An Introduction to Bayesian System Identification, and my work about Positive System Identification
1. Introduction

The mathematical expression for a system

Figure 1. A mass-spring-damper system

Input: Force $F(t)$
Output: displacement $y(t)$

Suppose we can only observe input and output at $t = 1, 2 \ldots$.

(1) Discrete differential equations:

\[
y(t) + a_1 y(t - 1) + a_2 y(t - 2) = b F(t - 1)
\]

(2) Transfer function: z-transform

\[
Z(y(t - k)) = z^{-k} Y(z) \quad Z(y(t)) = Y(z)
\]

\[
G(z) = \frac{Y(z)}{F(z)} = \frac{b z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}}
\]

(3) State space equations:

\[
x(t) = \begin{bmatrix} y(t) \\ v(t) \end{bmatrix} \quad v(t) = \frac{dy}{dt} \approx \frac{y(t) - y(t - 1)}{1}
\]

\[
x(t + 1) = Ax(t) + BF(t) \quad y(t) = Cx(t)
\]

Those equations are necessary for analysis or designing a control law of a system.
1. Introduction

Transfer function: \( G(z) = \frac{Y(z)}{U(z)} \)

State space:

\[
\begin{align*}
    x(t + 1) &= Ax(t) + Bu(t) \quad & \text{State space:} \\
    y(t) &= Cx(t) \quad & \text{Differential equations:}
\end{align*}
\]

\[
y(t) + a_1 y(t - 1) + a_2 y(t - 2) = bF(t - 1)
\]
1. Introduction

**Category:**

- **System expression:**
  - Differential equations
  - Transfer function
  - State space equations

- **System classes:**
  - LTI (Linear time invariant) system
  - LPV (Linear parameter-varying) system
  - Nonlinear system
  - Block-oriented nonlinear system
1. Introduction

Category:

System expression

- Differential equations (system parameters)
- Transfer function
- State space equations

System classes

- LTI (Linear time invariant) system
- LPV (Linear parameter-varying) system
- Nonlinear system
- Block-oriented nonlinear system
1. Introduction

Problem formulation:

Discrete LTI system differential equations: FIR (Finite impulse response) model

\[ y(t) = \sum_{i=1}^{n} g(i)u(t - i) + v(t) \]

Given: \( \{ u(t), y(t), t = 1, \ldots, N \} \)

Unknown noise: \( v(t) \)

Estimate: system parameters \( g(i) = 1, \ldots, n \)

System parameters \( g(i) = 1, \ldots, n \) is equal to finite impulse responses of real system

Impulse response is

The output \( \{ y(t), t = 1, \ldots \} \)
if the input is a impulse signal \( u(t) = \begin{cases} 1, & t = 1 \\ 0, & \text{otherwise} \end{cases} \)
1. Introduction

A matrix form:

\[ Y_N = \Phi_N^T \theta + \Lambda_N \]

Where

\[ Y_N := (y(n + 1) \ldots y(N))^T \]
\[ \theta := (g(1) \ldots , g(N))^T \]
\[ \Lambda_N := (v(n + 1) \ldots v(N))^T \]

Objective: To estimate \( \theta \)
2. Bayesian system identification

Procedures:

Consider $\theta, \Lambda_N, Y_N$ to be random variables

1. Select a prior $p(\theta)$, assume the noise distribution $p(\Lambda_N)$.

$$Y_N = \Phi_T^N \theta + \Lambda_N \quad p(Y_N|\theta) = p(\Phi_T^N \theta + \Lambda_N)$$

2. Use Bayes’ Theorem to obtain posterior $p(\theta|Y_N)$.

$$p(\theta|Y_N) = \frac{p(\theta)p(Y_N|\theta)}{\int p(\theta)p(Y_N|\theta)d\theta}$$

3. Maximize posterior to obtain estimation of $\theta$.

$$\max_{\theta} : \quad p(\theta|Y_N)$$

2. Bayesian system identification

Examples:

(1)  Prior: $\theta \sim N(0, \xi I)$  Noise: $\Lambda_N \sim N(0, \sigma^2 I)$

$$\min_\theta : \frac{1}{\sigma^2} (Y_N - \Phi_N^T \theta)^T (Y_N - \Phi_N^T \theta) + \frac{1}{\xi} ||\theta||_{l2}^2$$

L2 regularization: Enforce smoothness in estimation

(2)  Prior: $p(\theta) \propto e^{-\xi ||\theta||}$  Noise: $\Lambda_N \sim N(0, \sigma^2 I)$

$$\min_\theta : \frac{1}{\sigma^2} (Y_N - \Phi_N^T \theta)^T (Y_N - \Phi_N^T \theta) + 2\xi ||\theta||_{l1}$$

L1 regularization: Enforce sparsity in estimation

Hyperparameters: $\xi, \sigma$  (decided in a preliminary step)

3. Positive system identification

Problem formulation:

\[
\{u(t), y(t), t = 1, \ldots, N\} \xrightarrow{\text{Estimation}} y(t) = \sum_{i=1}^{n} g(i)u(t - i) + v(t)
\]

\(v(t)\) is additive noise, we aim to estimate \(\{g(i), i = 1, \ldots, n\}\)

How to judge a externally positive system:

1. Impulse responses are all nonnegative.
2. Has nonnegative outputs if inputs are nonnegative.

Examples:

- Network flows: traffic, transport, communication
- Social science: population models
- Biology/Medicine: proteins
Gaussian prior is **not appropriate** because it contains \((-\infty, 0]\).

We consider to use a **maximum entropy prior** to derive a suitable prior for positive system identification.

**My Previous Result:**

A **truncated multivariate normal distribution** is the **maximum entropy prior** supported on \([0, \infty)\) with a known mean and a covariance.

\[
p(\theta) = \begin{cases} 
Ae^{-\theta^T\theta - \theta^T P^{-1} \theta}, & \theta \geq 0 \\
0, & otherwise
\end{cases}
\]

**Noise assumption:**
Assume the noise to be white Gaussian \(\Lambda_N \sim N(0, \sigma^2 I)\) and is independent with \(\theta\).
3. Positive system identification

Test systems:
In the following experiments, I use the following positive systems to generate data.

Fig. 3. Impulse responses of $F_1$

Transfer function:

$$F_1(z) = \frac{0.355z^2 + 0.02z + 1.2}{z^2 - 1.3z + 0.333}$$

Performance measure: $\text{fit}$

$$W(\hat{\theta}) = 100 \left( 1 - \frac{\sum_{t=1}^{M} (y_{test}(t) - \hat{y}(t|\hat{\theta}))^2}{\sqrt{\sum_{t=1}^{M} (y_{test}(t) - \overline{y}_{test})^2}} \right)$$

Fig. 3. Impulse responses of $F_1$
3. Positive system identification

Estimators:

G+DC (Normal Gaussian prior):

Prior: \( \theta \sim N(0, K) \)

Hyperparameters:

\[
K^{DC} = c_k \rho_k^{i+j/2} \lambda_k^{i-j} \quad c > 0, 1 > \rho, \lambda > 0
\]

Estimation:

\[
\min_{\theta} \quad \frac{1}{\sigma^2} ||Y_N - \Phi_N^T \theta||^2 + \theta^T K^{-1} \theta \\
\text{s.t.:} \quad \theta_i \geq 0, \quad i = 1, \ldots, n
\]
3. Positive system identification

Estimators:

**TG (Truncated normal prior): proposed**

Prior:

\[ p(\theta) = \begin{cases} 
Ae^{-a' T \theta - \theta^T P^{-1} \theta}, & \theta \geq 0 \\
0, & \text{otherwise}
\end{cases} \]

Hyperparameters:

\[ a_i' = c \rho^i \quad P_{ij} = c \rho^{\frac{i+j}{2}} \lambda^{i-j} \quad c > 0, \ 1 > \rho, \ \lambda > 0 \]

Estimation:

\[ \begin{align*}
\min_{\theta} \quad & \frac{1}{\sigma^2} ||Y_N - \Phi_N^T \theta||^2 + (\theta - \alpha')^T P(\theta - \alpha') \\
\text{s.t.} \quad & \theta_i \geq 0, \ i = 1, \ldots, n
\end{align*} \]
3. Positive system identification

Estimators:

**EXP (Exponential prior):**

Prior:

\[
p(\theta) = \begin{cases} 
\prod_{i=1}^{n} \frac{1}{a_i} e^{-\sum_{i=1}^{n} \frac{\theta_i}{a_i}}, & \theta_i \geq 0 \\
0, & \text{otherwise}
\end{cases}
\]

Hyperparameters:

\[a_i = c \alpha \rho_i^a\]

Estimation:

\[
\min_{\theta} \frac{1}{\sigma^2} ||Y_N - \Phi_N^T \theta||^2 + 2 \sum_{i=1}^{n} \frac{\theta_i}{a_i}
\]

s.t. \( \theta_i \geq 0, \ i = 1, \ldots, n \)
3. Positive system identification

Simulation results:

We use the three estimators for identifying $F_1$ 100 times.

Table 1. Mean and variance of fits for 100 identification by G+DC and TG

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>G+DC</td>
<td>66.6</td>
<td>486.2</td>
</tr>
<tr>
<td>TG</td>
<td>67.4</td>
<td>353.8</td>
</tr>
</tbody>
</table>

Performance:

$\text{TG} > \text{G+DC} > \text{EXP}$
4. Summary

1. I reviewed some fundamental knowledge in control system theory as well as system identification

2. I explained the Bayesian methods for identifying a linear time invariant system

3. I described my work for extending the Bayesian methods to identify positive systems and the simulations demonstrate the effectiveness of the proposed method.