227-0689-00L System Identification **21 November 2023**

Solution 8: Regularization

Problem 1:

It is desired to identify a parameter $\theta \in \mathbb{R}^n$ in a least-squares manner using the m-sample observation process (with regressor matrix $\Phi \in \mathbb{R}^{m\times n}$)

$$\beta = \Phi\theta + E, \tag{8.1}$$

The error vector $E \in \mathbb{R}^m$ is assumed to have mean zero and covariance $\sigma^2 I$.

a) Find an explicit expression for the least-squares estimate $\theta^{LS} = \operatorname{argmin}_\theta \|\Phi\theta - \beta\|_2^2$

b) Let $P \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix and let $\gamma \geq 0$ be a regularization constant. Find an expression for $\theta^P = \operatorname{argmin}_\theta \|\Phi\theta - \beta\|_2^2 + \gamma\theta^T P^{-1}\theta$

c) Prove that θ^{LS} and θ^P are related together by

$$\theta^P = ((\Phi^T\Phi + \gamma P^{-1})^{-1}\Phi^T\Phi)\theta^{LS}. \tag{8.2}$$

d) Calculate the bias vectors and the covariance matrices of θ^{LS} and θ^R.

Solution

a) The least squares estimate is formed by setting the gradient of $\|\Phi\theta - \beta\|_2^2$ to zero

$$0 = 2\Phi^T(\Phi\theta^{LS} - \beta), \qquad \theta^{LS} = (\Phi^T\Phi)^{-1}(\Phi^T\beta). \tag{8.3}$$

b) The regularized expression may be calculated via similar principles:

$$0 = 2\Phi^T(\Phi\theta^P - \beta) + 2P^{-1}\theta^P, \qquad \theta^P = (\Phi^T\Phi + P^{-1})^{-1}(\Phi^T\beta). \tag{8.4}$$

c) Equation (8.2) holds by considering

$$\theta^P = (\Phi^T\Phi + P^{-1})^{-1}(\Phi^T\beta) \tag{8.5a}$$
$$= (\Phi^T\Phi + P^{-1})^{-1}(\Phi^T\Phi)(\Phi^T\Phi)^{-1}(\Phi^T\beta) \tag{8.5b}$$
$$= (\Phi^T\Phi + P^{-1})^{-1}(\Phi^T\Phi)\theta^{LS}. \tag{8.5c}$$

d) The bias of the least squares estimator is

$$\begin{aligned}
\mathbb{E}[\theta^{LS}] - \theta &= \mathbb{E}[(\Phi^T\Phi)^{-1}\Phi^T(\beta)] - \theta && (8.6a)\\
&= \mathbb{E}[(\Phi^T\Phi)^{-1}\Phi^T(\Phi\theta + E)] - \theta && (8.6b)\\
&= \mathbb{E}[(\Phi^T\Phi)^{-1}(\Phi^T\Phi)\theta] - \mathbb{E}[(\Phi^T\Phi)^{-1}\Phi^T(-E)] - \theta && (8.6c)\\
&= \theta - \theta = 0, && (8.6d)
\end{aligned}$$

because the mean of E is zero.

The covariance of the least squares estimator is

$$\begin{aligned}
\mathrm{Cov}(\theta^{LS}) &= \mathrm{Cov}((\Phi^T\Phi)^{-1}\Phi^T(\beta)) && (8.7a)\\
&= \mathrm{Cov}((\Phi^T\Phi)^{-1}\Phi^T(\Phi\theta + E)) && (8.7b)\\
&= \mathrm{Cov}(\theta + (\Phi^T\Phi)^{-1}\Phi^T E)) && (8.7c)\\
&= \mathrm{Cov}((\Phi^T\Phi)^{-1}\Phi^T E)) && (8.7d)\\
&= (\Phi^T\Phi)^{-1}\Phi^T\mathrm{Cov}(E)((\Phi^T\Phi)^{-1}\Phi^T)^T && (8.7e)\\
&= \sigma^2(\Phi^T\Phi)^{-1}. && (8.7f)
\end{aligned}$$

The bias of the regularized estimator is

$$\begin{aligned}
\mathbb{E}[\theta^P] - \theta &= \mathbb{E}[((\Phi^T\Phi + \gamma P^{-1})^{-1}\Phi^T\Phi)\theta^{LS}] - \theta && (8.8a)\\
&= (I - ((\Phi^T\Phi + \gamma P^{-1})^{-1}\Phi^T\Phi))\theta. \qquad \neq 0 \text{ (in general)} && (8.8b)
\end{aligned}$$

justifying that the regularized estimator is biased.

The covariance of the regularized estimator is

$$\begin{aligned}
\mathrm{Cov}(\theta^P) &= \mathrm{Cov}((\Phi^T\Phi + \gamma P^{-1})^{-1}\Phi^T\Phi\theta^{LS}) && (8.9a)\\
&= (\Phi^T\Phi + \gamma P^{-1})^{-1}\Phi^T\Phi\mathrm{Cov}(\theta^{LS})((\Phi^T\Phi + \gamma P^{-1})^{-1}\Phi^T\Phi)^T && (8.9b)\\
&= (\Phi^T\Phi + \gamma P^{-1})^{-1}(\Phi^T\Phi)[\sigma^2(\Phi^T\Phi)^{-1}](\Phi^T\Phi)(\Phi^T\Phi + \gamma P^{-1})^{-1} && (8.9c)\\
&= \sigma^2(\Phi^T\Phi + \gamma P^{-1})^{-1}(\Phi^T\Phi)(\Phi^T\Phi + \gamma P^{-1})^{-1} && (8.9d)
\end{aligned}$$

Problem 2:

This problem will involve system identification of a linear system subject to abrupt transitions in dynamics.

Let T be a time horizon and $k \in 1..T$ be the time index. The order-r system model at time k will be expressed as $[a(k), b(k)]$ with

$$G(\mathrm{e}^{j\omega}; k) = \frac{\sum_{p=1}^{r} b_p(k)e^{pj\omega}}{1 + \sum_{p=1}^{r} a_p(k)e^{pj\omega}}. \tag{8.10}$$

The observations $\{u(k), y(k)\}_{k=-r+1}^{T}$ are collected from system (8.10).

a) Formulate a regressor matrix Φ, a vector β and a parameter vector θ such that the following least squares task has the cost function:

$$\theta^{LS} = \underset{\theta}{\operatorname{argmin}} \|\Phi\theta - \beta\|_2^2 = \underset{[a(k),b(k)]}{\operatorname{argmin}} \left(\sum_{k=1}^{T} (y(k) + \sum_{p=1}^{r} a_p(k) y(k-p) - \sum_{p=1}^{r} b_p(k) u(k-p) \right)^2 \tag{8.11}$$

b) Find an explicit expression for the least-squares parameter estimate θ^{LS} from (8.11)

c) A regularizer $R(\theta) = \sum_{k=1}^{T-1} \left((a(k+1) - a(k))^2 + (b(k+1) - b(k))^2 \right)$ is added to the least squares expression. Find an explicit expression for the regularized estimate θ^R from

$$\theta^R = \underset{\theta}{\operatorname{argmin}} \|\Phi\theta - b\|_2^2 + \gamma R(\theta) \tag{8.12}$$

d) What is the function of the regularizer R? What does the regularizer penalize? How does the regularizer help to detect abrupt transitions?

Solution

This question is based on the work in [1].

a) The parameter vector θ is the concatenation $[a(1); a(2); \ldots; a(T); b(1); b(2); \ldots; b(T)]$. The vector β is formed by $[y(1); y(2), \ldots, y(T)]$. The regressor matrix Φ has a structure similar to a block-diagonal and Toeplitz matrices. The regressor Φ may be composed of the horizontal concatenation $\Phi = [\Phi_a, \Phi_b]$, under the definitions

$$\Phi_a = \operatorname{blkdiag}([y(0), y(-1), \ldots, y(-r+1)], \ldots [y(T-1), y(T-2), \ldots, y(T-r)]) \tag{8.13}$$

$$\Phi_b = \operatorname{blkdiag}([u(0), u(-1), \ldots, u(-r+1)], \ldots [u(T-1), u(T-2), \ldots, u(T-r)]). \tag{8.14}$$

b) Under the definitions from part a), the least-squares estimate is $\theta^{LS} = (\Phi^T \Phi)^{-1} (\Phi^T \beta)$

c) Letting $\otimes$ be the Kronecker product operator, $\mathbf{1}_{1\times r}$ be a ones vector of size $1 \times r$, and I_n as an $n \times n$ identity matrix; we can define the differencing matrix D as

$$D = I_{T-1} \otimes (\mathbf{1}_{1\times r} \otimes [-1, 1]). \tag{8.15}$$

The regularizer $R(\theta)$ may be interpreted as

$$R(\theta) = \|D\theta\|_2^2. \tag{8.16}$$

The regularized system estimate is

$$\theta^R = (\Phi^T \Phi + \gamma D^T D)^{-1} (\Phi^T \beta). \tag{8.17}$$

d) The regularizer $R(\theta)$ may be interpreted as a sum-of-norms regularization penalty on the time-step-adjacent models. Such sum-of-norms regularizers tend to promote sparsity [2, 3] in the groups (e.g., group lasso). Under an appropriate choice of regularization parameter γ, only a small number of time-step differences will be nonzero (k such that $[a(k); b(k)] \neq [a(k-1); b(k-1)]$). The indices k with nonzero time-step differences will correspond to the identified abrupt transitions in system dynamics.

MATLAB **Exercise:**

This example will involve pulse-response estimation of an unknown linear system given observations inputs $u(k)$ and outputs $y(k)$. The ground truth (IIR) pulse response $g(k)$ will be approximated with an order-r FIR model $\hat{g}(k)$ under the relation (with error term $e(k)$):

$$y(k) = \sum_{i=0}^{\infty} g(i)u(k-i) + e(k) \tag{8.18}$$

$$y(k) \approx \sum_{i=0}^{r} \hat{g}(i)u(k-i) + e(k). \tag{8.19}$$

Assume that $u(k) = 0$ and $y(k) = 0$ for all $k \leq 0$.

Download the following file from Moodle to perform this exercise:

- `HS2023_SysID_Exercise_08_GenerateData.p`

The tasks of this exercise are as follows:

a) Acquire input-output observations by running `[u, y] = HS2023_SysID_Exercise_08_GenerateData(LegiNumber)`. The input is 10 periods of a length 255 PRBS signal.

b) Calculate an order $r = 20$ FIR model ($\hat{g}$) using least-squares on $\hat{g}^{LS} = \operatorname{argmin}_{\hat{g} \in \mathbb{R}^r} \sum_{k=0}^{T} (y(k) - \sum_{i=0}^{r} \hat{g}(i)u(k-i))$.

c) Perform regularized least squares estimation for $\hat{g}$ using the regularizer $R(\theta) = \gamma\theta^T P(\alpha)^{-1}\theta$, under the TC kernel $P_{ij}(\alpha) = \alpha^{\max(i,j)}$ with $\alpha = 0.5$ and $\gamma = 100$.

d) Split the input-output data into training/testing 70%/30% split (the first 7 periods are used for training, and the last 3 are for testing).

e) Use cross validation to find an optimal TC parameter α and system order r with $\gamma = 50$ (a grid search with $r \in 2, 3, \ldots, 30$ and $\alpha \in 0 : 0.05 : 0.95$ is permissible)

f) Compare the estimated impulse responses $\hat{g}$ from parts b, c, e (least squares, TC=0.5, TC cross validation) in a stem plot. Be sure to title the plot, label axes, and add a legend.

Figure 8.1 plots the recovered pulse response models for the system. Cross-validation returns an order-9 (10 element) pulse response with TC parameter $\alpha = 0.85$.

Figure 8.2 plots the cross-validation errors obtained by this estimation task.

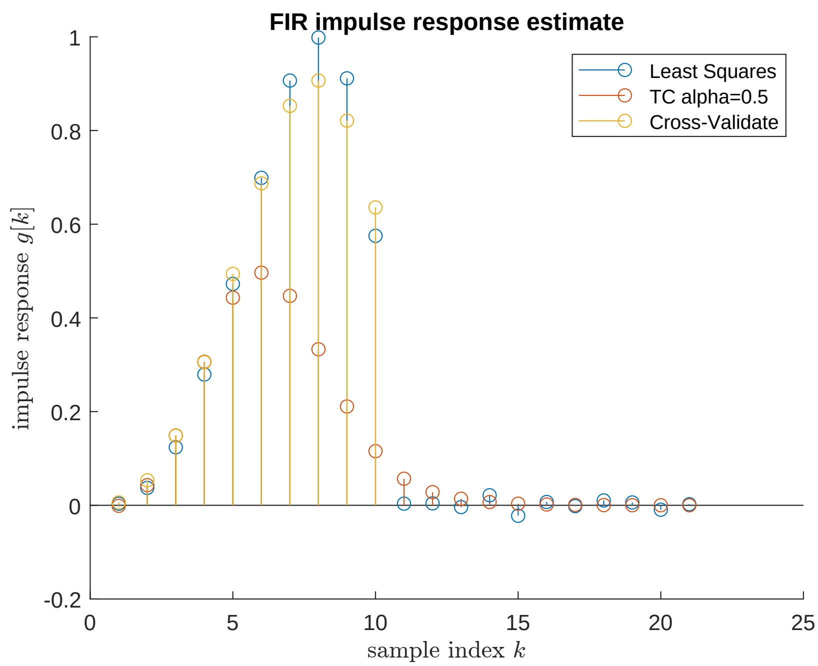

Figure 8.1: Recovered pulse response models for the system

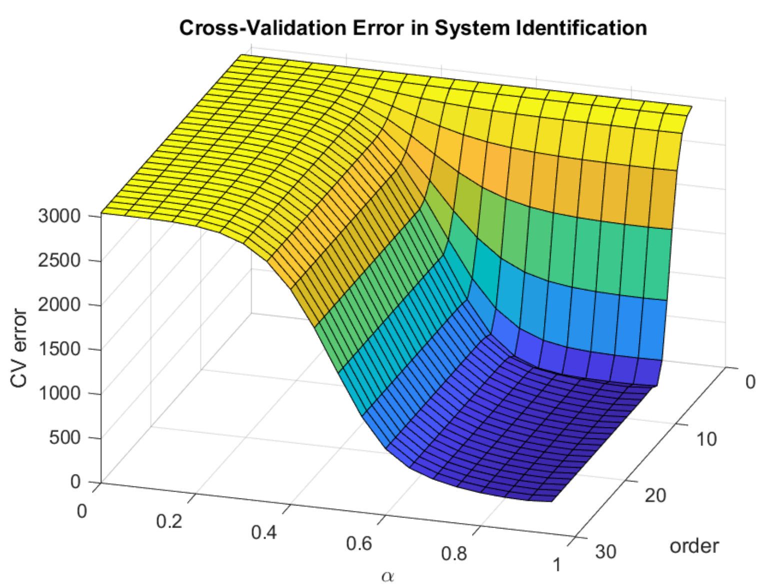

Figure 8.2: Cross-Validation error

References

[1] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of arx-models using sum-of-norms regularization," *Automatica*, vol. 46, no. 6, pp. 1107–1111, 2010. 8-3

[2] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group lasso and a sparse group lasso," *arXiv preprint arXiv:1001.0736*, 2010. 8-3

[3] F. Lindsten, H. Ohlsson, and L. Ljung, "Clustering using sum-of-norms regularization: With application to particle filter output computation," in *2011 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2011, pp. 201–204. 8-3