System Identification

GianAndrea Müller
Stefan Rickli

December 1, 2018

Contents

1 System Identification

2 Definitions

Definition 1. *A system is said to be **time invariant** if the response to a certain input is not depending on absolute time.*

Definition 2. *A system is said to be **linear** if its output response to a linear combination of inputs is the same as the linear combination of the output responses of the individual inputs.*

Definition 3. *A system is said to be **causal** if the output at a certain time depends on the input up to that time only.*

Definition 4. *A process is said to be **stationary** if it does not depend on time.*

3 Frequency Domain Methods

3.1 Sampling Operation

$\boxed{y(k) = y(t)|_{t=kT, k=0,1,2,\ldots}}$ Sampling with period T

3.2 Fourier Series of Periodic Signals

$$\boxed{X(e^{j\omega_m}) = \sum_{k=0}^{M-1} x(k) e^{-j\omega_m k}}$$

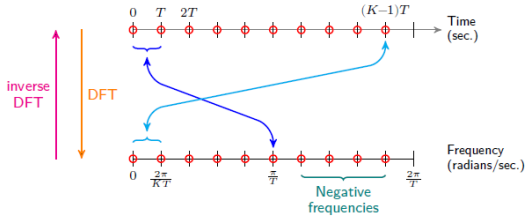
$$\omega_m = \frac{2\pi m}{M} = \omega_0$$

Non-negative frequencies are $m = 0$ to $m = M/s$.
They correspond to: $\omega_m = 0, \frac{2\pi}{M}, \frac{4\pi}{M}, \ldots, \frac{2\pi(M/2-1)}{M}, \pi$.

M	number of samples	
$\omega_0 = \frac{2\pi}{M}$	fundamental frequency ($y(k)$)	[rad]
T	sampling time	[s]
$\tau_p = MT$	period	[s]
$\omega_0 = \frac{2\pi}{\tau_p}$	fundamental frequency ($y(t)$)	[rad s^{-1}]

$$0, \underbrace{\frac{2\pi}{\tau_p}}_{\text{Fundamental frequency}}, \underbrace{2\left(\frac{2\pi}{\tau_p}\right), \ldots, \frac{M}{2}\left(\frac{2\pi}{\tau_p}\right)}_{\text{Harmonics}}$$

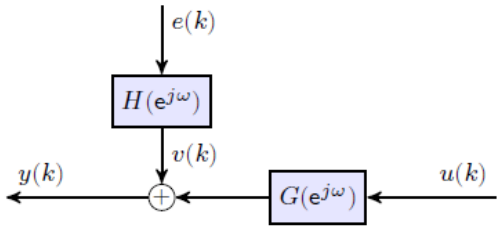
Definition 5. *The highest frequency $\omega_u = \omega_{M/2} = \frac{\pi}{T}$ is called the **Nyquist frequency**.*




```
%Non negative frequency vector
omega = omega_n(N,'p0');
%Discrete Fourier Transform
fft(u); %Matlab
DFT(u); %Lecture
```

Note that the definition of the fft in the Matlab documentation does not describe the actual implementation perfectly. The index actually runs from 0 to $N-1$ as in the definition made in class.

4 Spectral Estimation



$Y(j\omega) = G(j\omega)U(j\omega)$ Transfer function

$Y(e^{j\omega} = G(e^{j\omega}U(e^{j\omega})$ Discrete time TF

$$\frac{\omega_u}{2\pi} = \frac{r}{NT}$$

ω_u	input frequency	[rad s^{-1}]
N	calculation length	[ ]
T	experiment duration?	[s]
r	some integer	[ ]

$u(k) = \alpha\cos(\omega_u k),\ k = 0, 1, \ldots, K-1$ with $K \geq N$ Input

$y(k) = \alpha\left|G(e^{j\omega_u})\right|\cos(\omega_u k + \theta(\omega_u)) + v(k) + \text{transient}$ Output

where $\theta(\omega_u) = arg(G(e^{j\omega_u})$

4.1 Sinusoidal correlation methods

Correlation functions:

$$I_c(N) = \frac{1}{N}\sum_{k=0}^{N-1} y(k)\cos(\omega_u k)$$

$$I_s(N) = \frac{1}{N}\sum_{k=0}^{N-1} y(k)\sin(\omega_u k)$$

To calculate those from the data:

$$I_c(N) = \frac{\alpha}{2}\left|G(e^{j\omega_u})\right|\cos(\theta(\omega_u) + \frac{\alpha}{2}\left|G(e^{j\omega_u})\right|\frac{1}{N}\sum_{k=0}^{N-1}\cos(2\omega_u k + \theta(\omega_u)) + \frac{1}{N}\sum_{k=0}^{N-1} v(k)\cos(\omega_u k)$$

If the noise, $v(k)$ is sufficiently uncorrelated then the variance satisfies,

$$\lim_{N\to\infty} \text{var}\left\{\frac{1}{N}\sum_{k=0}^{N-1} v(k)\cos(\omega_u k)\right\} = 0$$

with a convergence rate of $1/N$.
Thus in the limit $N \to \infty$,

$$E\{I_c(N)\} \to \frac{\alpha}{2}\left|G(e^{j\omega_u})\right|\cos(\theta(\omega_u))$$
$$E\{I_s(N)\} \to -\frac{\alpha}{2}\left|G(e^{j\omega_u})\right|\sin(\theta(\omega_u))$$

and since $\lim_{N\to\infty} \text{var}\{I_c(N)\} = 0$, $\lim_{N\to\infty} \text{var}\{I_s(N)\} = 0$
The transfer function can be estimated via:

$$\boxed{\hat{G}_N(e^{j\omega_u}) = \frac{I_c(N) - jI_s(N)}{\alpha/2}}$$

- Advantages
  - Energy is concentrated at the frequencies of interest.
  - Amplitude of $u(k)$ can easily be tuned as a function of frequency.
  - Easy to avoid saturation and tune signal/noise (S/N) ratio.
- Disadvantages
  - A large amount of data is required.
  - Significant amount of time required for experiments.
  - Some processes won't allow sinusoidal inputs.

5 Frequency Domain Methods

$x(k),\ k = -\infty, \ldots, \infty$ Discrete-time domain signal

$$\boxed{X(e^{j\omega}) = \sum_{k=-\infty}^{\infty} x(k)e^{-j\omega k}}$$ Fourier Transform

- $X(e^{j\omega})$ is 2π periodic.
- If $\sum_{k=-\infty}^{\infty} |x(k)| < \infty$ then $X(e^{j\omega})$ converges.

$$\boxed{x(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega k}d\omega}$$ Inverse Fourier Transform

where $k = -\infty, \ldots, \infty$

5.1 Finite Energy Signal

5.1.1 Energy Spectral Density (Finite Energy Signal)

If $x(k)$ is a finite energy signal,

$$||x(k)||_2^2 = \sum_{k=-\infty}^{\infty} |x(k)|^2 < \infty$$

$$\boxed{S_x(e^{j\omega}) = |X(e^{j\omega})|^2}$$ Energy Spectral Density

- For finite energy signals the energy spectral density is easily calculated, for signals of infinite length and thus infinite energy however, this is not possible. For that reason the **power** spectral density is calculated instead!

5.1.2 Autocorrelation (Finite Energy Signal)

$$\boxed{R_x(\tau) = \sum_{k=-\infty}^{\infty} x(k)x(k-\tau), \quad \tau = -\infty, \ldots, 0, \ldots, \infty}$$

The spectral density is the Fourier Transform of the autocorrelation:

$$\sum_{\tau=-\infty}^{\infty} R_x(\tau)e^{-j\omega\tau} = S_x(e^{j\omega})$$

```
% autocorrelation for a non-periodic, finite energy signal
Correlation('finen',u); %Lecture
xcorr(u); %Matlab

% crosscorrelation for non-periodic, finite energy signals
Correlation('finen',u,y); %Lecture
xcorr(y,u); %Matlab
```

5.2 Discrete Periodic Signal

$$x(k) = x(k+M), \quad \forall\, k \in \{-\infty, \infty\} \quad \text{Periodic signal}$$

$$\omega_0 = \tfrac{2\pi}{M} \quad \text{Fundamental frequency}$$

- There are only M unique harmonics of the sinusoid $e^{j\omega_0}$.
- The non-negative harmonic frequencies are,

$$e^{jn\omega_0},\ n = 0, 1, \ldots, M/2$$

```
u_period = sqrt(variance)*randn(N,1)+bias;
u = repmat(u_period,periods,1);
```

5.2.1 Discrete Fourier Series (Discrete Periodic Signal)

$$\boxed{X(e^{j\omega_n}) = \sum_{k=0}^{N-1} x(k)e^{-j\omega_n k},\ \text{where } \omega_n = \tfrac{2\pi n}{N} = n\omega_0}$$

$$\boxed{x(k) = \tfrac{1}{N}\sum_{k=0}^{N-1} X(e^{j\omega_n})e^{j\omega_n k}} \quad \text{Inverse Transform}$$

5.2.2 Autocorrelation (Discrete Periodic Signal)

$$\boxed{R_x(\tau) = \tfrac{1}{N}\sum_{k=0}^{N-1} x(k)x(k-\tau)}$$

The Fourier transform of $R_x(\tau)$ is now defined as the **power spectral density**, since it is normalized with the signal length.

$$\boxed{\phi_x(e^{j\omega_n}) = \sum_{\tau=0}^{N-1} R_x(\tau)e^{-j\omega_n\tau} = \tfrac{1}{N}|X(e^{j\omega_n})|^2}$$

The energy in a single period is:

$$\sum_{k=0}^{N-1}|x(k)|^2 = \sum_{n=0}^{N-1}\phi_x(e^{j\omega_n})$$

```
Correlation('periodic',u);
```

5.2.3 Cross-Correlation (Discrete Periodic Signal)

$$\boxed{R_{yu}(\tau) = \tfrac{1}{N}\sum_{k=0}^{N-1} y(k)u(k-\tau)}$$

The Fourier transform of $R_{yu}(\tau)$ is now defined as the **cross-spectral density**.

$$\boxed{\phi_{yu}(e^{j\omega_n}) = \sum_{\tau=0}^{N-1} R_{yu}(\tau)e^{-j\omega_n\tau}) = \tfrac{1}{N}Y(e^{j\omega_n})U^*(e^{j\omega_n})}$$

```
Correlation('periodic',u,y);
```

5.3 Random Signal

Normally distributed noise:

$$\boxed{e(k) \in \mathcal{N}(0,\lambda) \Rightarrow \begin{cases}\mathbb{E}\left[e(k)\right] = 0 \text{ (zero mean)}\\ \mathbb{E}\left[|e(k)|^2\right] = \lambda \text{ (variance)}\end{cases}}$$

The $e(k)$ are independent and identically distributed (i.i.d.).

```
standard_deviation = 2;
variance = standard_deviation^2;
bias = 0;
N = 1024;
u = randn(N,1)*standard_deviation + bias;
```

5.3.1 Autocovariance (Random Signal)

$$\begin{aligned} R_x(\tau) &= \mathbb{E}\left[x(k)x(k-\tau)\right] \\ &= \mathbb{E}\left[x(k)x^*(k-\tau)\right] \text{ (in the complex case)} \\ &= \mathbb{E}\left[x(k)x^*(x-\tau)\right] \text{ (in the multivariable case)} \end{aligned}$$

General (non-stationary, non-zero mean) case:

$$\begin{aligned} R_x(s,t) &= \mathbb{E}\left[(x(s) - \mathbb{E}\left[x\right])(x(t) - \mathbb{E}\left[E\right])\right] \\ &= \mathbb{E}\left[x(s)x(t)\right] \text{ (if zero mean)} \\ &= R_x(s-t) \text{ (if stationary)} \end{aligned}$$

Further properties are

- $R_x(-\tau) = R_x^*(\tau)$
- $R_x(0) \geq |R_x(\tau)| \; \forall \tau > 0$

```
xcorr(u)/N; %Lecture 3.37
```

5.3.2 Power Spectral Density (Random Signal)

$$\boxed{\phi_x(e^{j\omega}) := \sum_{\tau=-\infty}^{\infty} R_x(\tau)e^{-j\omega\tau} \text{ where } \omega \in [-\pi, \pi)}$$

For a zero-mean random signal:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} |x(k)|^2 = \text{Var}\,(x(k)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_x(e^{j\omega}) d\omega$$

Further properties are

- $\phi_x(e^{j\omega}) \in \mathbb{R}$
- $\phi_x(e^{j\omega}) \geq 0 \; \forall \; \omega$
- $\phi_x(e^{j\omega}) = \phi_x(e^{-j\omega})$ for all real-valued $x(k)$

```
fft(xcorr(u)/N)) %Lecture 3.37 (?)
```

5.3.3 Cross-Covariance (Random Signal)

$$\boxed{R_{yu}(\tau) = \mathbb{E}\left[(y(k) - \mathbb{E}\left[y(k)\right])(u(k-\tau) - \mathbb{E}\left[u(k)\right]\right]}$$

For zero mean signals:

$$R_{yu}(\tau) = \mathbb{E}\left[y(k)u(k-\tau)\right]$$

Joint stationarity is required to make the definition dependent on τ only.
If $R_{yu}(\tau) = 0$ for all τ then $y(k)$ and $u(k)$ are uncorrelated.

```
xcorr(u,y)/N; %Lecture 3.37 (?)
```

5.3.4 Cross Power Spectral Density (Random Signal)

$$\boxed{\phi_{yu}(e^{j\omega}) = \sum_{\tau=-\infty}^{\infty} R_{yu}(\tau)e^{-j\omega\tau}, \; \omega \in [-\pi, \pi)}$$

The inverse is,

$$R_{yu}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{yu}(e^{j\omega})e^{j\omega\tau} d\omega$$

```
fft(xcorr(y,u)/N); %Lecture 3.37 (?)
```

5.4 Finite Length Signal

5.4.1 Discrete-Fourier Transform (Finite Length Signal)

$$\boxed{X_N(e^{j\omega_n}) = \sum_{k=0}^{N-1} x(k)e^{-j\omega_n k}, \text{ where } \omega_n = \frac{2\pi n}{N}}$$

The inverse DFT is

$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X_N(e^{j\omega_n})e^{j\omega_n k}, \quad k = 0, \ldots, N-1$$

5.4.2 Periodogram (Finite Length Signal)

$$\boxed{\frac{1}{N} \left|V_N(e^{j\omega})\right|^2}$$

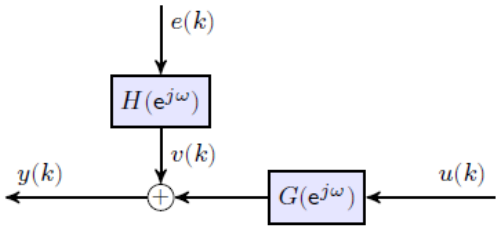
An asymptotically unbiased estimator of the spectrum is

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N}|V_N(e^{j\omega}|^2\right] = \phi_v(\omega)$$

This assumes that the autocorrelation decays quickly enough:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\tau=-N}^{N} |\tau R_v(\tau)| = 0$$

6 ETFE



Linear, time-invariant system, $g(l)$:

$$y(k) = \sum_{l=0}^{\infty} g(l)u(k-l) + v(k), \quad k = 0, 1, \ldots$$

Assumptions:

1. causal system: $g(l) = 0, \forall l < 0$
2. noise: $E\{v(k)\} = 0$, zero mean, stationary

Given $\{u(k), y(k)\}$ find an estimate $\hat{G}(e^{j\omega})$ such that it fits the $G(e^{j\omega})$.

$$\boxed{\text{Bias}(\hat{G}))G - E\{\hat{G}\}} \text{ Bias}$$

$$\boxed{\text{var}((\hat{G}) = E\left\{|\hat{G} - E\{\hat{G}\}|^2\right\}} \text{ Variance}$$

$$\boxed{\text{MSE}(\hat{G}) = E\left\{|G - \hat{G}|^2\right\}} \text{ Mean-square error}$$

Note that MSE($\hat{G}$) =var($\hat{G}$)+Bias^2($\hat{G}$).

6.1 Input-output relationship

For finite energy signals:

$$y(k) = \sum_{l=0}^{\infty} g(l)u(k-l) + v(k)$$

$$Y(e^{j\omega}) = G(e^{j\omega})U(e^{j\omega}) + V(e^{j\omega})$$

which in the idealized case leads to:

$$\boxed{\frac{Y(e^{j\omega})}{U(e^{j\omega})} = G(e^{j\omega}) + \frac{V(e^{j}\omega)}{U(e^{j\omega})} \approx G(e^{j\omega})}$$

In reality we only have N samples:

$$\underbrace{Y_N(e^{j\omega_n})}_{\text{length-N DFT}} = \sum_{k=0}^{N-1} y(k)e^{-j\omega_n k} \approx \sum_{k=-\infty}^{\infty} y(k)e^{-j\omega_n k} = Y(e^{j\omega_n}$$

$$\underbrace{U_N(e^{j\omega_n})}_{\text{length-N DFT}} = \sum_{k=0}^{N-1} u(k)e^{-j\omega_n k} \approx \sum_{k=-\infty}^{\infty} u(k)e^{-j\omega_n k} = U(e^{j\omega_n})$$

$$\boxed{\hat{G}_N(e^{j\omega_n}) := \frac{Y_N(e^{j\omega_n})}{U_N(e^{j\omega_n})}} \text{ ETFE}$$

6.2 Periodic input case

Period M inputs: $u(k) = u(k+M)$
If $sM = N$ for an integer s, the fourier series over N samples is equal to the real fourier series!

$$U_N(e^{j\omega_n}) = U(e^{j\omega_n}) \forall \omega_n = \tfrac{2\pi n}{N}, \; n = 0, \ldots, N-1$$

Then

$$Y_N(e^{j\omega_n}) = G(e^{j\omega_n})U_N(e^{j\omega_n}) + V_N(e^{j\omega_n})$$

$$\hat{G}_N(e^{j\omega_n}) = G(e^{j\omega_n}) + \tfrac{V_N(e^{j\omega_n})}{U_N(e^{j\omega_n})}$$

Bias:

$$E\{\hat{G}_N(e^{j\omega_n})\} = G(e^{j\omega_n}) + E\left\{\tfrac{V_N(e^{j\omega_n})}{U_N(e^{j\omega_n})}\right\} = G(e^{j\omega_n})$$

when assuming zero mean noise. Thus for periodic inputs with N being an integer number of periods, **the ETFE is unbiased**.
Variance:
For the unbiased case:

$$E\left\{|\hat{G}_N(e^{j\omega_n}) - G(e^{j\omega_n})|^2\right\} = \tfrac{\phi_v(e^{j\omega_n}) + \frac{2}{N}c}{\frac{1}{N}|U_N(e^{j\omega_n})|^2}$$

where $|c| \leq C = \sum_{\tau=1}^{\infty} |\tau R_v(\tau)|$ is assumed to be finite.
For estimates at different frequencies ($\omega_n \neq \omega_i$):

$$E\left\{(\hat{G}_N(e^{j\omega_n}) - G(e^{j\omega_n}))(\hat{G}_N(e^{-j\omega_i}) - G(e^{-j\omega_i}))\right\} = 0$$

Transient responses:
Initial transient corrupts the measurement

$$y(k) = G(u_{periodic}(k)W_{[0,N-1]}(k)) + v(k)$$

with the window function:

$$W_{[0,N-1]}(k) = \begin{cases} 1 & \text{if } 0 \leq 0 < N \\ 0 & \text{otherwise} \end{cases}$$

For all outputs up to time $k = N - 1$

$$y(k) = Gu_{periodic}(k) - \underbrace{G(u_{periodic}W_{(-\infty,-1)})}_{r(k)} + v(k)$$

$$Y_N(e^{j\omega_n}) = G(e^{j\omega_n})U_N(e^{j\omega_n}) + R_N(e^{j\omega_n}) + V_N(e^{j\omega_n})$$

The input in negative time, which is present in a ideal periodic input, and missing in a real periodic input, has an influence on positive time, which is described by $r(k)$.

When using a periodic signal multiple times the resulting DFT does not contain more information, since in a periodic signal there are only a certain number of frequencies contained, but the energy in those frequencies increases!

Transient bias error:

$$\hat{G}(e^{j\omega_n}) = \frac{Y_N(e^{j\omega_n})}{U_N(e^{j\omega_n})} = G(e^{j\omega_n}) + \frac{R_N(e^{j\omega_n})}{U_N(e^{j\omega_n})} + \frac{V_N(e^{j\omega_n})}{U_N(e^{j\omega_n})}$$

For periodic $u(k)$
As $N = mM, m \to \infty$

$$|U_N(e^{j\omega_n})| = m|U_M(e^{j\omega_n})|$$

For random $u(k)$
As $N \to \infty$

$$E\{|U_N(e^{j\omega_n})|\} \to \sqrt{N}\sqrt{\phi_u(e^{j\omega_n})}$$

Thus

$$\boxed{\left|\frac{R_N(e^{j\omega_n})}{U_N(e^{j\omega_n})}\right| \to 0 \text{ with rate } \begin{cases} \frac{1}{N} & \text{for periodic input} \\ \frac{1}{\sqrt{N}} & \text{for random inputs} \end{cases}}$$

A fix for getting rid of the influence of the transient response: Get rid of the first period.

```
r = 4;
N_split = N/r;
u_re = reshape(u(N_split+1:end),N_split,[]);
y_re = reshape(y(N_split+1:end),N_split,[]);

U_re = fft(u_re);
Y_re = fft(y_re);

G_re = Y_re./U_re;

G_avg = mean(G_re,2);
```

6.3 Spectral Transformations

If $v(k) = 0$

$$\phi_y(e^{j\omega_n}) = G(e^{j\omega_n})\phi_u(e^{j\omega_n})G^T(e^{j\omega_n})$$

where $G^T(e^{j\omega_n})$ is the complex conjugate of $G(e^{j\omega_n})$.
If $v(k) \neq 0$ and uncorrelated

$$\phi_y(e^{j\omega_n}) = |G(e^{j\omega_n})|^2\phi_u(e^{j\omega_n}) + |H(e^{j\omega_n})|^2$$

But this approach has no more phase information. For that reason use the cross spectrum:

$$\phi_{yu}(e^{j\omega_n}) = G(e^{j\omega_n})\phi_u(e^{j\omega_n}) + \phi_{uv}(e^{j\omega_n}) = G(e^{j\omega_n})\phi_u(e^{j\omega_n})$$

if $u(k)$ and $v(k)$ are uncorrelated.

$$\boxed{\hat{G}(e^{j\omega}) = \frac{\hat{\phi}_{yu}(e^{j\omega_n})}{\hat{\phi}_u(e^{j\omega_n})}} \text{ Spectral estimation methods}$$

where

$$\phi_y(e^{j\omega_n}) = |G(e^{j\omega_n})|^2\phi_u(e^{j\omega_n}) + \phi_v(e^{j\omega_n})$$

$$\phi_{yu}(e^{j\omega_n}) = G(e^{j\omega_n})\phi_u(e^{j\omega_n})$$

6.4 Approaches to spectral estimation

6.4.1 Spectral estimation (Periodogram)

The periodogram is an asymptotically unbiased estimator of the spectrum given $\lim_{n\to\infty} \frac{1}{N} \sum_{\tau=-N}^{N} |\tau R_v(\tau)| = 0$

$$\frac{1}{N}|V_N(e^{j\omega_n})|^2 \quad \text{Periodogram}$$

$$\boxed{\lim_{N\to\infty} E\left\{\frac{1}{N}|V_N(e^{j\omega_n})|^2\right\} = \phi_v(e^{j\omega_n})}$$

which is under the assumption

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\tau=-N}^{N} |\tau R_v(\tau)| = 0$$

6.4.2 Spectral estimation (via covariances)

The autocovariance of the noise for stochastic $v(k)$ is described as:

$$\hat{R}_v(\tau) = \begin{cases} \frac{1}{N-|\tau|} \sum_{k=\tau}^{N_1} v(k)v(k-\tau), & \text{for } \tau \geq 0 \\ \frac{1}{N-|\tau|} \sum_{k=0}^{N+\tau-1} v(k)v(k-\tau), & \text{for } \tau < 0 \end{cases}$$

This is an unbiased estimator of $R_v(\tau)$: $\mathrm{E}\left\{\hat{R}_v(\tau)\right\} = R_v(\tau)$

$$\hat{\phi}_v(e^{j\omega_n}) = \sum_{\tau=-N+1}^{N-1} \hat{R}_v(\tau)e^{-j\omega\tau}$$

```
xcorr(a-mean(a))-xcov(a) == 0 %true
```

The functions above both calculate the zero-mean shifted autocorrelation of a signal which is **not** equivalent to the autocovariance of the signal! They lack the normalization by $N - |\tau|$!

```
Covariance('zero-mean',a);
Covariance('zero-mean',a,b);
```

6.4.3 Spectral estimation (periodic signals)

Periodic signal $x(k)$ with period M, $N = mM$ for some integer m

$$R_x(\tau) = \frac{1}{M} \sum_{k=0}^{M-1} x(k)x(k-\tau)$$

The power spectral density can be calculated and is equal to the periodogram:

$$\phi_x(e^{j\omega_n}) = \sum_{\tau=0}^{M-1} R_x(\tau)e^{-j\omega_n\tau} = \frac{1}{M}\left|X_M(e^{j\omega_n})\right|^2$$

```
Ru = Correlation('periodic',u);
```

6.4.4 Spectral estimation (more general case)

Alternative autocorrelation estimate:

$$\hat{R}_x(\tau) = \begin{cases} \frac{1}{N} \sum_{k=\tau}^{N-1} x(k)x(k-\tau), & \text{for } \tau \geq 0 \\ \frac{1}{N} \sum_{k=0}^{N+\tau-1} x(k)x(k-\tau), & \text{for } \tau < 0 \end{cases}$$

Periodic $x(k)$: unbiased (exact) if $N = mM$

Random $x(k)$ biased $\mathrm{E}\left\{\hat{R}_x(\tau)\right\} = \frac{N-|\tau|}{N}R_x(\tau)$.
asymptotically biased as $N \to \infty, \tau/N \to 0$

```
Ru = xcorr(u)/N;
```

- The bias when estimating R_x for a random signal has the form a triangular weighting across τ. A scaling of the signal would remove the bias, which is effectively done by the spectral estimation for via covariances 6.4.2.

7 Averaging and Smoothing

Multiple experiments $u_r(k), y_r(k), r = 1\ldots, R, k = 0, \ldots, K-1$

$$\hat{G}(e^{j\omega_n}) = \sum_{r=1}^{R} \alpha_r \hat{G}_r(e^{j\omega_n})$$

where $\sum_{r=1}^{G} \alpha_r = 1$ and for calculating the average $\alpha_r = \frac{1}{R}$.

The averaging can be optimized by selecting α_r such that the variance $\sigma_r^2(e^{j\omega_n})$ is minimized.

$$\mathrm{Var}\left(\hat{G}(e^{j\omega_n})\right) = \mathrm{Var}\left(\sum_{r=1}^{R} \alpha_r(e^{j\omega_n})\hat{G}_r(e^{j\omega_n})\right) = \sum_{r=1}^{R} \alpha_r^2\sigma_r^2(e^{j\omega_n})$$

This is minimized by

$$\alpha_r(e^{j\omega_n}) = \frac{1/\sigma_r^2(e^{j\omega_n})}{\sum_{r=1}^{T} 1/\sigma_r^2(e^{j\omega_n})}$$

Thus the signal is weighted inversely proportional to the variance.

Thus if $\mathrm{Var}\left(\hat{G}_r(e^{j\omega_n})\right) = \frac{\phi_v(e^{j\omega_n})}{\frac{1}{N}|U_r(e^{j\omega_n})|^2}$ then $\alpha_r(e^{j\omega_n}) = \frac{|U_r(e^{j\omega_n})|^2}{\sum_{r=1}^{R}|U_r(e^{j\omega_n})|^2}$.

The best result is obtained if the input is the same for all r, which will lead to a reduction of the variance as follows:

$$\mathrm{Var}\left(\hat{G}(e^{j\omega_n})\right) = \frac{\mathrm{Var}\left(\hat{G}_r(e^{j\omega_n})\right)}{R}$$

Biased estimates will reduce the improvement in variance.

- Since we are adding complex numbers the magnitude of the average is not equal to the average of the magnitudes r_i.

7.1 Bias-variance trade-offs in data record splitting

Divide a data record into smaller parts for averaging:

$$\{u(k), y(k)\}, k = 0, \ldots, K-1$$

Choose R records and calculation length N, such that $NR \leq K$:

$$u_r(n) = u(rN + n)$$

And average the resulting estimates:

$$\boxed{\hat{G}(e^{j\omega_n}) = \frac{1}{R}\sum_{r=0}^{R-1}\hat{G}_r(e^{j\omega_n}) = \frac{1}{R}\sum_{r=0}^{R-1}\frac{\hat{Y}_r(e^{j\omega_n})}{\hat{U}_r(e^{j\omega_n})}}$$

As R increases:

- The number of points calculated, N decreases.
- The variance decreases (by up to $1/R$).
- The bias increases (due to non-periodicity transients).

Mean-square error

- Transient bias grows linearly with the number of data splits.
- Variance decays with a rate of up to 1/(number of averages).

What if there is no option of running periodic input experiments? $\rightarrow$ exploit the assumed smoothness of the underlying system.

7.2 Smoothing the ETFE

Assume the true system to be close to constant for a range of frequencies: $G(e^{j\omega_{n+r}}) \approx G(e^{j\omega_n})$ for $r = 0, \pm 1, \ldots, \pm r$.
The minimum variance smoothed estimate is:

$$\tilde{G}_N(e^{j\omega_n}) = \frac{\sum_{r=-R}^{R}\alpha_r\hat{G}_N(e^{j\omega_{n+r}})}{\sum_{r=-R}^{R}\alpha_r}, \qquad \alpha_r = \frac{\frac{1}{N}|U_N(e^{j\omega_r+n})|^2}{\phi_v(e^{j\omega_{n+r}})}$$

The summation above can then be approximated by an integral:

$$\approx \frac{\int_{\omega_{n-r}}^{\omega_{n+r}}\alpha(e^{j\zeta})\hat{G}_N(e^{j\zeta})d\zeta}{\int_{\omega_{n-r}}^{\omega_{n+r}}\alpha(e^{j\zeta})d\zeta}, \qquad \text{with } \alpha(e^{j\zeta}) = \frac{\frac{1}{N}|U_N(e^{j\zeta})|^2}{\phi_v(e^{j\zeta})}$$

Which can be reformulated using a smoothing window:

$$\tilde{G}_N(e^{j\omega_n}) = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}W_\gamma(e^{j(\zeta-\omega_n)})\alpha(e^{j\zeta})\hat{G}_N(e^{j\zeta})d\zeta}{\frac{1}{2\pi}\int_{-\pi}^{\pi}W_\gamma(e^{j(\zeta-\omega_n)})\alpha(e^{j\zeta})d\zeta} \qquad \text{with } \alpha(e^{j\zeta}) = \frac{\frac{1}{N}|U_N(e^{j\zeta})|^2}{\phi_v(e^{j\zeta})}$$

7.2.1 Assumptions on $\phi_v(e^{j\omega})$

Assume $\phi_v(e^{j\omega})$ is also a smooth function of frequency.

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}W_\gamma(e^{j(\zeta-\omega_n)})\left|\frac{1}{\phi_v(e^{j\zeta})} - \frac{1}{\phi(e^{j\omega_n})}\right|d\zeta \approx 0$$

Then use,

$$\alpha(e^{j\zeta}) = \frac{\frac{1}{N}|U_N(e^{j\zeta})|^2}{\phi_v(e^{j\omega_n})}$$

to get

$$\boxed{\tilde{G}_N(e^{j\omega_n}) = \frac{\frac{1}{2\pi}\int_{-pi}^{\pi}W_\gamma(e^{-j(\zeta-\omega_n)})\frac{1}{N}|U_N(e^{j\zeta})|^2\hat{G}_N(e^{j\zeta})d\zeta}{\frac{1}{2\pi}\int_{-\pi}^{\pi}W_\gamma(e^{j(\zeta-\omega_n)})\frac{1}{N}|U_N(e^{j\zeta})^2d\zeta}}$$

The wider the frequency window (decreasing γ)

- the more adjacent frequencies included in the smoothness estimate.
- the smoother the result.
- the lower the noise induced variance.
- the higher the bias.

7.2.2 Characteristic windows

$$\boxed{W_\gamma(e^{j\omega}) = \frac{1}{\gamma}\left(\frac{\sin\gamma\omega/2}{\sin\omega/2}\right)^2} \text{ Bartlett}$$

$$\boxed{W_\gamma(e^{j\omega}) = \tfrac{1}{2}D_\gamma(\omega) + \tfrac{1}{4}D_\gamma(\omega-\pi/\gamma) + \tfrac{1}{4}D_\gamma(\omega+\pi/\gamma)} \text{ Hann}$$

where

$$D_\gamma(\omega) = \frac{\sin\omega(\gamma+0.5)}{\sin\omega/2}$$

Properties of window functions:

- $\frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j\zeta})d\zeta = 1$
- $\int_{-\pi}^{\pi} \zeta W_\gamma(e^{j\zeta})d\zeta = 0$
- $M(\gamma) := \int_{-\pi}^{\pi} \zeta^2 W_\gamma(e^{j\zeta})d\zeta$
- $\bar{W}(\gamma) := 2\pi\int_{-\pi}^{\pi} W_\gamma^2(e^{j\zeta})d\zeta$

Bartlett	$M(\gamma) = \frac{2.78}{\gamma}$,	$\bar{W}(\gamma) \approx 0.67\gamma$	(for $\gamma > 5$)
Hamming	$M(\gamma) = \frac{\pi^2}{2\gamma^2}$,	$\bar{W}(\gamma) \approx 0.75\gamma$	(for $\gamma > 5$)

- $M(\gamma)$ gives an idea of the bias effect.
- $\bar{W}(\gamma)$ gives an idea of the variance effect.

7.2.3 Asymptotic bias properties

$$\mathrm{E}\left\{\tilde{G}(e^{j\omega_n}) - \mathrm{E}\left\{G(e^{j\omega_n})\right\}\right\} = \mathrm{E}\left\{\tilde{G}(e^{j\omega_n}) - G(e^{j\omega_n})\right\} =$$
$$M(\gamma)\left(\tfrac{1}{2}\underbrace{G''(e^{j\omega_n})}_{\text{curvature}} + \underbrace{G'(e^{j\omega_n})}_{slope}\frac{\phi_u'(e^{j\omega_n})}{\phi_u(e^{j\omega_n})}\right) + H.O.T.$$

Increasing γ

- makes the frequency window smaller.
- averages over fewer frequency values.
- makes $M(\gamma)$ smaller
- reduces the bias of the smoothed estimate $\tilde{G}(e^{j\omega_n})$

7.2.4 Asymptotic variance properties

$$\mathrm{E}\left\{(\tilde{G}(e^{j\omega_n}) - \mathrm{E}\left\{\tilde{G}(e^{j\omega_n})\right\})^2\right\} = \tfrac{1}{N}\bar{W}(\gamma)\frac{\phi_v(e^{j\omega_n})}{\phi_u(e^{j\omega_n})} + H.O.T.$$

Increasing γ

- makes the frequency window narrower.
- averages over fewer frequency values.
- makes $\bar{W}_\gamma$ larger.
- increases the variance of the smoothed estimate $\tilde{G}(e^{j\omega_n})$.

7.2.5 Asymptotic MSE properties

$$\mathrm{E}\left\{|\tilde{G}(e^{j\omega_n}) - G(e^{j\omega_n})|^2\right\} \approx M^2(\gamma)|F(e^{j\omega_n})|^2 + \tfrac{1}{N}\bar{W}(\gamma)\frac{\phi_v(e^{j\omega_n})}{\phi_u(e^{j\omega_n})}$$

where

$$F(e^{j\omega_n}) = \tfrac{1}{2}G''(e^{j\omega_n}) + G'(e^{j\omega_n})\frac{\phi_u'(e^{j\omega_n})}{\phi_u(e^{j\omega_n})}$$

If $M(\gamma) = M/\gamma^2$ and $\bar{W}(\gamma) = \bar{W}\gamma$ then MSE is minised by:

$$\boxed{\gamma_{optimal} = \left(\frac{4M^2|F(e^{j\omega_n})|^2\phi_u(e^{j\omega_n})}{\bar{W}\phi_v(e^{j\omega_n})}\right)^{1/5} N^{1/5}}$$

and

$$\text{MSE at } \gamma_{optimal} \approx CN^{-4/5}$$

8 Windowing and Input Signals

$$\phi_{yu}(e^{j\omega}) = G(e^{j\omega})\phi_u(e^{j\omega})$$

$$\boxed{\hat{G}(e^{j\omega_n}) = \frac{\hat{\phi}_{yu}(e^{j\omega_n})}{\hat{\phi}_u(e^{j\omega_n})}}$$

Recall that the smoothed ETFE is:

$$\tilde{G}_N(e^{j\omega_n}) = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)})\frac{1}{N}|U_N(e^{j\zeta})|^2\hat{G}_N(e^{j\zeta})d\zeta}{\frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)})\frac{1}{N}|U_n(e^{j\zeta})|^2 d\zeta}$$

The denominator term approaches $\frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)})\phi(e^{j\omega_n})d\zeta$ as $N \to \infty$.

If in addition $W_\gamma(e^{j\omega})$ is concentrated around $\zeta = 0$ (i.e. $\gamma/N \to 0$) then the denominator term approaches $\phi_u(e^{j\omega_n})$ as $N \to \infty$.

This motivates the smoothed spectral estimate:

$$\boxed{\tilde{\phi}_u(e^{j\omega_n}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)})\frac{1}{N}|U_N(e^{j\omega})|^2 d\zeta}$$

Similarly the numerator approaches ϕ_{yu} as $N \to \infty$:

$$\boxed{\tilde{\phi}_{yu}(e^{j\omega_n}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)})\frac{1}{N}|U_N(e^{j\omega})|^2 \hat{G}_N(e^{j\omega}) d\zeta}$$

For this reason the smoothed ETFE is equal to the smoothed spectral estimate for $N \to \infty$.

8.1 Frequency domain smoothing in matlab

```
gamma = 80;
U = fft(u); Y = fft(y);
G_est = Y./U;
G_est_smooth = G_est*0;

[om,Wg] = WfHann(g,N);
%shift to start at zero
zidx = find(om == 0);
omega = [om(zidx:N);om(1:zidx-1)];
Wg = [Wg(zidx:N) Wg(1:zidx-1)];
%variance weighting
a = U.*conj(U);
```

```
for wn = 1:N
  %reset normalization
  Wnorm = 0;
    for xi = 1:N
      %wrap window index
        widx = mod(xi-wn,N)+1;
        G_est_smooth(wn) = G_est_smooth(wn) +...
            Wg(widx)*G_est(xi)*a(xi);
        Wnorm = Wnorm + Wg(widx)*a(xi);
    end
    %weigh normalisation
    G_est_smooth(wn) = G_est_smooth(wn)/Wnorm;
end
```

8.2 Time domain windows

Define, via the inverse Fourier transform a time domain window:

$$\omega_\gamma(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j\omega})e^{j\zeta\tau} d\zeta$$

Then the smoothed input spectral estimate $\tilde{\phi}_u(e^{j\omega_n})$ is:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)})\frac{1}{N}|U_N(e^{j\omega})|^2 d\zeta \approx \sum_{\tau-\infty}^{\infty} \omega_\gamma(\tau)\hat{R}_u(\tau)e^{-j\tau\omega_n}$$

where

$$\omega_\gamma = \begin{cases} 0 & \text{for } \tau < -\gamma \\ > 0 & \text{for } -\gamma \le \tau \le \gamma \\ 0 & \text{for } \tau > \gamma \end{cases}$$

where often $\gamma << N$, which enables the faster calculated redefinition:

$$\boxed{\frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)})\frac{1}{N}|U_N(e^{j\omega})|^2 d\zeta \approx \sum_{\tau-\gamma}^{\gamma} \omega_\gamma(\tau)\hat{R}_u(\tau)e^{-j\tau\omega_n}}$$

The cross spectral estimate can also be formulated as a convolution in the frequency domain which leads to the analogous formulation to the spectral estimate of u:

$$\boxed{\tilde{\phi}_u(e^{j\omega_n}) = \sum_{\tau=-\gamma}^{\gamma} \omega_\gamma(\tau)\hat{R}_u(\tau)e^{-j\tau\omega_n}}$$

$$\boxed{\tilde{\phi}_{yu}(e^{j\omega_n}) = \sum_{\tau=-\gamma}^{\gamma} \omega_\gamma(\tau)\hat{R}_{yu}(\tau)e^{-j\tau\omega_n}}$$

8.2.1 Time domain smoothing in matlab

```
gamma = 80;
[~,Wg] = WtHann(gamma,N);
R_u = xcorr(u,N/2)/N; %Lecture 3.37
R_yu = xcorr(y,u,N/2)/N;

omega = Omega_n(N);
phi_u = zeros(size(omega));
phi_yu = zeros(size(omega));

tau = -gamma:gamma;
ind = tau+N/2;
for i = 1:N
%Lecture 5.9 and 5.11
    phi_u(i) = sum(Wg(tau+N/2).*R_u(ind).*exp(-1j.*tau.'.*
        omega(i)));
    phi_yu(i) = sum(Wg(tau+N/2).*R_yu(ind).*exp(-1j*tau.'.*
        omega(i)));
end
G_est_smooth = phi_yu./phi_u;
```

8.2.2 Window characteristics

Decreasing γ: narrower $\omega_\gamma(\tau)$, wider $W_\gamma(e^{j\omega})$

- the more frequencies, $\hat{G}(e^{j\omega_n})$ included in the smoothing.
- the fewer $\hat{R}(\tau)$ estimates included in the smoothing.
- the smoother the result.
- the lower the noise induced variance.
- the **higher the bias**.

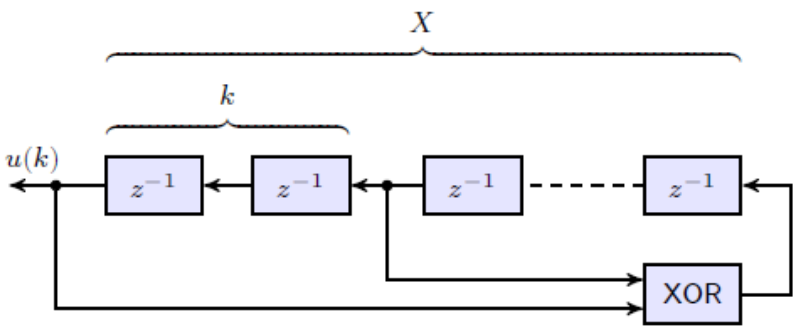
8.3 Input Signals

- Steps
- Doublet
- Sinusiods, Chirpts, Multi-Sines
- Filtered white noise
- Pseudo-Random Binary Signals (PRBS)

8.3.1 PRBS

$$u(k) = a \text{ or } -a$$

Shift-register generation



X	k
3	1 or 2
4	1 or 3
5	2 or 3
6	1 or 5
7	1 or 6
8	—
9	4 or 5
10	3 or 7

Periodicity

Periodic with period equal to at most $M = 2^X - 1$.

$$R_u(\tau) = \frac{1}{N}\sum_{k=0}^{N-1} u(k)u(k-\tau) = \begin{cases} a^2 & \text{if } \tau = 0 \\ \frac{-a^2}{2^X-1} & \text{if } \tau \neq 0 \end{cases}$$

Definition 6. ***Run length*** *defines how long the signal stays high.*

The run length distribution of $u(k)$ is then:

1/2 runs of length 1

1/4 runs of length 2

1/8 runs of length 3

⋮

Other properties:

- Equal energy at all frequencies.
- The maximum period of a PRBS signal can be found based on the discrete time dynamic system that generates the signal:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 2\gamma \end{bmatrix}$$

$$z = Ax(k)$$
$$x(k+1) = \mathrm{mod}_2(z(k))$$
$$y = Cx(k) - \gamma$$

where γ is the amplitude of the signal.

Now the maximum length period result from the idea that $x(k)$ and therefore the output $y(k)$ can only change so many times until it reaches a state it has already held before. From that point on, since the system is first order, the signal repeats. Thus in the longest case, all possible values contained in x are taken exactly once. Thus the maximum periodic length is equal to the number of possible binary numbers with 6 bits, where 6 is the order of the PRBS signal.

$\boxed{2^6 - 1}$ Maximum Length Period

8.3.2 PRBS in matlab

```
signal_order = 8;
singal_length = 2.^signal_order-1;
u = idinput(signal_length,'PRBS');
```

- Note that `idinput` only allows the choice of the total length of the signal and derives the fitting signal order such that the signal is at least of that chosen length. A warning is made if the chosen signal length is not equivalent to the run length.

8.3.3 Multi-sinusoidal signals

$$u(k) = \sum_{s=1}^{S} \sqrt{2\alpha_s} \cos(\omega_s k T + \phi_s)$$

where T is the sampling period, $\omega_s = \frac{2\pi}{T_P}$, $\frac{T_P}{T} = N$, $S \leq \frac{N}{2}$.

Choose N to be a power of 2 for efficient FFT calculations.

$$\sum_{s=1}^{S} \alpha_s = 1 \quad \text{Total signal power}$$

Schroeder phasing
Select the phases ϕ_s such that the minimize the peak amplitude:

$$\phi_s = 2\pi \sum_{j=1}^{s} j\alpha_s.$$

for equal power in each sinusoids:

$$\alpha_s = 1/S \text{ and } \phi_s = \tfrac{\pi(s^2+s)}{S}$$

9 Residual Spectra, Coherency, Aperiodicty, Offsets and Drifts

9.1 Residual Spectrum

9.1.1 Estimating $\phi_v(e^{j\omega_n})$

$$v(k) = y(k) - G(e^{j\omega})u(k)$$

$$\tilde{\phi}_v(e^{j\omega_n}) \approx \tfrac{1}{N}\tfrac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(e^{j(\zeta-\omega_n)}) \left| Y_N(e^{j\omega}) - \tilde{G}(e^{j\omega})U_N(e^{j\omega}) \right|^2 d\zeta \approx \tilde{\phi}_y(e^{j\omega_n}) - \tfrac{|\tilde{\phi}_{yu}(e^{j\omega_n})|^2}{\tilde{\phi}_u(e^{j\omega_n})}$$

How much energy is accounted for by the model? How much by noise?

$$\phi_v(e^{j\omega_n}) = \phi_y(e^{j\omega_n}) \left(1 - \tfrac{|\phi_{yu}((e^{j\omega_n}))|^2}{\phi_y(e^{j\omega_n})\phi_u(e^{j\omega_n})}\right)$$

$$\hat{\kappa}_{yu}(e^{j\omega_n}) = \sqrt{\tfrac{|\hat{\phi}_{yu}(e^{j\omega_n})|^2}{\hat{\phi}_y(e^{j\omega_n})\hat{\phi}_u(e^{j\omega_n})}}$$ Coherency Spectrum

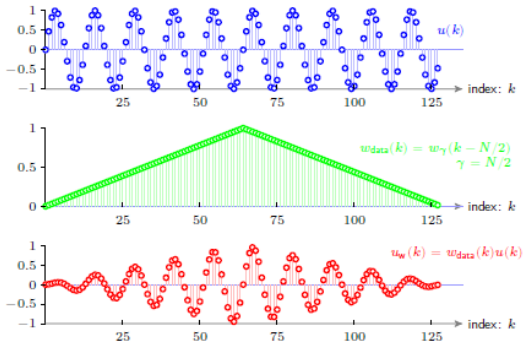
- If all of the energy in the output is due to the model for a frequency ω_n then $\hat{\kappa}_{yu}(e^{j\omega_n}) = 1$.
- This can be used as a measure of effectiveness of the modelling at a particular frequency.
- Theoretically, $0 \leq \hat{\kappa}_{yu}(e^{j\omega_n}) \leq 1$. One should aim to keep the coherency spectrum as high as possible. It can be adjusted by adjusting the smoothing.

9.2 Time-domain data windowing

Putting a time domain window directly on the data.

$$U_w(e^{j\omega_n}) = \sum_{k=0}^{N-1} w_{data}(k)u(k)e^{-jk\omega_n}$$

often with $w_{data}(k) = w_\gamma(k - N/2)$ (shifted to middle). Typically $\gamma = N/2$ such that all of the data is used.

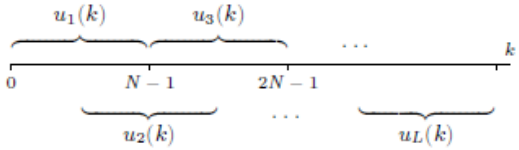


For an estimation of the periodogram **scaling is necessary!**

$$\frac{1}{E_{scl}}\frac{1}{N}|U_w(e^{j\omega_n})|^2 \quad \text{where } E_{scl} = \frac{\sum_{k=0}^{N-1}|w_{data}(k)u(k)|^2}{\sum_{k=0}^{N-1}|u(k)|^2} \approx \frac{1}{N}\sum_{k=0}^{N-1}|w_{data}(k)|^2 \quad \text{Periodogram}$$

- Transients influence the periodogram of a sequence, since by the application of the DFT, the signal is assumed to be periodic. Thus even if a sinusoid is sampled - if the periodic extension does not represent the same sinusoid - the DFT will show a range of frequencies, instead of a single distinct one.
- Time-domain data windowing can help reducing the influence of transients.

9.2.1 Welch's Method



1. Split the data record into L overlapping segments of length N.
2. $U_l(e^{j\omega_n}) = \sum_{k=0}^{N-1} w_{data}(k)u_l(k)e^{j\omega_n k}$
3. $\tilde{\phi}_u(e^{j\omega_n}) = \frac{1}{NLE_{scl}}\sum_{l=1}^{L}\left|U_l(e^{j\omega_n})\right|^2$

\+ Windowing can reduce transient response effects.

\+ Noise reduction from averaging and windowing.

\+ Variance error can be reduced.

\- Windowing can cause energy leakage to adjacent frequencies.

\- Frequency resolution deteriorates.

\- Bias error can be increased.

\- Noise on $u_l(k)$ and u_{l+1} is not uncorrelated.

- Do not use `welch()`, since it does not fit the definition here.

9.2.2 Drifts and Offsets

- Time domain windowing when an offset is present can lead to the introduction of additional frequencies close to zero, which is undesirable.
- Time domain windowing when a drift is present can lead to the introduction of additional frequencies close to the peak of the sampled sinusoid, which is undesirable as well.
- A possible solution is **preprocessing** the data via the assumption

  $$u_d(k) = u(k) - (alphak + \beta)$$

  where $\alpha k + \beta$ is the best linear fit to $u(k)$.

```
1 detrend(u);
```

\+ `detrend` can completely remove the effect of drifts and offsets.

\- It must be applied carefully, since if no drift/offset is present or if it is detected wrongly, large errors can be introduced.

10 Frequency Domain Subspace ID

$$x(k+1) = Ax(k) + Bu(k) \qquad A \in \mathbb{R}^{n_x \times n_x}$$
$$y(k) = Cx(k) + Du(k) \qquad D \in \mathbb{R}^{n_y \times n_u}$$

$$g(k) = \begin{cases} 0 & k = 0 \\ D & k = 0 \\ CA^{k-1} & k > 0 \end{cases} \quad \text{Pulse response coefficients}$$

$$G(e^{j\omega_n}) = \sum_{k=0}^{\infty} g(k)e^{-j\omega k} \quad 0 \le \omega \le \pi$$

Given

$$G(n) = G(e^{j\omega_n}) + V(e^{j\omega_n}), \quad n = 0, \ldots, N/2$$

Find

$$\hat{G}(e^{j\omega_n}) = \hat{C}\left(e^{j\omega}I - \hat{A}\right)^{-1}\hat{B} + \hat{D}$$

such that

$$\lim_{N\to\infty} ||\hat{G}(e^{j\omega_n}) - G(e^{j\omega_n})||_\infty = 0$$

$$\boxed{\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} \in \mathbb{R}^{n_y q \times n_x}} \text{ Extended observability}$$

$\text{rank}(\mathcal{O}) = n_x$ for all $n_y q \geq n_x$

$$\boxed{\mathcal{C} = \begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix} \in \mathbb{R}^{n_x \times n_u r}} \text{ Extended controllability}$$

$\text{rank}(\mathcal{C})=n_x$ for all $n_u r \geq n_x$
The ETFE matches the true transfer function only of there is no noise. In that case the inverse Fourier transform results in the time-aliased impulse response of the system h_k:

$$\boxed{h_k = \frac{1}{N}\sum_{n=0}^{N-1} G(e^{j\omega_n})e^{j2\pi kn/N} = CA^{k-1}\left(\sum_{l=0}^{\infty} A^{Nl}\right)B = CA^{k-1}\left(I - A^N\right)^{-1}B}$$

- We do not get the exact impulse response $g(k)$ since we do not have all the data until $N \to \infty$.

Now the **Hankel matrix** is formulated with the ultimate goal to extract estimates of the state space matrices.

$$H = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_r \\ h_2 & h_3 & \ddots & \ddots & h_{r+1} \\ h_3 & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & & \vdots \\ h_q & h_{q+1} & & \cdots & h_{q+r-1} \end{bmatrix} \quad \text{choose } q > n_x, r > n_x \text{ and } q + r - 1 \leq N - 1$$

which leads to:

$$\boxed{H = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix}\left(I - A^N\right)^{-1}\begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix}}$$

Next singular value decomposition can be used to decompose the Hankel matrix, which allows to calculate the **rank** of the state space, as the dimension of Σ_1.

$$H = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix}\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \quad \Sigma_1 \in \mathbb{R}^{n_x \times n_x}$$

Since U_2 is multiplied with $\mathbf{0}$ the $\text{range}(H) = \text{range}(U_1) = \mathcal{O}$. Which directly leads to $\text{span}(\{U_1\}) = \text{span}(\mathcal{O})$, which means that the observability matrix is some linear transform of U_1. Or differently said, for some choice of A they are equal. probably not correct!?

Next an estimate of $\hat{A}$ is made using the observability matrix, for which the selection matrix J is introduced:

$$J_1\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-2} \end{bmatrix} \text{ and } J_2\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} = \begin{bmatrix} CA \\ \vdots \\ CA^{q-1} \end{bmatrix}$$

Then $j_1\mathcal{O}A = J_2\mathcal{O}$
So $\hat{A}$ is the least squares solution to $J_1\hat{U}_1\hat{A} = J_2\hat{U}_1$.

And $\hat{C}$ can be found as: $J_e\mathcal{O} = C \implies J_3\hat{U}_1 = \hat{C}$.
Next, using the initial formulation:

$$\hat{G}(e^{j\omega_n}) = \hat{C}\left(e^{j\omega}I - \hat{A}\right)^{-1}\hat{B} + \hat{D}$$

$\hat{B}$ and $\hat{D}$ can be found as the solution to the minimization:

$$\boxed{\hat{B}, \hat{D} = \underset{B,D}{\text{argmin}} \sum_{n=0}^{N} ||G(e^{j\omega_n}) - D - \hat{C}(e^{j\omega_n}I - \hat{A})^{-1}B||_F^2}$$

10.1 Summarizing the subspace identification algorithm

Uniform data spacing case:

$$G(n), \quad \omega_n = \tfrac{\pi n}{N}, \quad n = 0, \ldots, N/2$$

1. Extend data to negative frequencies
$$G(n) = \bar{G}(N - n), \quad n = N/2 + 1, \ldots, N - 1$$
2. Calculate inverse DFT to obtain the time-aliased pulse response
$$\hat{h}_k = \tfrac{1}{N}\sum_{n=0}^{N-1} G(n)e^{j2\pi kn/N}, \quad k = 0, \ldots, N - 1$$
3. Form a block-Hakel matrix:

$$\hat{H} = \begin{bmatrix} \hat{h}_1 & \hat{h}_2 & \hat{h}_3 & \cdots \hat{h}_r & \\ \hat{h}_2 & \hat{h}_3 & \ddots & \ddots & \hat{h}_{r+1} \\ \hat{h}_3 & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & & \vdots \\ \hat{h}_q & \hat{h}_{q+1} & & \cdots & \hat{h}_{q+r+1} \end{bmatrix} \in \mathbb{R}^{n_y q \times \nu_r}$$

4. Calculate a singular value decomposition:

$$\hat{H} = \hat{U}\hat{\Sigma}\hat{V}^T$$

5. Select a model order, $\hat{n}_x$ and partition the SVD:

$$\hat{U}\hat{\Sigma}\hat{V}^T = \begin{bmatrix} \hat{U}_1 & \hat{U}_2 \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \hat{V}_1^T \\ \hat{V}_2^T \end{bmatrix}, \quad \hat{\Sigma} \in \mathbb{R}^{\hat{n}_x \times \hat{n}_x}$$

6. Estimate $\hat{A}$ via:

$$J_1 = \begin{bmatrix} I_{n_y(q-1)} & O_{n_y(q-1)\times n_y} \end{bmatrix}$$
$$J_2 = \begin{bmatrix} 0_{n_q(q-1)} \times n_y & I_{n_y(q-1)} \end{bmatrix}$$

Solve for $\hat{A}$ via LS:

$$\boxed{J_1\hat{U}_1\hat{A} = J_2\hat{U}_1}$$

7. Estimate $\hat{C}$ via

$$J_3 = \begin{bmatrix} I_{n_y} & 0_{n_y \times n_y(q_1)} \end{bmatrix}$$

$$\boxed{\hat{C} = J_3\hat{U}_1}$$

8. Find $\hat{B}$ and $\hat{D}$ via least squares:

$$\boxed{\hat{B}, \hat{D} = \underset{B,D}{\operatorname{argmin}} \sum_{n=0}^{N} ||G(e^{j\omega_n}) - D - \hat{C}(e^{j\omega_n} I - \hat{A})^{-1} B||_F^2}$$

9. Form the estimate

$$\boxed{\hat{G}(z) = \hat{D} + \hat{C}(zI - \hat{A})^{-1}\hat{B}}$$

- To get real valued B and D:

$$\begin{bmatrix} \text{real}\left((\hat{C}((e^{j\omega_n})I - \hat{A})^{-1}\right) & I \\ \text{imag}\left((\hat{C}((e^{j\omega_n})I - \hat{A})^{-1}\right) & 0 \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} \text{real}(G(n)) \\ \text{imag}(G(n)) \end{bmatrix}$$

10.1.1 Properties

- Asymptotic convergence

$$\lim_{N\to\infty} ||\hat{G}(e^{j\omega_n}) - G(e^{j\omega_n})||_\infty = 0, \quad n = 0, \ldots, N-1 \text{ w.p. } 1$$

- The algorithm is „correct“. If $V(e^{j\omega_n}) = 0$ then there exists a data length $N_0 < \infty$ such that:

$$||\hat{G}_N(e^{j\omega}) - G(e^{j\omega})||_\infty = 0 \text{ for all } N > N_0$$

\+ Time- and frequency-domain versions available. $N4SID$, etc

\+ Many variants which depend on weighting for noise.

\+ gives a state-space model directly.

\+ Can be effective in determining system order.

\+ Works equally well for MIMO systems.

\- Unusual noise weighting in frequency-domain case.

\- Truncated SVD reconstrutions are not Hankel.

\- $\hat{U}_1$ does not have the „shift“structure.

\- Least-squares noise assumptions are not correct.

\- Can give unstable models for stable systems.

10.2 Nonuniformly spaced frequencies

Time domain

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k) + Du(k)$$

Frequency domain

$$e^{j\omega}X(\omega) = AX(\omega) + BU(\omega)$$
$$Y(\omega) = CX(\omega) + BU(\omega)$$

For a specific frequency ω on each channel

$$U_i(\omega) = e_i \quad i = 1, \ldots n_u$$

Resulting system equations

$$e^{j\omega}X_i(\omega) = AX_i(\omega) + BU_i(\omega)$$
$$Y_i(\omega) = CX_i(\omega) + DU_i(\omega)$$

Defining

$$X_c(\omega) = \begin{bmatrix} X_1(\omega) & \cdots & X_{n_u}(\omega) \end{bmatrix}$$

and stacking the equations column-wise gives:

$$e^{j\omega} X_c(\omega) = AX_c(\omega) + B$$
$$G(\omega) = CX_c(\omega) + D$$

Multiplying by $e^{j\omega}$ and substituting, repeating and stacking row-wise:

$$\begin{bmatrix} G(e^{j\omega}) \\ e^{j\omega}G(e^{j\omega}) \\ \vdots \\ e^{j(q-1)\omega}G(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} X_c(\omega) + \Gamma \begin{bmatrix} I_{n_u} \\ e^{j\omega} I_{n_u} \\ \vdots \\ e^{j(q-1)\omega} I_{n_u} \end{bmatrix}$$

where

$$\Gamma = \begin{bmatrix} D & & & \\ CB & D & 0 & 0 \\ CAB & \ddots & \ddots & \\ \vdots & & \ddots & \\ CA^{q-2}B & \cdots & CB & D \end{bmatrix}$$

Repeat for all frequencies ω_i:

$$\mathcal{G} = \frac{1}{\sqrt{N}} \begin{bmatrix} G(e^{j\omega_1}) & \cdots & G(e^{j\omega_N}) \\ e^{j\omega_1}G(e^{j\omega_1}) & & e^{j\omega_N}G(e^{j\omega_N}) \\ \vdots & & \vdots \\ e^{j(q-1)\omega_1}G(e^{j\omega_1}) & \cdots & e^{j(q-1)\omega_N}G(e^{j\omega_N}) \end{bmatrix}$$

$$\mathcal{W} = \frac{1}{\sqrt{N}} \begin{bmatrix} I & \cdots & I \\ e^{j\omega_1} & & e^{j\omega_N} I \\ \vdots & & \vdots \\ e^{j(q-1)\omega_1} & \cdots & e^{j(q-1)\omega_N} I \end{bmatrix}$$

$$\mathcal{X}_c = \frac{1}{\sqrt{N}} \begin{bmatrix} X_c(\omega_1) \cdots & X_c(\omega_N) \end{bmatrix}$$

$$\boxed{g = \mathcal{O}\mathcal{X}_c + \Gamma\mathcal{W}}$$

Now as $\mathcal{O}$ and Γ are real valued

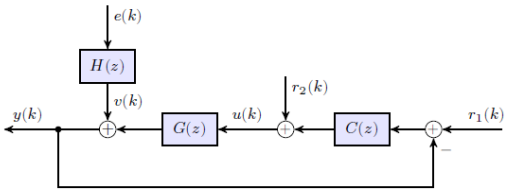
$$\underbrace{\begin{bmatrix} \text{real}(\mathcal{G}) & \text{imag}(\mathcal{G}) \end{bmatrix}}_{\mathcal{G}_r} = \mathcal{O} \underbrace{\begin{bmatrix} \text{real}(\mathcal{X}_c) & \text{imag}(\mathcal{X}_c) \end{bmatrix}}_{\mathcal{X}_{cr}} + \Gamma \underbrace{\begin{bmatrix} \text{real}(\mathcal{W}) & \text{imag}(\mathcal{W}) \end{bmatrix}}_{\mathcal{W}_r}$$

If $n_y q < n_u r$ then $\exists \mathcal{W}_r^\perp$ such that $\mathcal{W}_r \mathcal{W}_r^\perp$.

$$\mathcal{G}_r \mathcal{W}_r^\perp = (\mathcal{O}\mathcal{X}_{rc} + \Gamma\mathcal{W}_r)\mathcal{W}_r^\perp = \mathcal{O}\mathcal{X}_{cr}\mathcal{W}_r^\perp$$
$$\text{range}\left(\mathcal{G}_r \mathcal{W}_r^\perp\right) = \text{range}(\mathcal{O})$$

```
%To solve linear least squares problem: Ax = b
x = A\b;
```

11 Closed-Loop ID



$r_1(k), r_2(k)$ Excitation
$y(k), u(k), r_2(k), r_1(k)$ Measurements

$$r(k) = r_2(k) + C(z)r_1(k)$$

Motivation for closed-loop ID

- Unstable systems must be operated in closed-loop.
- Operational constraints may require closed-loop.
- Closed-loop controller maintains the system close to the operating point.
- Easier to focus identification on specific operation points.
- Will emphasize plant dynamics close to the cross-over frequency range.
- Closed-loop operation can remove a large-scale zero-frequency response.

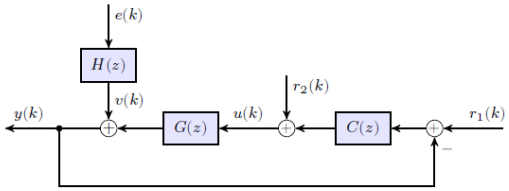
Methods overview

1. Direct (open-loop) methods:
$$\hat{G} = \hat{Y}_N/\hat{U}_N \text{ or } \hat{G} = \hat{\phi}_{yu}/\hat{\phi}_u$$
2. Indirect methods
$$y = \frac{G(z)}{(1+G(z)C(z))}r$$
3. Input-output methods:
$$y = \frac{G(z)}{(1+G(z)C(z))}r \text{ and } u = \frac{1}{(1+G(z)C(z))}r$$
4. Dual-Youla methods.

11.1 Direct Methods

$$\hat{\phi}_{yu}(e^{j\omega_n}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(\omega_n - \zeta)\frac{1}{N}Y_N(e^{j\zeta})\bar{U}_N(e^{j\zeta})d\zeta$$
$$\hat{\phi}_{u}(e^{j\omega_n}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_\gamma(\omega_n - \zeta)\frac{1}{N}|U_N(e^{j\zeta})|^2 d\zeta$$
$$\hat{G}(e^{j\omega_n}) = \frac{\hat{\phi}_{yu}(e^{j\omega_n})}{\hat{\phi}_{u}(e^{j\omega_n})}$$

Closed loop transfer functions:



$$y = v + G(r_2 + C(r_1 - y))$$
$$y = v + Gr_2 + GCr_1 - GCy$$
$$y(1+GC) = v + Gr_2 + GCr_1 \quad \text{where } S = \frac{1}{1+GC}$$
$$y = Sv + \underbrace{SGr_2 + SGCr_1}_{SGr} \quad \text{where } r = r_2 + Cr_1$$

$$u = r_2 + C(r_1 - (Gu + v))$$
$$u = r_2 + Cr_1 - CGu - Cv$$
$$u + CG = r_2 + Cr_1 - Cv$$
$$u = \underbrace{Sr_2 + SCr_1}_{Sr} - SCv$$

$$T_{yr} = SG$$
$$T_{ur} = S$$

```
Gdz = c2d(Gs,T_s,'zoh');
Cdz = ...;
Sdz = 1/(1+Gdz*Cdz);
Tdz_y_r1 = 1-Sdz;
Tdz_y_r = feedback(Gdz,Cdz);
Tdz_y_r = Sdz*Gdz;
T_u_r = Sdz;
%Use minreal() to reduce the TF
```

Further these identities can be shown if $r_2, v = 0$:

```
y = lsim(Tdz_y_r1,r_1,t);
y = lsim(Tdz_y_r,lsim(Cdz,r_1,t),t);
u = lsim(Sdz*Cdz,r_1,t);
u = lsim(Sdz,lsim(Cdz,r_1,t),t);
```

Assume that $\phi_{rv} = 0$

$$\hat{G} = \frac{\hat{\phi}_{yu}}{\hat{\phi}_u} \approx \frac{|S|^2 G\hat{\phi}_r - |S|^2\bar{C}\hat{\phi}_v}{|S|^2\hat{\phi}_r + |S|^2|C|^2\hat{\phi}_v} \approx \frac{G\hat{\phi}_r - \bar{C}\hat{\phi}_v}{\hat{\phi}_r + |C|^2\hat{\phi}_v}$$

- For system identification the used controller C should be not the optimal one for the use case, since a good controller suppresses information Find better explanation.
- Reducing the excitation to zero will only deliver information on C.

11.2 Input-Output Methods

Identification problems:
If $\phi_{vr} = 0$:

$$\hat{T}_{yr}(e^{j\omega_n}) = \frac{Y_N(e^{j\omega_n})}{R_N(e^{j\omega_n})} \text{ (asymptotically unbiased)}$$

$$\hat{T}_{ur}(e^{j\omega_n}) = \frac{U_n(e^{j\omega_n})}{R_N(e^{j\omega_n})} \text{ (asymptotically unbiased)}$$

Closed-loop identification approach:

$$\boxed{\frac{T_{yr}}{T_{ur}} = \frac{SG}{S} = G \longrightarrow \hat{G}(e^{j\omega_n}) = \frac{\hat{T}_{yr}(e^{j\omega_n})}{\hat{T}_{ur}(e^{j\omega_n})}}$$

- The estimates $\hat{T}_{yr}$ and $\hat{T}_{ur}$ may be unbiased. **Their ratio is not.**
- The estimated spectra are weighted by $S(e^{j\omega_n})$ or $S(e^{j\omega_n})C(e^{j\omega_n})$.
- The noise enters in a complicated manner.

11.2.1 Ratio distributions

Two normal substitutions $v \in \mathcal{N}(0,1)$ and $w \in \mathcal{N}(0,1)$.
The ratio $z = v/w$ is a stochastic variable with PDF:

$$\boxed{f_z(z) = \frac{1}{\pi}\frac{1}{1+z^2}} \text{ Cauchy Distribution}$$

- The variance of a Cauchy distribution is infinite.

11.2.2 Averaging closed-loop estimates

$$U_l(e^{j\omega_n}), Y_l(e^{j\omega_n})$$

$$\tilde{G}(e^{j\omega_n}) = \frac{1}{L}\sum_{l=1}^{L}\left(\frac{Y_l(e^{j\omega_n})}{U_l(e^{j\omega_n})}\right)$$

or

$$\tilde{G}(e^{j\omega_n}) = \frac{\tilde{Y}(e^{j\omega_n})}{\tilde{U}(e^{j\omega_n})}$$

- Either top or below is correct which is frequency dependent. Find better explanation.

11.3 Dual-Youla Methods

11.3.1 Youla Parametrizations

Coprime factorizations

$$G_0(s) = \frac{N_o(s)}{D_0(s)}$$

with $N_0(s), D_0(s)$ stable and coprime (no common zeros). **Coprime factorizations are not unique**.

Bezout identity
The transfer functions $N_0(s)$ and $D_0(s)$ are coprime iff there exists $U(s)$ and $V(s)$ such that

$$U(s)N_0(s) + V(s)D_0(s) = I$$

Normalized Coprime Factorizations
A coprime factorization is normalized if

$$D_0^*(s)D_0(s) + N_0^*(s)N_0(s) = I$$

```
sncfbal()
```

All stablizing controllers
If we have a controller C_0 which stablizes G_0 with

$$C_0 = \frac{X_0}{Y_0} \text{ } (X_0, Y_0 \text{ a coprime factorization})$$

then, all controllers, C, stablizing $G_0 = N_0/D_0$ have the form:

$$\boxed{C_Q = \frac{X_0 + QD_0}{Y_0 - QN_0}, \text{ with } Q \text{ stable}}$$

11.3.2 Dual-Youla methods

Control design:
Given $G(s)$ select $C(s)$ form the set of controllers stabilizing $G(s)$.
Closed-loop identification:
Given a particular controller $C(s)$ select $G(s)$ from the set of all plants stabilized by $C(s)$.
Both problems can be formulated as a search over stable $Q(s)$.

1. **Formulation**

   $$y = Gu + He \rightarrow Dy = Nu + Fe$$

   where $e \in \mathcal{N}(0,1)$ and D, N, F stable. Now find D, N and possibly F from the data, knowing $C_0 = X_0/Y_0$.

2. **Parametrization with R and F**

   $$G_R = \frac{N}{D} = \frac{N_0 + RY_0}{D_0 - RX_0} \text{ } R \text{ is stable}$$

   $$H_{R,F} = \frac{F}{D} = \frac{F}{D_0 - RX_0} \text{ } F \text{ is stable and stably invertible}$$

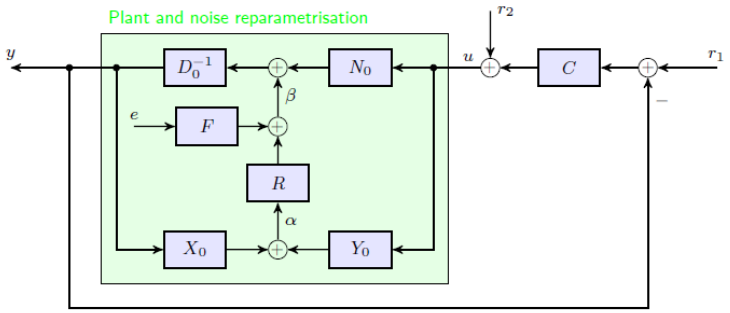
3. **Equivalent open-loop ID experiment:**

   $$(D_0 - RX_0)y = (N_0 + RY_0)u + Fe$$

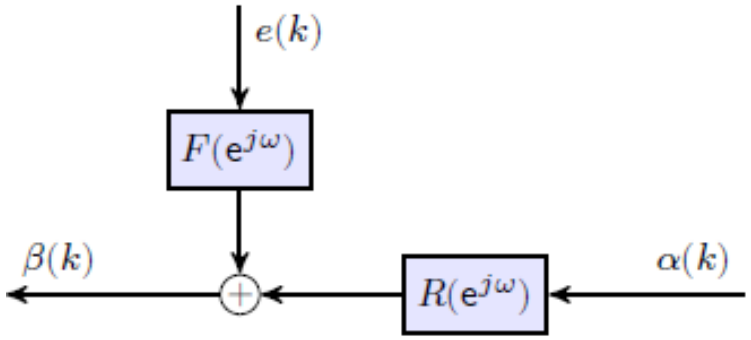
   and rearranging

   $$\underbrace{D_0 y - N_0 u}_{=:\beta} = R\underbrace{(X_0 y + Y_0 u)}_{=:\alpha} + Fe$$

   Open-loop system: $\beta = R\alpha + Fe$ with R and F stable.



- The transfer function from β to α is zero. Thus the system is actually open loop! No feedback!



$\beta = D_0 y - N_0 u$ (filtered input and output signals)

$$\begin{aligned}\alpha &= X_0 y + Y_0 u \\ &= X_0 y + Y_0 \left(r_2 + \frac{X_0}{Y_0}(r_1 - y) \right) = X_0 y + Y_0 \left(r - \frac{X_0}{Y_0} y \right) \\ &= Y_0 r \text{ (filtered excitation signal)}\end{aligned}$$

11.3.3 Summary

1. Factorise: $C_0 = X_0 / Y_0$.
2. Choose excitation: r (Note $\alpha = Y_0 r$ filtering).
3. Run closed-loop experiments with C_0, measuring y and u.
4. Choose an initial model, $P_0 = N_0 / D_0$ (must be stabilised by C_0).
5. Filter measurements, $\beta = D_0 y - N_0 u$ (time or frequency domain).
6. Filter excitation, $\alpha = Y_0 r$.
7. Esimate $\hat{R}$ (and $\hat{F}$) from $\beta = R\alpha + Fe$.
8. Calculate plan estimate, $\hat{G} = (N_0 + \hat{R} Y_0)/(D_0 - \hat{R} X_0)$.

12 Time-Domain Correlation Methods

12.1 Parametrised Model Sets

We are looking for the plant, G, in a parametrised set

$$\{G(\theta)\}$$

where $\theta \in \mathbb{R}^d$ is the parameter vector.

Model structure	Parameter vector $\theta \in \mathbb{R}^d$
Pulse response: $g(k)$	$\begin{bmatrix} g(0) & g(1) & \dots \end{bmatrix}$
Transfer function: $\frac{B(z)}{A(z)}$	$\begin{bmatrix} a_1 & \dots & b_1 & \dots \end{bmatrix}$
State-space: $\left[\begin{array}{c\|c} \text{A} & \text{B} \\ \hline \text{C} & \text{D} \end{array}\right]$	$\begin{bmatrix} A_{ij} & \dots & B_{ij} & \dots & C_{ij} \dots & D_{ij} & \dots \end{bmatrix}$

12.2 Identification Framework

$$Z_K = \{u(0), y(0), \dots, u(K-1), y(K-1)\} \quad \text{Measurement data}$$

$$J(\theta, Z_K) \quad \text{Objective}$$

$$\hat{\theta} = \arg\min J(\theta, Z_K) \quad \text{General optimization formulation}$$

12.2.1 Possible Objectives

Residual error objectives:

$$e(k, \theta) = y(k) - G(\theta) u(k) \quad \text{error}$$

$$J(\theta) = ||e(\theta)||_2^2 \text{ or } ||e(\theta)||_\infty \text{ or } ||e(\theta)||_1$$

Parametric error objective:

$$J(\theta) = ||\theta - \theta_0||_2 \text{ or } E\{\theta - \theta_0\}$$

Prediction error objective:

$$J(\theta) = E\{y(k+1) - \hat{y}(k+1, \theta | k)\}$$

12.3 Correlation-Based Methods

Input-Output relationship:

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

$$\boxed{R_{yu}(\tau) = g(k) * R_u(\tau)}$$

This can be written in matrix form (Toeplitz):

$$\begin{bmatrix} R_{yu}(0) \\ R_{yu}(1) \\ R_{yu}(2) \\ \vdots \end{bmatrix} = \begin{bmatrix} R_u(0) & R_u(-1) & R_u(-2) & \cdots \\ R_u(1) & R_u(0) & R_u(-1) & \\ R_u(2) & R_u(1) & \ddots & \ddots \\ \vdots & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \\ g(2) \\ \vdots \end{bmatrix}$$

Having only a finite data estimate:

$$\underbrace{\begin{bmatrix} \hat{R}_{yu}(0) \\ \vdots \\ \hat{R}_{yu}(N-1) \end{bmatrix}}_{\hat{R}_{yu}} = \underbrace{\begin{bmatrix} \hat{R}_u(0) & \cdots & \hat{R}_u(-(N-1)) \\ \vdots & & \vdots \\ \hat{R}_{yu}(N-1) & \cdots & \hat{R}_u(0) \end{bmatrix}}_{\bar{R}_N} \begin{bmatrix} \hat{g}(0) \\ \vdots \\ \hat{g}(N-1) \end{bmatrix}$$

- Note that $R_u(-\tau) = R_u(\tau)$.
- In the periodic (noise-free) and FIR case this is exact.
- $\hat{g}$ is uniquely determined, if $\bar{R}_N$ is invertible, which is given if $u(k)$ is persistently exciting.

12.4 Persistency of Excitation

A stationary input $u(k)$ is persistently exciting of order n if

$$\bar{R}_n = \begin{bmatrix} R_u(0) & \cdots & R_u(-(n-1)) \\ \vdots & & \vdots \\ R_u(n-1) & \cdots & R_u(0) \end{bmatrix}$$

is positive definite.

- This is sufficient to uniquely determine the first n coefficients of the pulse response, $\hat{g}k$) via the correlation approach.
- The definition also applies to deterministic signals.
- A signal is called **persistently exciting** if this holds for all n

Spectra of persistently exciting signals
$u(k)$ is persistently exciting of order n if $\phi_u(e^{j\omega}) \neq 0$ for at least n frequencies.
Moving average (MA) filtering

$$M_n(z) = m_1 z^{-1} + \cdots + m_n z^{-n} \quad n^{th} \text{ order MA filter}$$

If for all n^{th} order MA filters,

$$|M_n(e^{j\omega})|^2 \phi(u)(e^{j\omega}) = 0 \Rightarrow M(e^{j\omega}) = 0$$

then $u(k)$ is persistently exciting of order at least n.

- A step function is persistently exciting of order 1.
- A $PRBS$ signal is persistently exciting of order M.
- A sum of sinusoids $u(k) = \sum_{s=1}^{S} \alpha_s \cos(\omega_s k + \phi_s)$ is persistently exciting of order.

$$\begin{cases} 2S & \text{if } 0 < \omega_s < \pi,\ s = 1, \ldots, S \\ 2S-1 & \text{if } \omega = 0 \text{ or } \omega = \pi \in \{\omega_s, s = 1, \ldots, S\} \\ 2S-2 & \text{if } \omega = 0 \text{ and } \omega = \pi \in \{\omega_s, s = 1, \ldots, S\} \end{cases}$$

12.5 Autoregressive moving average models

Model form:

$$G(z) = \frac{b_1 z^{-1} + \ldots + b_m z^{-m}}{1 + a_1 z^{-1} + \ldots + a_n z^{-n}}$$

Input-Output relationship:

$$\begin{aligned} y(k) &= G(z)u(k) \\ &= -a_1 y(k-1) - \cdots - a_n y(k-n) + b_1 u(k-1) + \cdots + b_m u(k-m) \\ &= \phi^T(k)\theta \end{aligned}$$

where

$$\begin{aligned} \phi(k) &= \begin{bmatrix} -y(k-1) & \cdots & -y(k-n) & u(k-1) & \cdots & u(k-m) \end{bmatrix}^T \\ \theta &= \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_m \end{bmatrix}^T \end{aligned}$$

where $\phi(k)$ is called the **regressor vector** and θ is called the **parameter vector**.

12.5.1 ARX or ARMAX

Consider $y(k) = \phi^T(k)\theta$ for $k = 0, \ldots, N-1$

$$\underbrace{\begin{bmatrix} y(0) \\ \cdots \\ y(N-1) \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} \phi^T(0) \\ \cdots \\ \phi^T(N-1) \end{bmatrix}}_{\Phi} \theta$$

or in matrix form: $Y = \Phi\theta$
Least squares solution:

$$\hat{\theta} = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y$$

```
Theta = Phi\Y;
```

The least squares solution $\hat{\theta}$ solves the problem:

$$\begin{aligned} \underset{\theta}{minimise} &||\epsilon||_2 \\ \text{subject to} &Y = \Phi\theta + \epsilon \end{aligned}$$

Thus the cost function is defined as

$$J(\theta, Z_N) = ||Y - \Phi\theta||_2 = ||\epsilon||_2$$

12.5.2 Statistical Properties of the LS Estimate

$$Y = \Phi\theta + \epsilon, \quad \epsilon = \begin{bmatrix} \epsilon(0) \\ \vdots \\ \epsilon(N-1) \end{bmatrix} \quad \text{Model}$$

Error assumptions: $E\{e\} = 0$ and $E\{\epsilon\epsilon^T\} = \sigma^2 I$.

$$\boxed{E\{\hat{\theta}\} = \theta \text{ (unbiased estimator)}}$$

$$\boxed{\text{cov}\{\theta\} = E\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} = \sigma^2(\Phi^T\Phi)^{-1}}$$

For a model with correlated noise:

$$E\{\epsilon\epsilon^T\} = R$$

$$\boxed{E\{\hat{\theta}\} = \theta \text{ (unbiased estimator)}}$$

$$\boxed{\text{cov}\{\theta\} = E\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} = (\Phi^T\Phi)^{-1}\Phi^T R\Phi(\Phi^T\Phi)^{-1}}$$

Thus the variance of the error depends on the observations Φ!

12.6 Best linear unbiased estimator (BLUE or Markov estimator)

For a model with correlated noise:

$$E\{\epsilon\epsilon^T\} = R$$

Best linear estimator:

$$\hat{\theta} = Z^T Y \text{ where } Z = R^{-1}\Phi(\Phi^T R^{-1}\Phi)^{-1}$$

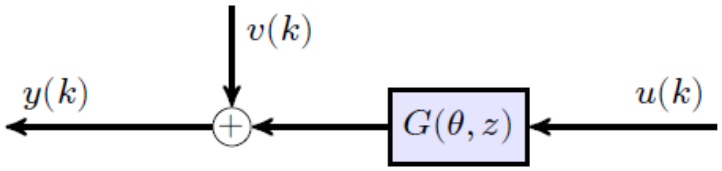
satisfies:

$$\boxed{E\{\hat{\theta}\} = \theta \text{ (unbiased estimator)}}$$

$$\boxed{\text{cov}\{\hat{\theta}_Z\} = (\Phi^T R{-}1\Phi)^{-1} \leq \text{cov}\{\hat{\theta}\} \text{ for any unbiased estimate}}$$

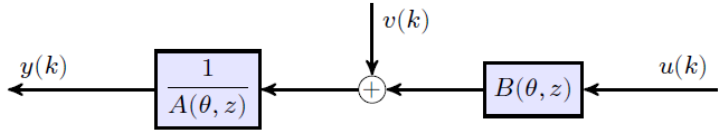
- BLUE requires knowledge of the error covariance, R, in order to reduce the variance of the error of the estimation.

12.7 Pitfall of Noisy Case



The setup above is only valid for the noise-free case. As soon as there is noise the equation derived from the setup above delivers biased results.

In actuality the problem can only be solved as a linear least squares problem if the setup can be described like this:



If this is not the case we get biased results. A possible remedy would be:

1. Assume the system to be of the form:

$$\begin{bmatrix} y(0) \\ \vdots \\ y(N-1) \end{bmatrix} = \begin{bmatrix} \phi^T(0) \\ \vdots \\ \phi^T(N-1) \end{bmatrix} \theta + \begin{bmatrix} v(0) \\ \vdots v(N-1) \end{bmatrix}$$

   where $\phi^T(k) = \begin{bmatrix} -y(k-1) & \cdots & -y(k-N) & u(k-1) & \cdots & u(k-m) \end{bmatrix} 1 \begin{bmatrix} -v(k-1) & \cdots \end{bmatrix}$

2. Assume zero noise influence in the regressor $\phi(k)$.
3. Solve with linear least squares.
4. Calculate the $\hat{V} = Y - \Phi\hat{\theta}$ as the residuals.
5. Recalculate taking into account the estimate of the noise $\hat{v}(k)$.
6. Repeat.

13 Prediction Error Methods

Given $Z_K = \{u(0), y(0), \ldots, u(K-1), y(K-1)\}$ what is the best estimate of $y(K)$?
Typical assumptions

- $G(z)$ and $H(z)$ stable.

- $H(z)$ is stably invertible (no zeros outside the unit disk).

  Given $v(k),\ k = 0, \ldots, K-1$ can we determine $e(k)$?

  $$H_{inv}(z): \quad e8k) = \sum_{i=0}^{\infty} h_{inv}(i)v(k-i) \quad \text{Inverse filter}$$

  Note that the inverse filter has to be causal: $h_{inv}(k) = 0, k < 0$ and stable: $\sum_{k=0}^{\infty} |h_{inv}(k)| < \infty$.

  If $H(z)$ has no zeros for $|z| \geq 1$ then $H_{inv}(z) = \frac{1}{H(z)}$.

- $e(k)$ has known statistics.

13.1 One step ahead prediction

Assume $H(z)$ is known, $H(z)$ monic ($h(0) = 1$).

$$\begin{aligned} v(k) &= \sum_{i=0}^{\infty} h(i)e(k-i) \\ &= \underbrace{e}_{h(0)v(k)}(k) + \underbrace{\sum_{i=1}^{\infty} h(i)e(k-i)}_{=m(k-1)\ \text{„observed“}} \end{aligned}$$

The above simplification of $h(0)v(k) = e(k)$ only applies if $H(z)$ is monic, i.e. $h(0) = 1$.

$$\hat{v}(k|k-1) \quad \text{Prediction based on measurement}$$

which leads to

$$\hat{v}(k|k-1) = m(k-1) = \sum_{i=1}^{\infty} h(i)e(k-i)$$

The error in this prediction is then $e(k)$ which cannot be reduced.

In general:

$$\text{Prob}\{x \leq e(k) \leq x + \delta x\} = \int_x^{x+\delta x} f_e(x)dx \approx f_e(x)\delta x$$

$$\text{Prob}\{x \leq v(k) \leq x + \delta x | v_{-\infty}^{k-1}\} \leq f_e(x - m(k-1))\delta x$$

which is equivalent to a shift of the distribution with $m(k-1)$. The expected value, the mean or the most likely value (which is not the same for an asymmetric distribution) of the shifted distribution would be a valid choice for $\hat{v}(k|k-1)$.

$$\boxed{\hat{v}(k|k-1) = m(k-1) = \sum_{i=1}^{\infty} h(i)e(k-i) = -\sum_{i=1}^{\infty} h_{inv}(i)v(k-i) \approx -\sum_{i=1}^{k} h_{inv}(i)v(k-i)}$$

13.1.1 Moving Average Model

$$v(k) = e(k) + ce(k-1) \Rightarrow H(z) = 1 + cz^{-1}$$

$$H_{inv}(z) = \frac{1}{1 + cz^{-1}} = \sum_{i=0}^{\infty} (-c)^i z^{-i}$$

where $|c| < 1$ for stable invertibility.

$$\boxed{\hat{v}(k|k-1) = cv(k-1) - c^2 v(k-2) + c^3 v(k-3) + \cdots + -(-c)^k v(0)}$$

Written recursively:

$$\begin{aligned} H(z)\hat{v}(k|k-1) &= (H(z) - 1)v(k) \\ \hat{v}(k|k-1) + c\hat{v}(k-1|k-2) &= cv(k-1) \\ \hat{v}(k|k-1) &= c\underbrace{(v(k-1) - \hat{v}(k-1|k-2))}_{\epsilon(k-1)} \\ &= c\epsilon(k-1) \end{aligned}$$

13.1.2 Autoregressive Noise Model

$$v(k) = \sum_{i=0}^{\infty} a^i e(k-i) \quad |a| < 1 \text{ for stablity}$$

$$H(z) = \sum_{i=0}^{\infty} a^i z^{-i} = \frac{1}{1 - az^{-1}}$$

$$H_{inv}(z) = 1 - az^{-1}$$

$$\boxed{\hat{v}(k|k-1) = (1 - H_{inv}(z))v(k) = av(k-1)}$$

13.2 Output Prediction

$$\begin{aligned} y(k) &= G(z)u(k) + v(k) \\ \hat{y}(k|k-1) = \text{E}\{y(k)|Z_K\} &= G(z)u(k) + \hat{v}(k|k-1) \\ &= G(z)u(k) + (1 - H_{inv}(z))v(k) \\ &= H_{inv}(z)G(z)u(k) + (1 - H_{inv}(z))y(k) \end{aligned}$$

Note that the last line is obtained by replacing $v(k) = y(k) - G(z)u(k)$.

The prediction error would then be:

$$\begin{aligned} y(k) - \hat{y}(k|k-1) &= -H_{inv}(z)G(z)u(k) + H_{inv}(z)y(k) \\ &= H_{inv}(z)(y(k) - G(z)u(k)) = H_{inv}(z)v(k) \\ &= e(k) \end{aligned}$$

The one-step ahead predictor and the prediction error are parametrised by θ. This leads to an optimisation problem:

$$J(\theta, Z_K) = \frac{1}{K}\sum_{k=0}^{K-1} l(\epsilon_F(k,\theta)) \text{ where typically } l(\epsilon_F(k,\theta)) = ||\epsilon_F(k,\theta)||_2$$
$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} J(\theta, Z_K)$$

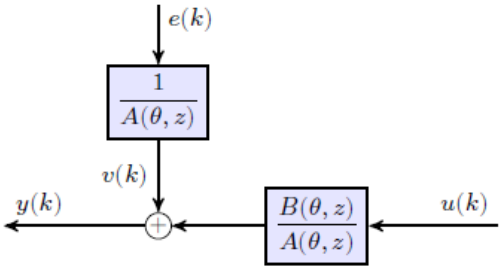
where ϵ_F is an optionally filtered error signal.

13.3 Different Models

Short name	Property	Example
AR	autoregressive	$A(z)y(t)$
X	extra	$B(z)u(t)$
MA	moving average	$C(z)e(t)$

Table 1: Model naming terminology

13.3.1 Equation Error Model Structure: ARX



$$\boxed{y(t) + a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a) = b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) + e(t)}$$

$$\theta = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n_a} & b_1 & \cdots & b_{n_b} \end{bmatrix}^T$$

\+ The one step ahead predictor directly defines a linear regression.

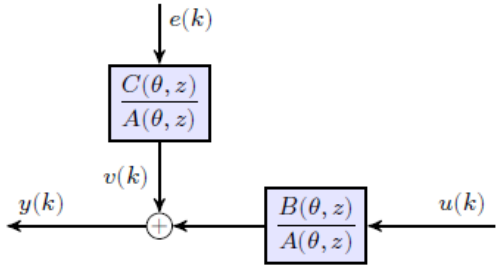
\- Lack of adequate freedom in describing the properties of the disturbance term.

$$\begin{aligned} \hat{y}(k|\theta) &= H_{inv}(\theta,z)G(\theta,z)u(k) + (1 - H_{inv}(\theta,z))y(k) \\ &= B(z)u(k) + (1 - A(z))y(k) \\ &= \theta^T\phi(k) = \phi^T(k)\theta \end{aligned}$$

$$\boxed{\text{vector of prediction errors}}\ Y - \phi\theta = \epsilon$$

- In this case the smallest error is also the smallest prediction error which is not generally the case.

13.3.2 Equation error structure: ARMAX



The ARMAX model structure expands the flexibility of the ARX by describing the equation error as a moving average of white noise.

$$\boxed{\begin{aligned} &y(t) + a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a) = b_1 u(t-1) + \\ &\cdots + b_{n_b} u(t-n_b) + e(t) + c_1 e(t-1) + \cdots + c_{n_c} e(t-n_c) \end{aligned}}$$

with $C(z) = 1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}$. The whole thing can be written as:

$$A(z)y(t) = B(z)u(t) + C(z)e(t)$$

$$\begin{aligned} \hat{y}(k|\theta) &= \frac{B(z)}{C(z)}u(k) + \left(1 - \frac{A(z)}{C(z)}\right)y(k) \\ C(z)\hat{y}(k|\theta) &= B(z)u(k) + (C(z) - A(z))y(k) \\ \hat{y}(k|\theta) &= B(z)u(k) + (1 - A(z))y(k) + (C(z) - 1)\underbrace{(y(k) - \hat{y}(k|\theta))}_{\epsilon(k)} \\ &= \begin{bmatrix} b_1 & \cdots & a_1 & \cdots & c_1 & \cdots \end{bmatrix} \\ &\quad \begin{bmatrix} u(k-1) & \cdots & -y(k-1) & \cdots & \epsilon(k-1) & \cdots \end{bmatrix} \\ &= \phi^T(\theta,k)\theta \end{aligned}$$

This is not linear in θ, since ϵ depends on the chosen parametrisation, but is called pseudo-linear.

Optimisation:

$$\underset{\theta,\epsilon}{\text{minimise}} ||\epsilon||_2 \quad \text{(or more generally, } l(\epsilon)\text{)}$$
$$\text{subject to } Y = \Phi(\epsilon)^T\theta + \epsilon \quad \text{(nonlinear equality constraint)}$$

Constrained minimisation code for ARMAX

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2}$$
$$B(z) = b_1 z^{-1} + b_2 z^{-2}$$
$$C(z) = 1 + c_1 z^{-1} + c_2 z^{-2}$$
$$\theta = \begin{bmatrix} b_1 & b_2 & a_1 & a_2 & c_1 & c_2 \end{bmatrix}^T$$

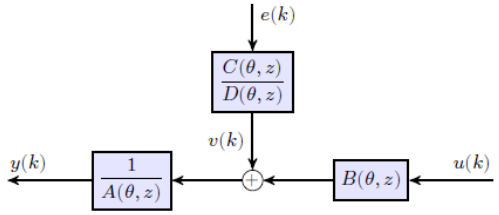
Fix code, probably not running!

```
%Create data part of regressor. Assume plant at rest
PhiTyu(1,:) = [0,0,0,0];
PhiTyu(2,:) = [u(1),0,-y(1),0];
for i = 3:K,
  PhiTyu(i,:) = [u(i-1),u(i-2),-y(i-1),-y(i-2)];
end

[x,fval] = fmincon(@(x)ARMAXobjective(x),x0,...
[],[],[],[],[],[],@(x)ARMAXconstraint(x,y,PhiTyu));
function [f] = ARMAXobjective(x) % x = [theta; e]
f = sqrt(x(7:end)'*x(7:end));
function [c,ceq] = ARMAXconstraint(x,y,PhiTyu)
e = x(7:end);
PhiTe = zeros(K,2);
PhiTe(2,1) = e(1);
for j = 3:K,
PhiTe(j,:) = [e(j-1), e(j-2)];
end
ceq = y - [PhiTyu, PhiTe] * theta - e; c = [];
```

13.3.3 Equation error structure: ARARMAX

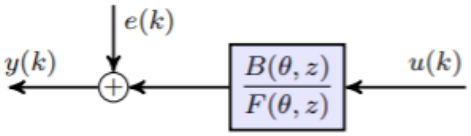
Instead of modelling the equation error as a moving average in can also be described as an autoregression, leading to a ARARX structure. A further generalization using an ARMA description of the equation error leads to the ARARMAX structure below:



$$G(\theta, z) = \frac{B(\theta, z)}{A(\theta, z)}$$
$$H(\theta, z) = \frac{C(\theta, z)}{A(\theta, z)D(\theta, z)}$$

13.3.4 Output Error Model Structure

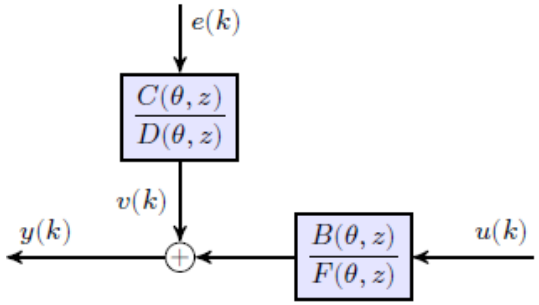
Contrary to the equation error model structure, this approach parametrizes the transfer functions independently.



$$\boxed{\begin{aligned} &w(t,\theta) + f_1 w(t-1,\theta) + \cdots + f_{n_f} w(t-n_f,\theta) \\ &= b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) \end{aligned}}$$

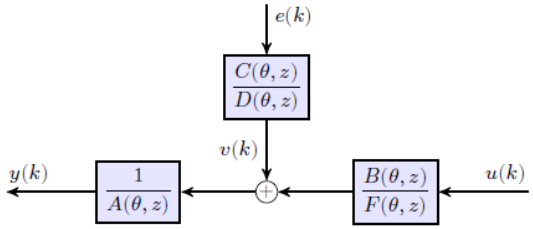
$$G(\theta, z) = \frac{B(\theta, z)}{F(\theta, z)}$$
$$H(\theta, z) = 1$$
$$\hat{y}(k|\theta) = \frac{B(\theta, z)}{F(\theta, z)} u(k) = \phi(k, \theta)^T \theta$$
$$\theta = \begin{bmatrix} b_1 & b_2 & \cdots & b_{n_b} & f_1 & f_2 & \cdots & f_{n_f} \end{bmatrix}$$

13.3.5 Box-Jenkins



$$G(\theta,z) = \frac{B(\theta,z)}{F(\theta,z)}$$
$$H(\theta,z) = \frac{C(\theta,z)}{D(\theta,z)}$$
$$\hat{y}(k|\theta) = \frac{D(z)}{C(z)}\frac{B(z)}{F(z)}u(k) + (1 - \frac{D(z)}{C(z)}y(k)$$

13.3.6 General model



$$G(\theta,z) = \frac{B(\theta,z)}{A(\theta,z)F(\theta,z)}$$
$$H(\theta,z) = \frac{C(\theta,z)}{A(\theta,z)D(\theta,z)}$$
$$\hat{y}(k|\theta) = \frac{D(z)}{C(z)}\frac{B(z)}{F(z)}u(k) + (1 - \frac{D(z)A(z)}{C(z)})y(k)$$

13.4 Known noise model (with ARMAX dynamics)

Assume known noise: $v(k) = L(z)e(k)$
Thus

$$A(z)y(k) = B(z)u(k) + L(z)e(k)$$

$$y_L(k) = L^{-1}(z)y(k)$$
$$u_L(k) = L^{-1}(z)u(k)$$

$$A(z)y_L(k) = B(z)u_L(k) + e(k)$$

for which least squares gives consistent estimates.

13.5 High-order model fitting

Assume an ARARX structure:

$$A(z)y(k) = B(z)u(k) + \tfrac{1}{D(z)}e(k) \quad e(k) \sim \mathcal{N}(0,\lambda)$$

Fitting a high order model (order of $D(z)$ is n_d)

$$A(z)D(z)y(k) = B(z)D(z)u(k) + e(k)$$

Least squares estimate with orders $n + n_d$ and $m + n_d$. This gives a consistent estimate of

$$\tfrac{B(z)D(z)}{A(z)D(z)} = \tfrac{B(z)}{A(z)}$$

This amounts to making the noise model sufficiently rich to capture additional autoregressive features in the noise. In practice though, the cancellation will not be exact. $\hat{A}(z)$ and $\hat{B}(z)$ will be high order.

14 Parameter estimation statistics

14.1 Model description

As found in chapter 4, Ljung:

$$\boxed{y(t) = G(z,\theta)u(t) + H(z,\theta)e(t), \quad f_e(x,\theta)}$$

where $f_e(x,\theta)$ is the PDF of $e(t)$ and $e(t)$ is assumed to be white noise.

Definition 7. *The **moments** of a PDF describe generalized concept of expected values:*

$$Ee^m = \in x^m f_e(x)dx = 0$$

where m is a positive integer. For $m = 1$, the first moment, is the expected value of a PDF, for $m = 2$, the second moment, is the variance of a PDF.

14.2 Basic setup

$$G = G(\theta, z), \quad H = H(\theta, z) \quad \text{Parametrised model}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} J(\theta, Z_K) \quad \text{Estimation}$$

Consider K observations. Each is a realisation of a random variable, with joint probability distribution

$$f(x_1, \ldots, x_K|\theta)$$

For independent variables

$$f(x_1, \ldots, x_K; \theta) = f_1(x_1; \theta) f_2(x_2; \theta) \cdots f_K(x_K; \theta) = \prod_{i=1}^{n} f_i(x_i; \theta)$$

Substituting the observation Z_K gives a function of θ

$$\mathcal{L}(\theta) = f(x_1, \ldots, x_k; \theta)|_{x_i = z_i, i=1,\ldots,K} \quad \text{Likelihood function}$$

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta) \quad \text{Maximum likelihood estimator}$$

The log-likelihood is often mathematically easier to consider

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \ln \mathcal{L}(\theta)$$

It will give the same $\hat{\theta}$ since ln is monotonic.

14.3 Bayesian approach

Consider θ to be a random variable with pdf: $f_\theta(x)$, where $x_i = z_i$. This is an **a priori** distribution assumed before the experiment. Thus we have some knowledge about the variable that is to be estimated.

$$\mathbb{P}(\theta|z_1, \ldots, z_K) = \frac{\mathbb{P}(Z_K \cap \theta)}{\mathbb{P}(Z_K)} \quad \text{Conditional Probability}$$

$$\mathbb{P}(Z_K \cap \theta) = \mathbb{P}(Z_K|\theta)\,\mathbb{P}(\theta) \quad \text{Since } \theta \text{ is not independent from } Z_K$$

which combined results in

$$\boxed{\mathbb{P}(\theta|z_1, \ldots, z_K) = \frac{\mathbb{P}(Z_K|\theta)\mathbb{P}(\theta)}{\mathbb{P}(Z_K)}} \text{ Bayes' Theorem}$$

where $\mathbb{P}(\theta)$ contains the prior knowledge.
So

$$\boxed{\underset{\theta}{\operatorname{argmax}}\ f(\theta|z_1, \ldots, z_K) = \underset{\theta}{\operatorname{argmax}}\ f(Z_K|\theta) f_\theta(\theta)}$$

where $f_\theta(\theta)$ weights the maximum likelihood estimator towards the expected/most realistic value.

14.4 Maximum a posteriori (MAP) estimation

The MAP is closely related to ML estimation but in addition employs an augmented optimization objective which incorporates a prior distribution that quantifies the additionally available information on the estimated quantity.
Given the data Z_K

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ f(Z_K|\theta) f_\theta(\theta)$$

The maximum likelihood estimator can be interpreted as:

$$\begin{aligned}\theta_{ML} &= \underset{\theta}{\operatorname{argmax}} f(x_1, \ldots, x_K|\theta)\Big|_{x_i = z_i, i=1,\ldots,K} \\ &= \underset{\theta}{\operatorname{argmax}} f(Z_K|\theta)\end{aligned}$$

This estimate coincides with the simple maximum likelihood estimate if θ is assumed to be uniformly distributed.

14.5 Cramér-Rao bound

The Cramér-Rao bound expresses a lower bound on the variance of unbiased estimators of a deterministic (fixed, though unknown) parameter.

$$P = \mathrm{E}\left\{(\hat{\theta}(Z_K) - \theta_0)(\hat{\theta}(Z_K) - \theta_0)^T\right\} \quad \text{Mean-square error matrix}$$

Assume $\mathrm{E}\left\{\hat{\theta}(Z_K)\right\} = \theta_0$ and $Z_K \subset \mathcal{R}^K$. Then $P \geq M^{-1}$ where M is the **Fischer information Matrix**

$$\begin{aligned}M &= \mathrm{E}\left\{(\frac{d}{d\theta} \ln f(Z_K|\theta))(\frac{d}{d\theta} \ln f(Z_K|\theta)^T\right\}\Big|_{\theta=\theta_0} \\ &= -\mathrm{E}\left\{\frac{d^2}{d\theta^2} \ln f(Z_K|\theta)\right\}\Big|_{\theta=\theta_0}\end{aligned}$$

Consider a parametrised family of pdfs $f(x_1, \ldots, x_K|\theta) = \prod_{i=1}^{K} f_i(x_i|\theta)$
Then

$$\lim_{K\to\infty} \hat{\theta}_{ML} \overset{w.p.\ 1}{\longrightarrow} \theta_0$$

and

$$\lim_{K\to\infty} \sqrt{K}(\hat{\theta}_{ML}(Z_K) - \theta_0) \sim \mathcal{N}(0, M^{-1})$$

14.6 Prediction error statistics

The prediction error framework is set up as follows:

$$\epsilon(k,\theta) = y(k) - \hat{y}(k,\theta)$$

where the error $\epsilon(k,\theta)$ is assumed to be independent and identically distributed variables with pfd $f_e(x;\theta)$.
The joint pdf writes as

$$f(X_K;\theta) = \prod_{k=1}^{K} f_e(\epsilon(k,\theta);\theta)$$

$$\begin{aligned}\hat{\theta}_{ML} &= \underset{\theta}{\text{argmax}}\ f(X_K;\theta)|_{X_K=Z_K} \\ &= \underset{\theta}{\text{argmax}}\ \mathcal{L}(\theta) \\ &= \underset{\theta}{\text{argmax}}\ \ln f(Z_K|\theta) \\ &= \underset{\theta}{\text{argmax}}\ \frac{1}{K}\sum_{k=1}^{K} \ln f_e(\epsilon(k,\theta);\theta)\end{aligned}$$

The prediction error methods give the same estimate if the prediction error cost function is chosen as

$$l(\epsilon,\theta) = -\ln f_\epsilon(\epsilon|\theta)$$

Thus

$$\boxed{\hat{\theta}_{PE} = \underset{\theta}{\text{argmin}}\ \tfrac{1}{K}\sum_{k=1}^{K} l(\epsilon(k,\theta);\theta) = \hat{\theta}_{ML}}$$

14.7 Linear regression statistics

One-step ahead predictor

$$\hat{y}(k|\theta) = \phi^T(k)\theta + \mu(k)$$

In the ARX case $\mu(k) = e(k)$, in other cases $\mu(k)$ can depend on Z_K.

$$\epsilon(k) = y(k) - \phi^t(k)\theta$$

A typical cost function is:

$$J(\theta, Z_K) = \tfrac{1}{K}\sum_{k=0}^{K-1} \tfrac{\epsilon(k)^2}{2}$$

Least-squares criterion:

$$\hat{\theta}_{LS} = \underbrace{\left(\frac{1}{K}\sum_{k=0}^{K-1}\phi(k)\phi^T(k)\right)^{-1}}_{R_K^{-1}\in\mathbb{R}^{d\times d}} \underbrace{\frac{1}{K}\sum_{k=0}^{K-1}\phi(k)y(k)}_{f_K\in\mathcal{R}^d}$$

Asymptotic bias:

$$\lim_{K\to\infty}\hat{\theta}_{LS} - \theta_0 = \lim_{K\to\infty} R_K^{-1}\tfrac{1}{K}\sum_{k=0}^{K-1}\phi^T(k)v(k) = (R^*)^{-1}f^*$$

$$R^* = \mathrm{E}\left\{\phi(k)\phi^T(k)\right\}, \quad f^* = \mathrm{E}\left\{\phi(k)v(k)\right\}$$

Under the assumption that $\lim_{K\to\infty}\hat{\theta}_{LS} = \theta_0$ we require $(R^*)^{-1}f^* = 0$
So

1. R^* must be non-singular. Persistency of excitation requirement.
2. $f^* = \mathrm{E}\left\{\phi(k)v(k)\right\} = 0$ This happens if either
   a) $v(k)$ is zero mean and independent of $\phi(k)$
   b) $u(k)$ is independent of $v(k)$ and G is FIR ($n = 0$)

This gives

$$\lim_{K\to\infty}\sqrt{K}(\hat{\theta}_{LS} - \theta_0) \sim \mathcal{N}(0, \sigma_0^2(R^*)^{-1})$$

14.8 Correlation methods

Ideally the sequence of prediction errors is white. Thus the prediction errors are uncorrelated with the data.
Approach:

1. Select a sequence $\zeta(k)$, derived from the past data Z_K
2. Require that the error $\epsilon(k,\theta)$ is uncorrelated with $\zeta(k)$
   $\frac{1}{K}\sum_{k=0}^{K-1}\zeta(k)\epsilon(k,\theta) = 0$ could also use $\alpha(\epsilon)$
3. The ID problem can be then viewed as finding θ such that this relationship is satisfied
   The values $\zeta(k)$ are known as **instruments**.
   Typically $\zeta(k) \in \mathbb{R}^{d\times n_y}$ where $\theta \in \mathbb{R}^d$ and $y(k) \in \mathcal{R}^{n_y}$.

Procedure:

1. Choose a linear fitter for the prediction errors
   $$\epsilon_F(k,\theta) = F(z)\epsilon(k,\theta)$$

2. Choose a sequence of correlation vectors $\zeta(k, Z_K, \theta)$ constructed from the data
3. Choose a function $\alpha(\epsilon)$ (default is $\alpha(\epsilon) = \epsilon$) Then

$$\hat{\theta} = \theta \text{ solving } f_K(\theta, Z_K) = \frac{1}{K} \sum_{k=0}^{K-1} \zeta(k, \theta)\alpha(\epsilon(k, \theta)) = 0$$

14.8.1 Pseudo-Linear Regressions

For ARX, ARMAX, etc., model structures we can write the predictor

$$\hat{y}(k|\theta) = \phi^T(k, \theta)\theta$$

This can be solved via LS methods, but alternatively correlation based solutions can be found:

$$\boxed{\hat{\theta}_{PLR} = \theta \text{ solving } \frac{1}{K} \sum_{k=0}^{K-1} \phi(k, \theta) \underbrace{(y(k) - \phi^T(k, \theta)\theta)}_{\text{prediction error}} = 0}$$

The prediction errors are orthogonal to the regressor, $\phi(k, \theta)$.

14.9 Instrumental Variable Methods

$$\hat{\theta}_{IV} = \theta \text{ solving } \frac{1}{K} \sum_{k=0}^{K-1} \zeta(k, \theta)(y(k) - \phi^T(k, \theta)\theta) = 0$$

This is solved by

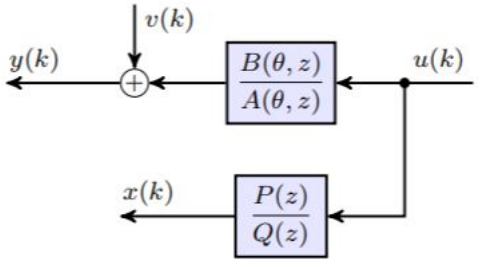
$$\hat{\theta}_{IV} = \left(\frac{1}{K} \sum_{k=0}^{K-1} \zeta(k)\phi^T(k) \right)^{-1} \frac{1}{K} \sum_{k=0}^{K-1} \zeta(k)y(k)$$

For consistency we require

$$\mathrm{E}\left\{\zeta(k)\phi^T(k)\right\} \quad \text{to be nonsingular}$$

and

$$\mathrm{E}\left\{\zeta(k)v(k)\right\} = 0 \quad \text{(uncorrelated w.r.t. prediction error)}$$



1. Estimate $\hat{\theta}_{LS}$ via linear regression.
2. Select $Q(z) = \hat{A}_{LS}$ and $P(z) = \hat{B}_{LS}$, since the procedure works well when P and Q are close to B and A.
3. Calculate $\hat{\theta}_{IV}$

- Variance and MSE depend on the choice of instruments.
- Consistency (asymptotically unbiased) is lost if:
  - Noise and instruments are correlated (for example, in closed-loop generating instruments from u)
  - Model order selection is incorrect.
  - Filter dynamics cancel plant dynamics.
  - True system is not in the model set.
- Closed-loop approaches: generate instruments from the excitation r.

15 Nomenclature

$y(k) = Gu(k)$	output signal	[ ]
$u(k)$	input signal	[ ]
G	plant	[ ]
$\hat{G} = \frac{y}{u}$	estimated plant	[ ]
$Y(e^{j\omega})$	output spectrum	[ ]
$U(e^{j\omega})$	input spectrum	[ ]
ZOH	zero order hold	
DAC	digital analog converter	
ADC	analog digital converter	

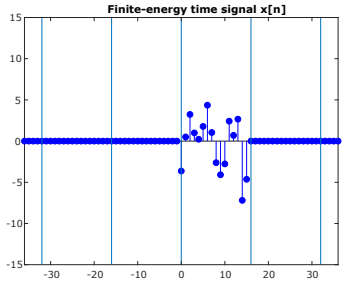
16 Overviews

16.1 Transfer Functions

Nonperiodic Time Domain Signal

$$x[n] = \int_0^1 \hat{x}(\theta) e^{2\pi i n \theta} d\theta$$

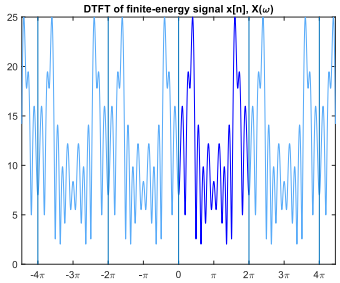
$$x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega k} d\omega$$



Discrete Time Fourier Transform (DTFT)

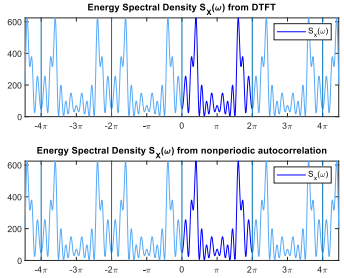
$$\hat{x}(\theta) = \sum_{n=-\infty}^{\infty} x[n] e^{-2\pi i n \theta}$$

$$X(\omega) = \sum_{k=-\infty}^{\infty} x(k) e^{-j\omega k}$$



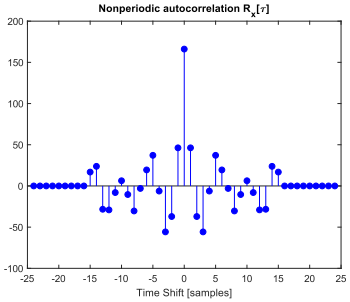
Energy Spectral Density

$$S_x(\omega) = |X(\omega)|^2 = \sum_{\tau=-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau}$$



Nonperiodic Autocorrelation

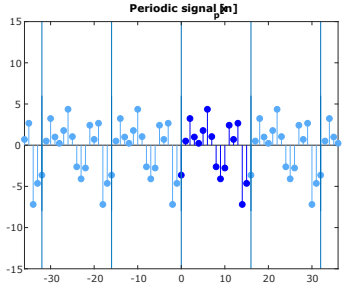
$$R_x(\tau) = \sum_{k=-\infty}^{\infty} x(k) x(k-\tau)$$



Periodic Time Domain Signal

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{x}[k] e^{2\pi i n \frac{k}{N}}$$

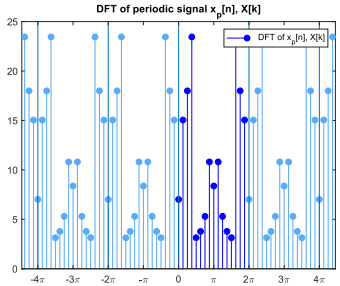
$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X(\omega_n) e^{j\omega_n k}, \ \omega_n = 2\pi \frac{n}{N}$$



Discrete Fourier Transform

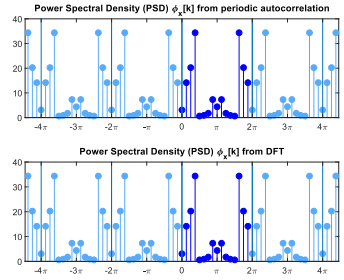
$$\hat{x}[k] = \sum_{n=0}^{N-1} x[n] e^{-2\pi i k \frac{n}{N}}$$

$$X(\omega_n) = \sum_{k=0}^{N-1} x(k) e^{-j\omega_n k}, \ \omega_n = 2\pi \frac{n}{N}$$



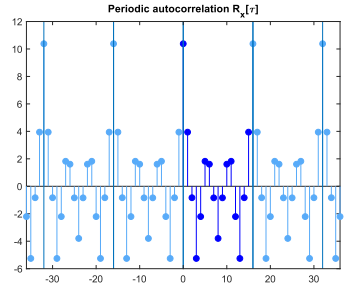
Power Spectral Density (PSD)

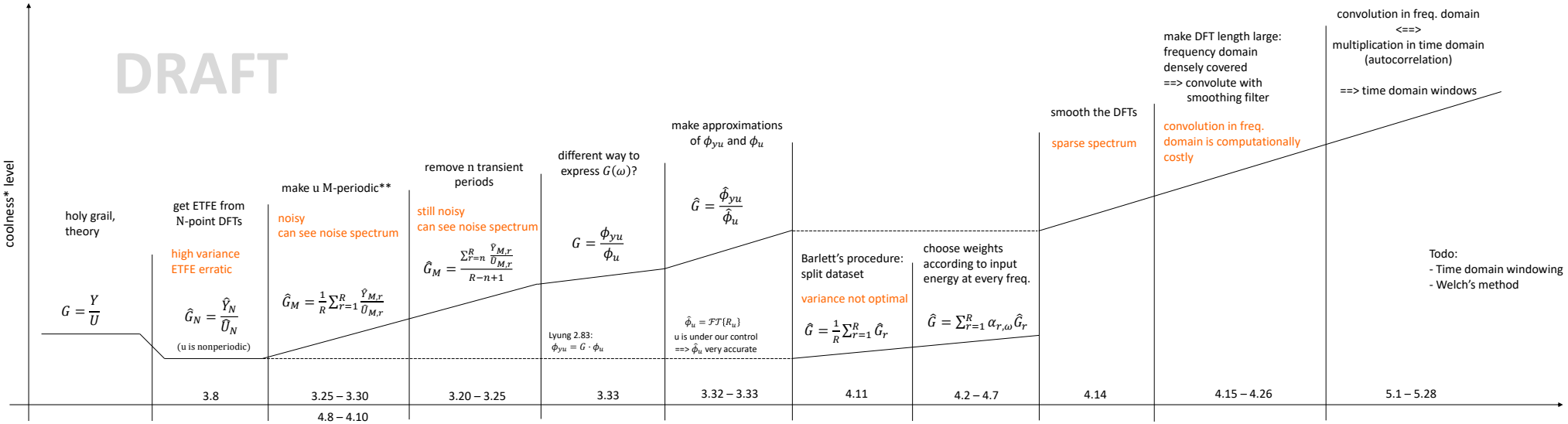
$$\phi(\omega_n) = \frac{1}{N} |X(\omega_n)|^2 = \sum_{\tau=0}^{N-1} R_x(\tau) e^{-j\omega_n \tau}$$



Periodic Autocorrelation

$$R_x(\tau) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) x(k-\tau)$$





*) ok, maybe better "usefulness"

**) Ch.3 uses DFTs of length $N \cdot M$, whereas Ch. 4 uses DFTs of length M, which is usually equivalent when averaged.