## 2.160 System Identification, Estimation, and Learning

# Lecture Notes No. 19

**May 1, 2006**

## 14 Experiment Design

### 14.1 Review of System ID Theories for Experiment Design

Key Requirements for System ID

- Consistent (unbiased) estimate* $\hat{\theta}_N \to \theta_0$
- Small covariance $\text{cov}\left(\hat{\theta}_N\right)$

*) Unbiased estimate and consistent estimate are equivalent if the system is ergodic.

(1) $E\left[\hat{\theta}_N\right] = \theta_0$ The ensemble mean of $\hat{\theta}_N$ is equal to the true value $\theta_0$ …..unbiased.

(2) $\lim_{N\to\infty} \hat{\theta}_N \to \theta_0$ The limit of $\hat{\theta}_N$ as $N$ tends to infinity is the true value $\theta_0$ ..consistent

**Major Theoretical Results**

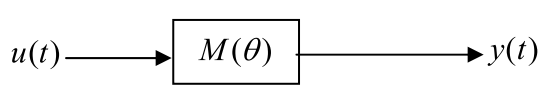
***i) Informative data sets***

Is data set $Z^\infty$ informative enough to distinguish any two different models $W_1(q)$ and $W_2(q)$ in the same model set $M(\theta)$?

$$\left(W_1(e^{i\omega}) - W_2(e^{i\omega})\right)Z(t) \equiv 0 \to W_1(e^{i\omega}) \equiv W_2(e^{i\omega}) \qquad (3)$$

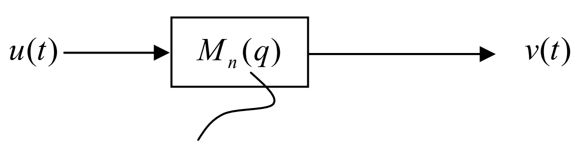
The data set is Informative (for any linear time-invariant model), if the Spectrum Matrix:

$$\Phi_z(\omega) = \begin{bmatrix} \Phi_u(\omega) & \Phi_{uy}(\omega) \\ \Phi_{yu}(-\omega) & \Phi_y(\omega) \end{bmatrix} \text{ : Positive definite for almost all } \omega. \quad (4)$$




***ii) Persistence of excitation***

Any (n-1)-st order ***MA*** filter cannot filter the input to zero.




$$M_n(q) = m_1 q^{-1} + \cdots + m_n q^{-n}$$

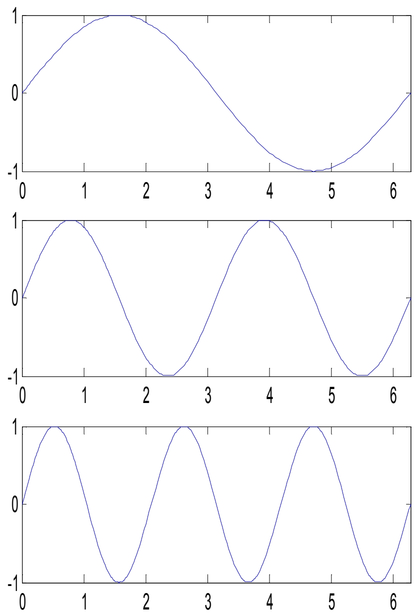
$$\text{Persistently Exciting of order } n \leftrightarrow \overline{R}_n = \begin{bmatrix} R_u(0) & R_u(1) & \cdots & R_u(n-1) \\ R_u(1) & R_u(0) & \cdots & R_u(n-2) \\ \vdots & & & \\ R_u(n-1) & & \cdots & R_u(0) \end{bmatrix} \quad (5)$$

is Non-singular

In particular for

$$G(q,\theta) = \frac{q^{-n_k}\left(b_1 + b_2 q^{-1} + \cdots + b_{n_b} q^{n_b+1}\right)}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}} \quad (6)$$

An open-loop experiment with an input sequence $u(t)$ that is persistently exciting of order $n_b + n_f$ is informative enough to distinguish any two transfer functions of the above form.




*n* sinusoids w/different frequencies persistently exciting of order *2n* → Informative enough w.r.t. $G(q,\theta)$ having at most *n* poles and *n* zeros

***iii) Signal to Noise Ratio***

The Prediction-Error Method (PEM) determines parameter vector $\theta$ in such a way that a frequency-weighted squared error norm be minimized. For a fixed noise model:

$$H(q,\theta) = H_{fixed}(q)$$

the following signal-to-noise ratio is the weight of the norm:

$$\hat{\theta} = \arg\min_{\theta} \int_{-\pi}^{\pi} \underbrace{\left|G_0(e^{i\omega}) - G(e^{i\omega},\theta)\right|^2}_{\text{Error of transfer function}} \cdot \underbrace{\frac{\Phi_u(\omega)}{\left|H_{fixed}(e^{i\omega})\right|^2}}_{\text{Signal to Noise Ratio}} d\omega \quad (7)$$

***iv) Asymptotic Variance***

Consider PEM with $V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\varepsilon^2(t,\theta)$. If the true model is involved in the model set $\theta_0 \in D_M$, the variance of parameter estimation error approaches the following Asymptotic Variance, as the number of data tends to infinity,

$$Cov\hat{\theta}_N \sim \frac{1}{N}P_\theta = \frac{\lambda_0}{N}\left[\overline{E}\left(\psi(t,\theta_0)\psi^T(t,\theta_0)\right)\right]^{-1} \tag{8}$$

which depends on:

- Number of data; $N$
- Noise level: $\lambda_0$
- Sensitivity: $\psi(t,\theta_0) = \frac{d}{d\theta}\hat{y}(t|\theta)\Big|_{\theta_0}$ (9)

The Punch Line

Using many parameters $\theta \in R^d$ (large $d$) may provide an accurate model with a small value of bias ($\theta_0 \in D_M$), but a large $d$ tends to worsen the variance and convergence speed, if some parameters in $\theta$ have small sensitivity values, $\frac{d}{d\theta}\hat{y}(t|\theta)$.

The above asymptotic variance can be expressed in the frequency-domain as

$$Cov\hat{\theta}_N \sim \frac{1}{N}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{\Phi_v(\omega)}T'(e^{i\omega},\theta_0)\Phi_{x_0}(\omega)T'^T(e^{i\omega},\theta_0)d\omega\right]^{-1} \tag{10}$$

where $T'(q,\theta) = \frac{d}{d\theta}T(q,\theta) = \left[G'(q,\theta), H'(q,\theta)\right]$

$\Phi_v(\omega) = \lambda_0\left|H(e^{i\omega},\theta_0)\right|^2$

$\lambda_0 = \Phi_{e_0}(\omega)$

## 14.2 Design Space of System ID Experiments

1) Data Acquisition

- Sensors: What to measure and how to measure
  - Output variables
  - Disturbances: If feasible, it is better to measure disturbances
- Actuators: What to manipulate
  - Inputs physical limit and constraints must be taken into account
- Sampling rate… 2.161 Signal Processing

High sampling rate means

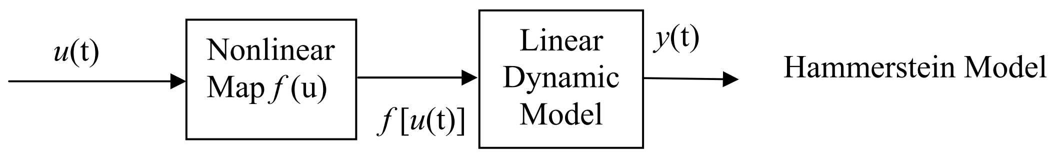
- More data, but
- Capturing noise rather than signals

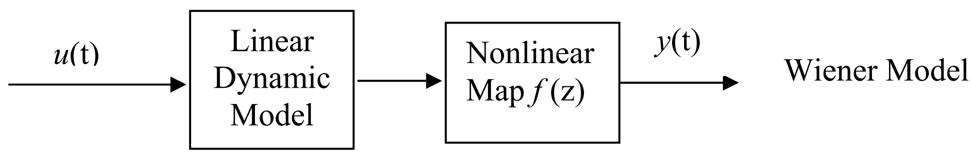
Anti-aliasing

Pre-filter design

2) Model Structure Selection

- Linear/Non-linear
  a) Linear parameterization of nonlinear systems
  b) Hammerstein and Wiener Models




- Linear Time-Invariant Models
  FIR, ARX, ARMA, OE, J-B, State Space
- Trade-off between bias (accuracy) and variance (reliability)
  System order

3) Input Signal Design

How do we design an input sequence (waveform) to be given to the system so that data are informative and that the covariance of estimate may be small?

<u>The Punch Line</u>

Both bias and variance conditions are provided in terms of the second-order properties of input signals. See eqs. (4), (5),(7), (8) and (10). They hinge on signal ***power spectra***, but signal waveforms are free to choose. This gives a great degree of freedom in designing an input sequence.

## 14.3 Input Design for Open-Loop Experiments

In an open-loop system, u(t) and $e_0$(t) are independent. Therefore, $\Phi_{ue_0} = 0$, $\Phi_{e_0 u} = 0$ in eq. (10).

$$Cov\hat{\theta}_N \sim \frac{1}{N}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{\Phi_v(\omega)}T'(e^{i\omega},\theta_0)\Phi_{x_0}(\omega)T'^T(e^{i\omega},\theta_0)d\omega\right]^{-1} = \frac{1}{N}P_\theta$$

This asymptotic variance $P_\theta$ can be broken down to the term comprising the input spectrum $\Phi_u(\omega)$ and the one having no input spectrum. Taking inverse of $P_\theta$,

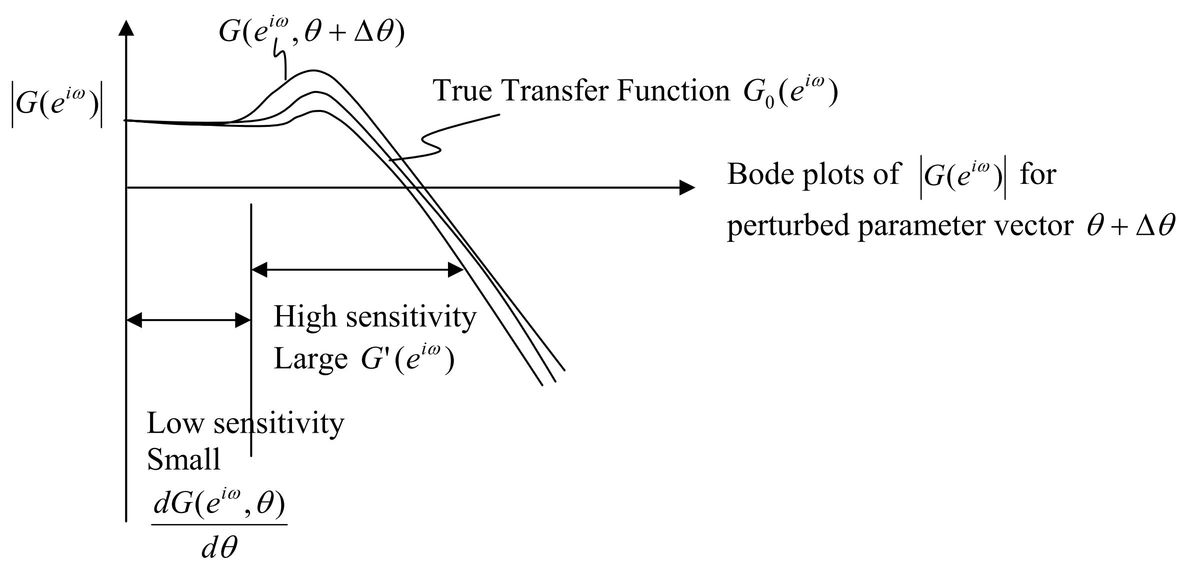
$$P_\theta^{-1} = \int_{-\pi}^{\pi} A(\omega)\Phi_u(\omega)d\omega + B \qquad (12)$$

where

$$A(\omega) = \frac{1}{2\pi\Phi_v(\omega)} G'(e^{i\omega}, \theta_0) G'^T(e^{i\omega}, \theta_0) \qquad (13)$$

$$B = \frac{\lambda_0}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\Phi_v(\omega)} H'(e^{i\omega}, \theta_0) H'^T(e^{i\omega}, \theta_0) d\omega \qquad (14)$$

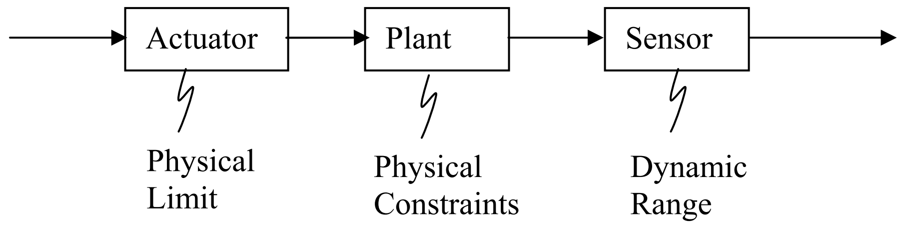
The asymptotic variance $P_\theta$ must be small for fast convergence. This can be achieved by allocating the input spectrum such that the first term in (12) becomes large. That is, the input spectrum should be large at frequencies where the squared parameter sensitivity of the transfer function, $\left|G'(e^{i\omega}, \theta_0)\right|^2$, is large. See the figure below.




To achieve fast convergence, choose the input spectrum $\Phi_u(\omega)$ such that $\Phi_u(\omega)$ has a large magnitude at frequencies where $A(\omega)$ in (13) has a large magnitude. Note $A(\omega)$ is proportional to $G'G'^T$; the sensitivity of the transfer function $G(e^{i\omega})$ to parameter changes $\theta$. Plotting $G(e^{i\omega}, \theta + \Delta\theta)$ for perturbed parameter vector $\theta$ depicts where to allocate large $\Phi_u(\omega)$.

This guide line for input design is only for the spectrum; the actual waveform can be determined freely by considering other requirements.

### 14.4 Practical Requirements for Input Design

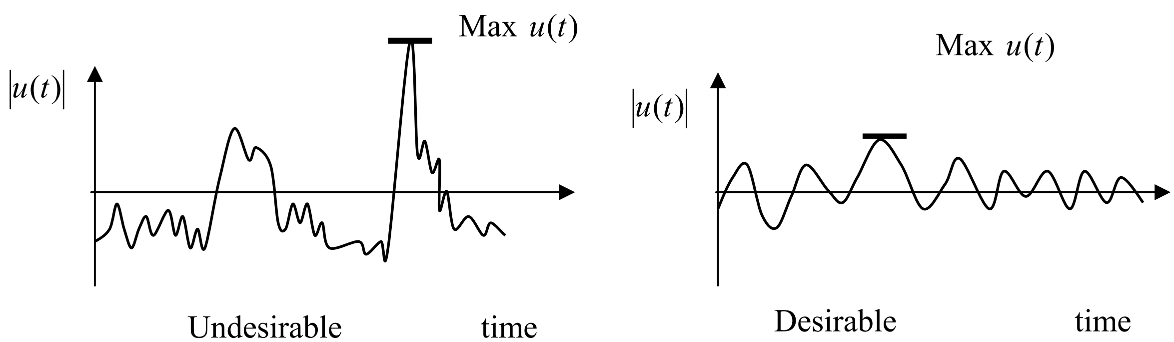



When conducting experiments, various constraints due to physical limits of the process must be satisfied. Among others, the input amplitude limit is a common constraint

$$\underline{u} \le |u(t)| \le \bar{u} \tag{15}$$

where $\underline{u}$ and $\bar{u}$ are lower and upper limits.

In general, the larger the input magnitude $|u(t)|$ becomes, the smaller the asymptotic variance $P_C$ becomes. Therefore the input sequence should have large amplitude most of the time.




To evaluate this aspect of input signal, the following crest factor is often used for zero-mean signals:

$$C_r^2 \equiv \frac{\max_{1\le t\le N} u^2(t)}{\frac{1}{N}\sum_{t=1}^{N} u^2(t)}, \tag{16}$$

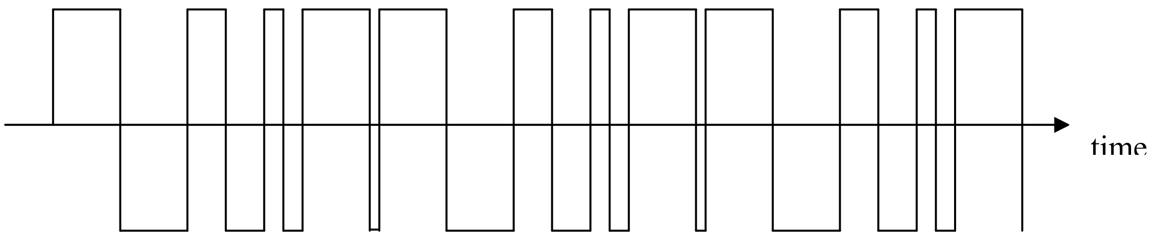
Note $C_r \ge 1$; **Smaller is better.**

### 14.5 System ID Using Random Signals

The best crest factor (1, the lowest value) is achieved with zero mean binary signals:

$u(t) = \pm\bar{u}$ . The signal shown below is an example of random binary sequence with an auto-correlation:

$$R_u(\tau) = E[u(t)u(t-\tau)] = 1 \cdot \delta(\tau) \tag{17}$$




In system identification, random signals are often used for input sequences because of their superb noise reduction properties. The following is to show this property.

Consider the model description based on impulse response given by

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) + v(t) \tag{18}$$

where $v(t)$ is noise, which is uncorrelated with a random input $u(t)$,

$$E[u(t)v(t-\tau)] \equiv 0 \text{ , for all } \tau . \tag{19}$$

Evaluating the correlation between the input and the output yields

$$\begin{aligned} R_{yu}(\tau) &= E[y(t)u(t-\tau)] = E[\{\sum_{k=1}^{\infty} g(k)u(t-k) + v(t)\}u(t-\tau)] \\ &= \sum_{k=1}^{\infty} g(k)E[u(t-k)u(t-\tau)] + \underbrace{E[v(t)u(t-\tau)]}_{0} \\ &= \sum_{k=1}^{\infty} g(k)R_u(k-\tau) = g(\tau) \end{aligned} \tag{20}$$

Note that the noise term is averaged out when the input-output correlation is computed. This is particularly useful for noisy systems. If the input is a random sequence with unit variance, the input-output correlation $\{R_{yu}(k), k = 1,2,\cdots\}$ gives the impulse response $\{g(k), k = 1,2,\cdots\}$ .

The advantage of the random input signal is clear when it is compared with a standard impulse response test using the following input:

$$u(t) = \begin{cases} \alpha, & t = 0 \\ 0, & t \neq 0 \end{cases} \tag{21}$$

The observed output is given by

$$y(t) = \alpha \cdot g(t) + v(t)$$

If one determines the impulse response from this output:

$$\hat{g}(t) = \frac{y(t)}{\alpha} = g(t) + \frac{v(t)}{\alpha} \tag{22}$$

Note that noise $v(t)$ shows up in the output $y(t)$. To eliminate the effect of $v(t)$, the magnitude of the input $\alpha$ must be very large, which is prohibited in most applications. Thus, the input-output correlation method using a random input sequence is a powerful system identification method particularly for noisy systems.

## 14.6 Pseudo-Random Binary Signal (PRBS)

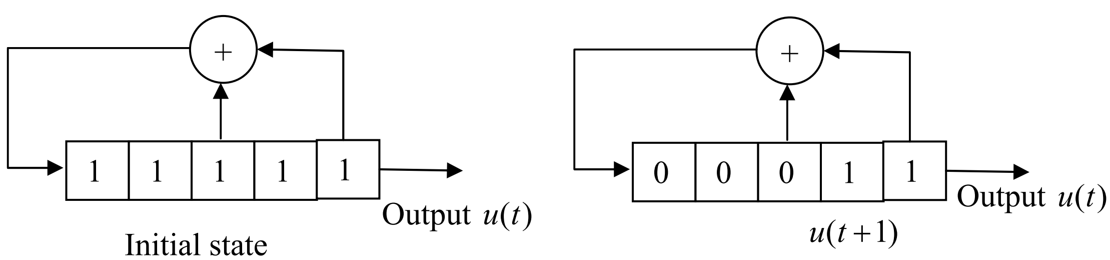
Now that random binary signals are useful for system identification, how can we generate them? A Pseudo-Random Binary Signal (PRBS) is a periodic, deterministic signal with white noise like properties. It has been widely used for system identification as well as for spread spectrum wireless communication and GPS.

PRBS can be generated easily with a shift register that circulates its output to the input gate and thereby generates a periodic, long-sequence binary signal. Mathematically, the signal is written as the following difference equation:

$$\begin{aligned} u(t) &= rem(A(q)u(t),2) \\ &= rem(a_1 u(t-1) + \cdots + a_n u(t-n),2) \end{aligned} \tag{23}$$

where *rem(x,2)* is the remainder as *x* is divided by 2, i.e. module 2 of *x*. The output is therefore binary.

***Example***: A 5 bit shift register




Binary shift register with feedback

Initial state

1 1 1 1 1 00011011101010000100101100 1111100011011101010000100101100

31 bits $= 2^5 - 1$ Repeat the same 31 bits

- The sequence repeats itself after every sequence of 31 bits.
- This sequence is uniquely determined by the feedback law A(q)
- $M = 2^n$-1 is the maximum-length sequence.

All the maximum-length Pseudo-Random Binary Signals (PRBS) have the following mean, covariance, and power spectrum.

$$\text{Mean: } \left| \frac{1}{M} \sum_{t=1}^{M} u(t) \right| = \frac{\bar{u}}{M} \tag{24}$$

$$\text{Covariance: } \frac{1}{M} \sum_{t=1}^{M} u(t) u(t+k) = \begin{cases} \bar{u}^2 & k = 0, \pm M, \pm 2M, \cdots \\ -\frac{\bar{u}^2}{M} & elsewhere \end{cases} \tag{25}$$

$$\text{Power Spectrum: } \Phi_u(\omega) = \frac{2\pi \bar{u}^2}{M} \sum_{k=1}^{M-1} \delta\left( \omega - \frac{2\pi k}{M} \right) \qquad 0 \le \omega \le 2\pi \tag{26}$$

Image removed for copyright reasons.

Features of PRBS

1) The power spectrum of $n$-th order PRBS with maximum length $M = 2^n$-1 has $M$-1 frequency peaks between $-\pi \le \omega \le \pi$, and is persistently exciting of order $M$-1; high-density, white noise.
2) Crest factor of 1; Maximum strength input sequence
3) Periodic
    - Quasi-stationary
    - Noise reduction by taking average of samples *y(t), y(t+T), y(t+2T),…,* which are outputs to the same input, *u(t), u(t+T),…*

An alternative to PRBS is to generate white, zero-mean Gaussian noise, and take the sign of the noise for generating a binary sequence. Compared with this regular random binary signal, PRBS has the following features

4) The regular random signal needs a long-sequence (n>>1) to secure the second order properties, while PRBS needs a much shorter bit length.
5) PRBS's spectrum has a special pattern, which facilitates analytic calculations.

PRBS has been used for

- Global Positioning System
- Spread-Spectrum, Code-Division Multiple Access (CDMA)

  …Cellular Phones, Bluetooth Wireless Networking etc.

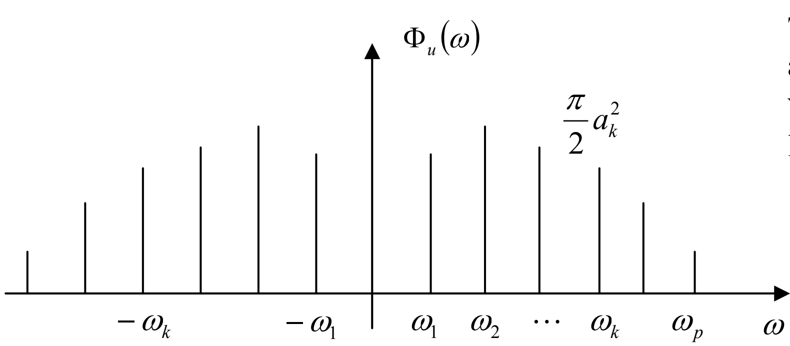
### 14.7 Sinusoidal Inputs.

Considering the requirements for input sequence design, a natural choice of input is to form it as a sum of sinusoids with different frequencies:

$$u(t) = \sum_{k=1}^{p} a_k \cos(\omega_k t + \phi_k) \tag{27}$$

At steady state ($t$: large) this multi-sine function has the following power spectrum:

$$\Phi_u(\omega) = 2\pi \sum_{k=1}^{p} \frac{a_k^2}{4} \left[\delta(\omega - \omega_k) + \delta(\omega + \omega_k)\right] \tag{28}$$




These should be placed at frequencies where $A(\omega)$ in (13) is large

For each frequency, a pair of non-zero components is generated at $\pm\omega_k$.
If $G(q,\theta)$ has at most $n$ poles and $n$ zeros, $p \geq n$ frequencies satisfy the persistently exciting condition. The problem with the multi-sin input is that, depending on phase angle $\phi_k$, the crest factor often becomes vary large. If all the sinusoids are in phase, $\phi_k = 0$,

$$C_r^2 = \sqrt{\frac{\max\limits_{1 \le t \le N} u^2(t)}{\frac{1}{N}\sum\limits_{t=1}^{N} u^2(t)}} = \frac{Amplitude \quad of \quad u(t)}{\sqrt{signal \quad power}} = \frac{\sum\limits_{k=1}^{p} a_k}{\sqrt{\sum\limits_{k=1}^{p} \frac{a_k^2}{2}}} \Rightarrow \sqrt{2p} \qquad for \quad a_k = a, \qquad k = 1, \cdots p \tag{29}$$

If all the sinusoids have the same amplitude, the crest factor is $\sqrt{2p}$ , which becomes large as $p$ increases.

Figure 13-5 from Ljung, Lennart. *System Identification: Theory for the User*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999, chapter 13. ISBN: 0136566952.
Image removed for copyright purposes.

Figure 1 Multi-Sine inputs

As shown in Figure 1 (Figure 13.5 in Ljung's book), the magnitude becomes very large 9a) when 10 sinusoids with zero phase angles $\phi_k = 0$ are superimposed; Figure(a). To overcome this problem, the phase angles $\phi_k$ ( $k = 1 \cdots p$ ) should be randomized (See Figure (c) ), or placed at particular angles that are as much out of phase as possible:

$$\phi_k = \phi_1 - \frac{k(k-1)}{p}\pi \qquad 2 \le k \le p \tag{30}$$
$$\phi_1 = \text{arbitrary}$$

This is called the ***Schroeder phase choice***.

***Swept Sinusoids (Chirp Signals)***

Continuously varying the frequency between $\omega_1$ and $\omega_2$ over a time period of $0 \le t \le T$ creates a Chirp Signal:

$$u(t) = A\cos\left(\omega_1 t + (\omega_{21} - \omega_1)\frac{t^2}{2T}\right) \tag{31}$$

The instantaneous frequency, $\omega_i$ , is obtained by differentiating $u(t)$

$$\omega_i = \omega_1 + \frac{t}{T}(\omega_2 - \omega_1) \tag{32}$$

This Chirp signal has the same crest factor as that of a single sinusoid, that is $\sqrt{2}$ . See Figure 2 below.

Figure 13-6 from Ljung, Lennart. *System Identification: Theory for the User*.
2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999, chapter 13. ISBN: 0136566952.
Image removed for copyright purposes.

Figure 2 Chirp Signal and its spectrum