n-1)$

$$\begin{aligned} x(k+1) &= Ax(k) \\ y(k) &= Cx(k) \\ Y(n-1) &= \{y(0), y(1), \dots, y(n-1)\} \end{aligned}$$

in which $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{p\times n}$

Definition (Observability)

(A, C) is *observable* if these n measurements *uniquely* determine the system's initial state $x(0)$.

- Observability is a property of the deterministic model equations

Observability Matrix

- For the n measurements, the system model gives

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n-1) \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{\mathcal{O}} x(0)$$

in which $\mathcal{O} \in \mathbb{R}^{np\times n}$ is the Observability Matrix

- (A, C) is *observable* if and only if $\mathrm{rank}(\mathcal{O}) = n$

Observability and Canonical Forms

- For the linear system

$$\begin{bmatrix} \\ x \\ \\ \end{bmatrix}_{k+1} = \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} \\ x \\ \\ \end{bmatrix}_k$$

$$y_k = \begin{bmatrix} & C & \end{bmatrix} \begin{bmatrix} \\ x \\ \\ \end{bmatrix}_k$$

- Find a similarity transformation T:

$$\begin{bmatrix} z_1 \\ \hline z_2 \end{bmatrix} = \begin{bmatrix} T_1 \\ \hline T_2 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} \Rightarrow \widetilde{A} = TAT^{-1}, \ \widetilde{C} = CT^{-1}$$

Canonical Forms: Observable and Unobservable Modes

- So the transformed system has the following canonical form

$$\begin{bmatrix} z_1 \\ \hline z_2 \end{bmatrix}_{k+1} = \underbrace{\left[\begin{array}{c|c} \widetilde{A}_{11} & 0 \\ \hline \widetilde{A}_{21} & \widetilde{A}_{22} \end{array}\right]}_{\widetilde{A}} \begin{bmatrix} z_1 \\ \hline z_2 \end{bmatrix}_k$$

$$y_k = \underbrace{\begin{bmatrix} \widetilde{C}_1 & 0 \end{bmatrix}}_{\widetilde{C}} \begin{bmatrix} z_1 \\ \hline z_2 \end{bmatrix}_k$$

where $(\widetilde{A}_{11}, \widetilde{C}_1)$ is observable

- In this structure
  - z_1 are the observable modes
  - z_2 are the unobservable modes
    - however, the system is still *detectable* if $\lambda(\widetilde{A}_{22}) \leq 1$

Detectability Property

Definition (Detectability)

A linear system is *detectable* when all the unobservable modes are stable

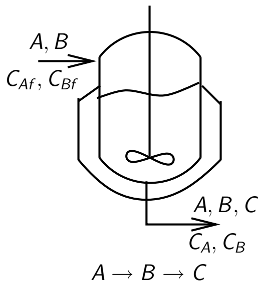
- This property is important for partially observable systems
- An observable system is also detectable

The property of detectability is important for control because one may successfully design a control system for an unobservable but detectable system so as to estimate and control the unstable modes.

Advanced Process Control.
W.H. Ray, 1981.

Illustrative Example: CSTR Reactor (Ray, 1981)

- Reaction: $A \rightarrow B \rightarrow C$
- Controlled variables: C_A, C_B
- Manipulated variables: C_{Af}, C_{Bf}
- State variables: dimensionless concentrations - $x_1(C_A)$, $x_2(C_B)$
- Input variables: dimensionless feeds - $u_1(C_{Af})$, $u_2(C_{Bf})$



$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} -A_1 & 0 \\ A_2 & -A_3 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{B} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$y = Cx$$

in which $A_1, A_2, A_3 \geq 0$ are constants

Solution for 2 Cases

1. Only x_1 is measured
   - $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$

$$\mathcal{O} = \begin{bmatrix} 1 & 0 \\ -A_1 & 0 \end{bmatrix} \Rightarrow \mathrm{rank}(\mathcal{O}) = 1$$

   - Unobservable system!
   - But it is still detectable: $\lambda(A) < 0$ (continuous system)
2. Only x_2 is measured
   - $C = \begin{bmatrix} 0 & 1 \end{bmatrix}$

$$\mathcal{O} = \begin{bmatrix} 0 & 1 \\ A_2 & -A_3 \end{bmatrix} \Rightarrow \mathrm{rank}(\mathcal{O}) = 2$$

   - Completely observable system!

Example Remarks

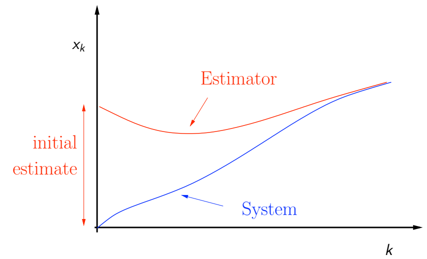
$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C} \qquad r_1 = k_1 C_A \quad r_2 = k_2 C_B$$

1. Physical Reasons
   - C_B depends on both C_A and C_B
   - C_A is independent of C_B
2. Consequences
   - By measuring $x_2(C_B)$ and knowing $\mathbf{u}$, $x_1(C_A)$ can be determined
   - By measuring $x_1(C_A)$ and knowing $\mathbf{u}$, $x_2(C_B)$ can take any value

Deterministic Stability of State Estimator

Definition (Asymptotic Stability of the State Estimator)

The estimator is asymptotically stable in sense of an observer if the estimator is able to "recover" from the incorrect initial value of state as data with no measurement noise are collected.



For example:

- assume an incorrect initial estimate
- the estimator converges (asymptotically) to the correct value

Litmus Test

State estimation: probabilistic optimality versus stability

Kalman filtering was first publicly presented (to somewhat more than polite applause) on April 1, 1959. But please note: Kalman filtering is not a triumph of applied probability: the theory has only a slight inheritance from probability theory while it has become an important pillar of system theory.
R. Kalman, 1994

As Kalman has often stressed the major contribution of his work is not perhaps the actual filter algorithm, elegant and useful as it no doubt is, but the proof that under certain technical conditions called "controllability" and "observability," the optimum filter is "stable" in the sense that the effects of initial errors and round-off and other computational errors will die out asymptotically.
T. Kailath, 1974

State Estimation: Optimality does not Ensure Stability

- For the linear system

$$x(k+1) = Ax(k)$$
$$y(k) = Cx(k)$$

- Consider the case when $A = I, C = 0$
  - Optimal estimate is $\widehat{x}(k) = \overline{x}(0)$ (for a chosen initial condition)
  - Estimator does not converge to the true state $x(0)$
    - unless we have luckily chosen $\overline{x}(0) = x(0)$
  - Unobservable (and undetectable) system

Cost Convergence and Stability Lemmas

Lemma (Convergence of estimator cost)

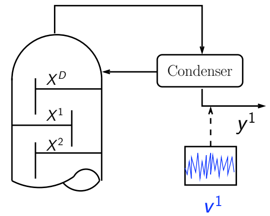
Given noise-free measurements $Y(T)$, the optimal estimator cost $\Phi^0(Y(T))$ converges as T increases, regardless of the system observability.

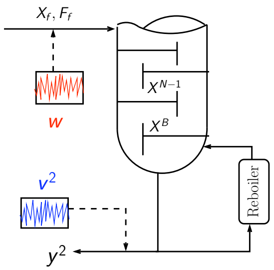
Lemma (Estimator stability - convergence to the true state)

For (A, C) observable and $Q, R > 0$ (positive definite), the optimal linear state estimator is asymptotically stable

$$\widehat{x}(T) \rightarrow x(T) \quad \text{as } T \rightarrow \infty$$

Obtaining Q and R from Data



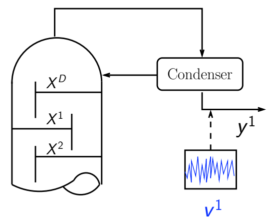


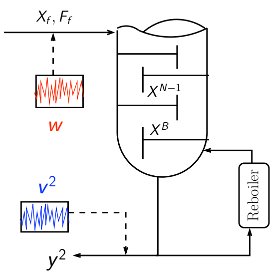
- Model discretized with $t_k = k\Delta t$:

$$x_{k+1} = f(x_k, u_k) + g(x_k, u_k)w_k$$
$$\begin{bmatrix} y^1 \\ y^2 \end{bmatrix}_k = \begin{bmatrix} 1 & \cdots & 0 \\ 0 & \cdots & 1 \end{bmatrix} x_k + \begin{bmatrix} v^1 \\ v^2 \end{bmatrix}_k$$

- Measurements are only X^D, X^B at the discretization times
- Noise w_k affects all the states
- Noise v_k corrupts the measurements

Motivation for Using Autocovariances





Idea of Autocovariances

- The state noise w_k gets propagated in time
- The measurement noise v_k appears only at the sampling times and is not propagated in time
- Taking autocovariances of data at different time lags gives covariances of w_k and v_k

Let w_k, v_k have zero means and covariances Q and R

Mathematical Formulation of the ALS

Linear State-Space Model:

$$\begin{aligned} x_{k+1} &= Ax_k + Gw_k \qquad & w_k &\sim N(0, Q) \\ y_k &= Cx_k + v_k \qquad & v_k &\sim N(0, R) \end{aligned}$$

- Model (A, C, G) known from the linearization, finite set of measurements: $\{y_0, \ldots, y_k\}$ given.
- Only unknowns are noises w_k and v_k.

- $y_k = Cx_k + v_k$
- $y_{k+1} = CAx_k + CGw_k + v_{k+1}$
- $y_{k+2} = CA^2x_k + CAGw_k + CGw_{k+1} + v_{k+2}$

- $E[y_k y_k^T] = R$
- $E[y_{k+2} y_{k+1}^T] = CAGQG^TC^T$

The Autocovariance Least-Squares (ALS) Problem

Skipping a lot of algebra, we can write:

Autocovariance Least Squares

$$\Phi = \min_{Q,R} \left\| \mathscr{A}_N \begin{bmatrix} (Q)_s \\ (R)_s \end{bmatrix} - \hat{b} \right\|^2$$

1. A least-squares problem in a vector of unknowns, Q, R
2. Form $\mathscr{A}_N$ from known system matrices
3. $\hat{b}$ is a vector containing the estimated correlations from data

$$\hat{b} = \frac{1}{T}\sum_{k=1}^{T} \begin{bmatrix} y_k y_k^T \\ \vdots \\ y_{k+N-1} y_k^T \end{bmatrix}_s$$

What about this idea?

Our new proposal!

- Choose a suboptimal state estimator gain L and apply state estimation to $\{y_k\}$ to obtain preliminary $\{\widehat{x}_k\}$.
- Obtain estimates of w_k and v_k from

$$G\hat{w}_k = \widehat{x}_{k+1} - A\widehat{x}_k$$
$$\hat{v}_k = y_k - C\widehat{x}_k$$

- Obtain estimates of Q and R from sample variances!

$$\widehat{Q} = \frac{1}{T}\sum_{k=1}^{T} \hat{w}_k \hat{w}_k{}^T \qquad \widehat{R} = \frac{1}{T}\sum_{k=1}^{T} \hat{v}_k \hat{v}_k{}^T$$

The bad news . . .

Unfortunately an estimate is not the same as the true noise

$$G\widehat{Q}G^T = AL(CSC^T + R)L^TA^T \quad \neq GQG^T$$
$$\widehat{R} = CSC^T + R \quad \neq R$$

in which S satisfies the Lyapunov equation

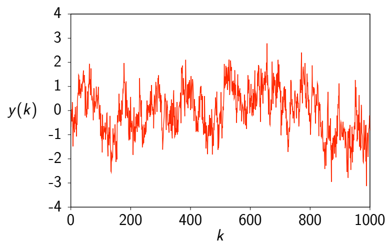
$$S = (A - ALC)S(A - ALC)^T + GQG^T + ALRL^TA^T$$

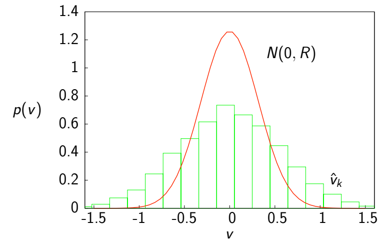
Maybe they are close?

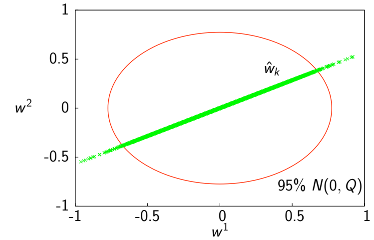
- Example:

$$A = \begin{bmatrix} 0.04 & 0.99 & 0.15 \\ 0.31 & 0.16 & 0.48 \\ 0.02 & 0.18 & 0.74 \end{bmatrix} \qquad C = \begin{bmatrix} 1.30 \\ 0.50 \\ 0.05 \end{bmatrix}^T$$

$$G = \begin{bmatrix} 0.51 & 0.00 \\ 0.00 & 0.92 \\ 0.30 & 0.00 \end{bmatrix} \qquad Q = 0.1I \qquad R = 0.1$$







Comparison of Results

- True values:

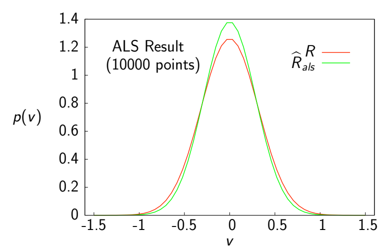
$$Q = \begin{bmatrix} 0.10 & 0.00 \\ 0.00 & 0.10 \end{bmatrix}, \quad R = 0.10$$

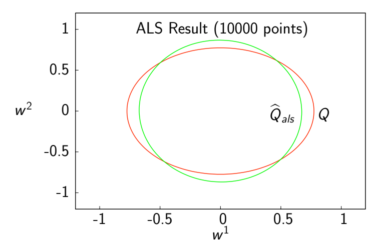
- Using $\hat{w}_k$, $\hat{v}_k$ samples:

$$\widehat{Q} = \begin{bmatrix} 0.07 & 0.04 \\ 0.04 & 0.02 \end{bmatrix}, \quad \widehat{R} = 0.32$$

- Using ALS:

$$\widehat{Q}_{als} = \begin{bmatrix} 0.08 & 0.00 \\ 0.00 & 0.13 \end{bmatrix}, \quad \widehat{R}_{als} = 0.08$$





Still more bad news

If you were lucky and somehow guessed (or estimated) the optimal L for processing the data...

Optimal Pre-filtering of the measurements

Still incorrect $\widehat{Q}$, $\widehat{R}$

$$G\widehat{Q}G^T = AL(CP^-C^T + R)L^TA^T \quad \neq GQG^T$$
$$\widehat{R} = CP^-C^T + R \quad \neq R$$

in which P^- satisfies the filtering Riccati equation

$$P^- = GQG^T + AP^-A^T - AP^-C^T(CP^-C^T + R)^{-1}CP^-A^T$$

Conclusions

Today we have learned ...

- Concepts of observability and detectability of linear systems
  - Illustrated through chemical reactor example
- Introduction to State Estimator Stability
  - Stability versus optimality
  - Cost convergence and estimator stability lemmas
- Obtaining Covariances from Data
  - Separating effects of Q and R in measurement y
  - Autocovariance Least-Squares (ALS) technique to estimate Q and R

Additional Reading

C.-T. Chen. *Linear System Theory and Design*. Oxford University Press, 3rd edition, 1999.

A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

B. J. Odelson, A. Lutz, and J. B. Rawlings. The autocovariance least-squares methods for estimating covariances: Application to model-based control of chemical reactors. *IEEE Ctl. Sys. Tech.*, 14(3):532–541, May 2006a.

B. J. Odelson, M. R. Rajamani, and J. B. Rawlings. A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308, February 2006b. URL `http://www.elsevier.com/locate/automatica`.

M. R. Rajamani and J. B. Rawlings. Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. Accepted for publication in *Automatica*, September, 2007.

M. R. Rajamani, J. B. Rawlings, and T. A. Soderstrom. Application of a new data-based covariance estimation technique to a nonlinear industrial blending drum. Submitted for publication in *IEEE Transactions on Control Systems Technology*, October, 2007.

W. H. Ray. *Advanced Process Control*. McGraw-Hill, New York, 1981.