Bayesian State Estimation: Review of Results for Discrete Time Linear Systems

Sachin C. Patawardhan
Department of Chemical Engineering
I.I.T. Bombay
Email: sachinp@iitb.ac.in
Outline

- Motivation and brief history of the problem
- Bayesian Formulation
- Optimal Recursive Estimation and Kalman Filtering
- Properties and Interpretations of Kalman Filtering
- Stationary Kalman Predictor and Time Series Models
- Optimization formulation
  - Dealing with constraints on states using batch data
  - Moving Horizon estimation
Motivation

- Quality variables: product concentration, average molecular weight, melt viscosity etc.
  - Costly to measure on-line
  - Measured through lab assays: sampled at irregular intervals
- Measurements available from wireless sensors are at irregular intervals due to packet losses
- For satisfactory control of such processes: Quality variable / efficiency parameters should be estimated at a higher frequency
- Remedy: Soft Sensing and State Estimation
Inferential Measurement: Basic Idea

Since fast sampled (primary) variables (temperatures, pressures, levels, pH) are correlated with the quality variables, can we infer values of the quality variables on-line at the fast rate from measurements of primary variables and some model that relates them with the primary measurements?

On line state estimation:
Relevant theoretical developments span over last five decades
Real-time implementation has become feasible only after availability of fast microprocessors
Model Based Soft Sensing

Fast-rate Low-cost measurements from Plant (Temperature / Pressure / Speed)

Irregularly / Slowly sampled Quality variables from Lab assays

Dynamic Model (ODEs/ PDEs)

On-line Fast Rate Estimates of Quality variables

Soft Sensing: Cost Effective Solution
Origin

- Earliest stimulus: study of motion of planets and comets using telescopic observations.

- Orbits of revolution of heavenly bodies can be characterized by 6 parameters, and inferring these parameters from measured data was investigated by Gauss (1795)

- First mathematical approach: Method of Least Squares, independently published by Gauss in (1809) and Lagrange (1806)

- Gauss: “...but since all our measurements are nothing but more than approximation to the truth, same must be true about the calculations resting upon them and the highest aim of all computations made concerning concrete phenomena must be to approximate, as nearly as practicable, to the truth. . . .”

- The quest for generating practicable approximations to the truth has now grown into a rich and vibrant area of research
Corner Stones

- Maximum likelihood estimation:
  - Later introduced by R. A. Fisher (1912) was anticipated by Gauss
  - Forms foundation of modern day state estimation theory

- Kolmogorov (1941) and Wiener (1942) independently developed linear minimum mean-square estimation (non-recursive formulations) and provided the foundation for Kalman Filter theory.
  - Proposed to construct the optimal estimate as a linear combination of measurements such that the mean square error is minimized.

- Development of Kalman filter (1960) major corner stone in the history state estimation
  - Problem solved in discrete and recursive settings suitable for online computer implementation.
Soft Sensing Approaches

- Static / Algebraic Correlations
  - Principle Components Analysis
  - Neural Networks

- Dynamic Model based State Estimation
  - Stochastic (e.g. Kalman filters)
  - Deterministic (e.g. Luenberger Observers)
Outline

- Motivation and brief history of the problem
- Bayesian Formulation
- Optimal Recursive Estimation and Kalman Filtering
- Properties and Interpretations of Kalman Filtering
- Stationary Kalman Predictor and Time Series Models
- Optimization formulation
  - Dealing with constraints on states using batch data
  - Moving Horizon estimation
Discrete Time Linear Systems

Consider Continuous Time Linear Perturbation Model obtained through linearization of a mechanistic model

\[
\frac{dx}{dt} = Ax(t) + Bu(t) + Hd(t)
\]

\[
y(t) = Cx(t)
\]

Perturbation variables

x(t) = X(t) - \bar{X} \quad ; \quad y(t) = Y(t) - \bar{Y}

u(t) = U(t) - \bar{U} \quad ; \quad d(t) = D(t) - \bar{D}

Computer Controlled Systems

Manipulated inputs are piecewise constant

\[ u(t) = u(k) \]

for \[ t = kT \leq t < (k + 1)T \]

Difficulty

Disturbance inputs d(t) are NOT piecewise constant functions!

How to develop a discrete time model?

5/30/2012
Unmeasured Disturbances

Simplifying Assumption 1:
Sampling interval \( T \) is small enough so that disturbance inputs can be approximated as piecewise constant functions during the sampling interval

\[ d(t) = d(k) \quad \text{for} \quad t = kT \leq t < (k+1)T \]

Simplifying Assumption 2:
\( d(k) \): zero mean white noise process with

\[ \text{Cov}[w(k)] = E[w(k)w(k)^T] = Q_d \]

Simplifying Assumption 3:
Measurements are corrupted with zero mean white noise process \( \{v(k)\} \) with
\[ \text{Cov}[v(k)] = E[v(k)v(k)^T] = R \]
Unmeasured Disturbances

Define \( w(k) = \Psi d(k) \)

\[
E\{w(k)\} = \Psi E\{d(k)\} = \bar{0}
\]

\[
Cov\{w(k)\} = E\{w(k)w(k)^T\} = \Psi E\{d(k)d(k)^T\} \Psi^T = \Psi Q_d \Psi^T
\]

Let \( Q = \Psi Q_d \Psi^T \)

\[
x(k+1) = \Phi x(k) + \Gamma u(k) + w(k)
\]

\[
y(k) = C x(k) + v(k)
\]

Given measurements \( \{y(k)\} \), inputs \( \{u(k)\} \) and the model, how to construct optimal state estimate?

**Primary Requirement**

Error between the state estimate and the true process state should be “as small a possible”
Thus, given stochastic state space model

\[ x(k+1) = \Phi x(k) + \Gamma u(k) + w(k) \]
\[ y(k) = C x(k) + v(k) \]

where \( w(k) \) and \( v(k) \) are uncorrelated (in time and with each other) random sequences with zero mean and known variances

\[ E[w(k)w(k)^T] = Q \]
\[ E[v(k)v(k)^T] = R \]

\( Q \) quantify uncertainties in state dynamics and/or modeling errors

\( R \) quantifies variability of measurement errors

How to design an optimal state estimator?
Bayesian Formulation

Since the sequences \{w(k)\} and \{v(k)\} are stochastic processes, the state sequence \{x(k)\} is also a stochastic process.

Define set of measurements up to instant \(k\)

\[ Y^k \equiv \{(y(0), u(0)), (y(1), u(1)), \ldots, (y(k), u(k))\} \]

Objective

Given the stochastic difference equation model and distribution of the initial state, find the conditional probability density function (PDF)

\[ p[x(k)|Y^k] \]

A suitable point estimate can then be constructed using the conditional density function.
Bayesian Formulation

Bayesian theory facilitates modelling of the uncertainty associated with a system and the outcomes of interest by incorporating prior knowledge and observational evidence.

Objective of Bayesian inference is to use priors and causal knowledge, quantitatively and qualitatively, to infer the conditional probability, given finite observations.

Conditional Density Function

\[
p(A | B) = \frac{p(A, B)}{p(B)}
\]
Recursive Bayesian filtering

Bayes’ Theorem

\[ p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)} \]

**Assumption 1:** The states follow a first-order Markov process

\[ p[x(k) \mid x(k - 1), x(k - 2), \ldots, x(0)] = p[x(k) \mid x(k - 1)] \]

**Assumption 2:** The observations are independent of the given states
Bayesian Estimation

Alternative Approaches

- **Sequential Estimation:** Methods that obtain the conditional density function by application of Bayes’ rule, and then obtain the estimate using one of the optimization criteria.

- **Direct Optimization:** Methods that assume a suitable form for the prior probability density function and convert the estimation problem directly into an optimization problem.
Sequential Bayesian Formulation

\[ p[x(k) \mid Y^k] = \frac{p[x(k), Y^k]}{p[Y^k]} = \frac{p[x(k), y(k), Y^{k-1}]}{p[y(k), Y^{k-1}]} \]

Applying Baye's theorem to the numerator

\[ p[x(k), y(k), Y^{k-1}] = p[y(k) \mid x(k), Y^{k-1}]p[x(k), Y^{k-1}] \]

\[ = p[y(k) \mid x(k), Y^{k-1}]p[x(k) \mid Y^{k-1}]p[Y^{k-1}] \]

Since \( v(k) = y(k) - Cx(k) \) is independent to \( Y^{k-1} \),

it follows that

\[ p[x(k), y(k), Y^{k-1}] = p[y(k) \mid x(k)]p[x(k) \mid Y^{k-1}]p[Y^{k-1}] \]
Sequential Bayesian Formulation

Similarly applying Baye's rule to the numerator, we have

\[ p[y(k), Y^{k-1}] = p[y(k) | Y^{k-1}] p[Y^{k-1}] \]

Combining these expressions, we have

\[
p[x(k) | Y^k] = \frac{p[x(k), y(k), Y^{k-1}]}{p[y(k), Y^{k-1}]} = \frac{p[y(k) | x(k)] p[x(k) | Y^{k-1}] p[Y^{k-1}]}{p[y(k) | Y^{k-1}] p[Y^{k-1}]}
\]

\[
p[x(k) | Y^k] = \frac{p[y(k) | x(k)] p[x(k) | Y^{k-1}]}{p[y(k) | Y^{k-1}] p[Y^{k-1}]}
\]

Likelihood Function

Prior Density

Posterior Density

Evidence
Sequential Bayesian Formulation

Prior
The prior density defines the knowledge of the model

\[
p[x(k) | Y^{k-1}] = \int p[x(k) | x(k-1)] p[x(k-1) | Y^{k-1}] dx(k-1)
\]

\( p[x(k) | x(k-1)] \): Transition Density of the state

Likelihood
\( p[y(k) | x(k)] \) determines the measurement noise model

Evidence
The denominator involves an integral

\[
p[y(k) | Y^{k-1}] = \int p[y(k) | x(k)] p[x(k) | Y^{k-1}] dx(k)
\]

5/30/2012 State Estimation 20
Sequential Bayesian Estimation

**Prediction step:** posterior density at previous time step is propagated into next time step through state transition density to compute prior

\[
p\left[x(k) | Y^{k-1}\right] = \int p\left[x(k) | x(k-1)\right] p\left[x(k-1) | Y^{k-1}\right] dx(k-1)
\]

**Update step:** Computation of posterior density from the prior

\[
p\left[x(k) | Y^k\right] = \frac{p\left[y(k) | x(k)\right]}{p\left[y(k) | Y^{k-1}\right]} \times p\left[x(k) | Y^{k-1}\right]
\]

The Posterior Density function constitutes the complete solution to the sequential estimation problem.
Optimal Point Estimates

Posterior density function, even when we can compute it exactly, cannot be used directly for process monitoring and control. For example, to implement a state feedback control law, we need to have a single vector as a point estimate of the state.

Such point estimates can be constructed in an optimal manner using the posterior density function and by defining a suitable optimization criterion.

Popularly used optimization criteria

**Minimum mean-squared error (MMSE):**

\[
E\left[\|x(k) - \hat{x}(k)\|^2 \mid Y^k\right] = \int \|x(k) - \hat{x}(k)\|^2 p[x(k) \mid Y^k] dx(k)
\]

The optimal estimate in this case is the conditional mean

\[
\hat{x}(k \mid k) \equiv E[x(k) \mid Y^k]
\]
Optimal Point Estimates

Maximum a Posteriori (MAP): It is aimed at finding the mode of posterior probability

$$\hat{x}_{MAP}(k) \equiv \max_{x(k)} p[x(k) | Y^k]$$

Maximum likelihood (ML) estimate: which reduces to a special case of MAP where the prior is neglected

Minimax: Aimed at finding the median of posterior density function
Outline

- Motivation and brief history of the problem
- Bayesian Formulation
- Optimal Recursive Estimation and Kalman Filtering
- Properties and Interpretations of Kalman Filtering
- Stationary Kalman Predictor and Time Series Models
- Optimization formulation
  - Dealing with constraints on states using batch data
  - Moving Horizon estimation
MMSE Estimate

- Under weak conditions, the best (i.e. optimal) estimate is the conditional (or a posteriori) mean

\[ \hat{x}(k | k) = E[x(k) | Y^k] \]

Aim: Given the model together with the state and measurement noise covariance matrices, find the conditional mean and its covariance matrix recursively

**Prediction Step**

\[
E[x(k) | Y^{k-1}] = E[\Phi x(k - 1) + \Gamma u(k - 1) + w(k - 1) | Y^{k-1}] \\
= \Phi E[x(k - 1) | Y^{k-1}] + \Gamma u(k - 1) + E[w(k - 1)]
\]

OR

\[ \hat{x}(k | k - 1) = \Phi \hat{x}(k - 1 | k - 1) + \Gamma u(k - 1) \]
MMSE Estimate

\[
\text{Cov}[x(k) \mid Y^{k-1}] = \mathbb{E}[(x(k) - \bar{x}(k))(x(k) - \bar{x}(k))^T \mid Y^{k-1}]
\]

\[
\bar{x}(k) = \mathbb{E}[x(k) \mid Y^{k-1}]
\]

Equation governing estimation error dynamics

\[
\varepsilon(k \mid k-1) = \Phi \varepsilon(k-1 \mid k-1) + \omega(k-1)
\]

Prediction Error

\[
\varepsilon(k \mid k-1) \equiv x(k) - \hat{x}(k \mid k-1)
\]

Estimation Error

\[
\varepsilon(k-1 \mid k-1) \equiv x(k-1) - \hat{x}(k-1 \mid k-1)
\]

Update Step

\[
\hat{x}(k \mid k) = \hat{x}(k \mid k-1) + L(k)e(k)
\]

\[
e(k) = [y(k) - \hat{y}(k \mid k-1)]
\]

(with an arbitrary gain matrix \(L(k)\))
Mean Values of Estimation Errors

Error Dynamics

\[
\varepsilon(k | k-1) = \Phi \varepsilon(k-1 | k-1) + w(k-1)
\]

\[
\varepsilon(k | k) = \left[ I - L(k)C \right] \varepsilon(k | k-1) - L(k)v(k)
\]

Combining

\[
\varepsilon(k | k) = \left[ I - L(k)C \right] \left[ \Phi \varepsilon(k-1 | k-1) + w(k-1) \right] - L(k)v(k)
\]

Simplifying Assumption 4

Initial State at \( k = 0 \) is a Random Variable such that

\[
E[x(0)] = E[\hat{x}(0 \mid 0)] \quad \text{Cov}[x(0)] = P(0)
\]

\[\Rightarrow E[x(0) - \hat{x}(0 \mid 0)] = E[\varepsilon(0 \mid 0)] = \overline{0}\]
Estimation Errors: Covariance Matrices

Define
\[
P(k | k - 1) \equiv \text{Cov}[\varepsilon(k | k - 1)] = E[\varepsilon(k | k - 1)\varepsilon(k | k - 1)^T]
\]
\[
P(k - 1 | k - 1) \equiv \text{Cov}[\varepsilon(k - 1 | k - 1)] = E[\varepsilon(k - 1 | k - 1)\varepsilon(k - 1 | k - 1)^T]
\]

\[
P(k | k - 1) = \Phi P(k - 1 | k - 1)\Phi^T + Q
\]
(Recursive equation for update of prediction covariance)

Defining
\[
P_{\varepsilon\varepsilon}(k) \equiv E[\varepsilon(k | k - 1)e(k)^T]
\]
\[
P_{\varepsilon\varepsilon}(k) = E[\varepsilon(k | k - 1)(C\varepsilon(k | k - 1) + v(k))^T] = P(k | k - 1)C^T
\]

we have
\[
P(k | k) = P(k | k - 1) + L(k)P_e(k)L(k)^T - L(k)P_{\varepsilon\varepsilon}(k)^T - P_{\varepsilon\varepsilon}(k)L(k)^T
\]
Minimum Variance Design

Find gain matrix \( L(k) \) such that estimation error variance in minimum

\[
\text{Min} \quad tr[\mathbf{P}(k | k)] \\
\mathbf{L}(k)
\]

Necessary Condition for Optimality

\[
\frac{\partial tr[\mathbf{P}(k | k)]}{\partial \mathbf{L}(k)} = [0]
\]

\[
\Rightarrow \mathbf{L}^*(k) = [\mathbf{L}(k)]_{OPT} = \mathbf{P}_{ee}(k)\mathbf{P}_e(k)^{-1}
\]

\[
\Rightarrow [\mathbf{P}(k | k)]_{OPT} = \mathbf{P}(k | k - 1) - \mathbf{L}^*(k)\mathbf{P}_e(k)^{-1}\mathbf{L}^*(k)^T = [\mathbf{I} - \mathbf{L}^*(k)\mathbf{C}]\mathbf{P}(k | k - 1)
\]
Kalman Filter: Summary

Prediction
\[
\hat{x}(k | k-1) = \Phi \hat{x}(k-1 | k-1) + \Gamma u(k-1)
\]
\[
P(k | k-1) = \Phi P(k-1 | k-1) \Phi^T + Q
\]

Kalman Gain Computation
\[
L^*(k) = P_{ee}(k)P_e(k)^{-1}
\]
\[
= P(k | k-1)C^T \left[ CP(k | k-1)C^T + R \right]^{-1}
\]

Update
\[
e(k) = [y(k) - C\hat{x}(k | k-1)]
\]
\[
\hat{x}(k | k) = \hat{x}(k | k-1) + L^*(k)e(k)
\]
\[
P(k | k) = [I - L^*(k)C]P(k | k-1)
\]
Outline

- Motivation and brief history of the problem
- Bayesian Formulation
- Optimal Recursive Estimation and Kalman Filtering
- Properties and Interpretations of Kalman Filtering
- Stationary Kalman Predictor and Time Series Models
- Optimization formulation
  - Dealing with constraints on states using batch data
  - Moving Horizon estimation
Interpretations

Covariance matrix quantifies uncertainty associated with the estimated state.

\[ P(k | k - 1) > P(k | k) \]

Prediction step:

\[
[P(k | k)] - P(k | k - 1) = -L^*(k)P_e(k)^{-1}L^*(k)^T
\]

\( L^*(k)P_e(k)^{-1}L^*(k)^T \) is positive definite matrix

\[ \Rightarrow P(k | k) < P(k | k - 1) \]

Update step reduces covariance associated with the estimate.
Why study multivariate Gaussian distribution?

- From Central Limit Theorem, it follows that sum of many independent and equally distributed random variables can be well approximated by Gaussian distribution. If unknown disturbances are assumed to be arising from many independent physical sources, then Gaussian distribution is appropriate for modeling their behavior.

- Attractive mathematical properties: linear transformations of Gaussian distributions are still Gaussian distributed.

- For Gaussian distributed random variables, optimal estimated have a simple form.
Gaussian Noise and KF

Let the process noise, the measurement noise and the initial state have Gaussian normal distributions, i.e.

\[ \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad \text{and} \quad \mathbf{x}(0) \sim \mathcal{N}(\hat{\mathbf{x}}(0 | 0), \mathbf{P}(0)) \]

then, from the properties of Gaussian distributions it follows that

\[
p[\mathbf{x}(k) | \mathbf{Y}^{k-1}] \equiv \mathcal{N}(\hat{\mathbf{x}}(k | k-1), \mathbf{P}(k | k-1))
\]

and

\[
p[\mathbf{x}(k) | \mathbf{Y}^{k}] \equiv \mathcal{N}(\hat{\mathbf{x}}(k | k), \mathbf{P}(k | k))
\]

Also, the innovation sequence is a Gaussian stochastic process

\[
p[e(k) | \mathbf{Y}^{k}] \equiv \mathcal{N}(\mathbf{0}, \mathbf{P}_{ee}(k))
\]

\[
\mathbf{P}_{ee}(k) = \mathbf{C} \mathbf{P}(k | k-1) \mathbf{C}^T + \mathbf{R}
\]
Gaussian Noise and KF

When the process noise, the measurement noise and the initial state have Gaussian normal distributions, it can be shown that

\[ \hat{x}(k | k) \text{ generated using Kalman filter maximizes } p[x(k) | Y^k] \]

i.e. it is a "maximum a posteriori" or MAP estimate

\[ \hat{x}(k | k) \text{ generated using Kalman filter maximizes } \log \text{ likelihood function i.e.} \]

\[ \log(p[x(k) | Y^k]) = \log(p[x(k), Y^k]) - \log(p[Y^k]) \]

In other words, KF generates solution that minimizes

\[ \hat{x}(k | k) = \arg\min_{x(k)} \|y(k) - Cx(k)\|_{R^{-1}}^2 + \|x(k) - \hat{x}(k | k-1)\|_{P(k|k-1)^{-1}}^2 \]

Thus, Kalman Filter is a "Maximum Likelihood" (ML) Estimator
MAP Estimation

Likelihood function, \( p[y(k) \mid x(k)] \), evaluation

\[
y(k) = Cx(k) + v(k)
\]

Since \( p[y(k) \mid x(k)] \) is Gaussian, it can be characterized by the mean and the covariance

\[
E[y(k) \mid x(k)] = Cx(k) \quad \text{and} \quad \text{Cov}[y(k) \mid x(k)] = R
\]

\[
p[y(k) \mid x(k)] = \beta_1 \exp\left[ -\frac{1}{2} (y(k) - Cx(k))^\top R^{-1} (y(k) - Cx(k)) \right]
\]

\[
\beta_1 = \frac{1}{(2\pi)^{n/2} |R|^{1/2}}
\]

Evaluation of prior density, \( p[x(k) \mid Y^{k-1}] \)

\[
x(k) = \Phi x(k - 1) + \Gamma u(k - 1) + w(k - 1)
\]
\[
E[x(k) | Y^{k-1}] = E[\Phi x(k-1) + \Gamma u(k-1) + w(k-1) | Y^{k-1}] \\
= \Phi \hat{x}(k-1 | k-1) + \Gamma u(k-1) \\
= \hat{x}(k | k-1)
\]

\[
Cov[x(k) | Y^{k-1}] = Cov[x(k) - \hat{x}(k | k-1)] \\
= P(k | k-1)
\]

\[
\rho[x(k) | Y^{k-1}] = \beta_2 \exp\left[-\frac{1}{2}(x(k) - \hat{x}(k|k-1))^T P(k | k-1)^{-1}(x(k) - \hat{x}(k|k-1))\right]
\]

\[
\beta_2 = \frac{1}{(2\pi)^{n/2}|P(k | k-1)|^{1/2}}
\]
MAP Estimate

Un-normalized density

\[
p[x(k) \mid Y^k] \propto \beta \exp \left[ -\frac{1}{2} (y(k) - Cx(k))^T R^{-1} (y(k) - Cx(k)) \right] \\
\times \exp \left[ -\frac{1}{2} (x(k) - \hat{x}(k|k-1))^T P(k \mid k-1)^{-1} (x(k) - \hat{x}(k|k-1)) \right]
\]

The MAP estimate is obtained by employing the necessary condition for optimality, i.e.

\[
\frac{\partial \log p[x(k) \mid Y^k]}{\partial x(k)} = 0
\]

\[
\hat{x}(k|k)_{MAP} = \left[ C^T R^{-1} C + P(k|k-1)^{-1} \right]^{-1} \\
\times \left[ P(k|k-1)^{-1} \hat{x}(k|k-1) + C^T R^{-1} y(k) \right]
\]
MAP Estimate

Using matrix inversion lemma

\[ [A + BMD]^{-1} = A^{-1} - A^{-1}B[M^{-1} + DA^{-1}B]^{-1}DA^{-1} \]

with

\[ A \equiv P(k | k - 1)^{-1}, B \equiv C^T, M = R^{-1} \text{ and } D = C \]

and rearranging

\[ \hat{x}_{MAP}(k | k) = \hat{x}(k | k - 1) + L^*(k)e(k) \]

\[ e(k) = [y(k) - C\hat{x}(k | k - 1)] \]

\[ L^*(k) = P_{ee}(k)P_e(k)^{-1} = P(k | k - 1)C^T[CP(k | k - 1)C^T + R]^{-1} \]
Kalman Filter: Advantages

- Generates the maximum likelihood (ML) and maximum a posteriori (MAP) estimates of the states when noises are Gaussian
- Can be derived without making any assumptions on distributions of noises as a minimum variance estimator
- Requires only first and second moments of conditional densities of the states and the innovations
- Relatively easy to adapt to multi-rate and irregular sampling scenario
- Much easier to design than the pole placement approach
Convergence of Estimation Errors

Consider a KF as implemented on a linear deterministic system of the form

\[ x(k+1) = \Phi x(k) + \Gamma u(k) \]

\[ y(k) = Cx(k) \]

which is free of the state uncertainty and measurement noise

Kalman Gain Computation using **Riccati Equations**

\[ P(k \mid k-1) = \Phi P(k-1 \mid k-1) \Phi^T + Q \]

\[ L^*(k) = P(k \mid k-1)C^T \left[ CP(k \mid k-1)C^T + R \right]^{-1} \]

\[ P(k \mid k) = \left[ I - L^*(k)C \right] P(k \mid k-1) \]

where \( Q > 0; R > 0 \) are tuning matrices
Convergence of Estimation Errors

Kalman Filter

\[ \hat{x}(k+1 \mid k) = \Phi \hat{x}(k \mid k) + \Gamma u(k) \]
\[ \hat{x}(k \mid k) = \hat{x}(k \mid k-1) + L^*(k)[y(k) - C\hat{x}(k \mid k)] \]

Under the nominal conditions, the only source of estimation error is the initial state \( \hat{x}(0 \mid 0) \)

Error Dynamics

\[ \varepsilon(k+1 \mid k) = \Phi \varepsilon(k \mid k) \]
\[ \varepsilon(k \mid k) = [I - L^*(k)C] \varepsilon(k \mid k-1) \]

Combining

\[ \varepsilon(k+1 \mid k) = \Phi[I - L^*(k)C] \varepsilon(k \mid k-1) \]........(3)
Convergence of Estimation Errors

Define matrices
\[ \Pi(k \mid k - 1) = [P(k \mid k - 1)]^{-1} \quad \text{and} \quad \Pi(k \mid k) = [P(k \mid k)]^{-1} \]

Using matrix inversion lemma
\[
[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1}
\]

and Riccati equations, the following inequality can be proved
\[
\Pi(k + 1 \mid k) \leq [\Phi_{c}(k)]^{-T} \Pi(k \mid k - 1)[\Phi_{c}(k)]^{-1}
\]
\[
- [\Phi_{c}(k)]^{-T} \left[ \Pi(k \mid k - 1)(\Pi(k \mid k) + \Phi^{T}Q^{-1}\Phi)^{-1}\Pi(k \mid k - 1) \right] [\Phi_{c}(k)]^{-1}
\]
\[
\Phi_{c}(k) = \Phi[I - L^*(k)C]
\]

......(4)
Convergence of Estimation Errors

Define Lyapunov function

\[ V(k) = \varepsilon(k | k - 1)^T \Pi(k | k - 1) \varepsilon(k | k - 1) \]

Combining equation (3) with inequality (4)

\[ V(k + 1) - V(k) \leq -\varepsilon(k | k - 1)^T \Omega(k) \varepsilon(k | k - 1)^T \]

\[ \Omega(k) = \left[ \Pi(k | k - 1)(\Pi(k | k) + \Phi^T Q^{-1} \Phi)^{-1} \Pi(k | k - 1) \right] \]

Since \( \Omega(k) \) is always +ve definite

\[ \varepsilon(k | k - 1)^T \Omega(k) \varepsilon(k | k - 1)^T > 0 \]

and error dynamics given by equation (3) is Lyapunov stable
Convergence of Estimation Errors

Assumption: There exists $p_L, p_H > 0$ such that

$$p_L \mathbf{I} \leq \mathbf{P}(k | k-1) \leq p_H \mathbf{I} \quad \text{and} \quad p_L \mathbf{I} \leq \mathbf{P}(k | k) \leq p_H \mathbf{I}$$

$$\frac{1}{p_H} \| \mathbf{e}(k | k-1) \| \leq V(k) \leq \frac{1}{p_L} \| \mathbf{e}(k | k-1) \|$$

$$\| \Omega(k) \| = \left\| \Pi(k | k-1) \left( \Pi(k | k) + \Phi^T \mathbf{Q}^{-1} \Phi \right)^{-1} \Pi(k | k-1) \right\| \leq \frac{1}{p_H^2 [p_H + \left( \| \Phi \|^2 / \| \mathbf{Q}^{-1} \| \right)]}$$

$$V(k+1) - V(k) \leq - \frac{1}{p_H^2 [p_H + \left( \| \Phi \|^2 / \| \mathbf{Q}^{-1} \| \right)]} \| \mathbf{e}(k | k-1) \|^2$$

Thus, estimation error dynamics is asymptotically stable.
Outline

- Motivation and brief history of the problem
- Bayesian Formulation
- Optimal Recursive Estimation and Kalman Filtering
- Properties and Interpretations of Kalman Filtering
- Stationary Kalman Predictor and Time Series Models
- Optimization formulation
  - Dealing with constraints on states using batch data
  - Moving Horizon estimation
Kalman Predictor: Summary

Initialization Step: Initial mean, \( \hat{X}(0 \mid -1) \),
Initial Covariance \( P(0 \mid -1) \)

At Instant ' \( k \)'

Step 1: Compute Kalman Gain \( L_p^*(k) \)
\[
L_p^*(k) = \Phi P(k \mid k - 1) C^T \left[ R + CP(k \mid k - 1) C^T \right]^{-1}
\]

Step 2: Recursive Prediction Estimator
\[
e(k) = [y(k) - C\hat{x}(k \mid k - 1)]
\]
\[
\hat{x}(k + 1 \mid k) = \Phi \hat{x}(k \mid k - 1) + \Gamma u(k) + L_p^*(k)e(k)
\]

Step 3: Update Covariance matrix
\[
P(k + 1 \mid k) = \Phi P(k \mid k - 1) \Phi^T + Q - L_p(k)CP(k \mid k - 1) \Phi^T
\]
"Steady State" Kalman Predictor

As $k \to \infty$, under weak conditions
the optimal estimator will be time invariant

Theorem
Assume pair $(\Phi, \sqrt{Q})$ is stabilizable and the pair $(\Phi, C)$ is detectable
Then the solution of the Riccati equation
$$P(k | k-1) \to P_\infty > 0$$
where $P_\infty$ denotes solution of the Algebraic Riccati Equation
$$P_\infty = \Phi P_\infty \Phi^T + Q - L^*_p C P_\infty \Phi^T$$
$$L^*_p = \Phi P_\infty C^T [R + C P_\infty C^T]^{-1}$$

Lemma
Assume pair $(\Phi, \sqrt{Q})$ is controllable and $R$ is non-singular
Then all eigen values of $(\Phi - L^*_p C)$ are inside the unit circle.
(Dynamics governing the estimation error $\varepsilon(k | k-1)$ is asymptotically stable)
"Steady State" Kalman Predictor

As $k \rightarrow \infty$, $P(k \mid k-1) \rightarrow P_\infty$

where $P_\infty$ denotes solution of the Algebraic Riccati Equation

$$P_\infty = \Phi P_\infty \Phi^T + Q - L_{p,\infty}^* C P_\infty C^T$$

$$L_{p,\infty}^* = \Phi P_\infty C^T [R + CP_\infty C^T]^{-1}$$

Recursive Prediction Estimator

$$e(k) = y(k) - C\hat{x}(k \mid k-1)$$

$$\hat{x}(k+1 \mid k) = \Phi \hat{x}(k \mid k-1) + \Gamma u(k) + L_{p,\infty}^* e(k)$$

The above "steady state observer" can be written as

$$\hat{x}(k+1 \mid k) = \Phi \hat{x}(k \mid k-1) + \Gamma u(k) + L_{p,\infty}^* e(k)$$

$$y(k) = C\hat{x}(k \mid k-1) + e(k)$$

$$E[e(k)] = 0 \text{ and } \text{Cov}[e(k)] = R + CP_\infty C^T$$
Connection with Time Series Models

Stationary form of Kalman predictor is also known as **Innovation form of State Space Model**

\[
x(k + 1) = \Phi x(k) + \Gamma u(k) + L e(k)
\]

\[
y(k) = C x(k) + e(k)
\]

\[
E[e(k)] = 0 \text{ and } Cov[e(k)] = P_e
\]

The above stationary form of state space model is equivalent to **Box-Jenkins type time series model**

\[
y(k) = G(q)u(k) + H(q)e(k)
\]

\[
G(q) = C[\varphi I - \Phi]^{-1} \Gamma; \quad H(q) = I + C[\varphi I - \Phi]^{-1} L
\]

Estimation of a time series model (ARX/ARMAX/BJ) from input output data is equivalent to identifying Stationary form of Kalman predictor
Connection with Time Series Models

Thus, stationary form of Kalman predictor can be identified directly from input output data using ARX / ARMAX / Box-Jenkins parameterization and converting into state space realization.

Advantage: No need to model the state noise, \( w(k) \), and the measurement noise, \( v(k) \).

Innovation form of state space model

\[
\begin{align*}
x(k+1) &= \Phi x(k) + \Gamma u(k) + w(k) \\
y(k) &= C x(k) + v(k) \\
w(k) &= L e(k) \quad \text{and} \quad v(k) = e(k)
\end{align*}
\]

\[
\begin{align*}
E[w(k)] &= E[e(k)] = 0 \\
\text{Cov}[w(k)] &= L P_e L^T \quad \text{and} \quad \text{Cov}[v(k)] = P_e \\
\text{Cov}[w(k), v(k)] &= E[w(k)v(k)^T] = L P_e
\end{align*}
\]
Outline

- Motivation and brief history of the problem
- Bayesian Formulation
- Optimal Recursive Estimation and Kalman Filtering
- Properties and Interpretations of Kalman Filtering
- Stationary Kalman Predictor and Time Series Models
- Optimization formulation
  - Dealing with constraints on states using batch data
  - Moving Horizon estimation
**Constrained State estimation**

- In most physical systems, **states are bounded**, which introduces constraints on state / parameter estimates.
  - Examples: Concentrations cannot be -ve, mole fractions cannot exceed 1 etc.

- In the presence of constraints
  - Even when the state and the measurement noise densities are Gaussian, the conditional densities of the states are not Gaussian.
  - Difficult to derive Constrained Recursive Formulations

- Possible Solution: Direct optimization based formulation
  - Batch Formulation
  - Moving horizon estimation (MHE)
Consider the conditional density function for the entire state trajectory

\[ p[X^k \mid Y^k] \]

\[ X^k \equiv \{x(0), x(1), \ldots, x(k)\} \]

Since \( x(0), \{w(k)\} \) and \( \{v(k)\} \) are independent RVs

\[
p[X^k \mid Y^k] = \frac{p[Y^k \mid X^k]p[X^k]}{p[Y^k]} = p(x(0)) \prod_{j=1}^{k} p(w(j)) p(v(j)) \]

\[
= \frac{p[Y^k]}{p[Y^k]}
\]
Optimization Formulation

\[
\log p[X^k | Y^k] = \log p(x(0)) + \sum_{j=1}^{k} \log p(w(j)) \\
+ \sum_{j=1}^{k} \log p(v(j)) - \log p[Y^k]
\]

The maximum likelihood estimate of the state trajectory can be obtained by finding the value of $X^k$ that maximizes the above density function by neglecting $\log p[Y^k]$

\[
X_{MLE}^k = \min_{X^k} \left\{ -\log p(x(0)) - \sum_{j=1}^{k} \log p(w(j)) \\
- \sum_{j=1}^{k} \log p(v(j)) \right\}
\]
Unconstrained Batch Estimation

When the state noise, the measurement noise and the initial state have Gaussian distributions

\[
\begin{align*}
\min_{x(0), \ldots, x(k)} & \quad \left\{ [x(0) - x(0 | 0)]^T P(0 | 0)^{-1} [x(0) - x(0 | 0)] \right. \\
& \quad \left. + \sum_{j=1}^{k-1} [w(j)]^T Q^{-1} w(j) + \sum_{j=0}^{k} [v(j)]^T R^{-1} v(j) \right\}
\end{align*}
\]

Subject to

\[
\begin{align*}
w(j) &= x(j + 1) - \Phi x(j) - \Gamma u(j) \\
v(j) &= y(j) - C x(j)
\end{align*}
\]

Solution yields smoothed and current state estimates,

i.e., \( \hat{x}(0 | k), \hat{x}(1 | k), \ldots, \hat{x}(k - 1 | k), \hat{x}(k | k) \)

which are identical to the estimates generated by the Kalman smoother and the Kalman filter.
Constrained Batch Estimation

When the state noise, the measurement noise and the initial state have Gaussian distributions

$$\begin{align*}
\min_{x(0), \ldots, x(k)} & \left\{ \left[ x(0) - x(0 \mid 0) \right]^T P(0 \mid 0)^{-1} [x(0) - x(0 \mid 0)] + \sum_{j=1}^{k-1} [w(j)]^T Q^{-1} w(j) + \sum_{j=0}^{k} [v(j)]^T R^{-1} v(j) \right\} \\
\text{Subject to} & \\
w(j) &= x(j + 1) - \Phi x(j) - \Gamma u(j) \\
v(j) &= y(j) - C x(j) \\
\text{Bounds on state} : & \quad x_L \leq x(j) \leq x_H
\end{align*}$$

Solution yields smoothed and current state estimates, i.e., $\hat{x}(0 \mid k), \hat{x}(1 \mid k), \ldots, \hat{x}(k-1 \mid k), \hat{x}(k \mid k)$ under the constraints (bounds on the states)
Moving Horizon Estimation

Difficulty: Batch estimation is not suitable for on-line computations as problem size grows with time.

Remedy: Formulate a sequence of optimization problems over a moving window \([k-N,k]\).

\[
\min_{x(k-N),\ldots,x(k)} \left\{ V_{k-N}[x(k-N)] + \sum_{j=k-N}^{k-1} [w(j)]^T Q^{-1} w(j) + \sum_{j=k-N}^{k} [v(j)]^T R^{-1} v(j) \right\}
\]

Subject to

\[
w(j) = x(j+1) - \Phi x(j) - \Gamma u(j)
\]

\[
v(j) = y(j) - C x(j)
\]

Bounds on state: \(x_L \leq x(j) \leq x_H\)
Moving Horizon Estimation

\[ V_{k-N}[x(k - N)]: \text{Arrival Cost} \]

\[
V_{k-N}[x(k - N)] = \left\{ \begin{array}{l}
\|x(0) - \hat{x}(0 | 0)\|^2_{P(0|0)^{-1}} \\
\quad + \sum_{j=1}^{k-N-1} [w(j)]^T Q^{-1} w(j) + \sum_{j=0}^{k-N-1} [v(j)]^T R^{-1} v(j) \\
\end{array} \right. 
= -\log p[x(k - N) \mid Y^{k-N}]
\]

Unconstrained Case

\[
V_{k-N}[x(k - N)] = -\log p[x(k - N) \mid Y^{k-N}]
= \|x(k - N) - \hat{x}(k - N \mid k - N)\|^2_{P(k-N|k-N)^{-1}}
\]

\[ P(k - N \mid k - N): \text{Computed Using Kalman Filter} \]
Arrival Cost in MHE

Constrained Case:
Difficult to estimate the arrival cost analytically

Unconstrained Arrival Cost estimate can be used as an approximation

Unconstrained Case

\[ V_{k-N}[x(k-N)] \approx \| x(k-N) - \hat{x}(k-N | k-N) \|_P^2 \]

\[ P(k-N | k-N) : \text{Computed Using Kalman Filter} \]

Longer window size has to be used to nullify the effect of wrong estimate of the arrival cost
Summary

- For discrete time linear dynamic systems, Kalman filter provides the optimal and recursive solution to the unconstrained Bayesian state estimation problem.

- If no assumptions are made regarding the PDFs associated with the state and the measurement noise, then the Kalman filter estimates are minimum variance estimates.

- If the state and the measurement noise are assumed to be Gaussian processes, then the Kalman filter estimates are MAP and maximum likelihood estimates.
Summary

- An attractive feature of the Kalman filter: works with propagation and update of only the first two moments of the conditional densities of the states.

- In presence of constraints on the states, moving horizon estimation (MHE) can be used to generate smoothed and current state estimates, which is a direct optimization based formulation.

- The problem of estimating conditional densities of the states is shifted to the arrival cost estimation in MHE formulation.

- For the constrained case, it is difficult to estimate the arrival cost and approximations to be developed that are consistent with constraints.
References

- Gelb, A. Applied Optimal Estimation, MIT Press, 1974