Part IV

ADVANCED ISSUES IN MPC

Contents

Chapter 1

STATE-SPACE MODEL PREDICTIVE CONTROL

1.1 SHORTCOMINGS OF CURRENT INDUSTRIAL MPC PRACTICE

- Truncated Step Response Model:
  - Many model coefficients have to be stored:

    Example) 5 x 5 system with 30 step response coefficients on each gives 750 coefficients.

    The problem is much worse for systems with mixed time scale dynamics (e.g. a high-purity distillation column) where sample time needs to be chosen according to the fast time-scale dynamics, but the settling time is determined by the slow time-scale dynamics.

    This limits the size of application.
  - Unstable systems cannot be handled.
  - Truncation error is unavoidable.

- Disturbance estimation:
  - Step disturbance is assumed and, thus, drift, ramp, oscillatory disturbances cause poor performance.
  - No cross channel update.
  - Unmeasured outputs are not updated.

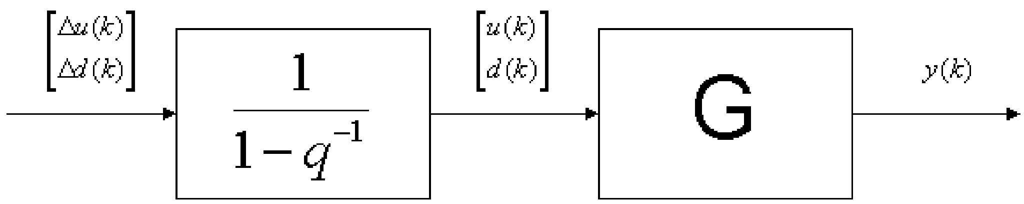
These shortcomings motivate development of

MPC based on a general state-space model.

1.2 STATE SPACE MPC

State Space Plant Model

Consider state space model of the plant obtained from either fundamental ODE's or system identification:



$$\begin{aligned} x(k+1) &= Ax(k) + B_u u(k) + B_d d(k) \\ y(k) &= Cx(k) \end{aligned}$$

$\Downarrow$ differencing

$$\begin{aligned} \Delta x(k+1) &= A\Delta x(k) + B_u \Delta u(k) + B_d \Delta d(k) \\ \Delta y(k) &= C\Delta x(k) \end{aligned}$$

$x(k)$: state
$u(k)$: control input
$y(k)$: measurement output
$d(k)$: measured disturbances

- The number of coefficients is reduced.

  Example) For 5 x 5 system with 10 states, only 200 coefficients need to be stored

- For appropriate choice of A, state space model can represent unstable process.
- No truncation error.

Prediction with State Space Plant Model

If we constrain that $\Delta u(k+m|k) = \cdots = \Delta u(k+p-1|k) = 0$,

$$\begin{bmatrix} \tilde{y}(k+1|k) \\ \tilde{y}(k+2|k) \\ \vdots \\ \tilde{y}(k+p|k) \end{bmatrix} = \begin{bmatrix} \Xi\Phi \\ \Xi\Phi^2 \\ \vdots \\ \Xi\Phi^p \end{bmatrix} X(k|k) + \begin{bmatrix} \Xi ,_d \\ \Xi\Phi ,_d \\ \vdots \\ \Xi\Phi^{p-1} ,_d \end{bmatrix} d(k)$$
$$+ \begin{bmatrix} \Xi ,_u & 0 & \cdots & 0 \\ \Xi\Phi ,_u & \Xi ,_u & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Xi\Phi^{p-1} ,_u & \Xi\Phi^{p-2} ,_u & \cdots & \Xi\Phi^{p-m} ,_u \end{bmatrix} \Delta\mathcal{U}(k)$$

Rewriting the above,

$$\Downarrow$$

$$\mathcal{Y}(k+1|k) = \mathcal{S}^X X(k|k) + \mathcal{S}^d \Delta d(k) + \mathcal{S}^{\mathcal{U}} \Delta\mathcal{U}(k)$$

1.3 DISTURBANCE ESTIMATION VIA STATE ESTIMATION

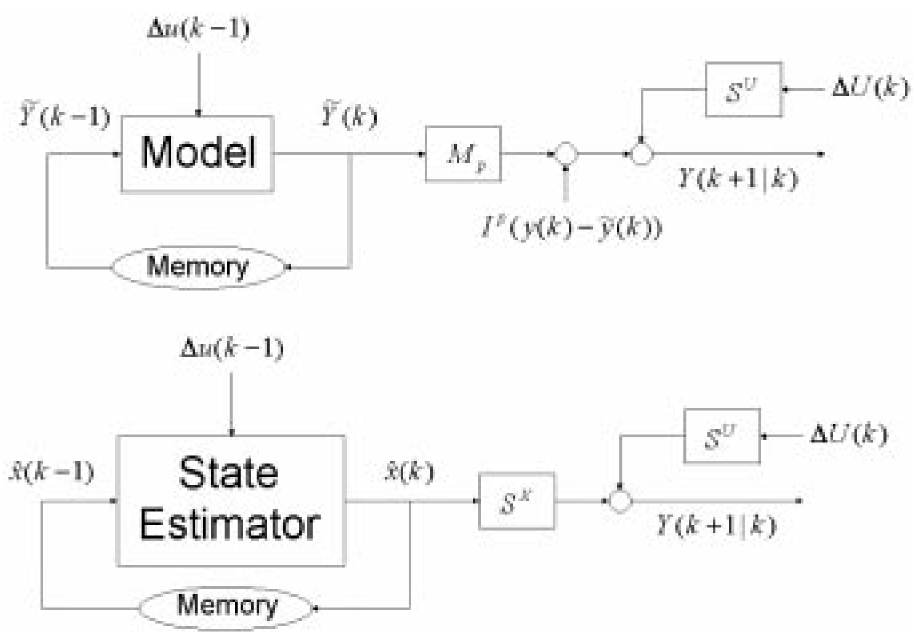
Motivation

In current industrial MPC algorithms,

- models are run open-loop
- feedback is entered into the prediction statically (no memory of the past feedbakc)

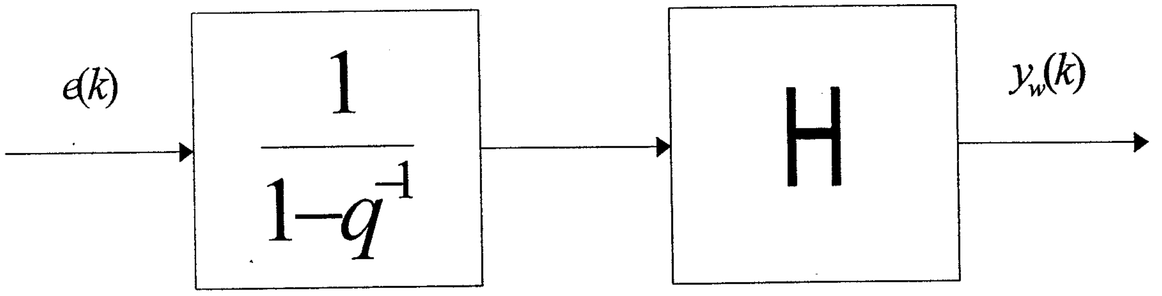
In disturbance estimation via state estimation

- unmeasured disturbance effects are included in the memory (state vector) and update is made directly to the states.
- fuller use of feedback measurement is allowed.



State Space Disturbance Model

State space disturbance model:



$$\begin{aligned} \Delta x_w(k+1) &= A_w \Delta x_w(k) + B_w e(k) \\ \Delta y_w(k) &= C_w \Delta x_w(k) + D_w e(k) \end{aligned}$$

State Space Disturbance Model Development

1. Assume somthing reasonable:

   - Step disturbance to output:

   $$\Delta y_w(k) = e(k)$$

   - Ramp disturbance to output:

   $$\begin{aligned} \Delta x_w(k+1) &= \Delta x_w(k) + e(k) \\ \Delta y_w(k) &= \Delta x_w(k) + e(k) \end{aligned}$$

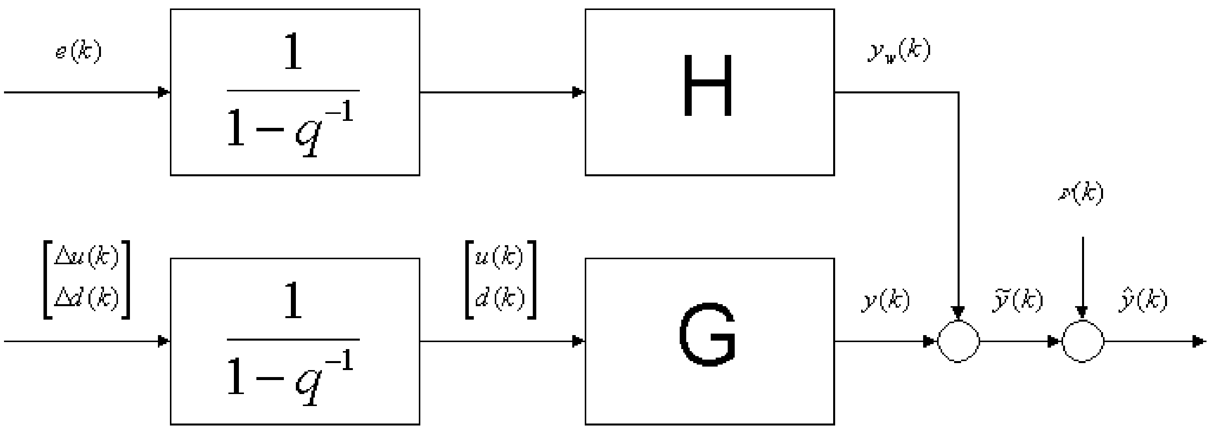
2. From funndamental ODE's: unmeasured disturbances in ODE's.

   $$\begin{aligned} \Delta x_w(k+1) &= A\Delta x_w(k) + B_w e(k) \\ \Delta y_w(k) &= C\Delta x_w(k) \end{aligned}$$

3. From Historical Plant Deta: Given historical plant deta, the stochastic state space model of the disturbance can be obtained using various techniques like spectral factorization and subspace identification.

Overall State Space Model

Plant with noise and state space disturbance model:



Overall model:

$$
\begin{bmatrix} \Delta x(k+1) \\ \Delta x_w(k+1) \\ e(k+1) \\ \tilde{y}(k+1) \end{bmatrix}
= \begin{bmatrix} A & 0 & 0 & 0 \\ 0 & A_w & 0 & 0 \\ 0 & 0 & 0 & 0 \\ CA & C_wA_w & C_wB_w & I \end{bmatrix}
\begin{bmatrix} \Delta x(k) \\ \Delta x_w(k) \\ e(k) \\ \tilde{y}(k) \end{bmatrix}
+ \begin{bmatrix} B_u \\ 0 \\ 0 \\ CB_u \end{bmatrix} \Delta u(k)
+ \begin{bmatrix} B_d \\ 0 \\ 0 \\ CB_d \end{bmatrix} \Delta d(k)
+ \begin{bmatrix} 0 \\ 0 \\ I \\ D_w \end{bmatrix} e(k+1)
$$

$$
\hat{y}(k) = [0 \; 0 \; 0 \; I] \begin{bmatrix} \Delta x(k) \\ \Delta x_w(k) \\ e(k) \\ \tilde{y}(k) \end{bmatrix} + \nu(k)
$$

Overall State Space Model (Continued)

$\Downarrow$ denote the above as

$$\begin{aligned} X(k+1) &= \Phi X(k) + ,_u \Delta u(k) + ,_d \Delta d(k) + ,_e e(k+1) \\ \hat{y}(k) &= \Xi X(k) + \nu(k) \end{aligned}$$

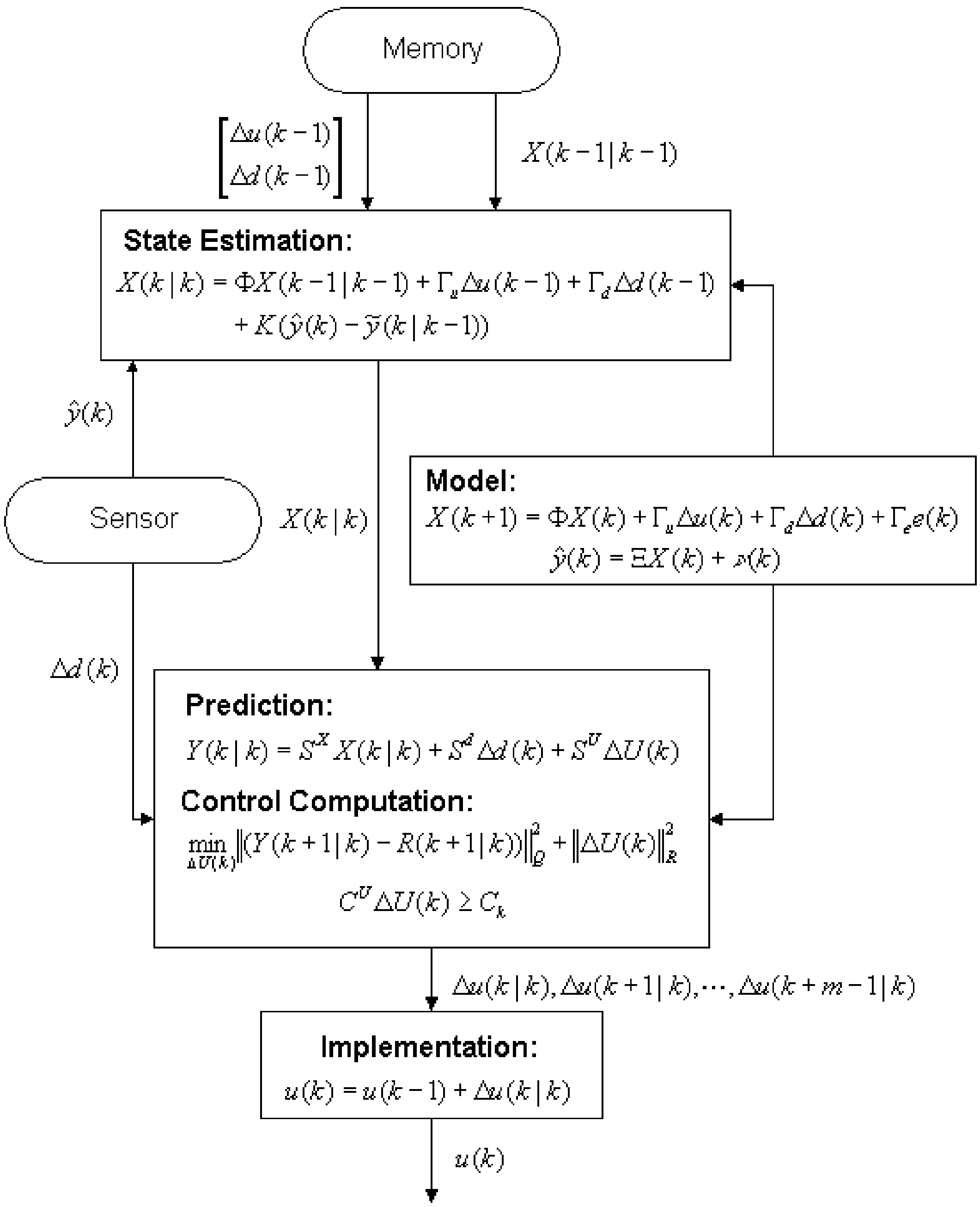
Given state space disturbance model, disturbance estimation can be done in a systematic way using well known Kalman filtering technique.

$$\begin{aligned} X(k|k-1) &= \Phi X(k-1|k-1) + ,_u \Delta u(k-1) + ,_d \Delta d(k-1) \\ X(k|k) &= X(k|k-1) + K(\hat{y}(k) - \tilde{y}(k|k-1)) \end{aligned}$$

- Recursive feedback update
- Various disturbance shape can be handled
- Cross-channel update

1.4 MPC FORMULATION USING STATE-SPACE MODEL

Overview



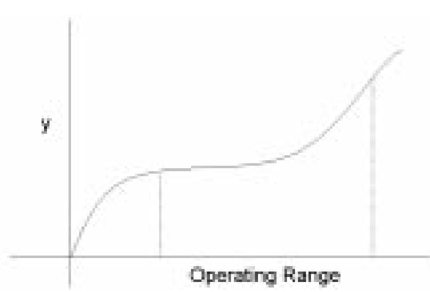
Chapter 2

NONLINEAR AND ADAPTIVE MODEL PREDICTIVE CONTROL

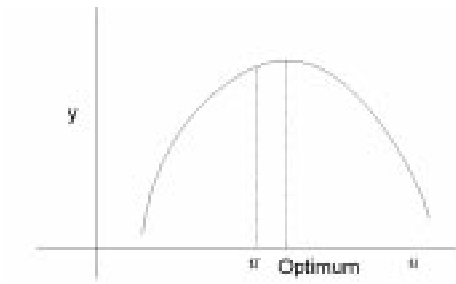
2.1 MOTIVATION

Why Nonlinear and Adaptive MPC?

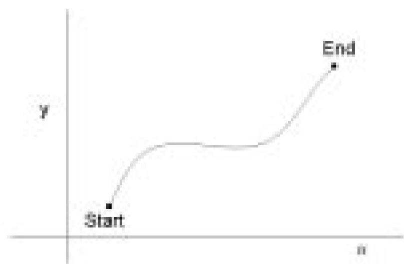
- Continuous processes with wide operating ranges



- Continuous processes with very strong nonlinearity (e.g., exothermic CSTR operated close to the optimum yield).



- Batch processes or other transition processes



These applications motivate development of

Nonlinear MPC or Adaptive MPC.

2.2 ISSUES IN NONLINEAR MPC

Issues

- **Nonlinear Models:** If first principle nonlinear ODE model is not available, do we have appropriate nonlinear system identification tools?
- **State Estimation:** At $t = k$,

  $$x(k-1|k-1), u(k-1), d(k-1), y(k-1) \quad \Longrightarrow \quad x(k|k)$$

  The *open-loop* model prediction can be done through nonlinear model integration. However, the measurement correction is much more difficult. For instance, is linear gain correction

  $$x(k|k) = x(k|k-1) + K(\hat{y}(k) - y(k|k-1))$$

  sufficient? Also, how should we choose the gain matrix K?
- **Control Computation:**
  The prediction equation is no longer linear in the future input moves, i.e.,

  $$\mathcal{Y}(k+1|k) = \tilde{\mathcal{F}}(x(k|k), d(k), \Delta\mathcal{U}(k))$$

  Since we have nonlinear prediction constraints, the optimization is no longer QP and can be computationally expensive and unreliable (e.g., local minima).

Nonlinear Models

- First principle nonlinear ODE models

$$\begin{aligned} \frac{dx}{dt} &= f(x, u, d, w) \\ \hat{y} &= g(x) \end{aligned}$$

  We will focus on this type of model in this lecture.

- Nonlinear difference equation model for nonlinear system identification

$$\begin{aligned} x(k+1) &= f(x(k), u(k), d(k)) \\ \hat{y}(k) &= g(x(k)) \end{aligned}$$

  - Artificial neural networks
  - Nonlinear series expansion models such as Volterra model and and NARX model.
  - Rectilinear models where g and/or g are piece-wise linear.
  - Linear model plus static nonlinearity
    - Hammerstein Model: input nonlinearity
    - Wiener Model: output nonlinearity

Generally, one shoud be very careful using a nonlinear model fitted to open loop data as the model can behave very differently when the loop is closed.

2.3 LINEARIZATION BASED NONLINEAR MPC

Standard Model

The standard model that we will use is of the following form:

$$\begin{aligned} \frac{dx}{dt} &= f(x,u,d,w) \\ \hat{y} &= g(x) + \nu \end{aligned}$$

We express the unmeasured disturbance w using the following stochastic equation driven by zero-mean white noise sequence $e(k)$:

$$\begin{aligned} x^e(k+1) &= A_e x^e(k) + B_e e(k) \\ w(k) &= C_e x^e(k) \end{aligned}$$

EXAMPLE: If $A_e = 1, B_e = 1, C_e = 1$, we have

$$w(k) = w(k-1) + e(k) \Longrightarrow \Delta w(k) = e(k)$$

This means $w(k)$ is a random step.

We will assume the above random step model for $w(k)$ **for simplicity**.

Standard Model (Continued) Combining the two equations give

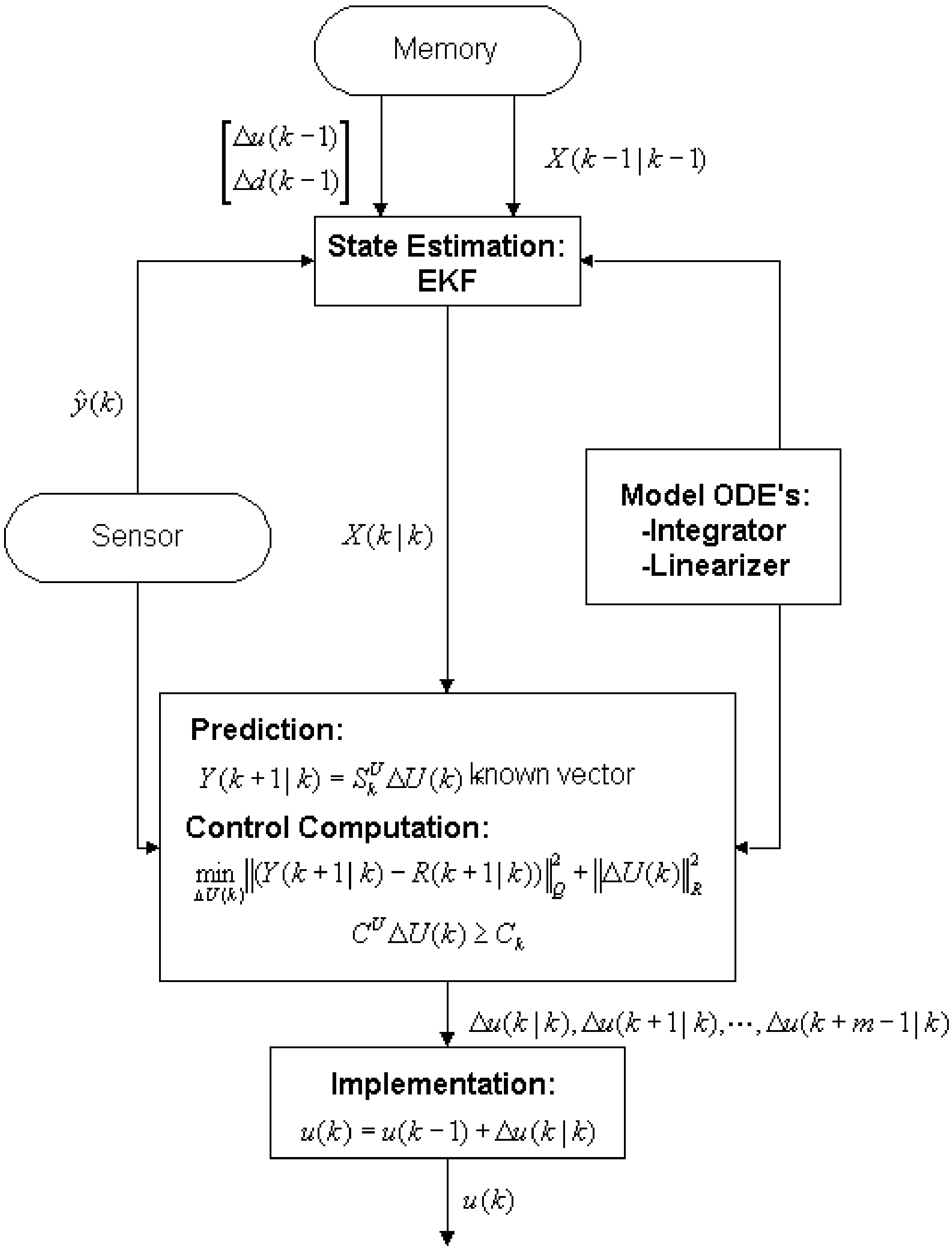
$$X(k+1) \triangleq \begin{bmatrix} x(k+1) \\ w(k+1) \end{bmatrix} = \begin{bmatrix} F_{t_s}(x(k), u(k), d(k), w(k)) \\ w(k) \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} e(k)$$

where $F_{t_s}(x(k), u(k), d(k), w(k))$ stands for the state vector resulting from integrating the ODE for one sample interval (from $t = k$ to $t = k+1$) with initial condition $x(k)$ and constant inputs of $u(t) = u(k)$ and $d(t) = d(k)$ and $w(t) = w(k)$.

We can also write the measurement equation as

$$\hat{y}(k) = g(x(k)) + \nu(k)$$

Overview of Linearization Based NLMPC



State Estimation

The following two steps are performed at $t = k$:

- **Model Prediction**

$$X(k|k-1) \triangleq \begin{bmatrix} x(k|k-1) \\ w(k|k-1) \end{bmatrix}$$

$$= \begin{bmatrix} F_{t_s}(x(k-1|k-1), u(k-1), d(k-1), w(k-1|k-1)) \\ w(k-1|k-1) \end{bmatrix}$$

Hence, this step involves nonlinear ODE integration for one sample interval.

- **Measurement Correction**

$$X(k|k) = X(k|k-1) + K_k(\underbrace{\hat{y}(k) - y(k|k-1)}_{\text{prediction error}})$$

where $y(k|k-1) = g(X(k|k-1))$.

K_k is the update gain ("filter gain"):

  - Linear update structure is retained (suboptimal).
  - The update gain needs to be varied with time due to the nonlinearity.
  - The gain matrix can be computed using the model linearized with respect to the current state estimate and using linear filtering theory $\Longrightarrow$ Extended Kalman Filter (see the attached paper by Lee and Ricker for details).

Prediction

One can follow the similar argument as before and construct

$$\begin{bmatrix} x(k+1|k) \\ x(k+2|k) \\ \vdots \\ x(k+p|k) \end{bmatrix} = \underbrace{\begin{bmatrix} F_{t_s}(x(k|k), u(k-1), d(k), w(k|k)) \\ F_{t_s}(x(k+1|k), u(k-1), d(k), w(k+1|k)) \\ \vdots \\ F_{t_s}(x(k+p-1|k), u(k-1), d(k), w(k+p-1|k)) \end{bmatrix}}_{\mathcal{F}:\text{ from ODE integration}}$$

$$+ \underbrace{\begin{bmatrix} B_k^u & 0 & \cdots & 0 \\ A_k B_k^u + B_k^u & B_k^u & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{j=1}^{p} A_k^{j-1} B_k^u & \Sigma_{j=1}^{p-1} A_k^{j-1} B_k^u & \cdots & \Sigma_{j=1}^{p-m+1} A_k^{j-1} B_k^u \end{bmatrix} \begin{bmatrix} \Delta u(k|k) \\ \Delta u(k+1|k) \\ \vdots \\ \Delta u(k+m-1|k) \end{bmatrix}}_{\mathcal{S}_k^{\mathcal{U}}:\text{ dynamic matrix}}$$

where $w(k+i|k) = w(k|k)$ and A_k and B_k^u are computed through

- *Linearization*

$$\tilde{A}_k = \left(\frac{\partial f}{\partial x}\right)_{x(k|k), u(k-1), d(k), w(k|k)}; \quad \tilde{B}_k = \left(\frac{\partial f}{\partial u}\right)_{x(k|k), u(k-1), d(k), w(k|k)}$$

- *Discretization*

$$A_k = \exp\left(\tilde{A}_k \cdot t_s\right); \qquad B_k^u = \int_0^{t_s} \left(\tilde{A}_k \cdot \tau\right) d\tau \cdot \tilde{B}_k^u$$

Denote the above as

$$\begin{aligned} \mathcal{X}(k+1|k) &= \mathcal{F}(x(k|k), u(k-1), d(k), w(k|k)) \\ &\quad + \mathcal{S}_k^{\mathcal{U}}(x(k|k), u(k-1), d(k), w(k|k)) \Delta\mathcal{U}(k) \end{aligned}$$

Summary

At $t = k$, we are given the previous estimate $(x(k-1|k-1), w(k-1|k-1))$, previous inputs $d(k-1), u(k-1)$, and new measurements $\hat{y}(k)$, $d(k)$. The following steps need to be performed:

1. **1-Step Model Integration:** Integrate the ODE for one time interval to obtain

$$X(k|k-1) \triangleq \begin{bmatrix} x(k|k-1) \\ w(k|k-1) \end{bmatrix}$$

$$= \begin{bmatrix} F_{t_s}(x(k-1|k-1), u(k-1), d(k-1), w(k-1|k-1)) \\ w(k-1|k-1) \end{bmatrix}$$

2. **Model Linearization:** Linearize the ODE and the measurement model with respsect to $X(k-1|k-1)$ and $X(k|k-1)$.

3. **Filter Gain Computation:** Obtain the filter gain matrix K_k using the linearized model matrices (see the details in the attached paper by Lee and Ricker).

4. **Measurement Update of $X(k)$:** Update the estimate for $X(k)$ based on the model prediction error:

$$X(k|k) = X(k|k-1) + K_k(\underbrace{\hat{y}(k) - y(k|k-1)}_{\text{prediction error}})$$

5. **Model Linearization:** Linearize the ODE model again with respect to the updated state $X(k|k)$.

6. **Dynamic Matrix Computation:** Use the linearized model matrices to construct the dynamic matrix $\mathcal{S}_k^{\mathcal{U}}(x(k|k), u(k-1), d(k), w(k|k))$ according to the formula given earlier.

7. **p-Step Model Integration:** Integrate the ODE model for p time steps starting from $x(k|k)$ and keeping inputs constant at $u(t) = u(k-1)$, $d(t) = d(k)$ and $w(t) = w(k|k)$ for $k \leq t < k+p$.

   The prediction equation is

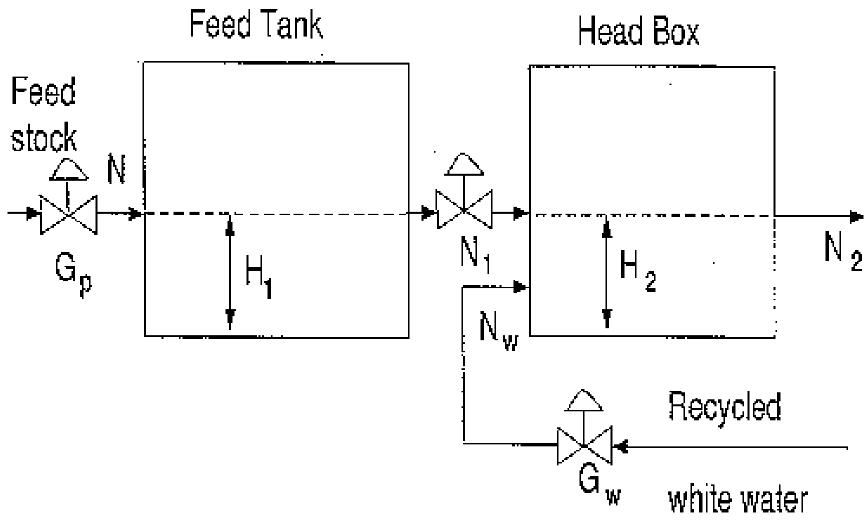
$$\begin{aligned} \mathcal{X}(k+1|k \; &= \; \mathcal{F}(x(k|k), u(k-1), d(k), w(k|k)) \\ &\quad + \mathcal{S}_k^{\mathcal{U}}(x(k|k), u(k-1), d(k), w(k|k))\Delta\mathcal{U}(k) \end{aligned}$$

8. **Input Computation:** Solve QP to find $\mathcal{U}(k)$.

9. **Input Implementation:** Implement $u(k) = u(k-1) + u(k|k)$.

2.4 EXAMPLE: PAPER MACHINE HEADBOX CONTROL

Problem Description



N: consistency of stock entering the feed tank.

N_w: consistency of recycled white water.

G_p: flowrate of stock entering the feed tank.

G_w: flowrate of recycled white water.

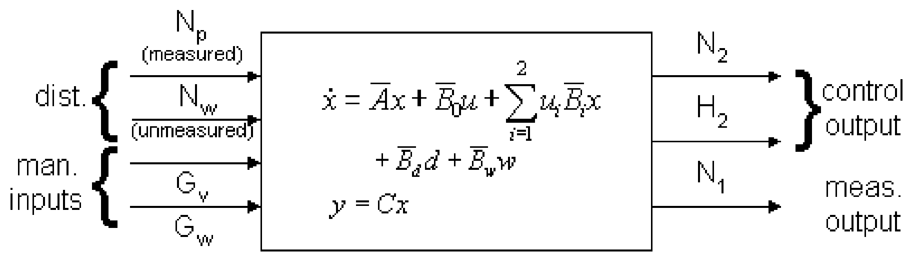
H_1: liquid level in the feed tank.

H_2: liquid level in the headbox.

N_1: consistency in the feed tank.

N_2: consistency in the headbox.

Some Specific Design Information



- ODE model for the above process is bilinear (see Lee and Ricker for model equations).
- We model N_w unmeasured disturbance as random walk, i.e.,

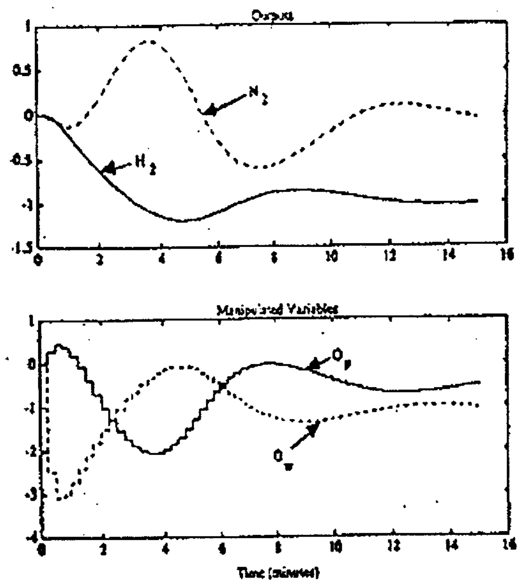
$$\begin{aligned} x^e(k+1) &= x^e(k) + e(k) \\ N^w(k) &= x^e(k) \end{aligned}$$

- We used the extended Kalman filter for state update.
- We used the following parameters for control computation:

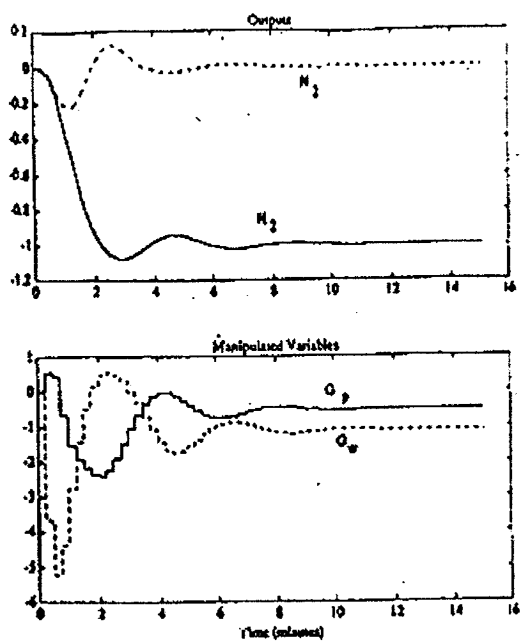
$$p = 5, \quad m = 3, \quad , {}^{y} = \text{diag}\{1,1,0\}, \quad , {}^{u} = \lambda \text{diag}\{1,1\}$$

Comparison Between Linear MPC and LB NLMPC

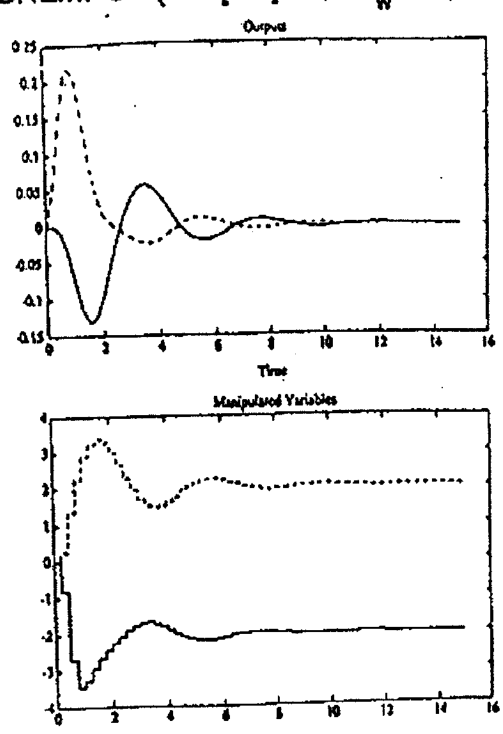
Linear MPC $(r = [0\ -1]^T\ , N_w = 0)$



LBNLMPC $(r = [0\ -1]^T\ , N_w = 0)$



LBNLMPC $(r = [0\ 0]^T\ , N_w = 10)$



2.5 ADDITIONAL ISSUES

Refinement

- **State Estimation:**
  - Computation of state estimates based on the full nonlinear model (e.g., moving horizon estimation) instead of the linearized model. Again, this requires NLP (computationally much more expensive).
- **Input Computation:**
  - Repetition of linearization and input trajectory calculation for better linearized model ($\Longrightarrow$ better dynamic matrix).
  - Replace linearized model based prediction equation

$$\begin{aligned}\mathcal{X}(k+1|k) &= \mathcal{F}(x(k|k), u(k-1), d(k), w(k|k)) \\ &\quad + \mathcal{S}_k^{\mathcal{U}}(x(k|k), u(k-1), d(k), w(k|k))\Delta\mathcal{U}(k)\end{aligned}$$

    with nonlinear algebraic constraints obtained from discretization (e.g., orthogonal collocation). This requires NLP instead of QP in control computation, however.

Alternatives

- **Gain scheduling:** separate model for different operating regimes.
- **Adaptive MPC:** recursive update of model parameters.

2.6 RECURSIVE PARAMETER ESTIMATION

Adaptation via Recursive Parameter Identification

State space representation of general model structure for parametric identification is

$$\begin{aligned} X(k+1) &= \Phi(\theta)X(k) + ,_u(\theta)\Delta u(k) + ,_d(\theta)\Delta d(k) + ,_e(\theta)e(k) \\ \hat{y}(k) &= \Xi X(k) + \nu(k) \end{aligned}$$

- Initiation Step: Initial parameter estimate, θ_0^* is obtained using I/O data from PRBS tests.

$$\Downarrow$$

$$\begin{aligned} X(k+1) &= \Phi(\theta_0^*)X(k) + ,_u(\theta_0^*)\Delta u(k) + ,_d(\theta_0^*)\Delta d(k) + ,_e(\theta_0^*)e(k) \\ \hat{y}(k) &= \Xi X(k) + \nu(k) \end{aligned}$$

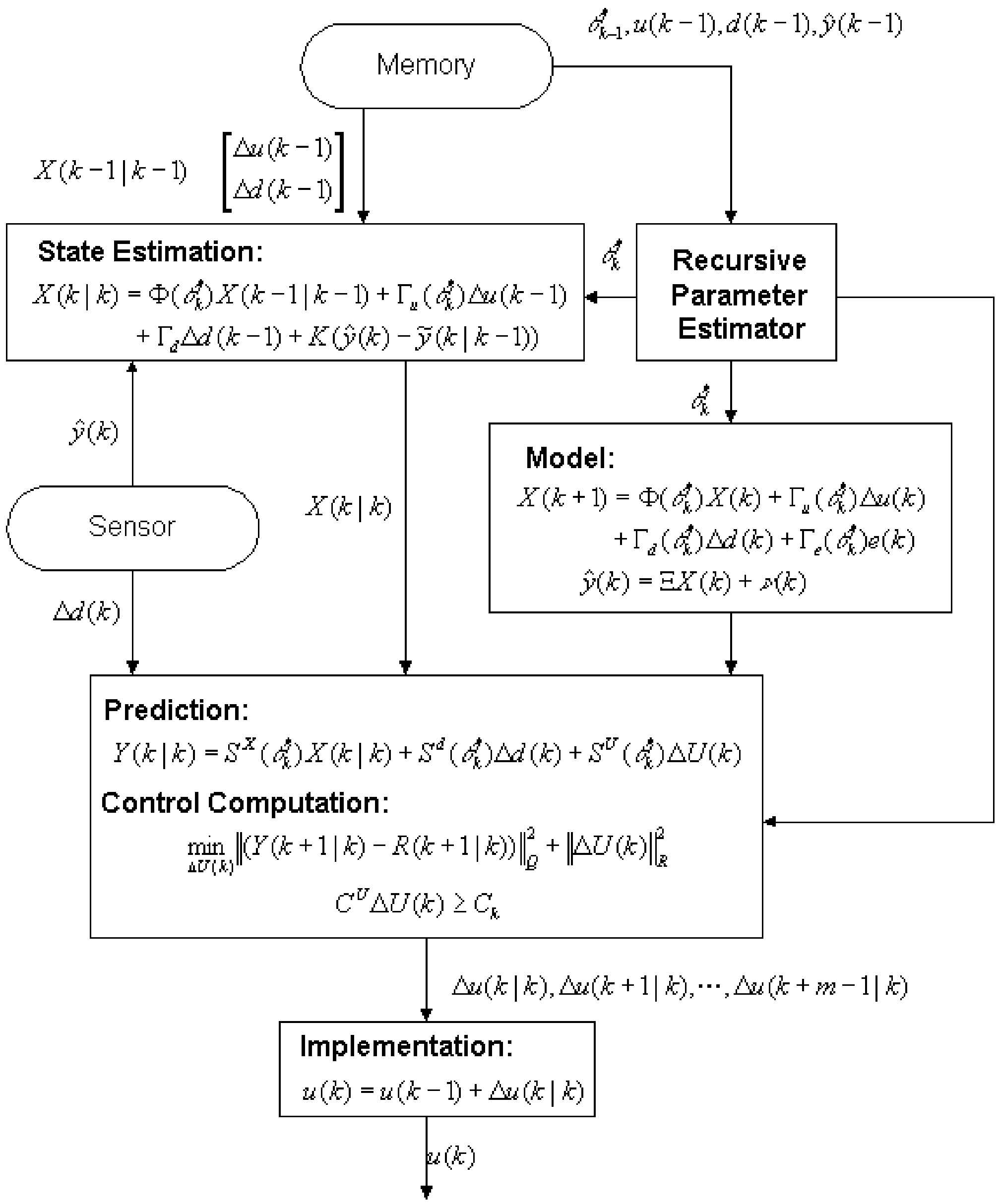
- kth Sampling time: Given $(k-1)$th parameter estimate θ_{k-1}^*, θ_k^* is obtained using $\theta_{k-1}^*, u(k-1), d(k-1), \hat{y}(k-1)$.

$$\Downarrow$$

$$\begin{aligned} X(k+1) &= \Phi(\theta_k^*)X(k) + ,_u(\theta_k^*)\Delta u(k) + ,_d(\theta_k^*)\Delta d(k) + ,_e(\theta_k^*)e(k) \\ \hat{y}(k) &= \Xi X(k) + \nu(k) \end{aligned}$$

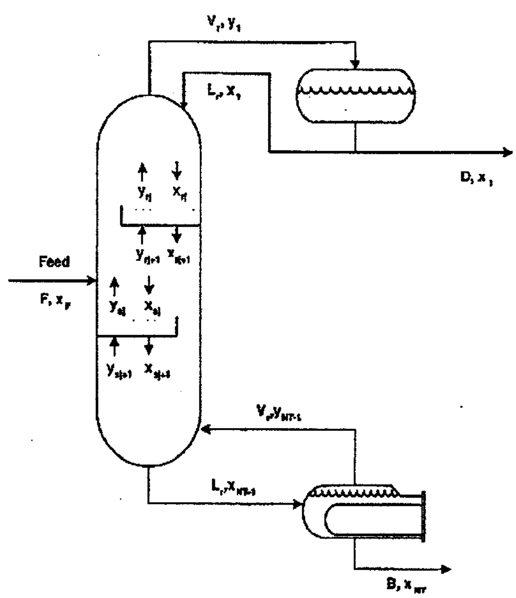
2.7 ADAPTIVE MPC FORMULATION

Overview



2.8 EXAMPLE: BINARY DISTILLATION COLUMN

Problem Description



x, y: liquid and vapor compositions.

D, B: overhead and bottom products.

L, V: liquid and vapor flow rate.

F, x_F: feed and feed composition.

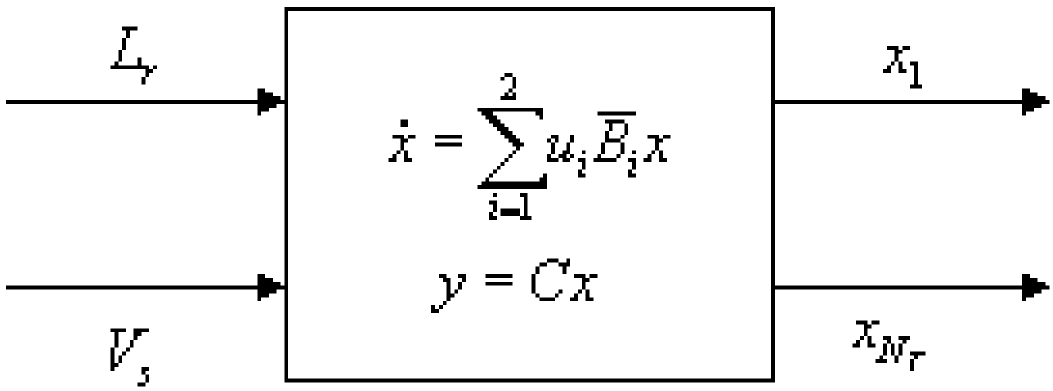
q: feed quality.

H: liquid holdup.

K: vapor-liquid equilibrium constant.

Control objective: regulation of overhead and bottom compositions x_1, x_{N_T}

Some Specific Design Information



- ODE model for the above process is bilinear (see Yoon et al. for model equations).
- Recursive least-square estimation technique is employed for parameter adpatation
- No input and state constraints are imposed.
- The following horizons for objective and control are used:

$$p = 22, \quad m = 20$$

Simulations

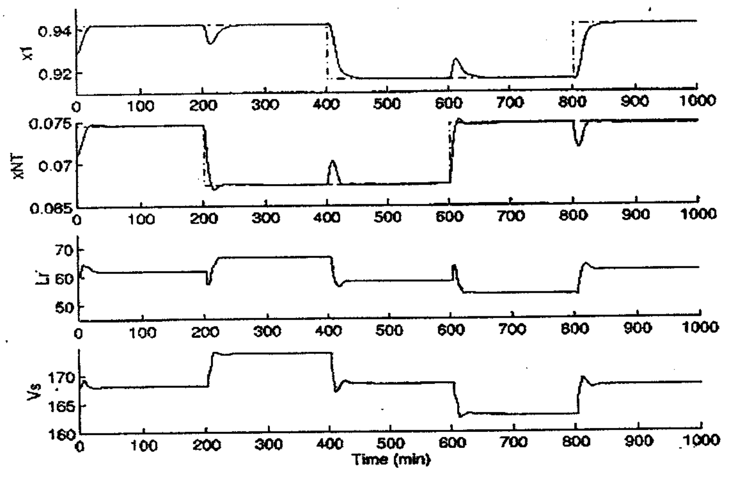


Figure 2.1: Responses to varying set-points

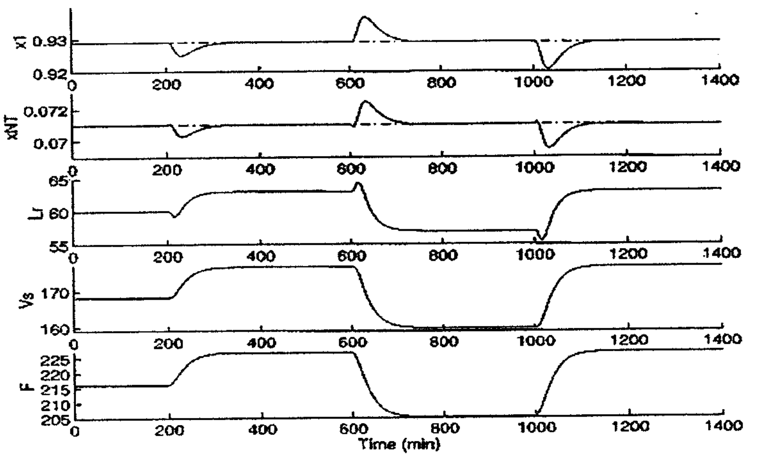


Figure 2.2: Responses to feed change

2.9 POTENTIAL IMPROVEMENTS IN SYSTEM IDENTIFICATION

Potential Improvements for Plant Testing

- *Simultaneous Excitation of Different Input Channels:* It is better to assimilate the situation a multivariable control system adjusts inputs altogether. An important issue then is

  how to coordinate inputs so that useful information is derived without causing problems to the operation.

  Perturbing each channel with an independent random signal seldom meets this requirement.

- *Control-Relevancy:* Since the ultimate purpose of a model is closed-loop control, the test should generate information that are important for control. Note that model uncertainty distribution is affected by information content in the data. Hence, the essence of the problem is

  how to distribute model uncertainty optimally for closed-loop control through data generation.

Potential Improvements for Plant Testing (Continued)

Plant-Friendliness: Plant tests during normal operations must be designed not to destroy the integrity of the on-going operation. This is particularly important when multiple input channels are to be excited simultaneously. In this case, one may lack intuition on how perturbations affect the key process variables. Plant friendliness can be achieved by incorporating

- input contraints (magnitude, rate, etc.)
- output constraints (formulated in a probabilistic manner)

Potential Improvements for Model Fitting

- *Deterministic vs. Stochastic Identification:* Deterministic identification is often adopted. However, including stochastic components can
  - improve the accuracy of the deterministic part.
  - yield a disturbance model useful for prediction.
- *SISO / MISO vs. MIMO Identification:*
  1. SISO or MISO identification is usually adopted but ignores the often-existing correlation among different output channels.
  2. MIMO identification
     - can potentially give a more accurate determinstic model, since disurbance effects are described more realistically.
     - allows *cross-channel* feedback update if the stochastic part is used for prediction.
     - is much more difficult in general.
     - suffers from identifiability problems and numerical difficulties (e.g., local minima) if time series models are used. There are so called subspace identification algorithms that allow direct construction of a state-space model in the following form:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + K\varepsilon(k) \\ y(k) &= Cx(k) + \varepsilon(k) \end{aligned}$$

Use of Historical Data

- Often times, disturbance model used for control is *assumed*. This model may not be useful for control if the assumed disturbance model is unrealistic.

- Evenif disturbance model used for control is identified from plant test data, it may not be useful for control provided that the data collected during the plant test do not contain the plant's representative disturbances.

- Plant's historical data are plenty and should contain the effect of various disturbances that enter the plant. Using such data and the deterministic system model, one can construct a stochastic model for residuals in the form of

$$\begin{aligned} x(k+1) &= Ax(k) + K\varepsilon(k) \\ y(k) &= Cx(k) + \varepsilon(k) \end{aligned}$$

  This can be combined with the deterministic model.