Regularization Methods for System Identification

Input Design

Biqiang MU

Academy of Mathematics and Systems Science, CAS

Table of contents

Introduction

Identification in automatic control

Build a mathematical model for a dynamic system by the data in automatic control

- model

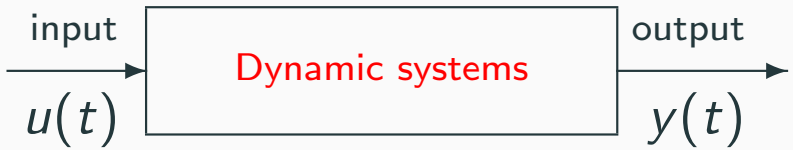
$$y(t) = f\big(x(t)\big) + v(t)$$

$$x(t) = \big[\underbrace{y(t-1), \cdots, y(t-n_y)}_{\text{delayed outputs}}, \underbrace{u(t-1), \cdots, u(t-n_u)}_{\text{delayed inputs}}\big]^T$$

- data

$$\{u(1), y(1), \cdots, u(N), y(N)\}$$

- goal: develop an estimate as well as possible



Two basic ways

1. estimation algorithm for given data
   - parametric models

$$y(t) = f(x(t), \theta) + v(t), \quad \widehat{\theta} = g(X, Y)$$

   - nonparametric models

$$y(t) = f(x(t)) + v(t), \quad \widehat{f}(x) = g(X, Y, x)$$

2. optimize the input for a chosen algorithm
   - parametric models (mean squared error)

$$\mathrm{MSE}(\widehat{\theta}) = E(\widehat{\theta} - \theta_0)(\widehat{\theta} - \theta_0)^T$$

$$U^* = \arg\min_U \ell\big(\mathrm{MSE}(\widehat{\theta})\big)$$

   - nonparametric models (goodness of fit)

$$\mathrm{GoF} = \frac{\sum_{t=1}^{N} \big(y(t) - \widehat{f}(x(t))\big)^2}{\mathrm{Var}(Y)}$$

$$U^* = \arg\min_U \mathrm{GoF}$$

Linear systems

The linear time-invariant (LTI) system identification is a classical and fundamental problem.

[1] Ljung, L. (1999). *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall.

Linear systems

The linear time-invariant (LTI) system identification is a classical and fundamental problem.

Output error systems (Ljung, 1999) [1]

$$y(t) = \sum_{k=1}^{\infty} g_k^0 u(t-k) + v(t), \;\; t = 1, 2, \cdots, \;\; v(t) \sim \mathscr{N}(0, \sigma^2)$$

[1] Ljung, L. (1999). *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall.

Linear systems

The linear time-invariant (LTI) system identification is a classical and fundamental problem.

Output error systems (Ljung, 1999) [1]

$$y(t) = \sum_{k=1}^{\infty} g_k^0 u(t-k) + v(t),\ t = 1, 2, \cdots,\ v(t) \sim \mathcal{N}(0, \sigma^2)$$

Impulse response functions

$$G(q, \theta_0) = \sum_{k=1}^{\infty} g_k^0 q^{-k},\ q^{-1} u(t+1) = u(t)$$

$$y(t) = G(q, \theta_0) u(t) + v(t),\ \theta_0 = [g_1^0, g_2^0, \cdots]^T$$

An LTI system is uniquely characterized by its impulse response.

It is an ill-posed problem

[1] Ljung, L. (1999). *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall.

Classical parametric methods

Parametric models

$$\sum_{k=1}^{\infty} g_k^0 q^{-k} = \frac{b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}}$$

- Model order selection: AIC, BIC, cross validation
- Parametric estimation methods: Maximum likelihood (ML), prediction error method (PEM), etc.

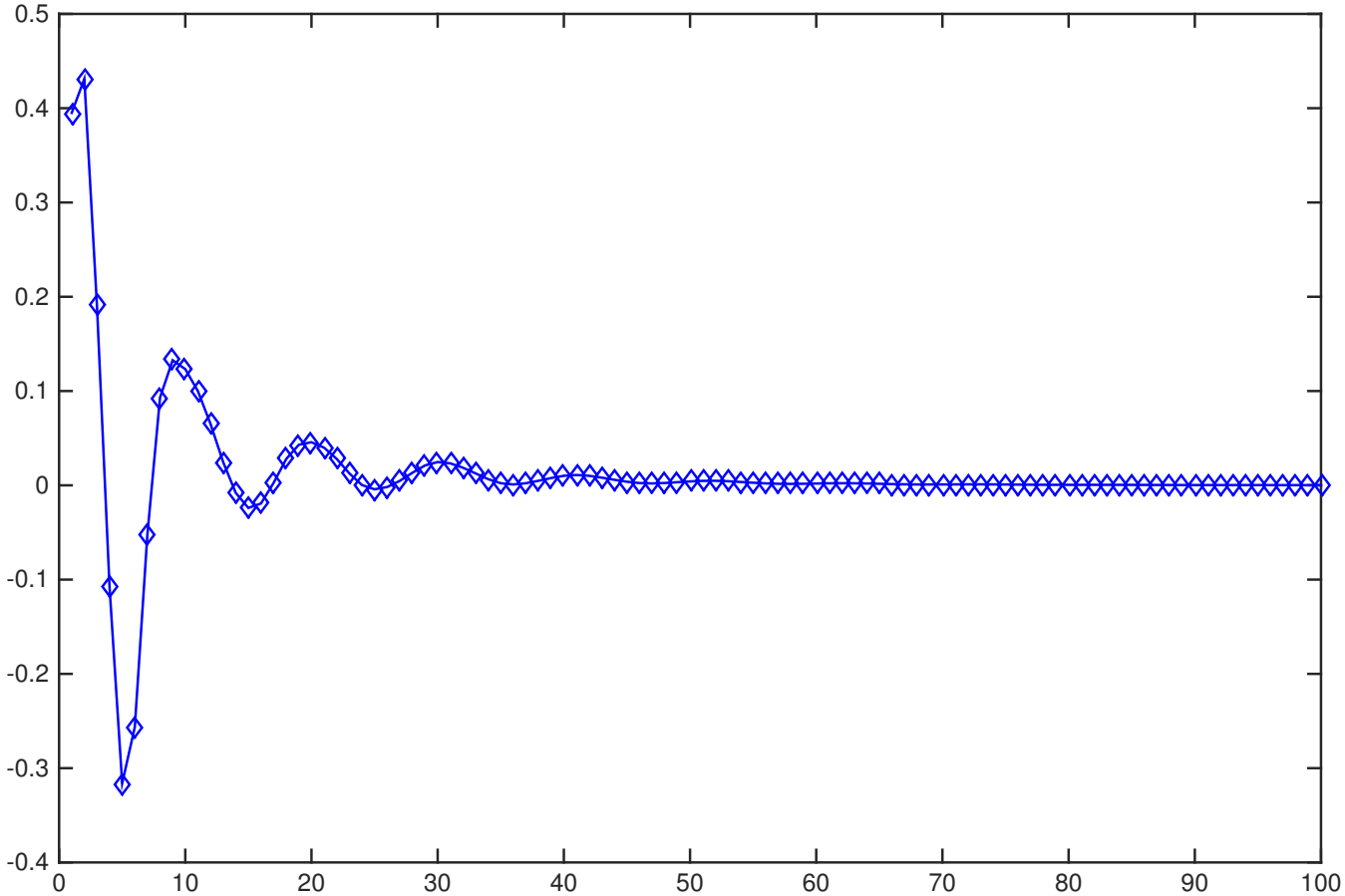
Asymptotic optimality

Regularization methods

Motivations

- small sample cases
- inputs with weakly persistent excitation
  - colored noise input
  - band-limited input

A typical impulse response sequence



A truncated high order finite impulse response system

$$\sum_{k=1}^{\infty} g_k^0 q^{-k} \to \sum_{k=1}^{n} g_k^0 q^{-k}$$

$$y(t) = \sum_{k=1}^{n} g_k^0 u(t-k) + v(t) = \varphi(t)^T \theta_0 + v(t)$$

$$Y = \Phi\theta_0 + V, \quad \theta_0 = [g_1^0, g_2^0, \ldots, g_n^0]^T$$

where

$$\Phi = \begin{bmatrix} u(0) & u(-1) & \ldots & u(-n+1) \\ u(1) & u(0) & \ldots & u(-n+2) \\ \vdots & \vdots & \ddots & \vdots \\ u(N-1) & u(N-2) & \cdots & u(N-n) \end{bmatrix}$$

$$Y = \begin{bmatrix} y(1) & y(2) & \ldots & y(N) \end{bmatrix}^T$$

$$V = \begin{bmatrix} v(1) & v(2) & \ldots & v(N) \end{bmatrix}^T$$

Least squares estimators

Least squares (LS) estimators

$$\widehat{\theta}^{\mathrm{LS}} \stackrel{\triangle}{=} \arg\min_{\theta \in \mathbb{R}^n} \|Y - \Phi^T \theta\|^2 = (\Phi^T \Phi)^{-1} \Phi^T Y$$

Mean squared error

$$\mathrm{MSE}(\widehat{\theta}^{\mathrm{LS}}) = E\left(\widehat{\theta}^{\mathrm{LS}} - \theta_0\right)\left(\widehat{\theta}^{\mathrm{LS}} - \theta_0\right)^T = \sigma^2 (\Phi^T \Phi)^{-1}$$

Regularization methods

The estimator:

$$\widehat{\theta}^{\mathrm{R}} = (\Phi^T \Phi + \sigma^2 K^{-1})^{-1} \Phi^T Y$$

Regularization methods

The estimator:

$$\widehat{\theta}^{\mathrm{R}} = (\Phi^T \Phi + \sigma^2 K^{-1})^{-1} \Phi^T Y$$

Three kinds of explanations

- Regularized least squares (RLS) estimators

$$\widehat{\theta}^{\mathrm{R}} \stackrel{\triangle}{=} \arg\min_{\theta \in \mathbb{R}^n} \|Y - \Phi\theta\|^2 + \sigma^2 \theta^T K^{-1} \theta$$

Regularization methods

The estimator:

$$\widehat{\theta}^{\mathrm{R}} = (\Phi^T \Phi + \sigma^2 K^{-1})^{-1} \Phi^T Y$$

Three kinds of explanations

- Regularized least squares (RLS) estimators

$$\widehat{\theta}^{\mathrm{R}} \stackrel{\triangle}{=} \arg \min_{\theta \in \mathbb{R}^n} \|Y - \Phi\theta\|^2 + \sigma^2 \theta^T K^{-1} \theta$$

- Guassian process (Bayesian explanation)

  Prior: $\theta_0 \sim \mathcal{N}(0, K)$ (K : Kernel matrix)

  Posterior: $\theta_0 | Y \sim \mathcal{N}(\widehat{\theta}^{\mathrm{R}}, \widehat{K}^{\mathrm{R}})$

Regularization methods

The estimator:

$$\widehat{\theta}^{\mathrm{R}} = (\Phi^T \Phi + \sigma^2 K^{-1})^{-1} \Phi^T Y$$

Three kinds of explanations

- Regularized least squares (RLS) estimators

$$\widehat{\theta}^{\mathrm{R}} \stackrel{\triangle}{=} \arg\min_{\theta \in \mathbb{R}^n} \|Y - \Phi\theta\|^2 + \sigma^2 \theta^T K^{-1} \theta$$

- Guassian process (Bayesian explanation)

Prior: $\theta_0 \sim \mathcal{N}(0, K)$ (K : Kernel matrix)

Posterior: $\theta_0 | Y \sim \mathcal{N}(\widehat{\theta}^{\mathrm{R}}, \widehat{K}^{\mathrm{R}})$

- Reproducing kernel Hilbert spaces (RKHSs)

$$\widehat{\theta}^{\mathrm{R}} \stackrel{\triangle}{=} \arg\min_{\theta \in \mathscr{J}} \|Y - \Phi\theta\|^2 + \gamma \|\theta\|^2_{\mathscr{J}}$$

A two step procedure

The seminal paper (Pillonetto & De Nicolao, 2010)[1]

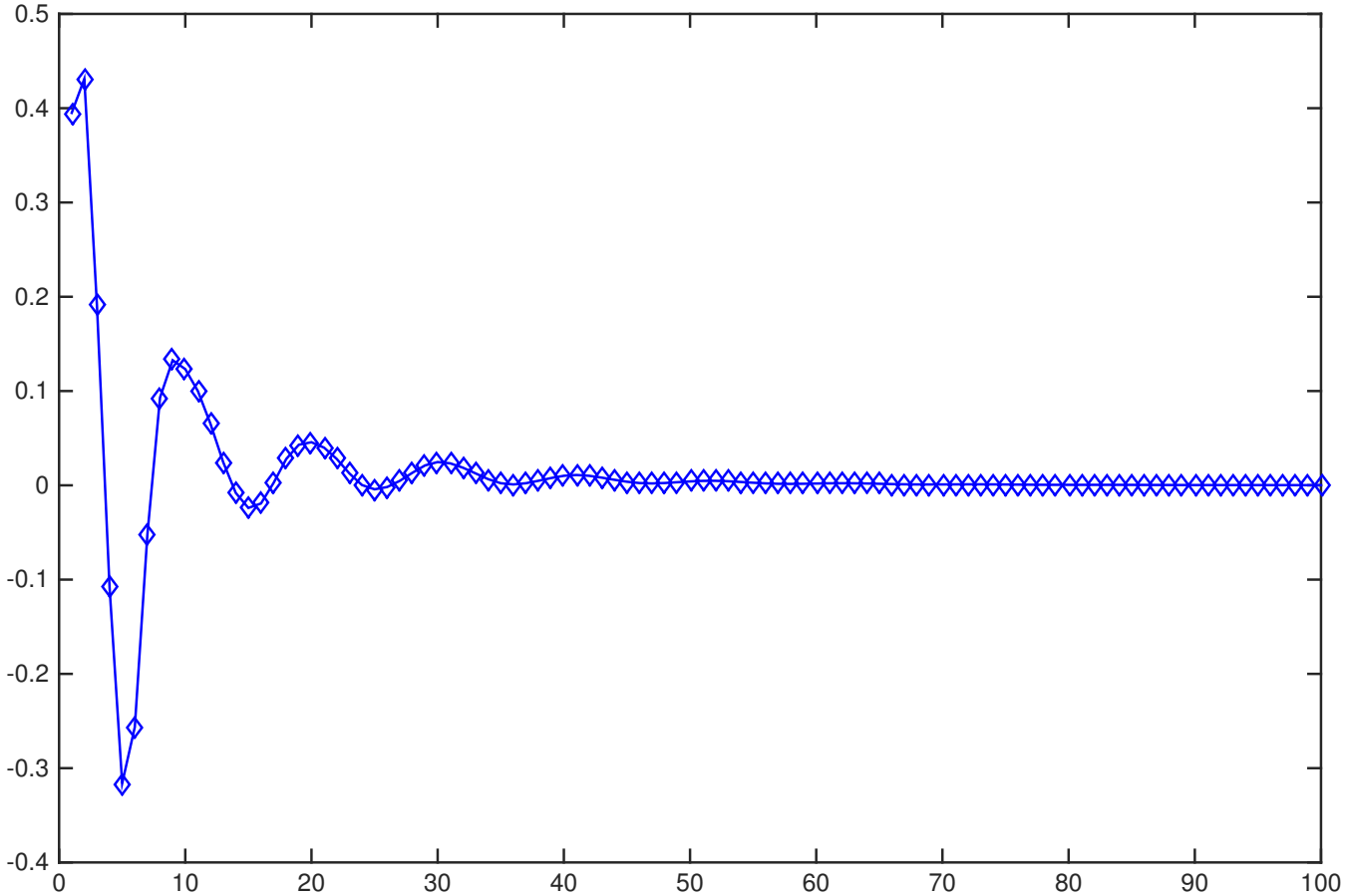
- Kernel design: parameterize K by using prior knowledge

$$K(\eta),\ \eta \text{ hyperparameter}$$

- Hyperparameter estimation: determine the hyperparameter by the data

[1] G. Pillonetto and G. De Nicolao. A new kernel-based approach for linear system identification. *Automatica, 46*, 81–93, 2010.

A typical impulse response sequence



Kernel design

TC kernel (Chen et al., 2012) [1]

$$K_{k,j}(\eta) = c\,\min(\lambda^k, \lambda^j),\ K(\eta) = c \begin{bmatrix} \lambda & \lambda^2 & \cdots & \lambda^n \\ \lambda^2 & \lambda^2 & \cdots & \lambda^n \\ \vdots & \ddots & \ddots & \vdots \\ \lambda^n & \lambda^n & \cdots & \lambda^n \end{bmatrix}$$

with hyperparameters $\eta = [c, \lambda]^T \in \Omega = \{c \geq 0, 0 \leq \lambda \leq 1\}$.

The estimator:

$$\widehat{\theta}^{\mathrm{R}}(\eta) = (\Phi^T \Phi + \sigma^2 K^{-1}(\eta))\Phi^T Y$$

[1] T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and Gaussian processes–Revisited. *Automatica, 48(8)*: 1525–1535, 2012.

Hyperparameter estimation

The goal

- estimate the hyperparameters based on the data

The essence

- tune model complexity in a continuous way

Some commonly used methods (Pillonetto et al., 2014) [1]

1. Empirical Bayes (EB)
2. Stein's unbiased risk estimator (SURE)
3. Cross validation (CV)

[1] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica, 50(3)*: 657–682, 2014.

Empirical Bayes

Gaussian prior

$$\theta \sim \mathcal{N}(0, K)$$
$$Y = \Phi\theta + V \sim \mathcal{N}(0, Q)$$
$$Q = \Phi K \Phi^T + \sigma^2 I_N$$

Empirical Bayes (EB)

$$\mathrm{EB} : \widehat{\eta}_{\mathrm{EB}} = \arg\min_{\eta \in \Omega} Y^T Q^{-1} Y + \log\det(Q)$$

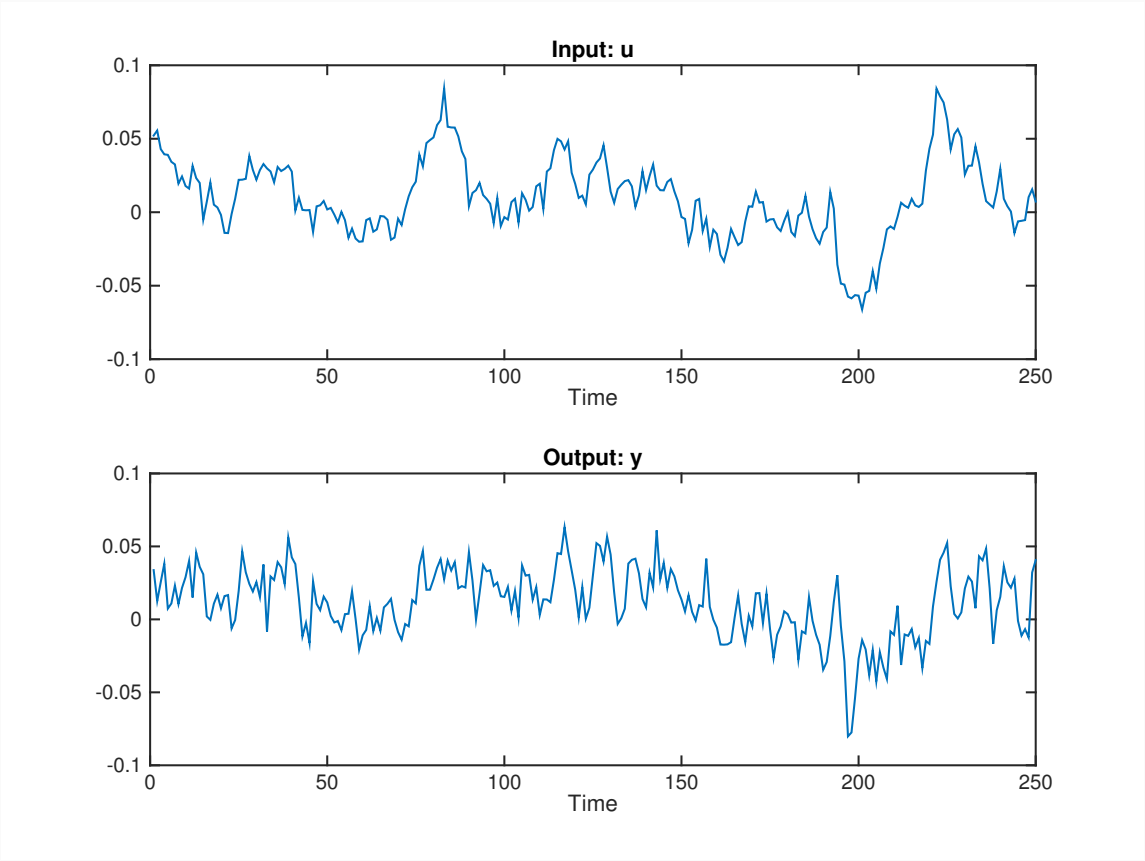
1. B. Mu, T. Chen and L. Ljung. On Asymptotic Properties of Hyperparameter Estimators for Kernel-based Regularization Methods. Automatica, 94: 381–395, 2018.
2. B. Mu, T. Chen and L. Ljung. Asymptotic Properties of Generalized Cross Validation Estimators for Regularized System Identification. Proceedings of the IFAC Symposium on System Identification, 203–205, 2018.
3. B. Mu, T. Chen and L. Ljung. Asymptotic Properties of Hyperparameter Estimators by Using Cross-Validations for Regularized System Identification. Proceedings of the IEEE Conference on Decision and Control, 644–649, 2018.

An example

Input-output data of a linear dynamic system:

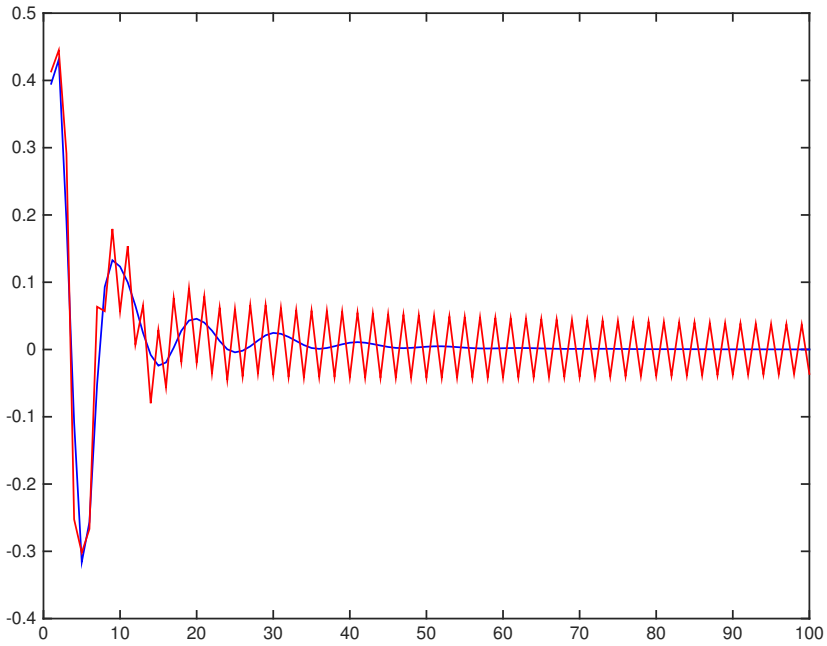
- Data size: 250
- Input: a filtered white noise
- Noise: a white noise with the signal to noise ratio 5.45

To estimate the first 100 impulse response coefficients

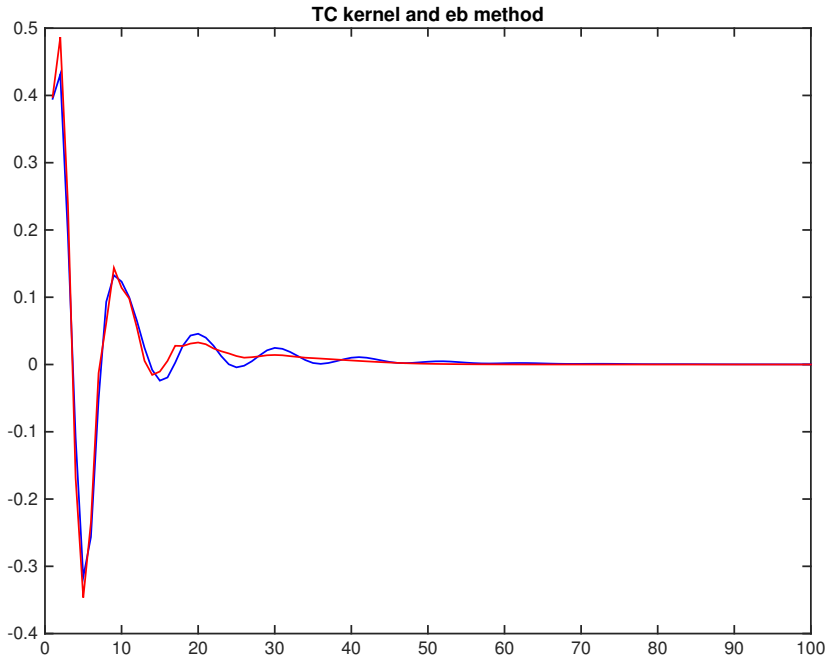


Estimation results

The OE-system of order 6 by CV



Regularization methods



Input design for regularization methods

What is input design?

Recall the truncated high order impulse response system

$$\begin{aligned} y(t) &= \sum_{k=1}^{n} g_k^0 u(t-k) + v(t) \\ &= \varphi(t)^T \theta_0 + v(t) \\ Y &= \Phi\theta_0 + V \end{aligned}$$

In particular

$$y(1) = g_1^0 u(0) + \cdots + g_n^0 u(-n+1) + v(t)$$

The goal

determine an input sequence

$$u_{-n+1}, \cdots, u_{-1}, u_0, u_1, \cdots, u_{N-1}$$

where, for simplicity, u_t is used to denote $u(t)$ hereafter, such that it

- minimizes some design criteria
- subject to certain constraints.

Input design in the ML/PEM framework

Least squares (LS) estimators

$$\widehat{\theta}^{\mathrm{LS}} \stackrel{\triangle}{=} \arg\min_{\theta} \|Y - \Phi^T\theta\|^2 = (\Phi^T\Phi)^{-1}\Phi^T Y$$

Mean squared error

$$\mathrm{MSE}(\widehat{\theta}^{\mathrm{LS}}) = E\left(\widehat{\theta}^{\mathrm{LS}} - \theta_0\right)\left(\widehat{\theta}^{\mathrm{LS}} - \theta_0\right)^T = \sigma^2(\Phi^T\Phi)^{-1}$$

Optimal inputs

$$\text{minimize } \det\left(\sigma^2(\Phi^T\Phi)^{-1}\right) \text{ subject to } \sum_{k=1}^{N} u_{k-i}^2 = \mathscr{E},\ i = 1, \cdots, n$$

The optimal inputs satisfy

$$\Phi^T\Phi = \mathscr{E} I_n$$

Several optimal inputs

- impulsive inputs $[\sqrt{\mathscr{E}}, 0, \cdots, 0]^T$
- white noise inputs

Problem formulation

The MSE matrix of the RLS estimate

$$M_N = \sigma^4 R^{-1} K^{-1} \theta_0 \theta_0^T K^{-1} R^{-1} + \sigma^2 R^{-1} \Phi^T \Phi R^{-1}$$
$$R = \Phi^T \Phi + \sigma^2 K^{-1}$$

Two methods

- replace θ_0 by a pilot estimate
- integrate out θ_0 under $\theta \sim \mathcal{N}(0, K)$

$$M_N = \sigma^2 (\Phi^T \Phi + \sigma^2 K^{-1})^{-1}$$

D-**optimality** (nonconvex)

$$\text{minimize } \det(\sigma^2 (\Phi^T \Phi + \sigma^2 K^{-1})^{-1})$$
$$\text{subject to } \sum_{k=1}^{N} u_{k-i}^2 = \mathscr{E}, \ i = 1, \cdots, n$$

Note that K in the objective function is assumed to be known here.

Periodic inputs

Suppose that the unknown inputs $u_{-n+1}, \cdots, u_{-1}$

$$u_{-i} = u_{N-i}, \quad i = 1, \cdots, n-1 \text{ (periodic)}$$

Thus the input design optimization becomes

$$u^* \stackrel{\triangle}{=} \arg\min_{u \in \mathscr{U}} \det(\sigma^2 R^{-1}), \; R = \Phi^T \Phi + \sigma^2 K^{-1}$$
$$\mathscr{U} = \{u \in \mathbb{R}^N | u^T u = \mathscr{E}\}, \; u = [u_0, u_1, \cdots, u_{N-1}]^T$$

Design matrix

The unknown inputs $u_{-n+1}, \cdots, u_{-1}$

$$u_{-i} = u_{N-i}, \quad i = 1, \cdots, n-1 \text{ (periodic)}$$

yield

$$\Phi = \begin{bmatrix} u_0 & u_{N-1} & \cdots & u_{N-n+2} & u_{N-n+1} \\ u_1 & u_0 & \cdots & u_{N-n+3} & u_{N-n+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ u_{n-2} & u_{n-3} & \cdots & u_0 & u_{N-1} \\ u_{n-1} & u_{n-2} & \cdots & u_1 & u_0 \\ u_n & u_{n-1} & \cdots & u_2 & u_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ u_{N-1} & u_{N-2} & \cdots & u_{N-n+1} & u_{N-n} \end{bmatrix}$$

the columns of which are circulant.

Convexity of design criteria

Toeplitz matrix

$$\Phi^T\Phi = \begin{bmatrix} r_0 & r_1 & \cdots & r_{n-2} & r_{n-1} \\ r_1 & r_0 & \ddots & r_{n-3} & r_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ r_{n-2} & r_{n-3} & \ddots & r_0 & r_1 \\ r_{n-1} & r_{n-2} & \cdots & r_1 & r_0 \end{bmatrix}$$

$$r_j = \sum_{k=0}^{N-1} u_k u_{k-j},\ j = 0, \ldots, n-1$$

is linear in the vector

$$r = [r_0, r_1, \cdots, r_{n-1}]$$

Convexity of design criteria

Toeplitz matrix

$$\Phi^T\Phi = \begin{bmatrix} r_0 & r_1 & \cdots & r_{n-2} & r_{n-1} \\ r_1 & r_0 & \ddots & r_{n-3} & r_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ r_{n-2} & r_{n-3} & \ddots & r_0 & r_1 \\ r_{n-1} & r_{n-2} & \cdots & r_1 & r_0 \end{bmatrix}$$

$$r_j = \sum_{k=0}^{N-1} u_k u_{k-j},\ j = 0, \ldots, n-1$$

is linear in the vector

$$r = [r_0, r_1, \cdots, r_{n-1}]$$

The cost function $\det(\sigma^2(\Phi^T\Phi + \sigma^2 K^{-1})^{-1})$ is convex in r.

Convexity of design criteria

Toeplitz matrix

$$\Phi^T\Phi = \begin{bmatrix} r_0 & r_1 & \cdots & r_{n-2} & r_{n-1} \\ r_1 & r_0 & \ddots & r_{n-3} & r_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ r_{n-2} & r_{n-3} & \ddots & r_0 & r_1 \\ r_{n-1} & r_{n-2} & \cdots & r_1 & r_0 \end{bmatrix}$$

$$r_j = \sum_{k=0}^{N-1} u_k u_{k-j},\ j = 0, \ldots, n-1$$

is linear in the vector

$$r = [r_0, r_1, \cdots, r_{n-1}]$$

The cost function $\det(\sigma^2(\Phi^T\Phi + \sigma^2 K^{-1})^{-1})$ is convex in r.

A quadratic transform from $u \in \mathbb{R}^N$ to $r \in \mathbb{R}^n$:

$$r = f(u) = [f_0(u), \cdots, f_{n-1}(u)]^T$$

Decomposition of the quadratic mapping

The decomposition of $f(\cdot)$:

$$f(u) = h_3(h_2(h_1(u))) \quad \text{with}$$

$$h_1(u) = W^T u, \quad h_2(z) = [z_0^2, z_1^2, \cdots, z_{N-1}^2]^T, \quad h_3(x) = Sx$$

where $x = [x_0, x_1, \cdots, x_{N-1}]^T$, $z = [z_0, z_1, \cdots, z_{N-1}]^T$, for even N,

Decomposition of the quadratic mapping

The decomposition of $f(\cdot)$:

$$f(u) = h_3(h_2(h_1(u))) \quad \text{with}$$
$$h_1(u) = W^T u, \quad h_2(z) = [z_0^2, z_1^2, \cdots, z_{N-1}^2]^T, \quad h_3(x) = Sx$$

where $x = [x_0, x_1, \cdots, x_{N-1}]^T$, $z = [z_0, z_1, \cdots, z_{N-1}]^T$, for even N,

$$W = \sqrt{\frac{2}{N}} \left[\frac{\xi_0}{\sqrt{2}}, \xi_1, \cdots, \xi_{\frac{N-2}{2}}, \frac{\xi_{\frac{N}{2}}}{\sqrt{2}}, \zeta_{\frac{N-2}{2}}, \cdots, \zeta_1 \right] \quad \text{(orthogonal matrix)}$$

$$S = \left[\xi_0, \xi_1, \cdots, \xi_{n-1} \right]^T$$

$$\xi_j = \begin{bmatrix} 1 \\ \cos(j\varpi) \\ \cos(2j\varpi) \\ \vdots \\ \cos((N-1)j\varpi) \end{bmatrix}, \quad \zeta_j = \begin{bmatrix} 0 \\ \sin(j\varpi) \\ \sin(2j\varpi) \\ \vdots \\ \sin((N-1)j\varpi) \end{bmatrix}, \quad j = 0, 1, \ldots$$

and $\varpi = 2\pi/N$.

Transformed input design problems

Convex optimization

$$r^* = \arg\min_{r \in \mathscr{F}} \log\det(\sigma^2 R^{-1})$$

where $\mathscr{F}$ is the image of $f(\cdot)$ under $\mathscr{U}$

$$\mathscr{F} \stackrel{\triangle}{=} \{f(u) | u \in \mathscr{U}\} \quad \text{(convex polytope)}$$
$$\mathscr{U} = \{u \in \mathbb{R}^N | u^T u \leq \mathscr{E}\}, \; u = [u_0, u_1, \cdots, u_{N-1}]^T$$
$$u_{-i} = u_{N-i}, \quad i = 1, \cdots, n-1$$

Finding optimal inputs

The mapping $f(\cdot)$

$$f(u) = h_3(h_2(h_1(u))) \quad \text{with}$$
$$h_1(u) = W^T u, \quad h_2(z) = [z_0^2, z_1^2, \cdots, z_{N-1}^2]^T, \quad h_3(x) = Sx$$

Finding optimal inputs

The mapping $f(\cdot)$

$$f(u) = h_3(h_2(h_1(u))) \text{ with}$$
$$h_1(u) = W^T u, \quad h_2(z) = [z_0^2, z_1^2, \cdots, z_{N-1}^2]^T, \quad h_3(x) = Sx$$

The inverse mapping of $f(\cdot)$

1. we find the inverse image of $h_3(\cdot)$ for $r \in \mathscr{F}$:

$$\mathscr{X}(r) \stackrel{\triangle}{=} \{x | Sx = r, x_i \geq 0\} \quad \text{(convex polytope)}$$

2. we find the inverse image of $h_2(\cdot)$ for $x \in \mathscr{X}(r)$:

$$\mathscr{Z}(r) \stackrel{\triangle}{=} \{z | h_2(z) \in \mathscr{X}(r)\} = \left\{[\pm\sqrt{x_0}, \cdots, \pm\sqrt{x_{N-1}}]^T | x \in \mathscr{X}(r)\right\}$$

3. we find the inverse image of $h_1(\cdot)$ for $z \in \mathscr{Z}(r)$:

$$\mathscr{U}(r) \stackrel{\triangle}{=} \left\{u | W^T u \in \mathscr{Z}(r)\right\} = \{Wz | z \in \mathscr{Z}(r)\}$$

Some special kernel matrices

Diagonal kernels:

- General diagonal kernel matrix $K = \text{diag}([\lambda_1, \lambda_2, \cdots, \lambda_n])$
  - Ridge kernel matrix $K = cI_n$, with $c > 0$
  - DI kernel matrix $K = \text{diag}(c[\lambda, \lambda^2, \cdots, \lambda^n]), c, \lambda > 0$.

$$r^* = [\mathscr{E}, \underbrace{0, \cdots, 0}_{n-1 \text{ zeros}}]^T \stackrel{\triangle}{=} r^\dagger, \quad \textit{namely}, \quad \Phi^T \Phi = \mathscr{E} I_n$$

- Typical "optimal" inputs:
  - Impulsive inputs: $u = [\sqrt{\mathscr{E}}, 0, \cdots, 0]^T$
  - White noise inputs asymptotically

TC kernel:

$$r^* \neq r^\dagger$$

An exmaple

system

$$y(t) = \sum_{k=1}^{n} g_k u(t-k) + v(t)$$
$$= \varphi(t)^T \theta + v(t)$$

setup

- the length of the data: $N = 50$
- the number of the parameters: $n = 50$

kernel

produced by a preliminary data

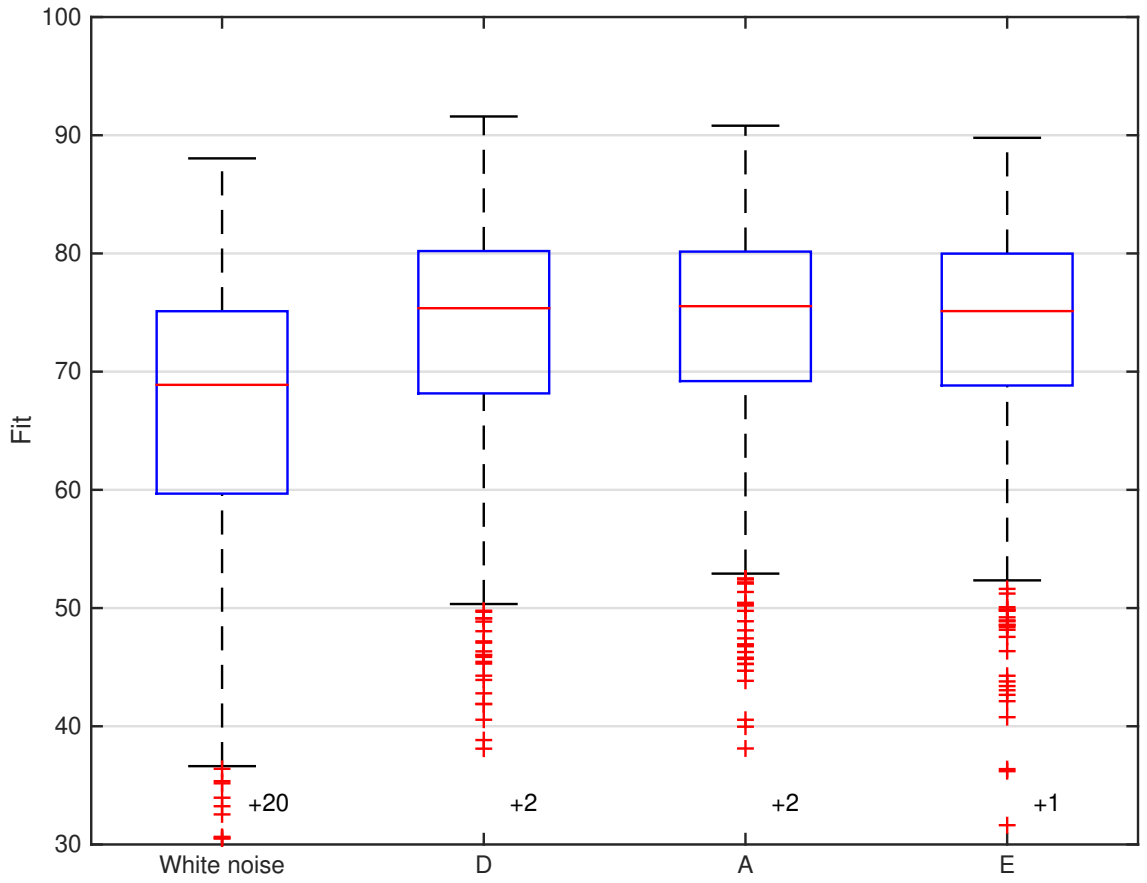
number of experiments

1000 runs

Performance measure

$$\text{Fit} = 100 \times \left(1 - \frac{\|\widehat{\theta}_{\text{im}} - \theta_0\|}{\|\theta_0 - \bar{\theta}_0\|}\right), \quad \bar{\theta}_0 = \frac{1}{n}\sum_{k=1}^{n} g_k^0$$

Simulation results



	White noise	D	A	E
Average fit	66.24	73.44	73.87	73.46

Conclusions

Conclusions

1. periodic inputs yield the input design suitable for finite sample size
2. formulate the input design for regularization methods as a nonconvex optimization problem in the Bayesian perspective
3. introduce a quadratic mapping to solve the optimization problem by two steps
4. transform the original nonconvex problem into a convex problem and to find the inverse of the quadratic mapping

Thanks for your listening

Questions?

References

Chen, T., Ohlsson, H., & Ljung, L. (2012). On the estimation of transfer functions, regularizations and gaussian processes–revisited. *Automatica*, *48*, 1525–1535.

Ljung, L. (1999). *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall.

Pillonetto, G., & De Nicolao, G. (2010). A new kernel-based approach for linear system identification. *Automatica*, *46*, 81–93.

Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., & Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, *50*, 657–682.