Determining Covariances from Data

James B. Rawlings
Department of Chemical and Biological Engineering
University of Wisconsin

ExxonMobil Workshop on
State Estimation and Nonlinear Control
Baytown, Texas

09 December 2004
Estimating Variances From Data

\[ x_k \]

\[ \bar{x}_j = \frac{1}{j} \sum_{k=1}^{j} x_k \]

\[ \sigma^2_j = \frac{1}{j-1} \sum_{k=1}^{j} (x_k - \bar{x}_j)^2 \]
Multidimensional Probability Distributions

ExxonMobil MHE
Evaluating Multivariate Probability Distributions

ExxonMobil MHE
Updating the Estimator from Data

Performance Objectives \((Q, R)\)

Constraints \((u_{\text{min},\text{max}}, y_{\text{min},\text{max}})\)

Model \((A, B, C, D)\)

Target calculation

Estimator

\(\hat{x}_k\)

\(d_k\)

Disturbance Model

Tuning \((Q_w, R_v)\)

Covariance estimator

\(y_k\)

\(u_k\)
Motivation

Why use data to compute noise covariances and the filter gain?

- Regulator penalties can come from business goals but estimator tuning is a major source of uncertainty.

- Industrial practitioners currently set covariances heuristically.

- Operators would have fewer tuning parameters to set.

- Covariances of disturbances are “measurable” quantities.

- Better state estimates lead to better control.
Autocovariance function

\[ x_{k+1} = Ax_k + Gw_k \quad w_k \sim N(0, Q_w) \]
\[ y_k = Cx_k + v_k \quad v_k \sim N(0, R_v) \]

- Model \((A, B, C, G)\) known, finite set of outputs \(y_k\) given
- Only unknowns are noises \(w_k\) and \(v_k\). \(w_k\) propagates through states, enabling distinction between \(Q_w, R_v\)
  - \(y_k = Cx_k + v_k\)
  - \(y_{k+1} = CAx_k + Cw_k + v_{k+1}\)
  - \(y_{k+2} = CA^2x_k + CAw_k + Cw_{k+1} + v_{k+2}\)
  - \(y_{k+3} = CA^3x_k + CA^2w_k + CAw_{k+1} + Cw_{k+2} + v_{k+3}\)
  - \(y_{k+4} = CA^4x_k + CA^3w_k + CA^2w_{k+1} + CAw_{k+2} + Cw_{k+3} + v_{k+4}\)

- Correlations between outputs will give noise covariances
Building a Least-Squares Problem

Given some arbitrary (stable) estimator, $L_i$, compute correlation of innovations process

$$y_k = y_k - C\hat{x}_{k|k-1}$$

Define: $y_j = E[y_k y_{k+j}^T]$

$$R(N) = \begin{bmatrix} y_0 & \cdots & y_{N-1} \\ \vdots & \ddots & \vdots \\ y_{N-1}^T & \cdots & y_0 \end{bmatrix} = \begin{bmatrix} C \\ C\tilde{A} \\ \vdots \\ C\tilde{A}^{N-1} \end{bmatrix} P\Theta^T + \begin{bmatrix} R_v & 0 & 0 \\\ 0 & \ddots & 0 \\\ 0 & 0 & R_v \end{bmatrix} + \Psi \begin{bmatrix} R_v & 0 & 0 \\\ 0 & \ddots & 0 \\\ 0 & 0 & R_v \end{bmatrix}$$

$$\tilde{A} = (A - AL_i C), \quad \tilde{G} = [G - AL_i], \quad \Psi = f(-AL_i)$$
Building a Least-Squares Problem

Express a weighted least-squares problem in a vector of unknowns, $Q_w, R_v$

$$\min_{Q_w, R_v} \Phi = \| A \begin{bmatrix} Q_w \\ R_v \end{bmatrix}_s - b \|_W^2$$

Given the estimated covariances:

1. Solve the steady-state Riccati equation

2. Compute the new filter gain

The objective is to minimize variance the estimate error.

$$\min E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T] = [P^-]_{k\to\infty}$$
Example

- 2 input, 2 output
- 4 state
- Covariances unknown
- Active input constraints

\[ G(z) = \begin{bmatrix} \frac{z}{2z-1} & \frac{z}{2.5z-1.5} \\ \frac{0.5z}{2z-1} & \frac{1.5z}{2.5z-1.5} \end{bmatrix} \]
Integrating Disturbance Models

Why use a disturbance model?

- Offset-free control
- Model mismatch
- Nonlinearities

\[
\begin{align*}
x_{k+1} &= Ax_k + Bu_k + B_d d_k + Gw_k \\
d_{k+1} &= d_k + \xi_k \\
y_k &= Cx_k + C_d d_k + v_k \\
\xi_k &\sim N(0, Q_\xi)
\end{align*}
\]

- Pure output disturbance model \((B_d = 0, C_d = I)\)
- Pure input disturbance model \((B_d = B, C_d = 0)\)
Integrating Disturbance Models

We don’t expect to see an integrated white noise disturbance in the plant

\[ E[d_k d_k^T] = kQ_\xi \]

- Slow-drift disturbance
- Plant/model mismatch!

- Use ALS to estimate the covariances, \( Q_w, R_v, Q_\xi \)
Regulator Payoff

\[
\Phi = \frac{1}{N} \sum_{j=k}^{k+\text{N}-1} \left\| y_j - r_j \right\|_Q^2 + \left\| u_j - u_{j-1} \right\|_S^2
\]

What does a reduction in average regulator cost mean?

1. Better tracking (translates to more pounds, quality, etc.)
2. Less control (translates to a reduction in consumables, utilities, etc.)
Case Study - Simulation

Ill-Conditioned Distillation Column - Zafiriou and Morari (1988)

- Structurally ill-conditioned LV distillation column
- Sensitive to input uncertainty
- Model mismatch in the unfavorable direction
- System destabilizes with 16.8% uncertainty in the input

\[
G_{\text{plant}}(s) = G(s) \begin{bmatrix} 1 + \delta & 0 \\ 0 & 1 - \delta \end{bmatrix}
\]
Case Study - Distillation Column

With model mismatch, it is possible to destabilize the system with a poor choice of estimator gain.

- 15% uncertainty ($\delta = 0.15$)
- Choosing $Q_w = \hat{Q}_w, Q_\xi = \hat{Q}_\xi, R_v = \hat{R}_v$ destabilizes the system!
- A careful industrial approach might be covariance matching

Given an arbitrary stable filter gain, process the data and compute the covariances from the residuals

\[
\hat{v}_k = y_k - C\hat{x}_{k|k-1} - \hat{d}_{k|k-1}
\]
\[
\hat{w}_k = \hat{x}_{k+1|k} - A\hat{x}_{k|k-1} - Bu_k
\]
Case Study - Distillation Column

Pure Output Disturbance Model - A Common Industrial Choice

\[
\begin{align*}
\Phi_1 &= 6.771 \\
\Phi_2 &= 17.635 \\
\Phi_3 &= 6.544
\end{align*}
\]

ExxonMobil MHE
Case Study - The Effects of Model Mismatch

- The potential payoff of using ALS grows rapidly with model mismatch
**Eastman Process**

- Gas phase reactor from the Eastman Chemical Company

![Eastman Process Diagram]
Eastman Process

Challenges of applying the methods to industrial data

- Data is pre-filtered in the DCS
- Unmodeled deterministic disturbances
- Unmodeled nonlinearities
- Difficult to evaluate potential benefits *a priori*
Temperature Measurement Innovations

\[ y_k = y_k - C\hat{x}_{k|k-1} \]

- A well-parameterized estimator will yield smaller prediction errors
Eastman Prediction Error

- Prediction error is a substitute for estimate error

\[
\text{cov}(y_k) = \begin{bmatrix}
3.58 \times 10^{-3} & 1.67 \times 10^{-3} \\
1.67 \times 10^{-3} & 8.17 \times 10^{-3}
\end{bmatrix}
\]

\[
\text{cov}(y_k) = \begin{bmatrix}
1.00 \times 10^{-5} & 3.15 \times 10^{-6} \\
3.15 \times 10^{-6} & 7.24 \times 10^{-5}
\end{bmatrix}
\]
Composition Prediction Errors

Original

ALS

Day 1

Day 2

Day 3

ExxonMobil MHE
Consistency of the Eastman Results

ExxonMobil MHE
Consistency of the Eastman Results

- Outliers correspond to temperature spike
- A new disturbance?
Consistency of the Eastman Results

- Do the outliers suggest a deterministic disturbance?

- The outliers rejoin the group if the prediction error spike is neglected.
Eastman Results - Improved Tracking

- Improved tracking of setpoint with proposed method

- Achieves improved control objective compared to current practice
Modification for Non-linear plants

Nonlinear state-space model:

\[
x_{k+1} = F(x_k, u_k, w_k) \\
y_k = h(x_k) + \nu_k
\]

‘Open-loop’ state estimates:

\[
\hat{x}_{k+1} = F(\hat{x}_k, u_k, 0) \\
\hat{y}_k = h(\hat{x}_k)
\]

where, \( w_k \sim N(0, Q_w) \) and \( \nu \sim N(0, R_v) \).

Subtracting the estimates from the plant, we get an approximate time-varying linear model for the innovations: (\( \varepsilon_k = x_k - \hat{x}_k \) and \( y_k = y_k - C_k \hat{x}_k \))

\[
\varepsilon_{k+1} = A_k \varepsilon_k + G_k w_k \\
y_k = C_k \varepsilon_k + \nu_k
\]

\[
A_k = \frac{\partial F(x_k, u_k, w_k)}{\partial x_k^T} \\
G_k = \frac{\partial F(x_k, u_k, w_k)}{\partial w_k^T} \\
C_k = \frac{\partial h(x_k)}{\partial x_k^T}
\]
Modification for Non-linear plants

Allowing the model to evolve from the arbitrary initial condition $\varepsilon_0$:

$$y_k = C_k A_0 \cdots A_{k-1} \varepsilon_0$$
$$+ C_k (A_1 \cdots A_{k-1} G_0 w_0 + A_2 \cdots A_{k-1} G_1 w_1 + \cdots + G_{k-1} w_{k-1})$$
$$+ \nu_k$$

Assumptions:

• The model is correct $(A_i, C_i, G_i)$

• The plant is open-loop stable i.e. $A_0 \cdots A_k \to 0$ for $k \to \infty$

• ONLY UNKNOWNS in above equation are noise covariances $Q_w$ and $R_v$. 
Least-Squares Modification for Non-linear plants

Define: \( y_j = E[Y_k Y_{k+j}^T] \)

\[
R(N) = \begin{bmatrix}
  y_0 & \cdots & y_{N-1} \\
  \vdots & \ddots & \vdots \\
  y_{N-1}^T & \cdots & y_0
\end{bmatrix}
\]

- The Least Squares procedure can now be used on the nonlinear data:

\[
A \begin{bmatrix} Q_w \\ R_v \end{bmatrix}_{\text{stacked}} = R(N)_{\text{stacked}}
\]

\( A \) is a matrix made up of the time-varying system matrices.
Strategy for Estimation

STRATEGY:

1. Choose an initial window size $p$ such that $(A_0 \cdots A_p)$ is small.

2. Use ALS to get estimate of $Q_w$ and $R_v$ from the next $N$ values of $\mathbf{y}_k$’s.

3. Keep sliding the window and get estimates from each new window.

4. Take the mean of all the estimates.
Schematic diagram of Blending Drum

BLENDING DRUM

A+B

flow control valve

LC

Blend of A and B

FC
Blending Drum Details

• The blending drum model is nonlinear and based on mass balances.

• The states are the level in the drum and the concentrations of A and B.

• All three states are measured.

• Measurement noises enter each of the outputs.

• The only significant state noise is in the level of the blending drum.
Blending Drum Model

Flow rates: $f_i$, Weight fractions: $x_i$, Level: $h$

• Nonlinear Drum - Volume:

$$V = C_1 h^3 + C_2 h^2 + C_3 h + C_4$$

$$\frac{dV}{dh} = 3C_1 h^2 + 2C_2 h + C_3$$

• Equations for the states: Mass balances for concentration and level

$$\frac{dx_m}{dt} = \frac{1}{\rho V} (f_c - x_c (f_{dil} + f_m + f_c))$$

$$\frac{dx_c}{dt} = \frac{1}{\rho V} ((x_{dm} - x_m) f_{dil} + (1 - x_m) f_m - x_m f_c)$$

$$\frac{dh}{dt} = \frac{1}{\rho V} (f_{dil} + f_m + f_c - (f_{out} + f_{dist}))$$
• Non-linear state equation,

\[ \dot{x} = f(x, u) + Gw_k, \quad w_k = f_{\text{dist}} \]

• Measurements: All states measured but corrupted with noise

\[
[ y_k ] = \begin{bmatrix} h \\ x_m \\ x_x \end{bmatrix} + \begin{bmatrix} v_k \end{bmatrix} \\

y_k = x_k + v_k
\]

• Disturbances: \( w_k = f_{\text{dist}} \) and \( v_k \) with covariances \( Q_w \) and \( R_v \)
Results

Covariances of Simulated Data:

\[ Q_w = 2 \times 10^{-5} \]

\[ R_v = \begin{bmatrix}
2 \times 10^{-9} & 0 & 0 \\
0 & 3 \times 10^{-7} & 0 \\
0 & 0 & 3 \times 10^{-3}
\end{bmatrix} \]

Covariance estimates using ALS:

\[ \hat{Q}_w \]

\[ \hat{R}_v \]
Noise Shaping Matrix $G$

Matrix $G$ specifies the number of independent noises

$$\begin{bmatrix} x \end{bmatrix}^{k+1} = A \begin{bmatrix} x \end{bmatrix}^{k} + G \begin{bmatrix} w \end{bmatrix}^{k}$$

- Generally many states but only a few independent disturbances
Noise Shaping Matrix $G$

- It's unlikely to have information about $G$ in most real applications. So have to estimate $GQ_wG^T$ from data ($10 \times 10 = 100$ unknowns for a state of size 10!)

- The information contained in measurements is usually not enough to find $GQ_wG^T$ uniquely.

- Utility of Finding $G$:
  1. Moving Horizon Estimator (need $GQ_wG^T$ and $R_v$)
  2. Monitoring of the noise covariances in the plant
  3. $Q_w$ can be estimated uniquely with a low rank $G$
Issues in estimating the full $GQ_wG^T$ covariance

- **ISSUE 1:** Partial state measurement i.e. $C$ matrix is NOT full column rank. Not enough information to find $Q_w$. $\hat{Q}_w$ NOT a unique, positive-semidefinite solution to the ALS for $G = I$.

- **ISSUE 2:** Covariance estimates may not be positive-semidefinite (physically meaningless).
Non-unique estimate

\[ \Phi = \min_{Q_w, R_v} \left\| A \begin{bmatrix} Q_w \\ R_v \end{bmatrix}_s - b \right\|_W^2 \]

- Full set of nonunique solutions with same value of \( \Phi \) given by:

\[ \hat{Q} = \hat{Q}_0 + x_1 Q_1 + x_2 Q_2 + \cdots + x_r Q_r \]

- Which one of these infinite solutions do we choose?

- **MINIMUM RANK** \( \hat{Q} \) chosen. Reason: This gives minimum number of independent noises affecting state explained by data.

\[ \hat{Q} = \tilde{G} \tilde{G}^T \text{ with } \tilde{Q}_w = I \]
• Example: \( \hat{Q} = \tilde{G}\tilde{G}^T = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \). \( \hat{Q} \) has rank 1 and \( \tilde{G} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \). Rank of \( \hat{Q} \) is the number of columns of \( \tilde{G} \).

• Optimization problem:

\[
\begin{align*}
\min_{x_1, x_2 \cdots x_r} \text{rank } Q \\
Q = \hat{Q}_0 + x_1 Q_1 + x_2 Q_2 + \cdots + x_r Q_r \\
Q = Q^T \\
Q \geq 0
\end{align*}
\]

• Optimizing over the vector \( x \) to find minimum rank \( Q \) as above is NP-hard (rank can take only integer values). Use a heuristic for the rank.
Rank Heuristic

- **Rank** = number of nonzero eigenvalues of a matrix
- **Trace** = sum of eigenvalues. Trace is a good surrogate for the Rank.

\[
\text{Rank}(Q)_{\text{min}} \geq \frac{1}{\lambda_{\text{max}}(Q)} \text{Tr}(Q)
\]

- **Surrogate Objective** for rank: minimize \(\text{Tr}(Q)\)
- **Constraints**: \(Q\) Positive Semidefinite. Semidefinite Programming (SDP)?
ALS-SDP : Trace Minimization

\[ \Phi = \min_{x} \operatorname{Tr} (Q(x)) \]

\[ Q = Q_0 + \sum_{i=1}^{m} x_i Q_i \]

\[ Q \geq 0, \quad Q = Q^T \]

- Linear Matrix Inequality (LMI) Constraint. Convex.
- \( \operatorname{Tr} (Q(x)) \) convex in \( x \). Easily solved using standard Newton method.
- Semidefinite constraint is included in objective as a penalty function, which in this case is \( \log\det(Q(x)) \). Follows standard SDP optimization.
ALS-SDP : Trade-Off Curve Method

\[ \Phi_1 = \min_Q \frac{||A_1 Q - b_1||^2}{\Phi} + \rho \text{Tr}(Q) \]

\[ Q \geq 0, \quad Q = Q^T \]

This method gives a unique, feasible solution.

Trade-off between fit to data (\(\Phi\)) and \(\text{Tr}(Q)\) (rank). Parameter \(\rho\) chosen to get the trade-off.

Each Point on this curve represents a different \(\rho\).
Example:

Let the plant be simulated using the following state-space matrices.

\[
A = \begin{bmatrix}
0.733 & -0.086 \\
0.172 & 0.991 \\
\end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \quad Q_w = 0.5 \quad R_v = 1
\]

Assume \( G \) IS UNKNOWN

- Since \( C \) is not full column rank, the estimate of \( Q_w \) is not unique from ALS. Use the new ALS-SDP.

- Both the trace minimization by searching over \( x \) and the trade-off curve method were tested.
Results - ALS-SDP with $x$ as optimization variable

- ALS-SDP: Trace minimization is done by optimizing among all solutions giving the same fit to data i.e. same $\Phi$

$$\hat{Q} = \hat{Q}_0 + x_1Q_1 + x_2Q_2 + \cdots + x_rQ_r$$

![Graph showing the relationship between $\hat{Q}$ and $x$ with eigenvalues $\lambda_1$ and $\lambda_2$.

Optimum $x$ with Trace $\hat{Q}$.

ExxonMobil MHE
Results of the Trade-Off Curve method

Minimum rank choice using trade-off curve

The Trade-Off Curve

\[ \rho = 5.78 \]
Present Literature

The competing methods that claim to find covariances can be classified in 3 main categories:

1. Subspace ID methods give biased estimates for finite sample size

   ![Diagram showing biased estimates for finite sample size and unbiased estimates for infinite sample size.]

   ALS-SDP method gives unbiased estimates for finite sample size and variance goes to 0 in infinite sample sizes.
2. Maximum Likelihood Estimates

- Highly nonlinear and complex
- No guarantee of global optimum or positive definiteness and slow convergence

3. Correlation Techniques: Similar to the ALS, but no positive definiteness constraints and no guarantee of convergence of iterative procedures
Model Mismatch

- No model is perfect. The mismatch between plant and model is accounted for by adding a disturbance model.

\[
\begin{align*}
    x_{k+1} &= Ax_k + Bu_k + Gw_k + Bd_k \\
    d_{k+1} &= d_k + \xi_k \\
    y_k &= Cx_k + v_k
\end{align*}
\]

- If \( \xi_k \sim N(0, Q_\xi) \), then using the ALS-SDP technique, the covariance part of the internal model variable \( d_k \), can also be estimated in an optimal way.

- Thus an optimal filter to take care of model-mismatch may be designed from data.
Conclusions

1. Noise covariances can be estimated by correlating measurements at different times. The second moment of the noise statistics, covariances $Q_w$ and $R_v$, can be estimated as a simple least-squares problem (ALS).

2. The ALS is extended to nonlinear plants by linearization and using a time-varying version of the ALS.

3. A convex SDP problem can be solved to give a positive semidefinite covariance for any set of measurements (ALS-SDP).

4. The ALS-SDP method finds the minimum number of independent process disturbances required to fit the data.

5. The ALS-SDP method is a useful aid in identifying models for monitoring and state estimation tools. Monitoring and state estimation tools work best using models containing the minimum number of independent disturbances.
Acknowledgments

- Eric L. Haseltine, MHE, Ph.D. 2005, postdoc, systems biology
- Brian J. Odelson, covariance estimation, Ph.D. 2003, BP Amoco
- Murali Rajamani, covariance estimation, Ph.D. student
- Matthew J. Tenny, nonlinear MPC and MHE, Ph.D. 2002, ExxonMobil