# MM7. Kalman Filter (Part one)

**Reading page: Chapt 7, pp.406-419**

- Explain MM6 exercise
- 7.1 Introduction
- 7.2 An Intuitive Description of Kalman filter
- 7.3 Formal Description of Scale Kalman Filter

# What have we talked through MM6 – **Discrete-Time Wiener Filters**?

6.1 (Ideal) Noncausal Wiener Filters

6.2 Causal Wiener Filters

# Discrete-Time Wiener Filters

**Motivation:**

- Estimate a WSS random sequence Y(n) based on the observation of another sequence X(n).

- **Without loss of generality we assume that**

  $$E\{Y(n)\}=E\{X(n)\}=0$$

- The goodness of the estimator is described by MSE

$$E\{(Y(n)-\hat{Y}(n))^2\}$$

# 6.1 Ideal (Noncausal) Wiener Filters

- **Problem Formulation:**

Seek a linear filter ← **system**

$$\hat{Y}(n) = \sum_{m=-\infty}^{\infty} h(m) X(n-m) = h(n) * X(n)$$

$LMMSEE: \quad \hat{Y} = h_0 + \sum_{m=1}^{M} h_m X(m)$

↑ Signals' relationship

Which minimizes the MSE

$$E\{(\hat{Y}(n) - Y(n))^2\}$$

The filter reaching above requirement is called *ideal (noncausal) Wiener filter*

# 6.1.3 Wiener Filter via LMMSEE

- Wiener filter:

$$R_{XY}(k) = \sum_{m=-\infty}^{\infty} h(m) R_{XX}(k-m) = h(k) * R_{XX}(k)$$

$$´H(f) = \frac{S_{XY}(f)}{S_{XX}(f)}$$

- MSE residual:

$$E\{(Y(n) - \hat{Y}(n))^2\} = \sigma_Y{}^2 - \sum_{m=-\infty}^{\infty} h(m) R_{XY}(m)$$

$$E\{(Y(n) - \hat{Y}(n))^2\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} [S_{YY}(f)] - \frac{|S_{XY}(f)|2}{S_{XX}(f)}] df$$

- LMMSEE:

$$\mathbf{h} = [h_1 \quad \cdots \quad h_M]^T \qquad \mu_{\mathbf{X}} = [\mu_{\mathbf{X}(1)} \quad \cdots \quad \mu_{\mathbf{X}(M)}]^T$$

$$\sum{}_{\mathbf{X}Y} = \begin{bmatrix} \sum_{X(1)Y} \\ \vdots \\ \sum_{X(M)Y} \end{bmatrix}, \qquad \sum{}_{\mathbf{XX}} = \begin{bmatrix} \sum_{X(1)X(1)} & \cdots & \sum_{X(1)X(M)} \\ \vdots & \ddots & \vdots \\ \sum_{X(M)X(1)} & \cdots & \sum_{X(M)X(M)} \end{bmatrix},$$

$$\mathbf{h} = \left(\sum{}_{\mathbf{XX}}\right)^{-1} \sum{}_{\mathbf{X}Y}$$

$$h_0 = \mu_Y - \mathbf{h}^T \mu_{\mathbf{X}} = \mu_Y - \left(\sum{}_{\mathbf{X}Y}\right)^T \left(\sum{}_{\mathbf{XX}}\right)^{-1} \mu_{\mathbf{X}}$$

- MSE residual:

$$E\{(Y - \hat{Y})^2\} = E\{Y^2\} - E\{\hat{Y}^2\} = E\{(Y - \hat{Y})Y\}$$

$$E\{(Y - \hat{Y})^2\} = \sigma_Y{}^2 - \mathbf{h}^T \sum{}_{\mathbf{X}Y}$$

# 6.2.7 Finite Wiener Filter

$$\text{Finite Wiener filter}: \quad \hat{Y}(n) = \sum_{m=0}^{M} h(m) X(n-m)$$

*Wiener–Hopf Solution:*

$$\mathbf{h}^{\mathbf{T}} = [h(0)\ h(1) \cdots h(M)] = \mathbf{R_{XX}}^{-1}\mathbf{R_{XY}}$$

*provided* $\mathbf{R_{XX}}$ *is invertible,*

*where*

$$\mathbf{R_{XX}} = \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & \cdots & R_{XX}(M) \\ R_{XX}(1) & R_{XX}(0) & \cdots & R_{XX}(M-1) \\ \vdots & \vdots & \ddots & \vdots \\ R_{XX}(M) & R_{XX}(M-1) & \cdots & R_{XX}(0) \end{bmatrix}$$

$$\mathbf{R_{XY}}^{\mathbf{T}} = [R_{XY}(0)\ R_{XY}(1) \cdots R_{XY}(M)]$$

- LMMSEE:

$$\mathbf{h} = [h_1 \ \cdots \ h_M]^T \qquad \mu_{\mathbf{X}} = [\mu_{\mathbf{X}(1)} \ \cdots \ \mu_{\mathbf{X}(M)}]^T$$

$$\sum_{\mathbf{X}Y} = \begin{bmatrix} \sum_{X(1)Y} \\ \vdots \\ \sum_{X(M)Y} \end{bmatrix}, \qquad \sum_{\mathbf{XX}} = \begin{bmatrix} \sum_{X(1)X(1)} & \cdots & \sum_{X(1)X(M)} \\ \vdots & \ddots & \vdots \\ \sum_{X(M)X(1)} & \cdots & \sum_{X(M)X(M)} \end{bmatrix},$$

$$\mathbf{h} = \left(\sum_{\mathbf{XX}}\right)^{-1} \sum_{\mathbf{X}Y}$$

$$h_0 = \mu_Y - \mathbf{h}^T \mu_{\mathbf{X}} = \mu_Y - \left(\sum_{\mathbf{X}Y}\right)^T \left(\sum_{\mathbf{XX}}\right)^{-1} \mu_{\mathbf{X}}$$

- MSE residual:

$$E\{(Y-\hat{Y})^2\} = E\{Y^2\} - E\{\hat{Y}^2\} = E\{(Y-\hat{Y})Y\}$$

$$E\{(Y-\hat{Y})^2\} = \sigma_Y^{\ 2} - \mathbf{h}^T \sum_{\mathbf{X}Y}$$

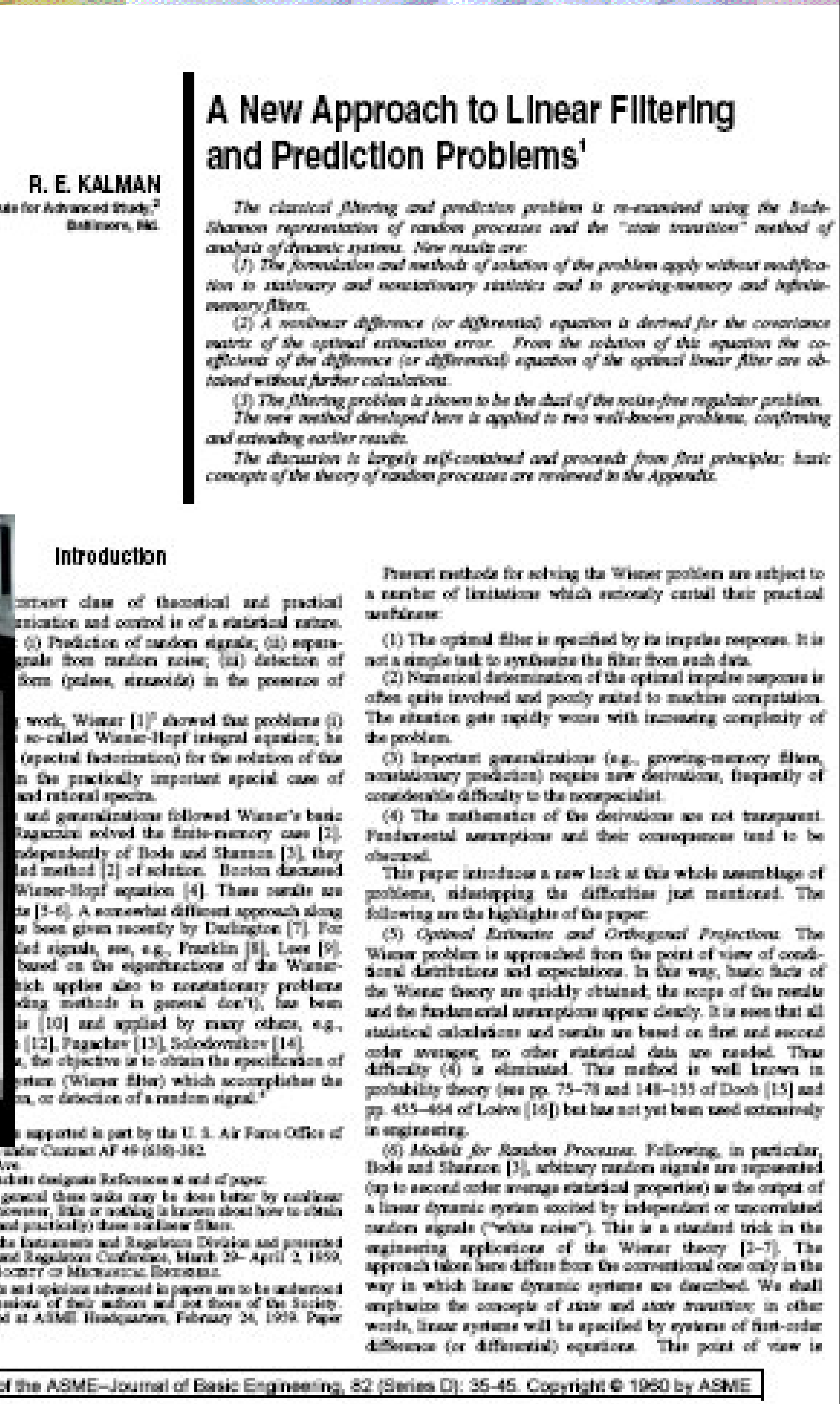
# Explain the MM6 Exercise!

# MM7. Kalman Filter (Part one)

## Reading page: Chapt 7, pp.406-419

- Explain MM6 exercise
- **7.1 Introduction**
- 7.2 An Intuitive Description of Kalman filter
- 7.3 Formal Description of Scale Kalman Filter

# 7.1 Introduction



R. E. KALMAN
Research Institute for Advanced Study,[2]
Baltimore, Md.

## A New Approach to Linear Filtering and Prediction Problems[1]

*The classical filtering and prediction problem is re-examined using the Bode-Shannon representation of random processes and the "state transition" method of analysis of dynamic systems. New results are:*

*(1) The formulation and methods of solution of the problem apply without modification to stationary and nonstationary statistics and to growing-memory and infinite-memory filters.*

*(2) A nonlinear difference (or differential) equation is derived for the covariance matrix of the optimal estimation error. From the solution of this equation the coefficients of the difference (or differential) equation of the optimal linear filter are obtained without further calculations.*

*(3) The filtering problem is shown to be the dual of the noise-free regulator problem.*

*The new method developed here is applied to two well-known problems, confirming and extending earlier results.*

*The discussion is largely self-contained and proceeds from first principles; basic concepts of the theory of random processes are reviewed in the Appendix.*

### Introduction

Present methods for solving the Wiener problem are subject to a number of limitations which seriously curtail their practical usefulness:

(1) The optimal filter is specified by its impulse response. It is not a simple task to synthesize the filter from such data.

(2) Numerical determination of the optimal impulse response is often quite involved and poorly suited to machine computation. The situation gets rapidly worse with increasing complexity of the problem.

(3) Important generalizations (e.g., growing-memory filters, nonstationary prediction) require new derivations, frequently of considerable difficulty to the nonspecialist.

(4) The mathematics of the derivations are not transparent. Fundamental assumptions and their consequences tend to be obscured.

This paper introduces a new look at this whole assemblage of problems, sidestepping the difficulties just mentioned. The following are the highlights of the paper:

(5) *Optimal Estimates and Orthogonal Projections.* The Wiener problem is approached from the point of view of conditional distributions and expectations. In this way, basic facts of the Wiener theory are quickly obtained; the scope of the results and the fundamental assumptions appear clearly. It is seen that all statistical calculations and results are based on first and second order averages; no other statistical data are needed. Thus difficulty (4) is eliminated. This method is well known in probability theory (see pp. 75–78 and 148–155 of Doob [15] and pp. 455–464 of Loève [16]) but has not yet been used extensively in engineering.

(6) *Models for Random Processes.* Following, in particular, Bode and Shannon [3], arbitrary random signals are represented (up to second order average statistical properties) as the output of a linear dynamic system excited by independent or uncorrelated random signals ("white noise"). This is a standard trick in the engineering applications of the Wiener theory [2–7]. The approach taken here differs from the conventional one only in the way in which linear dynamic systems are described. We shall emphasize the concepts of *state* and *state transition*; in other words, linear systems will be specified by systems of first-order difference (or differential) equations. This point of view is

[2] 7212 Bellona Ave.

[3] Numbers in brackets designate References at end of paper.

[4] Of course, in general these tasks may be done better by nonlinear filters. At present, however, little or nothing is known about how to obtain (both theoretically and practically) these nonlinear filters.

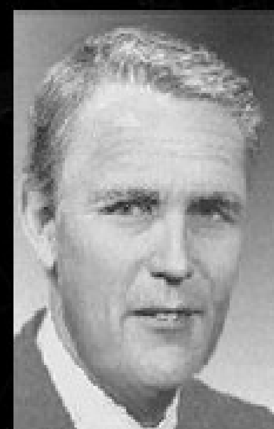
Contributed by the Instruments and Regulators Division and presented at the Instruments and Regulators Conference, March 29– April 2, 1959, of The American Society of Mechanical Engineers.

Note: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, February 24, 1959. Paper No. 59—IRD-11.

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

## Rudolf Emil Kalman

- Born 1930 in Hungary
- BS and MS from MIT
- PhD 1957 from Columbia
- Filter developed in 1960-61
- Now retired



# 7.1.1 What's Kalman Filter?

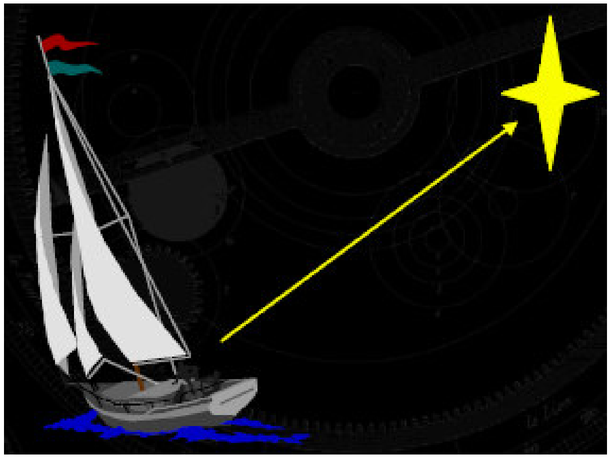
- One of the most well-known and often-used math. Tools for stochastic estimation from noisy measurements
- Rudolph E. Kalman in 1960 published his famous paper decribing a recursive solution to discrete-time linear filtering problem

**Features**

- Just some applied mathematics
- A linear system
- Noisy data in → hopefully less noisy output
- Delay is the price for filtering

# 7.1.2 What is it used for?

- Target (missiles etc) tracking
- Navigation
- Feedback control
- Computer vision
- Economics



- An example is estimating the position and velocity of a satelite from radar data. There are 3 components of position and 3 of velocity so there are at least 6 variables to estimate. These variables are called state variables. With 6 state variables the resulting Kalman filter is called a 6 dimensional Kalman filter.

# 7.1.3 Kalman Filter Formulation-I

$$x_k = Ax_{k-1} + Bu_k + w_{k-1}, \qquad z_k = Hx_k + v_k.$$

$$p(w) \sim N(0, Q)$$

$$p(v) \sim N(0, R).$$

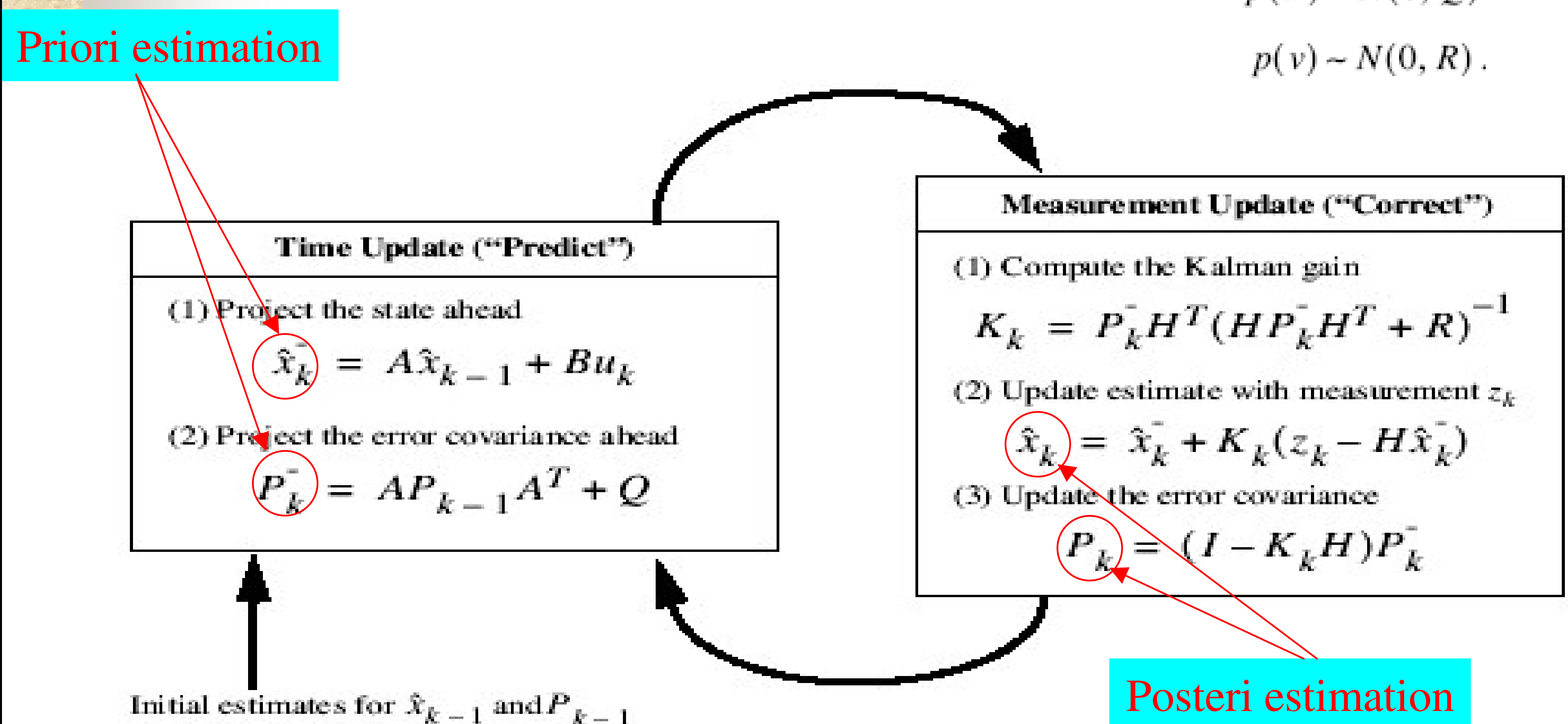



**Figure 4.2:** A complete picture of the operation of the Kalman filter, combining the high-level diagram of Figure 4.1 with the equations from table 4.1 and table 4.2.

# 7.2 An Intuitive Explanation

**INTRODUCTORY LESSON**

**The one dimensional Kalman Filter**

**PDJoseph@compuserve.com**

# 7.2.1 Assumptions

- Suppose we have a random variable x(t) whose value we want to estimate at certain times t0 ,t1, t2, t3, etc. Also, suppose we know that x(tk) satisfies a linear dynamic equation

  x(tk+1) = Ax(tk) + w(k) (the dynamic equation)

- F is a known number. In order to work through a numerical example let us assume A= 0.9
- Kalman assumed that w(k) is a random number selected by picking a number from a hat. Suppose the numbers in the hat are such that the mean of w(k) = 0 and the variance of w(k) is Q. we will take Q=100 for example.
- w(k) is called white noise, which means it is not correlated with any other random variables and most especially not correlated with past values of w.

# 7.2.2 Starting the KF Procedure

A Kalman filter needs an initial estimate to get started. It is like an automobile engine that needs a starter motor to get going. Once it gets going it doesn't need the starter motor anymore. Same with the Kalman filter. It needs an initial estimate to get going. Then it won't need any more estimates from outside. In later lessons we will discuss possible sources of the initial estimate but for now just assume some person came along and gave it to you.

- Now suppose that at time t0 someone came along and told you he thought x(t0) = 1000 but that he might be in error and he thinks the variance of his error is equal to P. Suppose that you had a great deal of confidence in this person and were, therefore, convinced that this was the best possible estimate of x(t0). This is the initial estimate of x. It is sometimes called the a ***priori estimate***.

# 7.2.3 First Step (State-Prediction)

- So we have an estimate of x(t0),which we will call xe = 1000.
- The variance of the error in this estimate is defined by $P = E[(x(t0) - xe)^2]$, e.g., we will take P = 40,000

- Now we would like to estimate x(t1):

  $x(tk+1) = Ax(tk) + w(k) \rightarrow x(t1) = Ax(t0) + w(0)$
- Dr. Kalman says our new best estimate of x(t1) is given as

  $newxe = Axe$ (equation 1)

  in our numerical example 900
- Why Dr. Kalman is right: We have no way of estimating w(0) except to use its mean value of zero!

# 7.2.3 Second Step (Variance-Prediction)

- What is the variance of the error of this estimate?

$$newP = E\,[(x(t1) - newxe)^2]$$

- Substitute the above equations in for x(t1) and newxe, then

$$newP = E\,[(Ax(t0) + w - Axe)^2]$$

$$= E[A^2(x(t0) - xe)^2] + E\,w^2 + 2F\,E\,(x(t0) - xe)*w]$$

- The last term is zero because w is assumed to be uncorrelated with x(t0) and xe. Then,

$$newP = PA^2 + Q \qquad \text{(Equation 2)}$$

- For our example, we have

$$newP = 40{,}000 \text{ X } .81 + 100 = 32{,}500$$

# Kalman Filter Formulation

$$x_k = Ax_{k-1} + Bu_k + w_{k-1}, \qquad z_k = Hx_k + v_k.$$

$$p(w) \sim N(0, Q)$$

$$p(v) \sim N(0, R).$$

Priori estimation

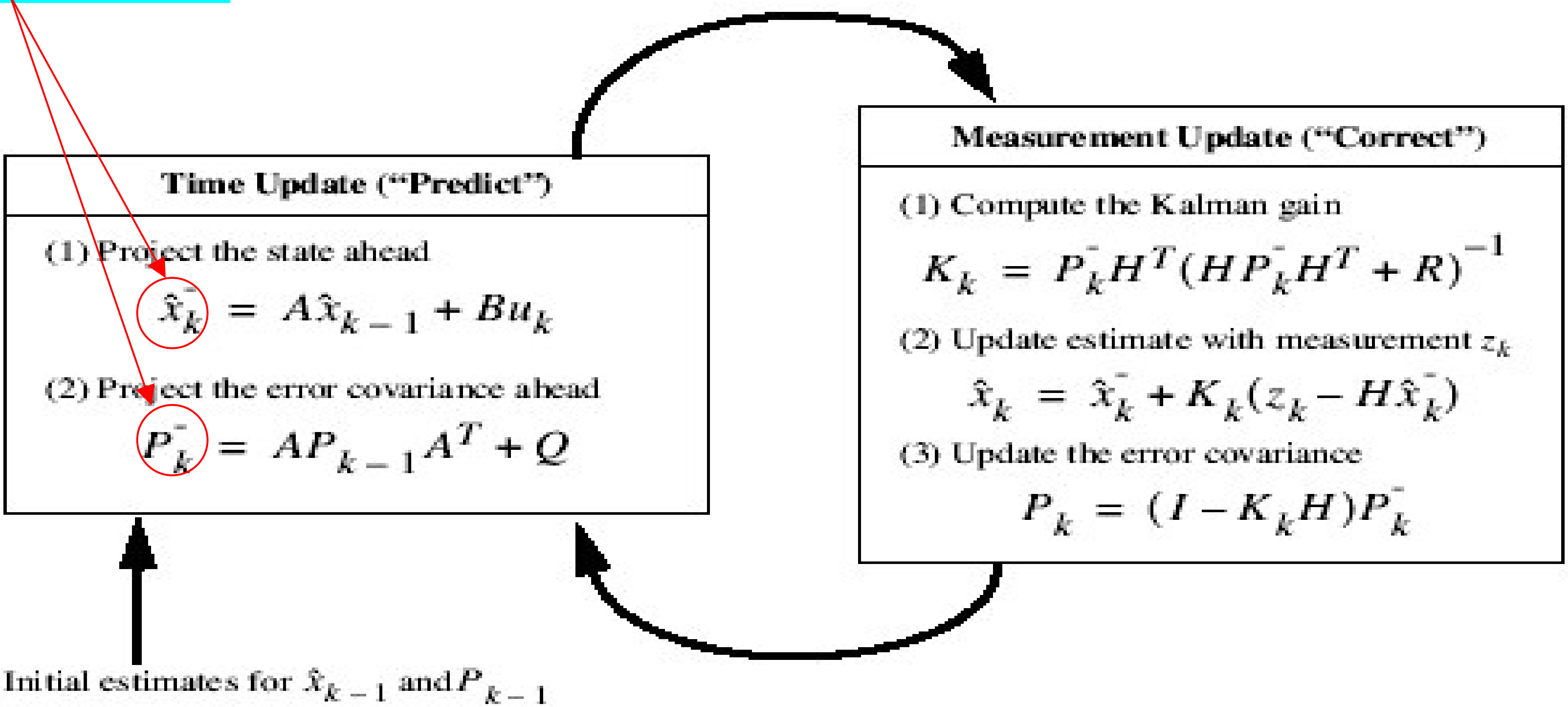



**Figure 4.2:** A complete picture of the operation of the Kalman filter, combining the high-level diagram of Figure 4.1 with the equations from table 4.1 and table 4.2.

# 7.2.4 Third Step (Measurement)

- Now, let us assume we make a noisy measurement of x. Call the measurement y and assume y is related to x by a linear equation. (Kalman assumed that all the equations of the system are linear. This is called linear system theory.)

$$y(1) = Hx(t1) + v(1)$$

where v is white noise with the variance denoted as R.

- H is some number whose value we know. We will use for our numerical example H = 1 , R= 10,000 and y(1) = 1200
- Notice that if we wanted to estimate y(1) before we look at the measured value we would use

$$ye = H*newxe$$

- for our numerical example we would have ye = 900

# 7.2.5 Fourth Step (State-Updating)

- Dr. Kalman says the new best estimate of x(t1) is given by

  newerxe = newxe + K*(y(1) -H*newxe)

  = newxe + K*(y(1) - ye) (equation 3)

- where K is a number called the Kalman gain.
- Notice that y(1) - ye is just our error in estimating y(1). For our example, this error is equal to plus 300. Part of this is due to the noise, v and part to our error in estimating x.
- If all the error were due to our error in estimating x, then Setting K=1 would correct our estimate by the full 300. But since some of this error is due to v, we will make a correction of less than 300 to come up with newerxe. We will set K to some number less than one.

# 7.2.6 Fifth Step (Variance-Updating)

- What value of K should we use? Before we decide, let us compute the variance of the resulting error

$$E\,(x(t1) - newerxe)^2 = E\,[x - newxe - K(y - H\,newxe)]^2$$
$$= E\,[(x - newxe - K(Hx + v - H\,newxe)]^2$$
$$= E\,[\{(1 - KH)(x - newxe)2 + Kv\}]^2$$
$$= newP(1 - KH)^2 + R^2$$

- where the cross product terms dropped out because v is assumed to be uncorrelated with x and newxe. So the newer value of the variance is now given by

$$newerP = newP(1 - KH)^2 + R(K^2) \qquad \text{(equation 5)}$$

# 7.2.7 Sixth Step (Kalman Gain)

- If we want to minimize the estimation error we should minimize newerP. We do that by differentiating newerP withrespect to K and setting the derivative equal to zero and then solving for K. A little algebra shows that the optimal K is given by

  $K = H\ newP/[newP(H^2) + R]$ (Equation 4)

- For our example,
  - K = .7647
  - newerxe = 1129
  - newerP = 7647
  - Notice that the variance of our estimation error is decreasing

# 7.2.8 Summary- Kalman Formulation-I

- x(tk+1) = Ax(tk) + w(k)    (Q)
- y(k) = Hx(tk) + v(k)    (R)

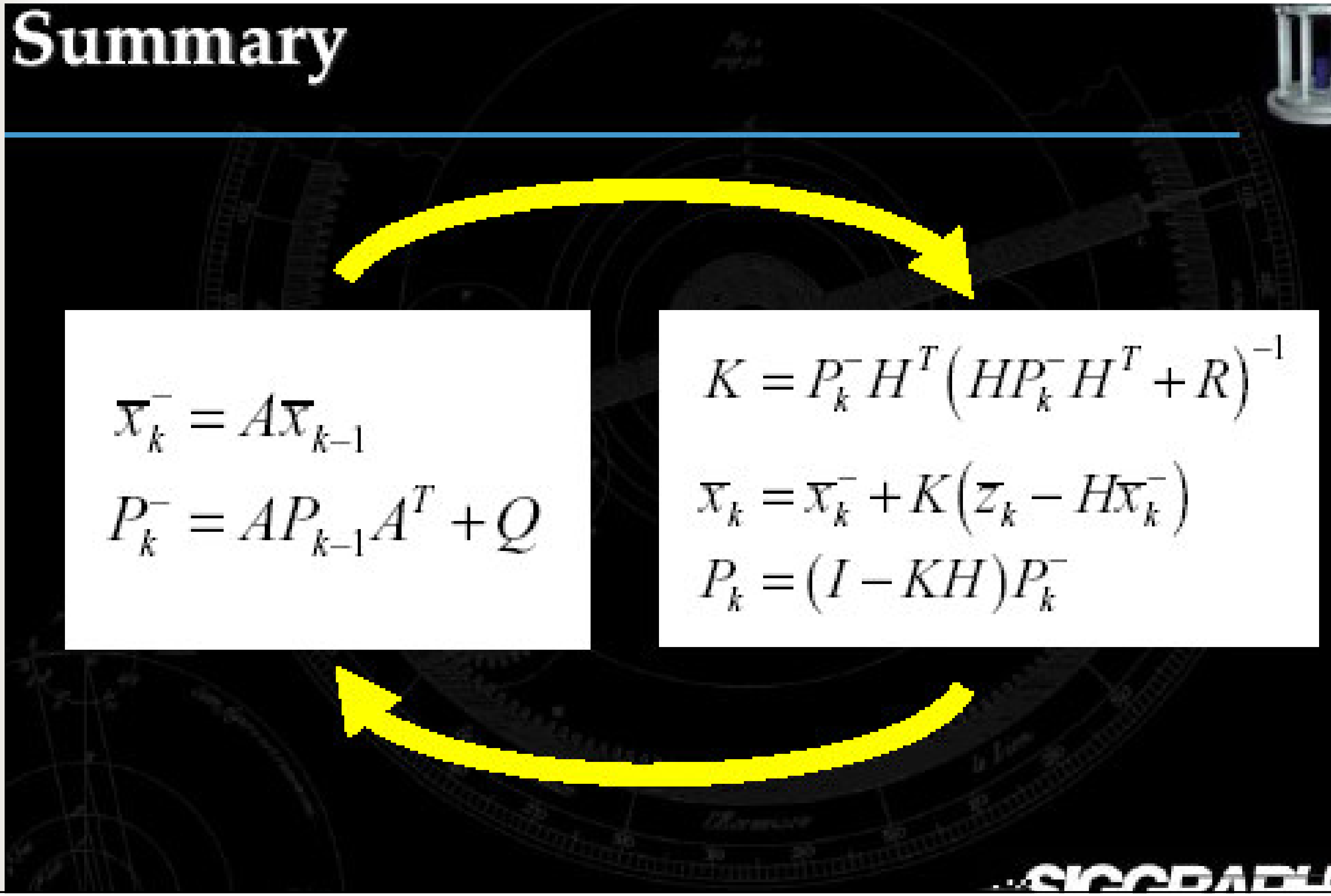
**Prediction:**

- newxe = Axe    (equation 1)
- newP = PA$^2$ + Q    (Equation 2)

**Updating:**

- newerxe = newxe + K*(y - ye)    (equation 3)
- K = H newP/[newP(H$^2$) + R]    (Equation 4)
- newerP = newP(1 - KH)$^2$ + R(K$^2$)    (equation 5)

# 7.3 Kalman Formulation-II



# 7.3 Kalman Formulation-III

## Signalmodel

$$s(n) = A(n)s(n-1) + w(n)$$
$$x(n) = H(n)s(n) + v(n)$$

Antagelser:

$$w(n) \in \ \mathrm{NID}(\underline{0}, Q(n))$$
$$v(n) \in \ \mathrm{NID}(\underline{0}, R(n))$$
$$s(1) \in \ \mathrm{N}(\underline{0}, P(1))$$
$$w(n), v(n) \text{ og } s(1) \text{ er uafhængige}$$

### Kalman filter rekursionsligningerne - prediktionsform

Notation:

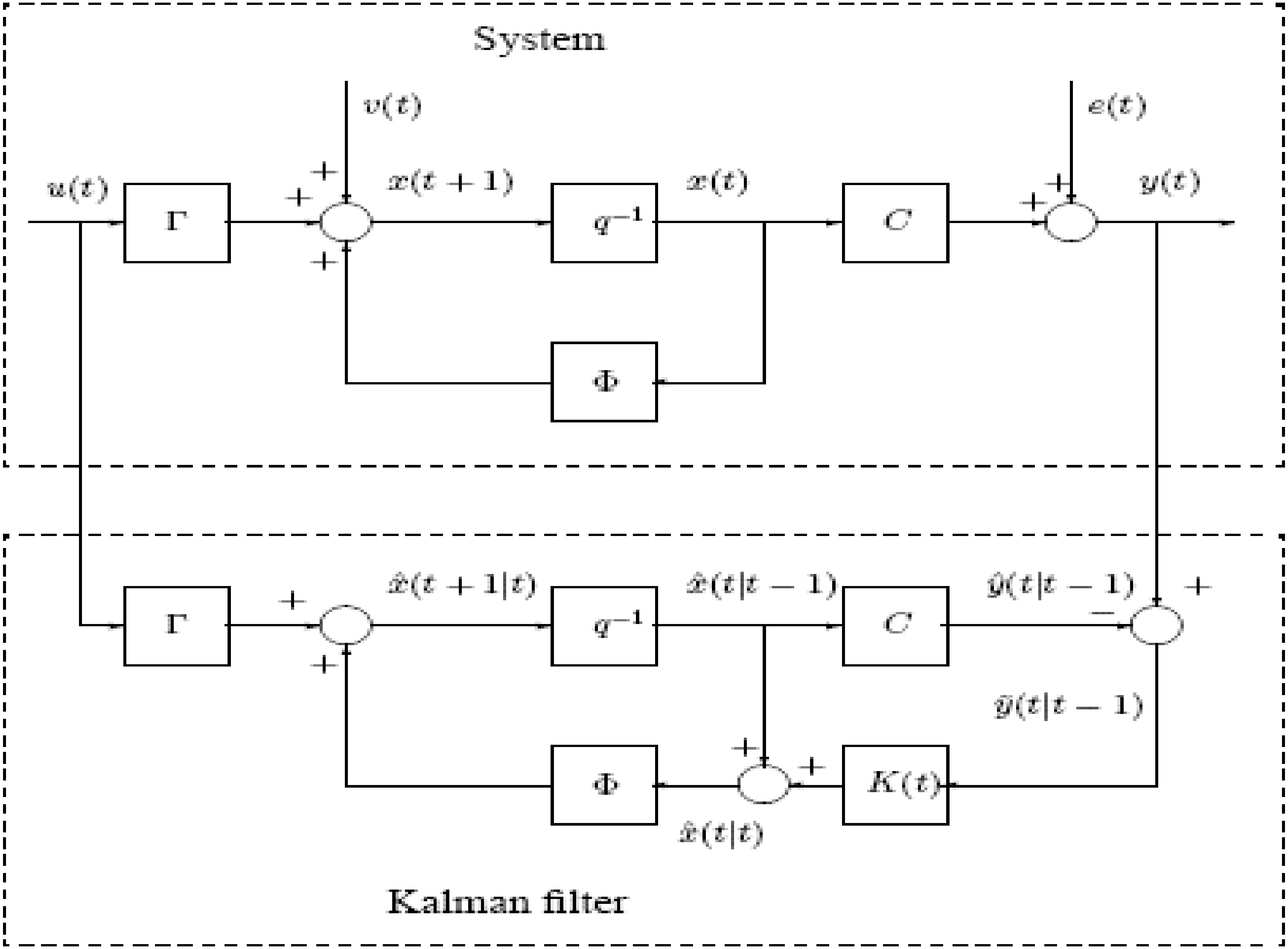
$\hat{s}(n+1|n) = \mathrm{E}(s(n+1)|x(n), \ldots, x(1))$ : *tilstands prediktion* af $s(n+1)$ baseret på målinger op til og med $n$.

Initial værdier:

$$\hat{s}(1|0) = \underline{0}$$
$$P(1)$$

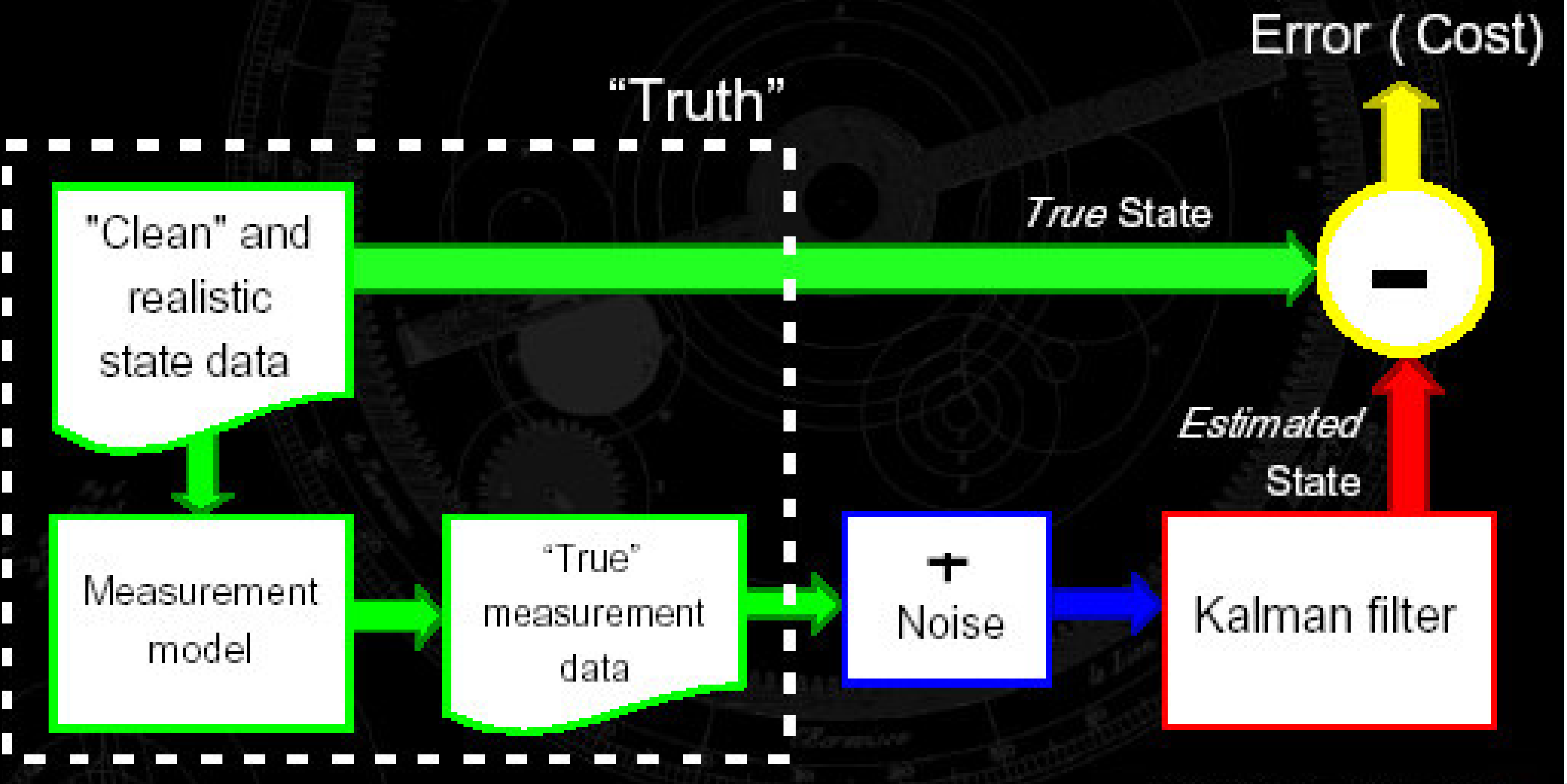
Rekusion:

$$K(n) = P(n)H(n)^T(H(n)P(n)H(n)^T + R(n))^{-1} \quad (3)$$
$$\hat{s}(n+1|n) = A(n+1)[\hat{s}(n|n-1) + K(n)(x(n) - H(n)\hat{s}(n|n-1))] \quad (4)$$
$$P(n+1) = A(n+1)[(I - K(n)H(n))P(n)]A(n+1)^T + Q(n) \quad (5)$$

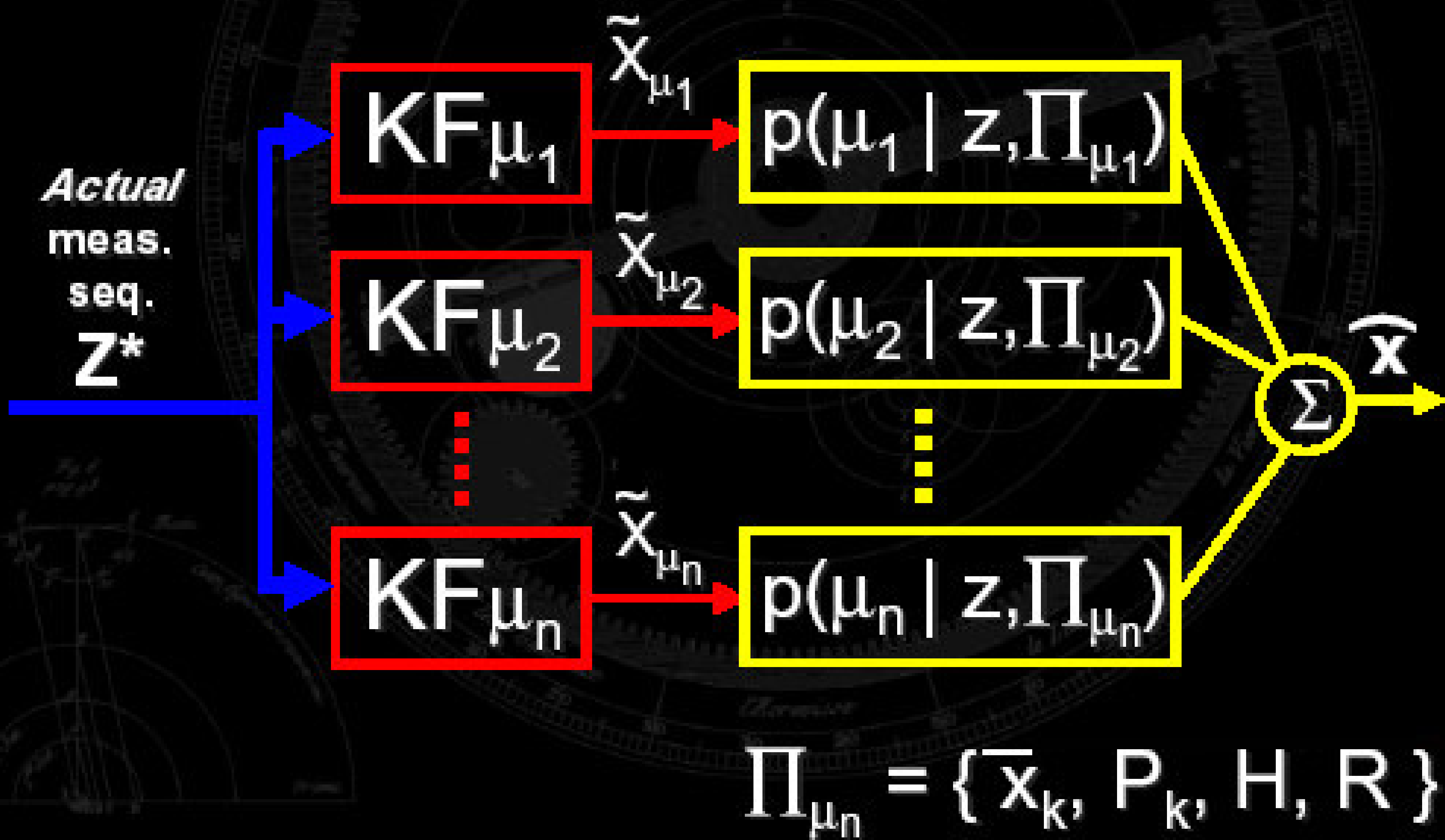



Figur 1: Blokdiagram for system og Kalmanfilter

# Parameter Optimization



# On-Line Multiple-Model Estimation



# Probability of Model $\mu$

For model $\mu$ with $\Pi_\mu = \{x, P, H, R\}$

$$p(\mu | z, \Pi_\mu) = \frac{1}{(2\pi |C|)^{\frac{n}{2}}} e^{-\frac{1}{2}(z-Hx)^T C^{-1} (z-Hx)}$$

where $C = HPH^T + R$

# Final Combined Estimate

$$\hat{x} = \sum_{\mu} \hat{y}_{\mu} \frac{p(\mu | z, \Pi_{\mu})}{\sum_{\nu} p(\nu | z, \Pi_{\nu})}$$