MULTI-OBJECTIVE OPTIMIZATION APPROACH TO OPTIMAL INPUT DESIGN FOR AUTOREGRESSIVE MODEL IDENTIFICATION

Katsuji Uosaki [*,1] **Toshiharu Hatanaka** [*]

[*] *Department of Information and Physical Sciences*
Graduate School of Information Science and Technology
Osaka University
Suita, Osaka 565-0871, Japan
Phone: +81-6-6879-7833 *Fax: +81-6-6879-7836*
E-mail: uosaki@ist.osaka-u.ac.jp

Abstract: Optimality criteria of the experimental design in dynamic system identification may sometimes be conflicting for structure determination and parameter estimation steps. This leads the necessity of the tradeoffs between the performances in these two steps. Here, the optimal input design in system identification is investigated as a multi-objective optimization problem. The Pareto-optimal set of inputs is derived and how to use it in system identification is discussed.

Keywords: System identification, Optimal experimental design, System structure, Parameter estimation, Multi-objective optimization, Pareto-optimal solution.

1. INTRODUCTION

System identication deals with the problems of mathematical model building of the dynamic systems based on observed input/output data from the systems. Based on the idea originally developed for static regression analysis (Fedorov 1972; Silvey 1980; Pukelsheim 1993), optimal experimenal design for system identification such as optimal design of input, sampling intervals, pre-filters, etc., has been extensively investigated (Mehra 1974, Goodwin and Payne 1977, Zarrop 1979, Forssell and Ljung 2000, and references therein) to extract the maximum information about the system. The most studies on this aspect were for optimal input design for accurate parameter estimation within a specified model structure or without some constraints on input or output, assuming the precise knowledge of the underlying model structure of the data generating processes. However, in many cases such knowledge is not available and hence the analysis of the data should be performed in two steps: identification of an appropriate model structure from a given class of competing models; and parameter estimation in the specified model structure. Despite the universal recognition of the importance of the first step in system identification, the studies on the optimal input design for this step is quite few, see Kabaila (Goodwin and Payne 1977) and Uosaki *et al.* (1984, 1987). And it is recognized that, the optimal experimental design for one of these steps may be highly inefficient for the other. This leads the necessity of the compromises in the design, i.e., the tradeoffs between the performances in these two steps should be introduced. Recognizing the existence of the conflicting optimality criteria, we investigate in this paper the optimal input design in system identification from the viewpoint of multi-objective optimization problem (Keeney and Raiffa 1976, Steuer 1986, Miettinen 1999). The Pareto-optimal set of inputs is derived and how it is used in system identification is discussed.

---

[1] This work is partially supported by the Grant-in-Aid for Scientific Research from the Japan Society of Promotion of Sciences (C)(2)14550447.

2. D_S- AND D-OPTIMAL INPUT DESIGN

Consider the following autoregressive model with exogeneous input (ARX model):

$$\begin{aligned} y_t &= a_1 y_{t-1} + \cdots + a_n y_{t-n} + b u_{t-1} + \varepsilon_t \\ &= \boldsymbol{a}_n^T y_{t-1}^{t-n} + b_1 u_{t-1} + \varepsilon_t \end{aligned} \quad (1)$$

where $\boldsymbol{a}_n = (a_1, \cdots, a_n)^T$, $y_{t-1}^{t-n} = (y_{t-1}, \cdots, y_{t-n})^T$, ε_t is independently normally distributed with zero mean and constant variance σ^2, and input u_t is zero-mean weak stationary. Furthermore, we assume all the roots of the equation

$$1 - a_1 z^{-1} - \cdots - a_n z^{-n} = 0 \quad (2)$$

are inside of the unit circle, then the output y_t is also zero-mean weak stationary. From the practical point of view, we consider here the variance of the system output sequence $Y = \{y_t, (t = 1, \ldots, N)\}$ cannot exceed some given value, say W, i.e.,

$$W \geq E[y_t^2] = \frac{1}{2} \int_{-\pi}^{\pi} dF_y(\omega) \quad (3)$$

where $F_y(\omega)$ is the spectral distribution function of the output sequence Y. Define the average information matrix M_{θ_n} for the parameter vector $\theta_n = (\sigma^2, b, a_1, \cdots, a_n)^T$ by

$$\begin{aligned} M_{\theta_n} = E_{Y|\theta_n} \Bigg\{ & \frac{1}{N\sigma^2} \sum_{t=1}^{N} \left(\frac{\partial \nu_t}{\partial \theta_n} \right)^T \left(\frac{\partial \nu_t}{\partial \theta_n} \right) \\ & + \frac{2}{\sigma^2} \left(\frac{\partial \sigma^2}{\partial \theta_n} \right)^T \left(\frac{\partial \sigma^2}{\partial \theta_n} \right) \Bigg\} \end{aligned} \quad (4)$$

where $\{\nu_t\}$ is the residual sequence given by

$$\nu_t = y_t - a_1 y_{t-1} - \cdots - a_n y_{t-n} - b u_{t-1}. \quad (5)$$

Since

$$\begin{aligned} \frac{\partial \nu_t}{\partial a_k} &= -y_{t-k}, \\ \frac{\partial \nu_t}{\partial b} &= -u_{t-1} \end{aligned} \quad (6)$$

we can write the average information matrix as

$$M_{\theta_n} = \frac{1}{N\sigma^2} E \sum_{t=1}^{N} \begin{bmatrix} \frac{2}{\sigma^2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & u_{t-1}^2 & u_{t-1}y_{t-1} & \cdots & u_{t-1}y_{y-(n-1)} & u_{t-1}y_{t-n} \\ 0 & y_{t-1}u_{t-1} & y_{t-1}^2 & \cdots & y_{t-1}y_{t-(n-1)} & y_{t-1}y_{t-n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & y_{t-(n-1)}u_{t-1} & y_{t-(n-1)}y_{t-1} & \cdots & y_{t-(n-1)}^2 & y_{t-(n-1)}y_{t-n} \\ 0 & y_{t-n}u_{t-1} & y_{y-n}y_{t-1} & \cdots & y_{t-(n-1)}y_{t-n} & y_{t-n}^2 \end{bmatrix}$$

$$= \frac{1}{\sigma^2} \begin{bmatrix} \frac{2}{\sigma^2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & H_n & h_1 & \cdots & h_{n-1} & h_n \\ 0 & h_1 & \rho_0 & \cdots & \rho_{n-2} & \rho_{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & h_{n-1} & \rho_{n-2} & \cdots & \rho_0 & \rho_1 \\ 0 & h_n & \rho_{n-1} & \cdots & \rho_1 & \rho_0 \end{bmatrix} \quad (7)$$

where

$$\begin{aligned} & E[y_t y_s] = \rho_{|t-s|}, \\ & r_n = (\rho_1, \rho_2, \cdots, \rho_n)^T, \\ & h_k = (\rho_k - a_1 \rho_{k-1} - \cdots - a_n \rho_{|k-n|})/b, \\ & \qquad (k = 1, \ldots, n) \\ & F_n = \begin{bmatrix} \rho_0 & \cdots & \rho_{n-1} \\ \vdots & \ddots & \vdots \\ \rho_{n-1} & \cdots & \rho_0 \end{bmatrix}, \end{aligned} \quad (8)$$

$$\begin{aligned} H_n &= (\rho_0 - 2\sum_{k=1}^{n} a_k \rho_k + \sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \rho_{|j-k|} - \sigma^2)/b^2 \\ &= (\rho_0 - 2\boldsymbol{a}_n^T r_n + \boldsymbol{a}_n^T F_n \boldsymbol{a}_n - \sigma^2)/b^2. \end{aligned}$$

The information matrix can be partitioned as

$$M_{\theta_n} = \frac{1}{\sigma^2} \begin{bmatrix} \frac{2}{\sigma^2} & 0 & 0 \\ 0 & H_n & G_n^T \\ 0 & G_n & F_n \end{bmatrix} \quad (9)$$

or

$$M_{\theta_n} = \frac{1}{\sigma^2} \begin{bmatrix} M_{\theta_{n-1}} & R_n \\ R_n^T & \rho_0 \end{bmatrix} \quad (10)$$

corresponding to the parameter partition $(\sigma^2, b, \boldsymbol{a}_n^T)$ or (θ_{n-1}, a_n), where

$$\begin{aligned} G_n &= (h_1, \cdots, h_n)^T = r_n - F_n \boldsymbol{a}_n, \\ R_n &= (0, h_n, \rho_{n-1}, \cdots, \rho_1)^T. \end{aligned} \quad (11)$$

Based on the average information matrix, we can derive the optimal inputs for competing autoregressive model discrimination and for accurate parameter estimation in autoregressive model (1).

2.1 Optimal input design for model discrimination

Now we consider the problem to find an optimal input which determines the adequate model structure. Various criteria for structure determination have been proposed such as the hypothesis testing approach (Atkinson and Cox 1974, Dette 1995) and the information criterion approach (Akaike 1974). In the hypothesis testing approach, we will find an optimal input which determines if the order of the autoregressive model (1) is n or $n-1$ under the output constraint (3). By using the parameter vector θ_n, the order determination problem can be stated as the testing hypothesis problem with null hypothesis:

$$H_0: \quad \theta_n = (\theta_{n-1}^T, 0)^T = \theta^{(0)} \tag{12}$$

against

$$H_1: \quad \theta_n = (\theta_{n-1}^T, a_n)^T = \theta^{(1)}, \;\; a_n \neq 0. \tag{13}$$

It is known that, under the alternative hypothesis H_1, a random variable

$$-2\log\lambda(Y) = -2\log\frac{\sup_{\theta^{(0)}}\; p(Y|\theta^{(0)})}{\sup_{\theta^{(1)}}\; p(Y|\theta^{(1)})} \tag{14}$$

converges in law to a noncentral χ^2 distribution with degree of freedom 1 and noncentrality parameter

$$\phi = N a_n^2(\rho_0 - R_n^T M_{\theta_{n-1}}^{-1} R_n) \tag{15}$$

by using the matrix partition (10). Since the power of the test becomes large for specified significant level when the noncentrality parameter is large, we will find the input sequence maximizing the D_s-criterion function (Silvey 1980)

$$J_s = \rho_0 - R_n^T M_{\theta_{n-1}}^{-1} R_n \tag{16}$$

subject to the constraint on output variance (3). Since $M_{\theta_{n-1}}$ is positive definite, $R_n^T M_{\theta_{n-1}}^{-1} R_n > 0$ unless $R_n = 0$. Therefore, in order to maximize the D_s-criterion function, ρ_0 should be maximized and $R_n = 0$ satisfying the constraint (3). This can be fulfilled when

$$\begin{aligned} &\rho_0 = W, \\ &\rho_1 = \rho_2 = \cdots = \rho_{n-1} = 0, \\ &\rho_n = a_n W. \end{aligned} \tag{17}$$

In this case, the maximum value of J_s is W.

2.2 Optimal input design for parameter estimation

Consider the problem to find an optimal input which gives an accurate estimate of the parameters $\theta_n = (\sigma^2, b, a_1, \cdots, a_n)^T = (\sigma^2, b, \boldsymbol{a}_n^T)^T$ subject to (3). Since the Cramer-Rao lower bound for the covariance of an unbiased estimate $\hat{\theta}_n$ is given by the inverse of the average information matrix M_{θ_n}, various scalar measures for the estimation performance based on M_{θ_n} and $M_{\theta_n}^{-1}$ have been considered.

(1) A-optimality: Minimize the trace of $M_{\theta_n}^{-1}$, i.e., minimize the average variance of the parameter estimate.
(2) E-optimality: Minimize the maximum eigenvalue of $M_{\theta_n}^{-1}$.
(3) D-optimality: Minimize the determinant or the generalized variance of $M_{\theta_n}^{-1}$.
(4) G-optimality: Minimize the maximum of the covariance of the output $y(t|\theta_n)$.

Among these, D-optimality has an advantage of the invariance property under scale changes of parameters, and it also implies G-optimality. So, we employ here

$$J_D = \det M_{\theta_n} \tag{18}$$

as the optimality measure. Using the matrix partition (9), we can write $\log J_D$ as

$$\begin{aligned} \log J_D &= \log\det M_{\theta_n} \\ &= \log 2 - (n+3)\log\sigma^2 + \log\det F_n \\ &\quad + \log(H_n - G_n^T F_n^{-1} G_n) - 2\log b \\ &= \log 2 - (n+3)\log\sigma^2 + \log\det F_n \\ &\quad + \log(\rho_0 - \sigma^2 - r_n^T F_n^{-1} r_n) - 2\log b \end{aligned} \tag{19}$$

Since F_n is positive definite, $r_n^T F_n^{-1} r_n > 0$ unless $r_n = 0$. This condition also maximizes $\det F_n$ since $\det F_n$ achieves its maximum if F_n is diagonal. Subject to the constraint (3) the maximum value of $\det F_n$ is achieved with $\rho_0 = W$. Summarizing the above, the D-optimal input should satisfy the following condition:

$$\begin{aligned} &\rho_0 = W, \\ &\rho_1 = \rho_2 = \cdots = \rho_{n-1} = \rho_n = 0. \end{aligned} \tag{20}$$

By this choice, $\log\det M_{\theta_n}$ achieves its maximum $2(W-\sigma^2)W^n/(\sigma^{2(n+3)}b^2)$.

3. OPTIMAL INPUT DESIGN AS A MULTI-OBJECTIVE OPTIMIZATION PROBLEM

It is shown in the previous section that D_s-optimal and D-optimal inputs are similar as $(\rho_0, \rho_1, \cdots, \rho_{n-1}) = (W, 0, \cdots, n)$ is common, but the difference is in ρ_n, which may affect the criterion functions of D- and D_s-optimality. After some manipulations, D_s- and D-optimality criteria can be expressed as functions of ρ_n such that

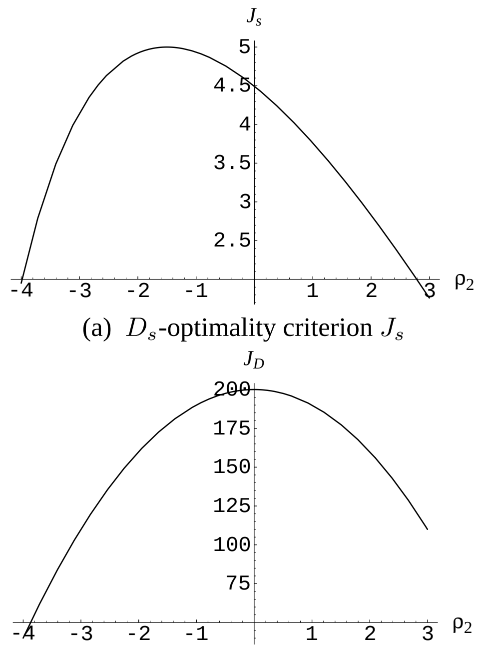
$$J_s = W - \frac{b^2(\rho_n - a_n W)^2}{(1+a_n^2)W - \sigma^2 - 2a_n\rho_n}, \tag{21}$$

$$J_D = \frac{2W^n(W - \sigma^2 - \dfrac{\rho_n^2}{W})}{\sigma^{2(n+3)}b^2} \tag{22}$$

with $(\rho_0, \rho_1, \cdots, \rho_{n-1}) = (W, 0, \cdots, 0)$.
Based on (21) and (22), the criteria J_D and J_s are evaluated as in Fig.1 for the ARX model,

$$y_t = 0.5y_{t-1} - 0.3y_{t-2} + u_{t-1} + \varepsilon_t \tag{23}$$

where ε_t is independently normally distributed with zero mean and unit variance, and the output variance $E[y_t^2]$ less than $W = 5$. This indicates that the optimal input for D-optimality criterion ($\rho_n = 0$) deteriorates the D_s-optimality criterion, and vice versa.
It is usual as shown above that we cannot find the optimal input that maximizes simultaneously both of two criteria J_D and J_s, i.e., two criteria may be conflicting. We will consider this as so-called *multi-objective* optimization problem (Keeney and Raiffa 1976, Steuer 1986, Miettinen 1999).



(a) D_s-optimality criterion J_s

(b) D-optimality criterion J_D

Fig. 1. D- and D_s-optimality criteria

3.1 Multi-objective optimization problem

Multiple, often conflicting objectives are common in real-world optimization problems. They are, without loss of generality, all to be maximized (or minimized) and all equally important, i.e., no additional knowledge about the problem is available. We assume that a solution to this problem can be described in terms of a vector of n decision variables, $\boldsymbol{x} = (x_1, \ldots, n)^T$ in decision space $\mathcal{D}$ as follows:

$$\begin{aligned} &\text{Maximize } f_k(\boldsymbol{x}), \ \ (k = 1, \ldots, k^*) \\ &\text{subject to } \boldsymbol{x} \in \mathcal{D} \end{aligned} \tag{24}$$

where the decision space $\mathcal{D}$ is determined by constraints on $\boldsymbol{x}$ such that

$$\begin{aligned} &g_\ell(\boldsymbol{x}) \geq 0, \quad (\ell = 1, \ldots, \ell^*) \\ &h_m(\boldsymbol{x}) = 0, \quad (m = 1, \ldots, m^*) \\ &x_i^{(L)} \leq x_i \leq x_i^{(U)}, \quad (i = 1, \ldots, n) \end{aligned} \tag{25}$$

For multiple conflicting objectives, each objective corresponds to a different optimal solutions, and a set of optimal solutions can be constructed by making a trade-offs between these solutions. However, the best solution in the set is uncertain with respect to all of the objectives.
In order to deal with this, the concept of dominance is used for most multi-objective optimization. In the optimization algorithms, two solutions are considered on the basis of whether one dominates the other or not. A solution $\boldsymbol{x}$ is said to dominate the other solution $\boldsymbol{x}'$ ($\boldsymbol{x} \succ \boldsymbol{x}'$) if both of the following conditions are satisfied.

(1) $\boldsymbol{x}$ is no worse than $\boldsymbol{x}'$ in all objectives, i.e., $f_k(\boldsymbol{x}) \geq f_k(\boldsymbol{x}')$ for all $k = 1, \ldots, k^*$
(2) $\boldsymbol{x}$ is strictly better than $\boldsymbol{x}'$ in at least one objective, i.e., $f_k(\boldsymbol{x}) > f_k(\boldsymbol{x}')$ for at least one $k = 1, \ldots, k^*$

For a given finite set of solutions, we can carry out all pair-wise comparisons and find which solution dominates and which solutions are non-dominated with respect to each other. At the end, we may obtain a set of solutions such that any two of the solutions in the set do not dominate each other and that a solution dominates any solutions outside of the set can be found in the set. The non-dominated set $\mathcal{P}^*$ of solutions that are not dominated by any member of the set $\mathcal{D}$ is called the Pareto-optimal set.
Figure 2 shows a Pareto-optimal set for the following two-objective optimization problem:

$$\begin{aligned} \text{Maximize } \ & f_1(x_1, x_2) = 10000 - x_1^2 - x_2, \\ & f_2(x_1, x_2) = 10000 - x_1 - x_2^2 \\ \text{subject to } \ & x_1 \in [-10, 10], \ x_2 \in [-10, 10] \end{aligned} \tag{26}$$

Pareto-optimal solutions (•) lie on the Pareto-optimal front (Pareto-optimal set), while dominated solutions are plotted by circles.
There are two principal challenges in Pareto set:

(1) Populating the Pareto set or finding Pareto solutions.
(2) Selecting from among Pareto solutions.

These are analogous to determining potential solutions and selecting from among the solutions.
Here, a number of the approaches to find the Pareto-optimal set are given.

(1) Weight sum method: It scalarizes a set of objectives into a single objective by pre-multiplying each objective with a user applied weight such that

$$\text{Maximize } \ F(\boldsymbol{x}) = \sum_{k=1}^{k^*} \alpha_k f_k(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{f}(\boldsymbol{x})$$

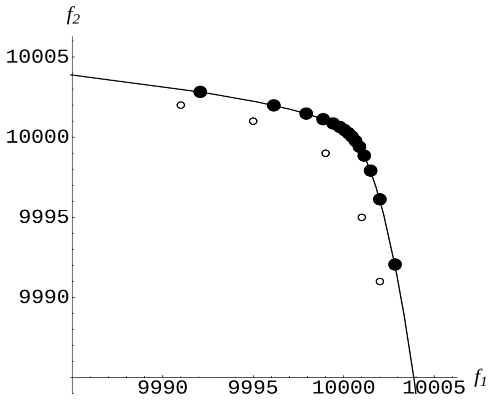


Fig. 2. Pareto-optimal solutions

$$\text{subject to } \boldsymbol{x} \in \mathcal{D} \quad (27)$$

where $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \cdots, f_{k^*}(\boldsymbol{x}))^T$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{k^*})^T$ with $\alpha_k \in (0,1)$, $\sum_{k=1}^{k^*} \alpha_k = 1$.

(2) ϵ-constraint method: It keeps one of the objectives and restricts the rest of the objectives within user-specified values ϵ_k.

(3) Weighted metric methods: It uses weighted metrics such as ℓ_p distance metrics to combine multiple objectives into a single objective instead of using a weighted sum of the objectives, where weighted ℓ_p distance metric of a solution $\boldsymbol{x}$ from the ideal solution $\boldsymbol{x}^* = (x_1^*, \ldots, x_{k^*}^*)^T$ is defined by $(\sum_{k=1}^{k^*} \alpha_k |f_k(\boldsymbol{x}) - x_k^*|^p)^{1/p}$ and parameter p can take any value between 1 and ∞.

(4) Value function method: A mathematical value function (or utility function) $\mathcal{U}$: $R^{k^*} \to R$ is introduced, and the maximum of $\mathcal{U}(\boldsymbol{f}(\boldsymbol{x}))$ is searched instead of the maximum of $\boldsymbol{f}(\boldsymbol{x})$.

Among these, the weighted sum method is the simplest. The concept is intuitive and easy. Moreover, it guarantees finding solutions on the Pareto-optimal set by the following theorems (Miettinen 1999).

Theorem 1

The solutions to the multi-objective optimization problem (27) are Pareto-optimal if the weights are positive for all matrices.

Theorem 2

If $\boldsymbol{x}^*$ is a Pareto-optimal solution of a convex multi-objective optimization problem, then there exists a non-zero positive weight vector $\boldsymbol{\alpha}$ such that $\boldsymbol{x}^*$ is a solution to the problem (27).

Then, the next problem is to determine which input in the Pareto-optimal set is preferable. This is done by using the user's higher-level information about the problem such as relative preference factor among the objectives. In the following, optimal input design problem will be formulated as a multi-objective optimization problem.

3.2 Pareto-optimal input for model discrimination and parameter estimation

Since two criteria J_D for model discrimination and J_s for parameter estimation are conflicting as shown above, we find the Pareto-optimal set of inputs for the two objectives J_D and J_s instead of finding the optimal input that maximizes simultaneously both of these two criteria. Applying theorems 1 and 2, Pareto-optimal set of inputs is obtained by solving

$$\text{Maximize} \;\; J = \alpha J_D + (1-\alpha) J_s, \quad \alpha \in (0,1) \quad (28)$$

subject to the constraint on output variance (3).
Following the argument in Section 2, the solution is given by

$$\begin{aligned} \rho_0 &= W, \\ \rho_1 &= \rho_2 = \cdots = \rho_{n-1} = 0, \\ \rho_n &= a_n \alpha W, \qquad \alpha \in (0,1) \end{aligned} \quad (29)$$

The Pareto-optimal set satisfying (29) for the ARX model (23) is shown by the heavy line in Fig.3. Since

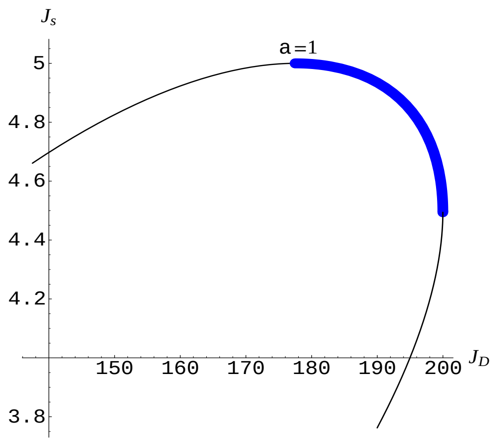


Fig. 3. Pareto-optimal set

the Pareto-optimal set is convex, the input which deteriorates J_D criterion $100\gamma\%$ from its maximum does not deteriorates J_s criterion $100(1-\gamma)\%$ but less than it, and vice versa. Fig. 4, which shows the deterioration rate, how much decrease the criterion from its maximum value. We can see that the input on the Pareto-optimal set does not deteriorate both J_s and J_D criteria so much, and hence we can use them as the input for identification.

Input satisfying condition (29) can be realized as follows:
It is known that the maximum number of input frequencies required is $n(n+1)/2$ or $(n+1)(n+2)/2$ by the theorem of Caratheodory (Fedorov 1972), and that the existence of the input sequence satisfying

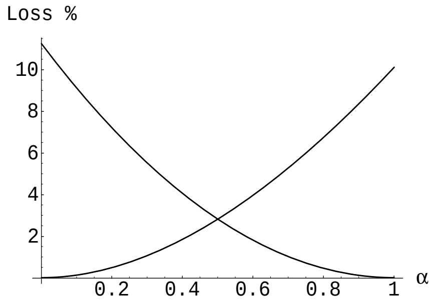


Fig. 4. Deterioration rate

the above conditions (17) can be realized by Zarrop's Chebyshev system approach consisting of $(n+1)/2$ or $(n+2)/2$ frequencies (Zarrop 1979). Hence, the optimal input can be chosen in the following form:

$$u_t = \sum_{\ell=1}^{p} m_\ell \cos(\omega_\ell t + \phi_\ell) \tag{30}$$

where $\omega_\ell \in (0, \pi)$, $\omega_k \neq \omega_\ell$ $(k \neq \ell)$, $m_\ell > 0$, $p = (n+1)/2$ or $(n+2)/2$, and ϕ_ℓ $(\ell = 1, 2, \ldots, p)$ are independently uniformly distributed on $[0, 2\pi]$. Following Ng and Qureshi's frequency domain design approach (Ng and Qureshi 1974), the amplitudes m_ℓ and frequencies ω_ℓ are satisfying the following system of nonlinear equations.

$$\begin{aligned}
\rho_0 &= \sum_{\ell=1}^{p} \frac{m_\ell b^2}{f(\omega_\ell)} + C_0 = W, \\
\rho_k &= \sum_{\ell=1}^{p} \frac{m_\ell b^2}{f(\omega_\ell)} \cos(k\omega_\ell) + C_k = 0, \\
&\qquad\qquad (k = 1, \ldots, n-1) \qquad (31) \\
\rho_n &= \sum_{\ell=1}^{p} \frac{m_\ell b^2}{f(\omega_\ell)} \cos(n\omega_\ell) + C_n = a_n, \alpha W
\end{aligned}$$

where

$$\begin{aligned}
&f(\omega) = A(e^{i\omega})A(e^{-i\omega}), \quad (\ell = 1, 2, \ldots, n) \\
&C_k = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-ik\omega}}{f(\omega)} d\omega, \quad (k = 1, 2, \ldots, n) \\
&C_0 \geq C_j, \;\; W - C_0 \geq |C_j|, \quad (j = 1, 2, \ldots, n) \\
&A(z^{-1}) = 1 - a_1 z^{-1} - \cdots - a_n z^{-n}. \qquad (32)
\end{aligned}$$

The system of equations can be solved in recursive manner.

4. CONCLUSION

Since the optimality criteria in input design for system identification are sometimes conflicting in model structure determination and parameter estimation, the trade-off between these two criteria should be introduced. To deal with this trade-off, multi-objective optimization approach is applied to optimal input design. How to derive the Pareto-optimal set of inputs and how to use it in system identification are discussed. This approach will become more important for optimal input design when other criterion corresponding to the connection between model uncertainty obtained by system identification and robust controllers (Forsell and Ljung, 2000) is introduced.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, **AC-19** (6), 716–722.

Atkinson, A. C. and D. R. Cox (1974). Planning experiments for discriminating between models (with Discussion). *J. Royal Statistical Society*, **B36** (3), 321–348.

Dette, H. (1995). Optimal design for identifying the degree of polynomial regression. *Annals of Statistics*, **23** (4), 1248-1266.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, NY.

Forsell, U. and L. Ljung (2000). Some results on optimal experiment design. *Automatica*, **36** (5), 749–756.

Goodwin, G. C. and R. L. Payne (1977). *Dynamic System Identification : Experiment Design and Data Analysis*. Academic Press, NY.

Keeney, R. L. and H. Raiffa (1976). *Decisions and Multiple Objectives: Preferences and Value Tradeoffs*. J. Wiley, NY.

Mehra, R. K. (1974). Optimal input signals for parameter estimation – survey and new results. *IEEE Trans. on Automatic Control*, **AC-19** (6), 753–768.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Kluwer, Boston.

Ng, T. S. and Z. H. Qureshi (1981). Optimal experiment design for autoregressive model with output power constraint. *IEEE Trans. on Automatic Control*, **AC-26** (6), 739–742.

Pukelsheim, F. (1993). *Optimal Design for Experiments*. J. Wiley, NY.

Silvey, S. D. (1980). *Optimal Design*. Chapman & Hall, London.

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation and Applications*. J. Wiley, NY.

Uosaki, K., I. Tanaka and H. Sugiyama (1984). Optimal input design for autoregressive model discrimination with constrained output variance. *IEEE Trans. on Automatic Control*, **AC-29** (4), 348–350.

Uosaki, K., I. Tanaka and H. Sugiyama (1987). Optimal input for discrimination of autoregressive models under output amplitude constraints. *International J. Systems Sciences*, **18**, 323–332.

Zarrop, M. B. (1979). *Optimal Experiment Design for Dynamic System Identification*. Springer-Verlag, NY.