Lectures notes "Signal theory: Part 1"

Ivan Markovsky
Department ELEC, VUB

1 Introduction

General information about the course

- **Lecturers:**
  - Leo Van Biesen (part 2, weeks 5, 6, 7)
  - Ivan Markovsky (part 1, weeks 2, 3, 4)
- **Homework for part 1:**
  - on Tuesday due Friday, the same week
  - on Friday due Tuesday, the week after
- **MATLAB:**
  - homework involve reading assignment, exercise problems, and numerical experiments with MATLAB
  - if you are not familiar with MATLAB, follow the optional course given by Peter Zoltan
- **Evaluation:**
  - 25% homework
  - 75% open-book exam on problems similar to the exercises

Part 1 topics

1. Signals and systems (mostly review from a new perspective)
   - expansion of a signal in a basis; orthonormal basis
   - linear time-invariant systems; behavioral approach
   - representations of LTI systems
     - convolution
     - differential/difference equation
     - transfer function
     - state-space representation
   - the realization problem
2. Random signals
   - covariance function and spectrum
   - Wiener-Khintchine theorem
   - stochastic systems
   - stochastic realization

3. Least-squares estimation
   - underdetermined and overdetermined systems of linear equations
   - least-norm solution and least-squares approximation
   - recursive least-squares approximation
   - Kalman filtering
4. Subspace methods
   - deterministic subspace algorithms
   - stochastic subspace algorithms
5. Compressive sensing
   - ℓ_0 vs ℓ_1-norm minimization
   - LASSO

Teaching materials for Part 1

- in pointcarre: notes, homework, lecture slides
- library references

Demo noise filtering

Demo singular spectrum analysis

2 Signals and systems

Classification of signals

- examples of "physical" signals
  - sound (speech, music, noise, ...)
  - image (black/white, gray scale — integer 0–255, color)
  - video (sequence of images)
  - daily exchange rates of one currency into another
- abstract representation as a function
  - notation $f : \mathscr{X} \to \mathscr{Y}$ (function from $\mathscr{X}$ to $\mathscr{Y}$)
    * $\mathscr{X}$ — domain (where the function argument, say x, takes its values)
    * $\mathscr{Y}$ — image (where the function values belong)
  - $y = f(x)$ — value of the function at the point $x \in \mathscr{X}$
  - note that the function f is not defined for values of x outside the domain, *i.e.*, for $x \notin \mathscr{X}$
- scalar (single-channel) vs vector (multi-channel)
- real-valued vs complex-valued
- continuous/analog (the domain is $\mathbb{R}$) vs discrete/sampled (the domain is $\mathbb{Z}$)
- periodic vs non-periodic
- one-dimensional (*e.g.*, sound) vs multi-dimensional (*e.g.*, image and video)

Basic signals and operations with them

- Basic signals
  - Dirak (continuous-time) and Kroneker (discrete-time) delta

$$\delta(t) = \begin{cases} 1, & t = 0 \\ 0, & \text{otherwise} \end{cases}$$

  - unit step

$$s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

  - constant (or DC) signal: a, where $a \in \mathbb{R}$
  - rectangular pulse:

$$p_{t_1,t_2}(t) = \begin{cases} 1, & t_1 \leq t \leq t_2 \\ 0, & \text{otherwise} \end{cases}$$

    (Represent the pulse p_{t_1,t_2} by a linear combination of shifted steps s.)
  - sinusoidal signal: $a\cos(\omega t + \phi)$
  - complex exponentials

$$e^{(\alpha+\beta t)+\mathbf{i}(\omega t+\phi)} = e^{\alpha} e^{\beta t} \big(\cos(\omega t + \phi) + \mathbf{i}\sin(\omega t + \phi)\big)$$

$$\begin{array}{lcl} A := e^{\alpha} & \text{—} & \text{amplitude} \\ \beta & \text{—} & \text{damping coefficient} \\ \omega & \text{—} & \text{frequency} \\ \phi & \text{—} & \text{initial phase} \\ \omega t + \phi & \text{—} & \text{instantaneous phase} \end{array}$$

    (link to a response of a second order system)

- Signal transformations
  - addition $x + y$
  - multiplication (modulation) xy
  - amplitude scaling ax, $a \in \mathbb{R}$
    * $|a| > 1$ — gain factor (ideal linear amplifier)
    * $|a| < 1$ — attenuation factor
  - time scaling $x(at)$, $a \in \mathbb{R}$
    (Exercise: sketch by hand and plot with MATLAB)
  - shift in time

$$x(t + \tau) =: (\sigma^{\tau} x)(t), \quad \tau \in \mathbb{R}$$

  - sampling (discretization in time, analog to digital conversion)
    * uniform sampling

$$\text{sample}_{t_s} : x_c \mapsto x_d, \quad x_d(t) = x_c(t_s t), \quad t \in \mathbb{Z}$$

      · t_s — sampling time
    * loss of information (aliasing)
    * Nyquist–Shannon sampling theorem (for band limited signals, critical frequency)

        * reconstruction of the continuous-time signal from the discrete-time samples via sinc interpolation is a convolution of of x_d with sinc function

$$f_\mathrm{c}(t) = \left(\sum_{\tau=-\infty}^{\infty} f_\mathrm{d}(\tau)\delta(t-\tau t_\mathrm{s})\right) \star \mathrm{sinc}(t)$$

        * non-uniform sampling (sparsity, compressive sampling)
    - interpolation $x_\mathrm{d} \mapsto x_\mathrm{c}$
        * digital to analog conversion
        * zero order hold
        * linear interpolation
        * polynomial interpolation
        * sinc interpolation (for band limited signals)
    - extrapolation: predict the signal in the future
    - quantization (discretization in value) $\mathrm{quant}_\Delta : x_\mathrm{c} \mapsto x_\mathrm{q},\ x_\mathrm{q}(t) = \left\lceil \frac{x_\mathrm{c}(t)}{\Delta} \right\rceil \Delta$
        * Δ — quantization level
        * loss of information
        * "quantization noise"

- Signal expansion in a basis
    - spaces and subspaces
    - basis of a subspace

- Transform techniques
    - integral/sum of Dirak/Kroneker deltas

$$x(t) = \cdots + x(-1)\delta(t+1) + x(0)\delta(t) + x(1)\delta(t-1) + \cdots = \sum_{\tau=-\infty}^{\infty} x(\tau)\delta(t-\tau) \qquad (\delta\text{-train})$$

      No computation required!
    - (discrete) Fourier transform

$$x(t) = \frac{1}{T}\sum_{k=0}^{T-1} X(k)e^{\mathbf{i}\frac{2\pi k}{T}t} \qquad \text{(DFT)}$$

      Computation is required.

- Measuring the size of a signal (x with domain $[0,\infty)$)
    - total energy: $\int_0^\infty x^2(\tau)\,d\tau$
    - peak value: $\max_{t\geq 0}|x|(t)$
    - root-mean-square (RMS) value

$$\sqrt{\lim_{t\to\infty}\frac{1}{t}\int_0^t x^2(\tau)}\,d\tau$$

Classification of systems

- examples of "physical" systems
  - electrical circuits
  - chemical processes
  - stock markets
  - the solar system
  - behavioral approach
    * manifest variables w
    * the universal set $\mathscr{U}$
    * behavior: the set of (allowed) signals $\mathscr{B}$; the system is a prohibition rule
    * example: the solar system, the prohibition rule are Kepler's laws
      (gravitational law + Newton's laws $\quad\leftrightarrow\quad$ Kepler's laws $\quad\leftarrow\quad$ observations)
  - a "signal processor" view of a system: map from an input signal to an output signal (an operator)
    * notation: $y = S(u)$

$$\begin{array}{ccl} u & \text{—} & \text{input} \\ y & \text{—} & \text{output} \\ S & \text{—} & \text{system} \end{array}$$

    * block diagram
      · building complicated systems with interconnections (summation and multiplication)
    * examples
      · scaling system: $y(t) = au(t)$, a is called the gain
      · amplifier if $|a| > 1$
      · attenuator if $|a| < 1$
      · inverter if $a < 0$
      · differentiator: $y(t) = u'(t)$
      · integrator: $y(t) = \int_0^t u(\tau)d\tau$
      · time delay: $y(t) = u(t-T)$
      · convolution system: $y(t) = \int u(t-\tau)h(\tau)d\tau$, h is a given function

- static (memory-less) vs dynamic (with memory)
- causal vs non-causal
- linear vs nonlinear
  - classical definition of a linear system:
    1. homogeneity: $S(au) = aS(u)$ (scaling before or after is the same, illustrate in block diagram)
    2. superposition: $S(u_1 + u_2) = S(u_1) + S(u_2)$ (summing before or after is the same, illustrate)
- time-invariant vs time-varying
- inputs and outputs (filters)
- response of a system

$$y = S(u) = S\left(\sum_{\tau=-\infty}^{\infty} x(\tau)\delta(t-\tau)\right) = \sum_{\tau=-\infty}^{\infty} x(\tau)S\big(\delta(t-\tau)\big) = \sum_{\tau=-\infty}^{\infty} x(\tau)h(t-\tau)$$

Homework **HW**

Optional reading assignments

- section 1.1 (classification of signals), 1.2 (classification of systems), and chapter 2 (representation of signals and systems) from [**?**]

Problems

- **System classification**
  - Give specific examples of:
    - linear static system
    - nonlinear static system
    - linear time-invariant dynamical systems
      - finite impulse response (FIR)
      - infinite impulse response (IIR)
      - scalar
      - multivariable
    - linear time-varying dynamical systems
    - nonlinear time-invariant dynamical systems
    - nonlinear time-varying dynamical systems

- **Response of an LTI system**

  1. Find analytically the response of 1st and 2nd order continuous-time linear time-invariant autonomous systems.
  2. Write a function that computes the response.
  3. Test the function on a numerical example and compare the result with the one obtained via the function `lsim` from Control Toolbox of Matlab.
  4. Generalize 1–3 for a general nth order linear time-invariant autonomous systems.

3 Representations of LTI systems

Review of matrix algebra

Representation of static systems

- function vs relation
- image representation
- kernel representation
- input/output representation

Representation of linear time-invariant systems

- difference/differential equations

$$\sum_{\tau=1}^{n} a_\tau y(t-\tau) = \sum_{\tau=1}^{m} b_\tau u(t-\tau)$$

  special case

$$y(t) = \sum_{\tau=1}^{m} b_\tau u(t-\tau)$$

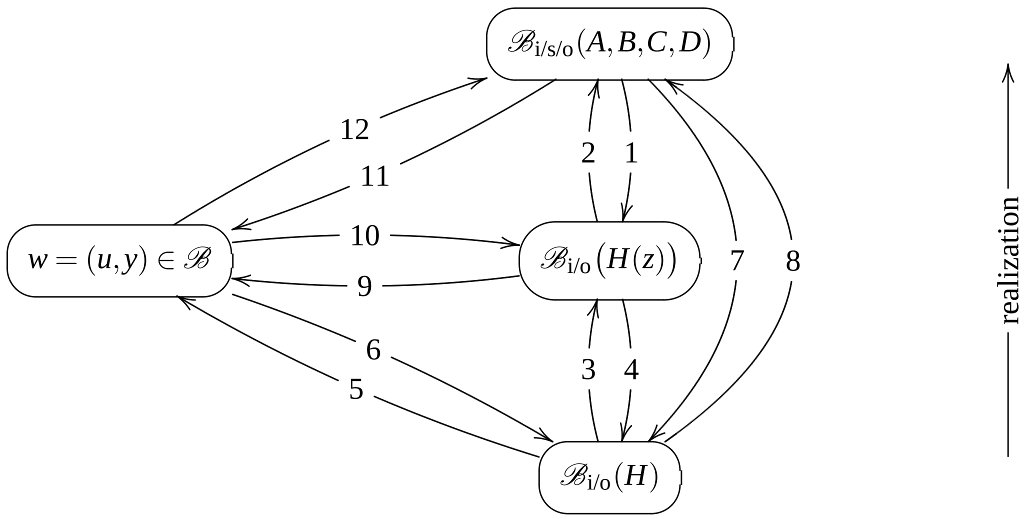
  - the differentiation (continuous-time systems) / shift (discrete-time systems) operator σ
  - autonomous systems, see Section 3.2 in [**?**]
  - initial conditions (free response)
  - systems with inputs
  - linear constant coefficient ordinary differential equation
  - example mass-spring-damper
- transfer function (forced response)
  - rational = finite order
- state space
  - nonuniqueness
  - realization problem
  - matrix exponential
- convolution

$$(x \star h)(t) = \sum_{\tau=-\infty}^{\infty} x(\tau)h(t-\tau)$$

  - properties (show in block diagrams)
    * $x \star h = h \star x$
    * $(x \star h_1) \star h_2 = x \star (h_1 \star h_2)$
    * $x \star (h_1 + h_2) = x \star h_1 + x \star h_2$
  - the deconvolution problem
- convolution systems
  - impulse response and step response
  - convolution theorem: $y = h \star u \leftrightarrow Y = HU$
    $$\begin{array}{ccc} u & \xrightarrow{h} & y \\ {\scriptstyle F}\downarrow & & \uparrow{\scriptstyle F^{-1}} \\ U & \xrightarrow{H} & Y \end{array}$$
  - infinite impulse response (IIR) systems, fading memory
  - finite impulse response (FIR) systems
  - DC gain $H(0) = \int_0^\infty h(\tau)d\tau = \lim_{\tau\to\infty} s(\tau)$ or $H(1) = \sum_0^\infty h(\tau)d\tau$

Links among representations

data ——— identification ——→ *model*



1. $H(z) = C(Iz - A)^{-1}B + D$
2. realization of a transfer function
3. Z or Laplace transform of $H(t)$
4. inverse transform of $H(z)$
5. convolution $y_{\text{d}} = H \star u_{\text{d}}$
6. exact identification
7. $H(0) = D$, $H(t) = CA^{t-1}B$ (discrete-time), $H(t) = Ce^{At}B$ (continuous-time), for $t > 0$
8. realization of an impulse response
9. simulation with input u_{d} and $x(0) = 0$
10. exact identification
11. simulation with input u_{d} and $x(0) = x_{\text{ini}}$
12. exact identification

Realization theory

- Sections 2.2 and 3.1 from [**?**]
- Sections 6.5–8 from [**?**]

Homework **HW**

Additional reading

Chapters 1 (behavioral models) and 4 (state-space representation) from (Polderman and Willems)

Problems

- **Matrix representation of the discrete Fourier transform**

  1. Find a matrix representation F of the discrete Fourier transform
  $$\widehat{x}(k) = \sum_{t=0}^{T-1} x(t)e^{-\mathbf{i}2\pi tk/T}, \qquad \text{for } k = 0,1,\dots,T-1$$
  of signal $x = \big(x(0),\dots,x(T-1)\big)$.
  2. What is the number of multiplications needed to compute $\widehat{x} = Fx$ by matrix-vector multiplication? Compare this number with the $T\log_2(T)$ multiplications needed for the same computation by the fast Fourier transform.

3. Using Matlab's `tic` and `toc` functions measure the computation times of the matrix-vector multiplication and the fast Fourier transform methods for the computation of the discrete Fourier transform of x. (Use a random x.) Repeat the experiment for different size n of x and plot the results.

4. Do the empirical observations match the theoretical predictions? Justify your answer by fitting the observed computation times to the analytical expressions.

- **Matrix representation of the convolution operation**

1. Find a matrix representation M_h of the convolution operator

$$(h \star x)(t) := \sum_{\tau=1}^{n} h(\tau)x(t-\tau), \qquad \text{for } t = 1, \dots, T$$

of the signals

$$h = \big(h(1), \dots, h(n)\big) \quad \text{and} \quad x = \big(x(-n+1), \dots, x(0), x(1), \dots, x(T)\big).$$

2. What is the number of multiplications needed to compute $y = M_h x$ by matrix-vector multiplication?

3. Propose a method for convolution based on the fast Fourier transform. What is the computational cost of this method?

4. Can you propose another fast method for convolution in the case when h is a sum of exponentials?

4 Stochastic models

Convolution of probabilities

- consider discrete random variables X
- let $p_i = \text{prob}(X = i)$ be the probability of the event that $X = i$
- we have that $0 \leq p_i \leq 1$ and $\sum p_i = 1$
- p is called the *probability density function (pdf)* of X
- consider *independent identically distributed (i.i.d.)* random variables X_1 and X_2 with pdf's $p^{(1)}$ and $p^{(2)}$
- problem: find the pdf of $X_1 + X_2$, *i.e.*, find the probabilities $\text{prob}(X_1 + X_2 = k)$
- example: dice trowing $p_i^{(j)} = 1/6$, for $j = 1,2$ and $i = 1, \dots, 6$
- we have $\text{prob}(X_1 = i \text{ and } X_2 = j) = p_i^{(1)} p_j^{(2)}$
- $X_1 + X_2 = k$ is a union of mutually exclusive events: $X_1 = i$ and $X_2 = k - i$, which probability is $p_i^{(1)} p_{k-i}^{(2)}$
- therefore
$$\text{prob}(X_1 + X_2 = k) = \sum p_i^{(1)} p_{k-i}^{(2)} = (p^{(1)} \star p^{(2)})(k)$$
- back to the dice example:
$$\frac{1}{6}(1,1,1,1,1,1) \star \frac{1}{6}(1,1,1,1,1,1) = \frac{1}{36}(1,2,3,4,5,6,5,4,3,2,1)$$
- note: a box convolved with a box results to a hat
- convolution of the box $N \geq 2$ times with itself

```
n = 7; T = 20; N = 5; box = zeros(1, T); box((round(T/2) - n):(round(T/2) + n))
p = box; figure(1), stem(p), pause(1), print_fig('conv0')
for i = 1:N, p = conv(p, box); stem(p), pause(1), print_fig(['conv' int2str(i)])
```

convergence to Gaussian distribution

- central limit theorem: sum of independently distributed random variables converges to a normally distributed random variable
- another solution using the *generating function*

$$p(z) = p_1 z + p_2 z^2 + \cdots + p_n z^n$$

- the generating function of $X_1 + X_2$ is $p^{(1)} p^{(2)}$
- product of polynomials is a convolution of their coefficients

Gaussian distribution (second order process)

- notation: $x \sim \mathbf{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right)$$

- Gaussian random vector $x \sim \mathbf{N}(\mu, V)$

$$p(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(V)}} \exp\left(-\frac{1}{2}(x-\mu)^\top V^{-1}(x-\mu)\right)$$

  - diagonal V — uncorrelated RV

- affine transformation of a Gaussian random vector $y = Ax + b$, $y \sim \mathbf{N}(A\mu + b, AVA^\top)$
  - of particular interest is a transformation A that de-correlates (whitens) x

$$AVA^\top = D \qquad D \text{ — diagonal}$$

  - Cholesky decomposition of a positive definite matrix $V = LDL^\top$, L — lower triangular with ones on the diagonal, then $A = L^{-1}$

Random processes

- a discrete-time random process (signal) y is a sequence of random variables
  - example white Gaussian process
  - Markov chains
- the mean of y is the sequence of the means
- covariance matrix (assuming zero mean process)

$$R_y = \big[\underbrace{\mathbf{E}\big(y(i)y(j)\big)}_{r_y(i,j)}\big]_{i,j=1}^{m} = \mathbf{E}\begin{bmatrix} y(1) \\ y(0) \\ \vdots \\ y(-m) \end{bmatrix} \begin{bmatrix} y(1) & y(0) & \cdots & y(-m) \end{bmatrix}$$

- property: positive semi-definite matrix
- intuition: rate of variation of the signal (show diagrams)

- *stationary process (signal):*
  - strict sense stationary
  - wide sense stationary: $\mathbf{E}(y)$ is a constant and $r_y(i,j)$'s depends only on $\tau = i - j$
    * then R_y is a Toeplitz structured matrix

$$r_y(\tau) = E\big(y(t)y(t-\tau)\big)$$

    * properties of correlation function
      · $r_y(\tau) = r_y(-\tau)$
      · $r_y(0) \geq |r_y(\tau)|$, for all τ
    * link to convolution: $r_y = y \star \mathrm{rev}(y)$, where "rev" is the time reversal

$$\big(\mathrm{rev}(y)\big)(t) = y(-t)$$

    * power spectral density
    * note that with probability one a realization of a random process has no finite energy and therefore no Fourier transform
    * random processes however usually have finite average power and can be characterized by the average power spectral density
    * Fourier transform of the covariance function

$$\phi_y(\omega) = \sum_{k=-\infty}^{\infty} r_y(k) e^{-\mathbf{i}\omega k}$$

    * properties
      · $\phi_y(\omega) \geq 0$, for all ω
      · if y is real, $\phi_y(\omega) = \phi_y(-\omega)$
  - Parseval's theorem: energy preservation in time and frequency domain

$$\sum_{t=-\infty}^{\infty} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_y(\omega)d\omega$$

  - ergodic process: expectation can be computed by time averaging

$$r_y(\tau) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} y(t)y(t-\tau)$$

- transformation of the power spectral density by an LTI system

$$\phi_y(\omega) = |H(\omega)|^2 \phi_u(\omega)$$

  - show first that

$$r_y(\tau) = \int\int h(\alpha)h(\beta)r_u(\tau+\alpha-\beta)d\alpha d\beta$$

- using $\phi_y(\omega) = \int r_y(\tau) e^{-\mathbf{i}\omega\tau} d\tau$ and the change of variables $\lambda = \tau + \alpha - \beta$, we have

$$\begin{aligned}
\phi_y(\omega) &= \int\int\int h(\alpha)h(\beta)r_u(\tau+\alpha-\beta)e^{-\mathbf{i}\omega\tau}d\alpha d\beta d\tau \\
&= \int\int\int h(\alpha)h(\beta)r_u(\tau+\alpha-\beta)e^{-\mathbf{i}\omega(\lambda-\alpha+\beta)}d\alpha d\beta d\lambda \\
&= \int h(\alpha)e^{-\mathbf{i}\omega\alpha}d\alpha \int h(\beta)e^{-\mathbf{i}\omega\beta}d\beta \int r_u(\lambda)e^{-\mathbf{i}\omega\lambda}d\lambda \\
&= H(-\omega)H(\omega)\phi_u(\omega) \\
&= |H(\omega)|^2\phi_u(\omega)
\end{aligned}$$

- another formula $\phi_{uy}(\omega) = H(\omega)\phi_u(\omega)$
- white Gaussian noise
  - power spectrum $\phi_u(\omega) = 1$, for all ω
- shaping filter
- spectral factorization problem
- parametrized random processes
  1. moving average (MA)
  2. auto regressive (AR)
  3. auto regressive moving average (ARMA)
  4. auto regressive moving average exogenous (ARMAX)
- realization of stochastic systems

Wiener-Khintchine theorem

- consider a random process y
- Fourier transform of the autocorrelation function r is the power spectral density ϕ_y
  - definition: $\phi_y := F(y)F^*(y) = |F(y)|^2$
  - theorem: $\phi_y = F(r_y)$
  - the proof is based on the following properties:
    * $F(y \star y) = F(y)F(y)$
    * $F\big(\mathrm{rev}(y)\big) = F^*(y)$
    * $y \star \mathrm{rev}(y) = r_y$

$$\phi_y = F(y)F^*(y) = F(y)F\big(\mathrm{rev}(y)\big) = F\big(y \star \mathrm{rev}(y)\big) = F(r_y)$$

- note: direct computation requires $O(n^2)$ flops, FFT computation requires $O(n\log(n))$

Estimation principles

- conditional expectation
  - consider two RV x and y with joint pdf $p_{x,y}$
  - we want to infer x from observations of y
  - estimator: $\widehat{x} = h(y)$
  - if x and y are independent, nothing can be said about x, seeing y
  - optimal estimator: $\widehat{x} = \mathbf{E}(x\bar{y})$ — conditional expectation of x, given y
  - important special case: linear estimators and Gaussian pdf
- maximum likelihood: parameter dependent pdf $p(x,\theta)$
  - substitute the observed data in $p(x,\theta)$ and view the resulting function $L(\theta)$ in θ as the "likelihood" for the occurrence of the data, given the parameters θ
  - maximize the likelihood for all admissible parameter values $\theta \in \Theta$
- minimization of the MSE: $\mathbf{E}\big((\theta - \widehat{\theta})^2\big)$

Spectral estimation

- basic problem: determine the spectral content of a signal (random process) based on finite set of observations
- PSD function describes the distribution of power of the signal with frequency
- "physical way of doing it": 1) filter with a sufficiently narrow band-pass filter centered at $f = f_0$ and measure the power of the output. 2) divide the power by the filter width, 3) repeat the process for different f_0
- traditionally based on the Fourier transform
- since 1980 new "modern"/"high resolution" approaches were developed
- in the last 10 years compressive sensing is taking over
- it is important to have good estimates for small number of data points, due to the fact that many signals are only "locally WSS", e.g., speech (WSS for about 20–80 ms)
- PSD is a function of an infinite number of ACF values, the task of estimating the PSD based on a finite date set is an impossible one. at best we can only estimate a subset of the most significant ACF values, which are assumed to be for $k = 0,\dots,M$. If the process exhibits strong correlations for $k > M$, the results are heavily biased.
- the "impossibility problem" can be resolved by postulating specific forms of the PSD, e.g., rational PSD. The process is parameterized by a finite (and small) number of coefficients. The spectral estimation problem becomes the one of parameter estimation. It is important however that the model is an accurate representation of the true PSD.
- nonparameteric methods
  - periodogram

$$\widehat{\phi}(\omega) = \frac{1}{N}\left|\sum_{t=1}^{T} y(t)e^{-\mathbf{i}\omega t}\right|^2$$

    Square the absolute value of the DFT on the process.

- correlogram

$$\widehat{\phi}(\omega) = \sum_{k=-(N-1)}^{N-1} \widehat{r}_y(k) e^{-\mathbf{i}\omega k}$$

Apply the DFT on the estimated autocorrelation sequence

$$\widehat{r}_y(k) = \frac{1}{T} \sum_{t=k+1}^{N} y(t)y(t-k), \qquad \text{for } 0 \leq k \leq T-1$$

For negative lags $\widehat{r}_y(-k) = \widehat{r}_y(k)$, $k = 0, \dots, T-1$.

- maximum likelihood

Stochastic realization

Homework HW

Reading assignment

- notes Leo part 1 sections 1.5–1.8 and notes Leo part 2

Problems

- **Wiener-Khintchine theorem** For a discrete-time signal y, let
  - $\phi_y := |F(y)|^2$, where $F(y)$ be a Fourier transform of y, and
  - $r_y := \sum_{t=1}^{T} y(t)y(t-\tau)$.

  Show that $\phi_y = F(r_y)$.

5 Least-squares estimation

Homework HW

Reading assignment

- notes Leo part 3, sections 3.1–3.3

Problems

- **Orthogonality principle for least-squares estimation**

  Show that

  1. $\widehat{x}$ being a least squares approximate solution of the system $Ax = b$, and
  2. $\widehat{x}$ being such that $b - A\widehat{x}$ is orthogonal to the span of the columns of A,

  are equivalent. (This result is known as the orthogonality principle for least squares approximation.)

- **Weighted least-squares approximate solution**

  For a given positive definite matrix $W \in \mathbb{R}^{m\times m}$, define the weighted 2-norm

  $$||e||_W = e^\top W e.$$

  The weighted least-squares approximation problem is

  $$\text{minimize} \quad \text{over } \widehat{x} \in \mathbb{R}^n \quad ||A\widehat{x} - b||_W. \tag{WLS}$$

  When does a solution exist and when is it unique? Under the assumptions of existence and uniqueness, derive a closed form expression for the least squares approximate solution.