メディア処理基礎 / Fundamentals of Media Processing

Fundamentals of Signal Processing
Part 2

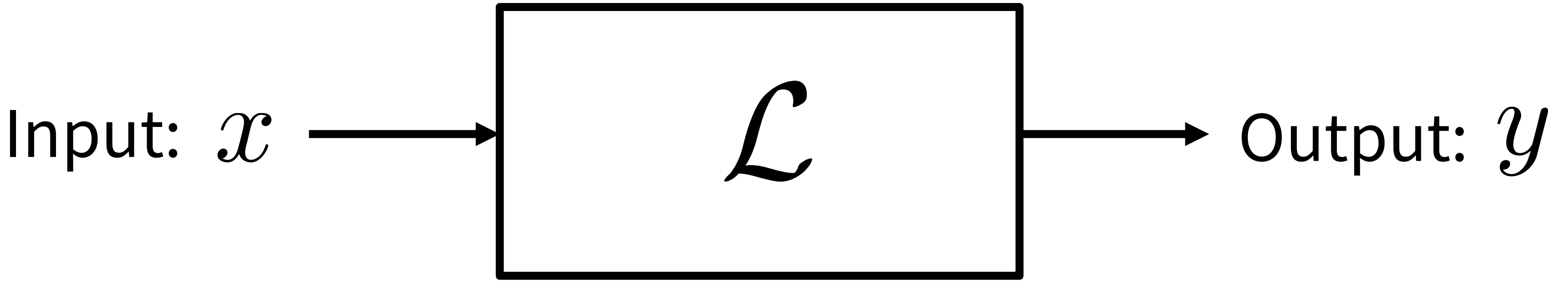
小山 翔一 / Shoichi Koyama

国立情報学研究所 / 総合研究大学院大学　准教授

Associate Professor, National Institute of Informatics / SOKENDAI

What is Signal Processing?

- Techniques for analyzing, modifying, and synthesizing signals, such as sound, images, and others



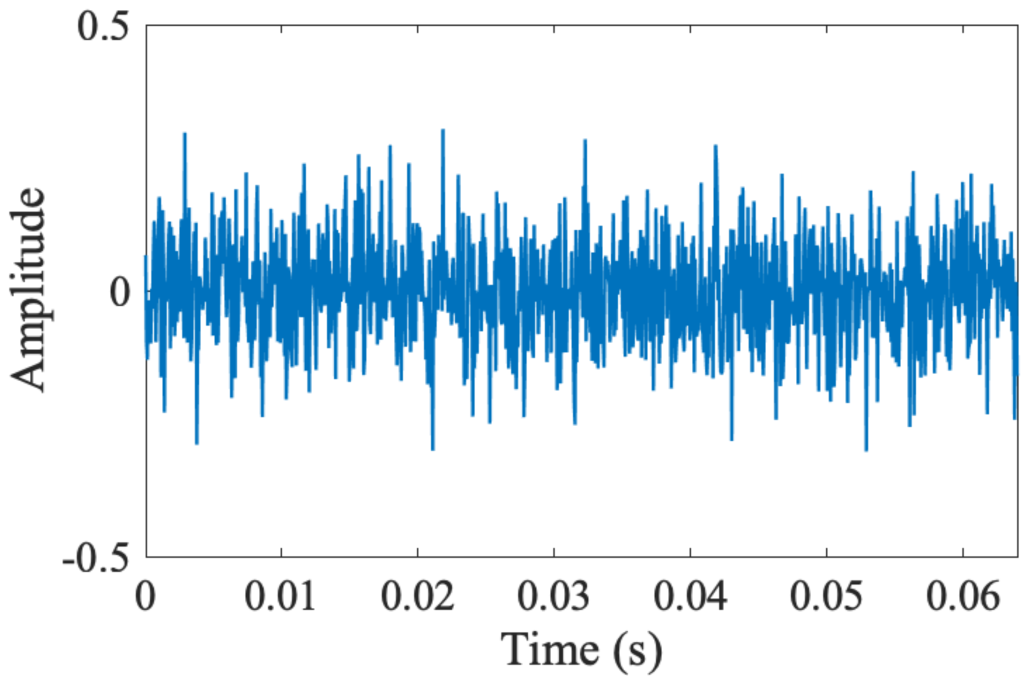
Generating output y by processing input x using mapping $\mathcal{L}$

See also https://youtu.be/R90ciUoxcJU

From deterministic to statistical signal processing

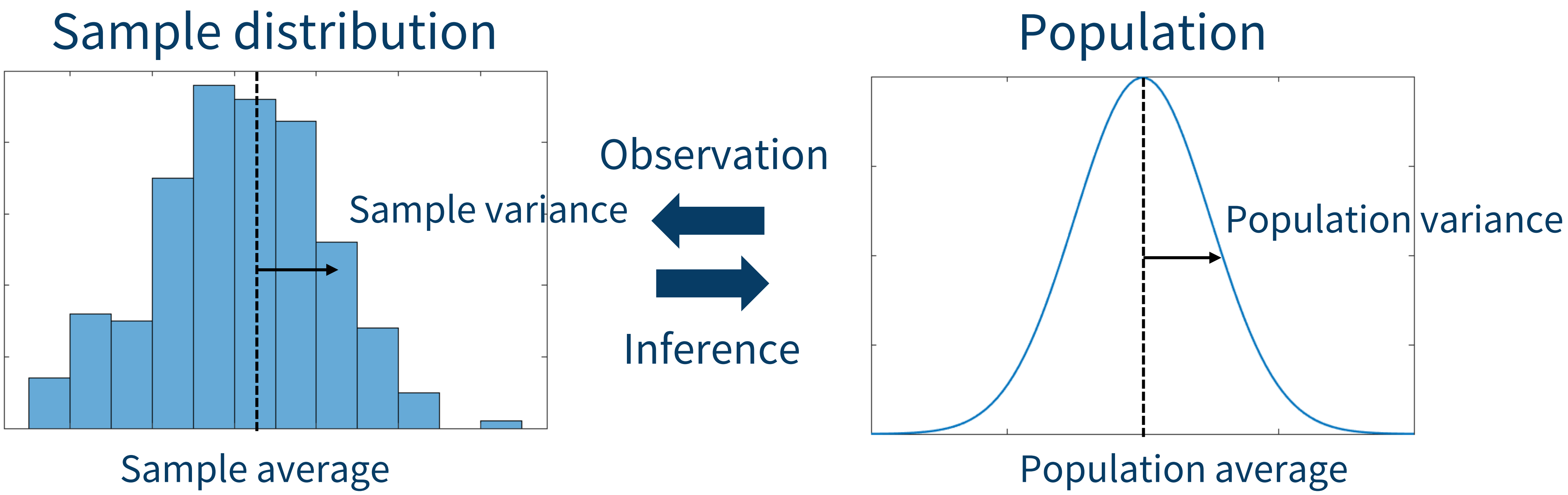
- A signal is considered to be deterministic in the classical signal processing
- In statistical signal processing, the observed signal is considered to be a stochastic process, i.e., random signal

➡ Gain insight into statistical property of signals



From deterministic to statistical signal processing

➢ Basic concept of statistical inference



Estimating some amounts of statistics of population
(e.g., population average/variance) from observed samples

BASICS OF STATISTICS/STOCHASTIC PROCESS

Probability density function

- Random variable X taking continuous value holds

$$P(a \leq X \leq b) = \int_a^b f(x)\mathrm{d}x$$

Probability that random variable X takes the value of $[a, b]$

- $f(x)$ is called probability density function (or p.d.f.) of X satisfying

$$f(x) \geq 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

- Cumulative distribution function is probability of $X \leq x$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$$

Expected value and Variance

- Expected value $E[X]$ and variance $V[X]$ are given by

Expected value: $$E[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

Variance: $$V[X] = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \mathrm{d}x$$

where $\mu = E[X]$

- Relationship between expected value and variance

$$V[X] = E[(X-\mu)^2] = E[X^2] - E[X]^2$$

Skewness and Kurtosis

- Skewness and Kurtosis are also statistics to characterize p.d.f.
- Skewness: measure for asymmetric diversity of p.d.f

$$\frac{1}{\sigma^3}E[(X-\mu)^3] = \frac{1}{\sigma^3}\int_{-\infty}^{\infty}(x-\mu)^3 f(x)\mathrm{d}x$$

- Kurtosis: measure for flatness of p.d.f.

$$\frac{1}{\sigma^4}E[(X-\mu)^4] = \frac{1}{\sigma^4}\int_{-\infty}^{\infty}(x-\mu)^4 f(x)\mathrm{d}x$$

where $\sigma^2 = V[X]$

- In general, moment of n th order is defined as

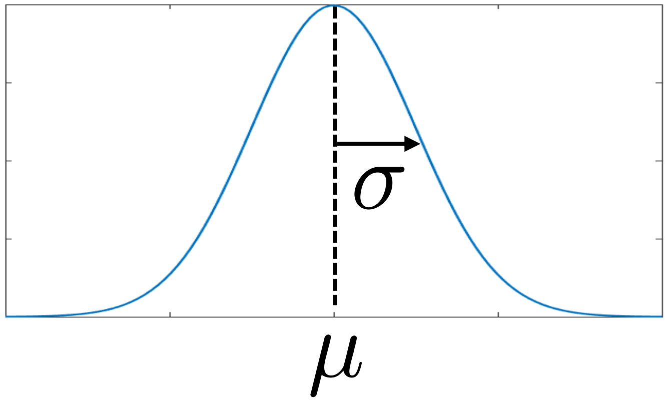
$$E[(x-\mu)^n]$$

Shape of p.d.f. is determined if all moments are determined

Normal distribution (Gaussian distribution)

- P.d.f. of normal distribution of mean μ and variance σ^2

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$
$$:= \mathcal{N}(x; \mu, \sigma^2)$$



$\left[\text{Random variable } X \text{ follows normal distribution} \Rightarrow X \sim \mathcal{N}(x; \mu, \sigma^2)\right]$

- Useful for modeling various phenomena (central limit theorem)
- Normal distribution is determined only by 1st- and 2nd-order moments
- P.d.f. having larger skewness than normal distribution is called super-Gaussian function, and having less skewness is sub-Gaussian function

Multivariate distribution

- Consider joint probability distribution for more than one random variables
- Joint probability density function $f(x, y)$ for random variables X and Y satisfies

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) \mathrm{d}x \mathrm{d}y$$

- Marginalization means eliminating one variable by integration of joint p.d.f.

$$g(x) = \int_{-\infty}^{\infty} f(x, y) \mathrm{d}y \ , \qquad h(y) = \int_{-\infty}^{\infty} f(x, y) \mathrm{d}x$$

Multivariate distribution

- In multivariate case, random variables are sometimes represented by vectors
- Representing N random variables by column vector $\boldsymbol{x} = [x_1, \ldots, x_N]^\mathsf{T}$
- Mean:

$$\begin{aligned}\boldsymbol{\mu} &= E[\boldsymbol{x}] \\ &= [E[x_1], \ldots, E[x_N]]^\mathsf{T} \\ &= [\mu_1, \ldots, \mu_N]^\mathsf{T}\end{aligned}$$

- Covariance matrix:

$$\begin{aligned}\boldsymbol{\Sigma} &= E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\mathsf{T}] \\ &= \begin{bmatrix} E[(x_1 - \mu_1)^2] & \cdots & E[(x_1 - \mu_1)(x_N - \mu_N)] \\ \vdots & \ddots & \vdots \\ E[(x_N - \mu_N)(x_1 - \mu_1)] & \cdots & E[(x_N - \mu_N)^2] \end{bmatrix}\end{aligned}$$

Multivariate normal distribution

- P.d.f. of N dimensional normal distribution:

$$
\begin{aligned}
f(\boldsymbol{x}) &= \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \\
&= \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})
\end{aligned}
$$

- Vector of random variables $\boldsymbol{x}$ following normal distribution is represented as

$$
\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})
$$

Statistics of random signal

- If statistical properties of signal is constant irrespective of time, that signal is called stationary signal. If not, that signal is called non-stationary signal.
  - Weakly stationary: Mean and variance of signal are constant irrespective of time
  - Strongly stationary: Higher-order statistics (including skewness and kurtosis), i.e., p.d.f of signal, are constant irrespective of time

Statistics of random signal

- Statistics of stationary signal $x(t)$
  - Ensemble mean:

$$\eta = E[x(t)] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

  - Autocorrelation function:

$$R(\tau) = E[x(t+\tau)x(t)] = \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; \tau)$$

Dependent only on time difference

  - Crosscorrelation function:

$$R_{xy}(\tau) = E[x(t+\tau)y(t)] = R_{yx}(-\tau)$$

  - Autocovariance function:

$$C(\tau) = E[(x(t+\tau)-\eta)(x(t)-\eta)] = R(\tau) - \eta^2$$

  - Crosscovariance function:

$$C_{xy}(\tau) = E[(x(t+\tau)-\eta_x)(y(t)-\eta_y)] = R_{xy}(\tau) - \eta_x \eta_y$$

Statistics of random signal

- Time average:

$$\bar{x}(t) = \lim_{T\to\infty} \int_{-T/2}^{T/2} x(t)\mathrm{d}t$$

- Ergodicity:
  - When time average and ensemble mean of stationary signal are equal, the signal has ergodicity
  - Stationary signal does not necessarily have ergodicity

Statistics of random signal

- Uncorrelated

$$E[x(t_1)y(t_2)] = E[x(t_1)]E[y(t_2)]$$

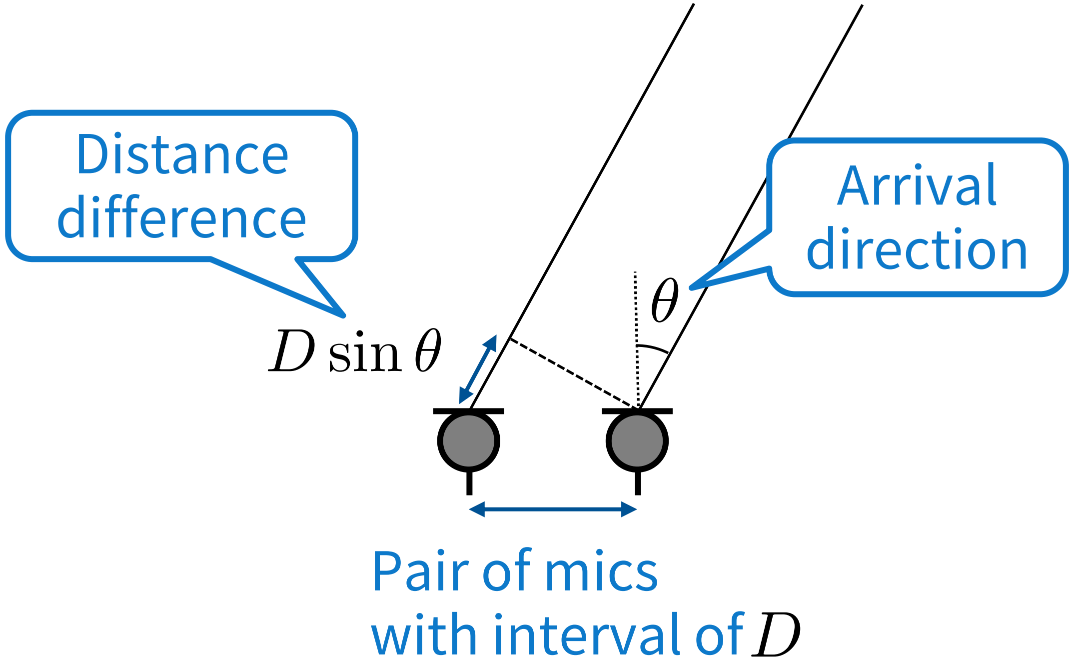
- Independent

$$f_{xy}(x, y; t_1, t_2) = f_x(x; t_1)f_y(y; t_2)$$

- Independent signals are uncorrelated, but uncorrelated signals are not necessarily independent

Example: direction-of-arrival estimation

- Estimating direction of sound source by using two mics based on time difference estimation



$$\begin{cases} s_1(t) = x(t - \tau_0) + n_1(t) \\ s_2(t) = x(t) + n_2(t) \\ \tau_0 = D \sin\theta / c \end{cases}$$

Peak of crosscorrelation function corresponds to τ_0

$$R_{12}(\tau) = E[s_1(t+\tau)s_2(t)]$$

STATISTICAL MODEL AND ESTIMATION

Parameter estimation

- Suppose to estimate unknown parameter $\boldsymbol{x} = [x_1, \ldots, x_N]^\mathsf{T} \in \mathbb{R}^N$ from observed signal $\boldsymbol{y} = [y_1, \ldots, y_M]^\mathsf{T} \in \mathbb{R}^M$ $(M \geq N)$
- Three representative parameter estimation methods (w/ p.d.f. $p(\cdot)$)
  - Maximum likelihood (**ML**) estimation:

$$\underset{\boldsymbol{x}}{\text{maximize}}\, p(\boldsymbol{y}; \boldsymbol{x})$$

  - Maximum a posteriori (**MAP**) estimation:

$$\underset{\boldsymbol{x}}{\text{maximize}}\, p(\boldsymbol{x}; \boldsymbol{y})$$

  - Minimum mean-square error (**MMSE**) estimation:

$$\underset{\hat{x}}{\text{minimize}}\, E[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2; \boldsymbol{y}]$$

Linear model

- Linear model is widely-used measurement model:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

- $\boldsymbol{A}$ is matrix of $M \times N$ having real-valued elements (i.e., $\boldsymbol{A} \in \mathbb{R}^{M \times N}$)
- Each element is given by

(m,n)th element of $\boldsymbol{A}$

$$y_m = \sum_{n=1}^{N} a_{mn} x_n$$

- Here, $\boldsymbol{A}$ is assumed to be given

ML estimation

- Observation values $x_1, \ldots, x_N$ are obtained as realization of random variables $X_1, \ldots, X_N$ by N observations
- In maximum likelihood principle, the measured observation is considered to be realization of the maximum probability, i.e., the most likely occurrence
- P.d.f. that observation x_n follows is denoted with paramater θ of p.d.f. as $f(x_n|\theta)$
- When observed signals are independent and identically distributed (i.i.d.),

$$
\begin{aligned}
P(X_1 = x_1, \ldots, X_N = x_N) &= \prod_{n=1}^{N} P(X_n = x_n) \\
&= \prod_{n=1}^{N} f(x_n|\theta)
\end{aligned}
$$

ML estimation

- Likelihood function $\mathcal{L}(\theta)$ is function of θ represented by p.d.f. of observed signal

$$\mathcal{L}(\theta) = \prod_{n=1}^{N} f(x_n|\theta)$$

- In ML estimation, estimate is given as θ such that likelihood function is maximized

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta)$$

- In many cases, log likelihood function (natural logarithm of $\mathcal{L}(\theta)$) is used because of simplicity of computation

log is monotonically increasing function

$$\hat{\theta} = \arg\max_{\theta} \log \mathcal{L}(\theta)$$

ML estimation in linear model

- When additive noise $\boldsymbol{n} \in \mathbb{R}^M$ is superimposed on the observation,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$$

- Noise is assumed to follow multivariate normal distribution of mean 0 and variance σ^2, i.e., $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{n}; \boldsymbol{0}, \sigma^2 \boldsymbol{I})$

$$p(\boldsymbol{n}) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{n}\|^2\right)$$

- P.d.f. of observation $p(\boldsymbol{y})$ becomes

$$p(\boldsymbol{y}) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2\right)$$

ML estimation in linear model

- Suppose to infer $\boldsymbol{x}$ by ML estimation. Log likelihood function becomes

$$\log \mathcal{L}(\boldsymbol{x}) = \log p(\boldsymbol{y}) = -\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + C$$

Constant not including $\boldsymbol{x}$

- ML solution $\hat{\boldsymbol{x}}$ is obtained by solving

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2$$

- Correspond to least-squares method that estimates unknown variable by minimizing cost function of squared errors

ML estimation in linear model

- By differentiating cost function by $\boldsymbol{x}$ and setting it to 0,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{x}}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2 &= \frac{\partial}{\partial \boldsymbol{x}}\left(\boldsymbol{y}^\top\boldsymbol{y}-\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{y}-\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{x}+\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\right)\\ &= -\boldsymbol{A}^\top\boldsymbol{y}-\left(\boldsymbol{y}^\top\boldsymbol{A}\right)^\top+2\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\\ &= 2\left(-\boldsymbol{A}^\top\boldsymbol{y}+\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\right)\\ &= 0\end{aligned}$$

- Cf., differential formula for vectors

$$\frac{\partial \boldsymbol{x}^\top\boldsymbol{b}}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{b}^\top\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{b}$$

$$\frac{\partial \boldsymbol{x}^\top\boldsymbol{B}\boldsymbol{x}}{\partial \boldsymbol{x}}=(\boldsymbol{B}+\boldsymbol{B}^\top)\boldsymbol{x}$$

ML estimation in linear model

- Optimal solution $\hat{\boldsymbol{x}}$ is

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^{\top}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{y}$$

- In linear model, ML solution with Gauusian assumption for noise corresponds to least-squares solution

Bayesian estimation

- In Bayesian estimation, unknown variable $\boldsymbol{x}$ of linear model is also regarded as random variable ➡ Main difference from ML estimation

- Based on Bayesian theorem, posterior probability ditribution, i.e., p.d.f. of $\boldsymbol{x}$ given $\boldsymbol{y}$ is represented as

$$p(\boldsymbol{x};\boldsymbol{y}) = \frac{p(\boldsymbol{y};\boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{y};\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}} \propto p(\boldsymbol{y};\boldsymbol{x})p(\boldsymbol{x})$$

– where $p(\boldsymbol{x})$ is prior distribution, and $p(\boldsymbol{y};\boldsymbol{x})$ is likelihood

MAP estimation

- In MAP estimation, the estimate is obtained by

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x};\boldsymbol{y})$$
$$= \arg\max_{\boldsymbol{x}} p(\boldsymbol{y};\boldsymbol{x})p(\boldsymbol{x})$$

Bayesian theorem

- When there is no prior information, $p(\boldsymbol{x})$ becomes constant (non-informative prior distribution), then, MAP estimate corresponds to ML estimate
- If prior distribution is conjugate prior of likelihood, posterior distribution can be simply calculated. E.g., when both $p(\boldsymbol{x})$ and $p(\boldsymbol{y};\boldsymbol{x})$ are Gaussian, posterior distribution $p(\boldsymbol{x};\boldsymbol{y})$ also becomes Gaussian

MAP estimation in linear model

- Assume Gaussian prior $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}; \mathbf{0}, \sigma_x^2 \boldsymbol{I})$

$$p(\boldsymbol{x}) = \frac{1}{(2\pi\sigma_x^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_x^2}\|\boldsymbol{x}\|^2\right)$$

- MAP estimate is obtained as

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} \left[-\log p(\boldsymbol{y}; \boldsymbol{x}) - \log p(\boldsymbol{x})\right] \\
&= \arg\min_{\boldsymbol{x}} \left[\frac{1}{\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \frac{1}{\sigma_x^2}\|\boldsymbol{x}\|^2\right] \\
&= \left(\boldsymbol{A}^\top \boldsymbol{A} + \frac{\sigma^2}{\sigma_x^2}\boldsymbol{I}\right)^{-1} \boldsymbol{A}^\top \boldsymbol{y}
\end{aligned}$$

➡ Corresponds to regularized least-squares solution

MMSE estimation

- In MMSE estimation, expected value of square error between ground truth $\boldsymbol{x}$ and estimate $\hat{\boldsymbol{x}}$ is minimized to obtain the estimate $\hat{\boldsymbol{x}}$

$$\hat{\boldsymbol{x}} = \arg\min_{\hat{\boldsymbol{x}}} E[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2]$$

- Suppose to estimate by using linear MMSE estimator $\boldsymbol{H}$ as $\hat{\boldsymbol{x}} = \boldsymbol{H}\boldsymbol{y}$

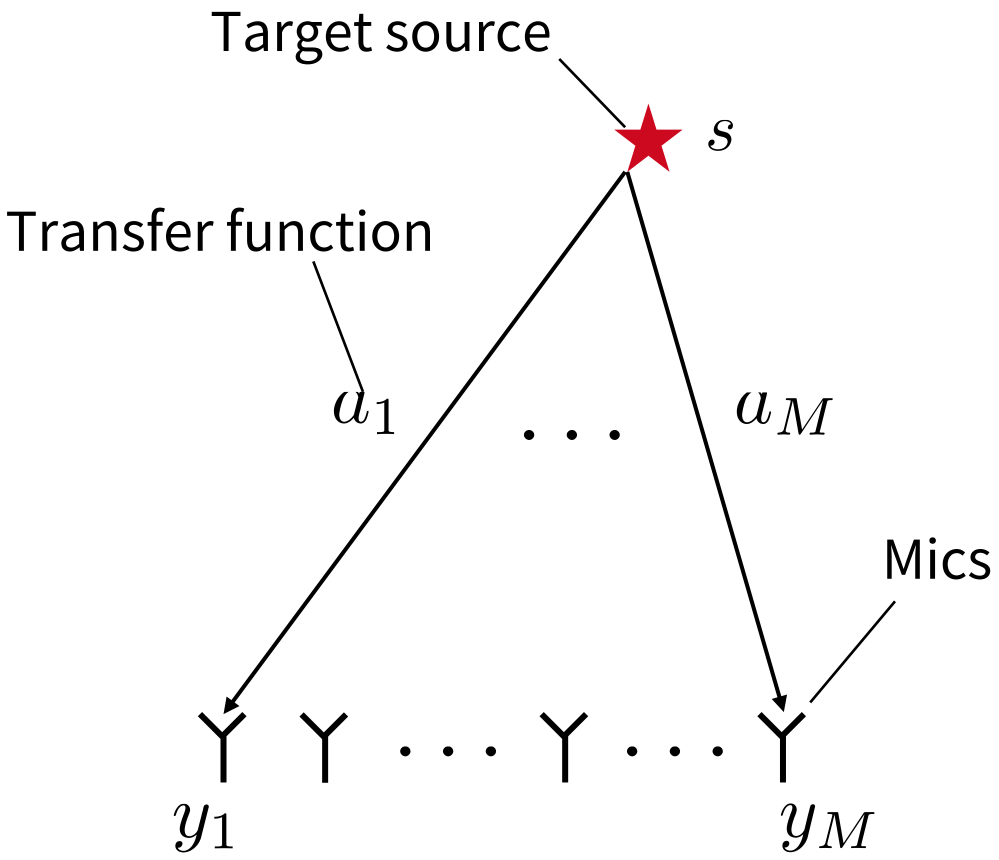
$$\underset{\boldsymbol{H}}{\text{minimize}}\, E[\|\boldsymbol{H}\boldsymbol{y} - \boldsymbol{x}\|^2]$$

  – Then, MMSE estimation is reduced to obtain the optimal $\boldsymbol{H}$

APPLICATION EXAMPLES OF STATISTICAL ESTIMATION

Beamforming

- Enhancing target source signal by multiple mics



Linear measurement model

$$
\begin{aligned}
\boldsymbol{y}(\omega) &= \begin{bmatrix} y_1(\omega) \\ \vdots \\ y_M(\omega) \end{bmatrix} \\
&= \begin{bmatrix} a_1(\omega) \\ \vdots \\ a_M(\omega) \end{bmatrix} s(\omega) + \begin{bmatrix} n_1(\omega) \\ \vdots \\ n_M(\omega) \end{bmatrix} \\
&= \boldsymbol{a}(\omega) s(\omega) + \boldsymbol{n}(\omega)
\end{aligned}
$$

Noise

P.d.f. for noise: Complex Gaussian

$$\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{n}; \boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

Beamforming

- Likelihood function

$$p(\boldsymbol{y}|s) = \frac{1}{\det(\pi\sigma^2\boldsymbol{I})} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{a}s\|^2\right)$$

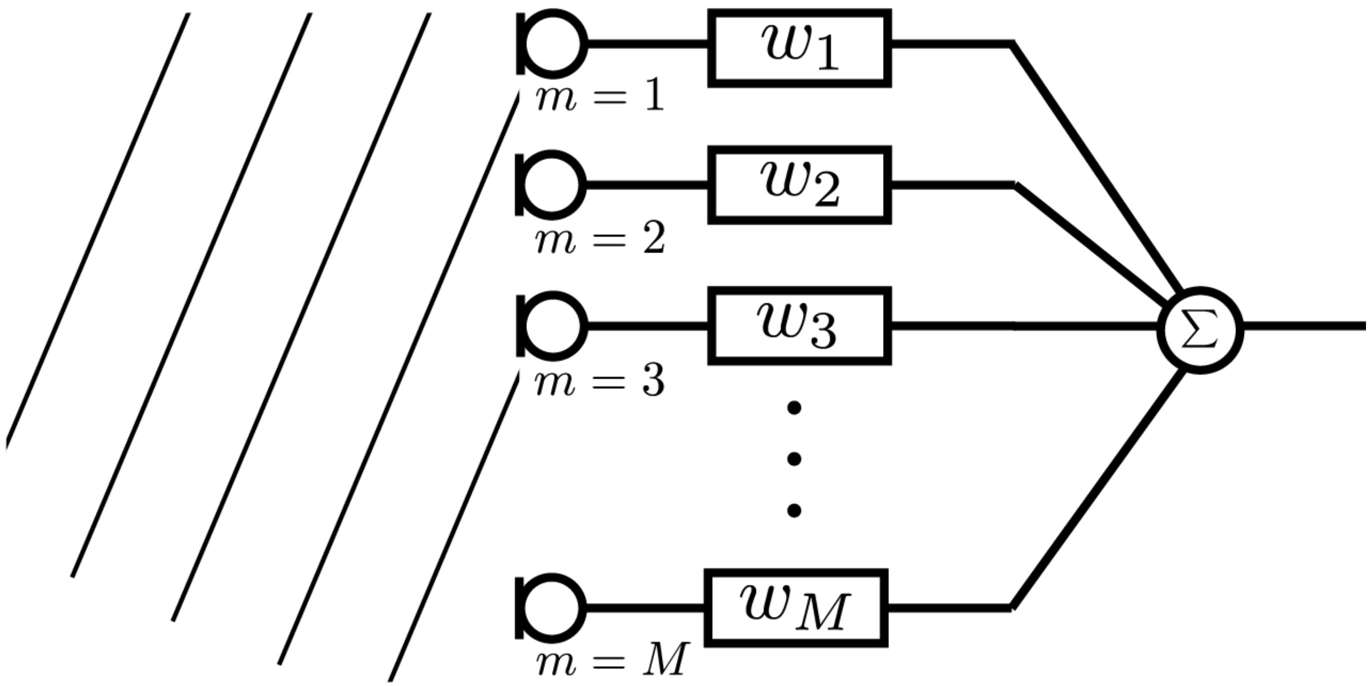
- ML estimate of target source signal

$$\begin{aligned}\hat{s} &= \arg\max_s p(\boldsymbol{y}|s) \\ &= \arg\min_s \|\boldsymbol{y} - \boldsymbol{a}s\|^2 \\ &= \frac{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{y}}{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{a}}\end{aligned}$$

➡ Corresponds to delay-and-sum beamformer

Delay-and-sum beamformer

- Time-delay of each mic is compensated, then in-phase signals are summed up to enhance sound from target direction



Array manifold vector: $\boldsymbol{a} = \left[\mathrm{e}^{-\mathrm{j}\omega\tau_1}, \quad \ldots, \quad \mathrm{e}^{-\mathrm{j}\omega\tau_M} \right]^{\mathsf{T}}$

Linear prediction

- Linear prediction: Predicting current signal value by linear combination of past signal values
- When using L samples of past signal values,

$$\hat{x}[n] = \sum_{l=1}^{L} h_l x[n-l]$$

- Expected value of square prediction error E_L is formulated as

$$\begin{aligned} E_L &= E[(x[n] - \hat{x}[n])^2] \\ &= E\left[\left(x[n] - \sum_{l=1}^{L} h_l x[n-l]\right)^2\right] \end{aligned}$$

➡ MMSE estimation to obtain h_l

Linear prediction

- Need to obtain coefficients h_l so that E_L is minimized. Such coefficients are obtained by setting differential of E_L by h_l to 0 as

$$
\begin{aligned}
\frac{\partial E_L}{\partial h_k} &= E\left[-2x[n-k]\left(x[n]-\sum_{l=1}^{L} h_l x[n-l]\right)\right] \\
&= 2\sum_{l=1}^{L} h_l E\left[x[n-k]x[n-l]\right] - 2E\left[x[n-k]x[n]\right] \\
&= 0
\end{aligned}
$$

- By denoting autocorrelation function of $x[n]$ as $\varphi[k]=\varphi[-k]=E[x[n]x[n-k]]$

$$
\sum_{l=1}^{L} h_l \varphi[l-k] = \varphi[k]
$$

Linear prediction

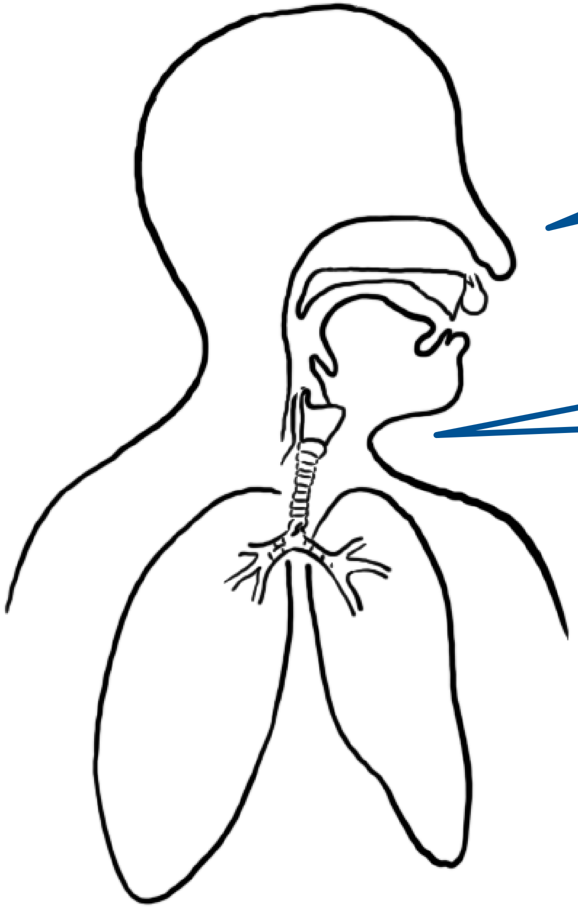
- Matrix-vector representation

$$\begin{bmatrix} \varphi[0] & \varphi[1] & \cdots & \varphi[L-1] \\ \varphi[1] & \varphi[2] & \cdots & \vdots \\ \vdots & \ddots & \ddots & \varphi[1] \\ \varphi[L-1] & \cdots & \varphi[1] & \varphi[0] \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_L \end{bmatrix} = \begin{bmatrix} \varphi[1] \\ \varphi[2] \\ \vdots \\ \varphi[L] \end{bmatrix}$$

- Optimal linear prediction coefficients are obtained by solving the above equation, which is called Yule-Walker equations
- Since the left side is Toeplitz matrix, an efficient technique, which is called Levinson—Durbin algorithm, can be used
- This signal model is also called auto-regressive (AR) model

Speech generation model

- Vocal tract system



Vocal tract characteristics
→ Spectral shape

Opening-and-closing of vocal folds
→ Cyclic pulse train

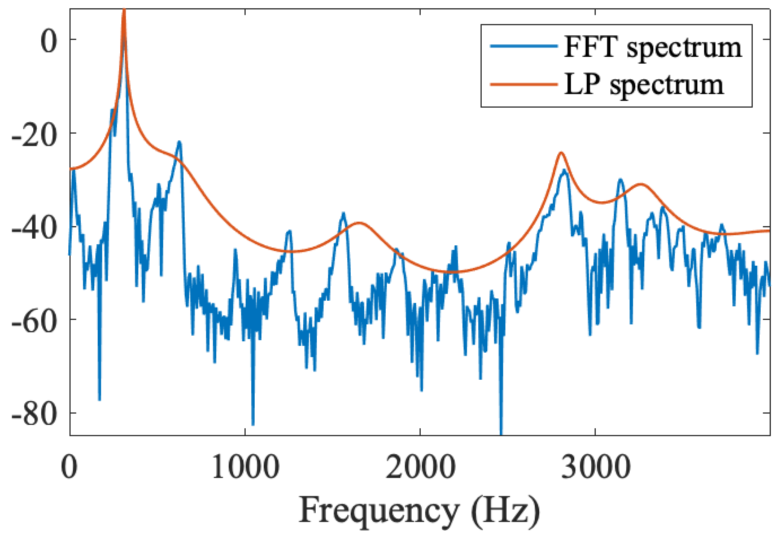
Speech can be modeled by convolution of cyclic pulse train and vocal tract filter

➡ <u>Source-filter model</u>

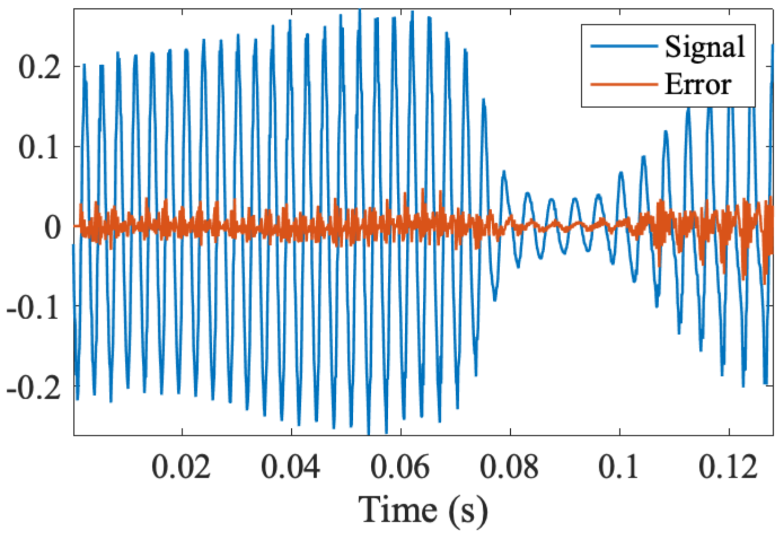
Speech generation model

➢ Vocal tract characteristics is well approximated by linear prediction

Spectral envelope obtained by linear prediction



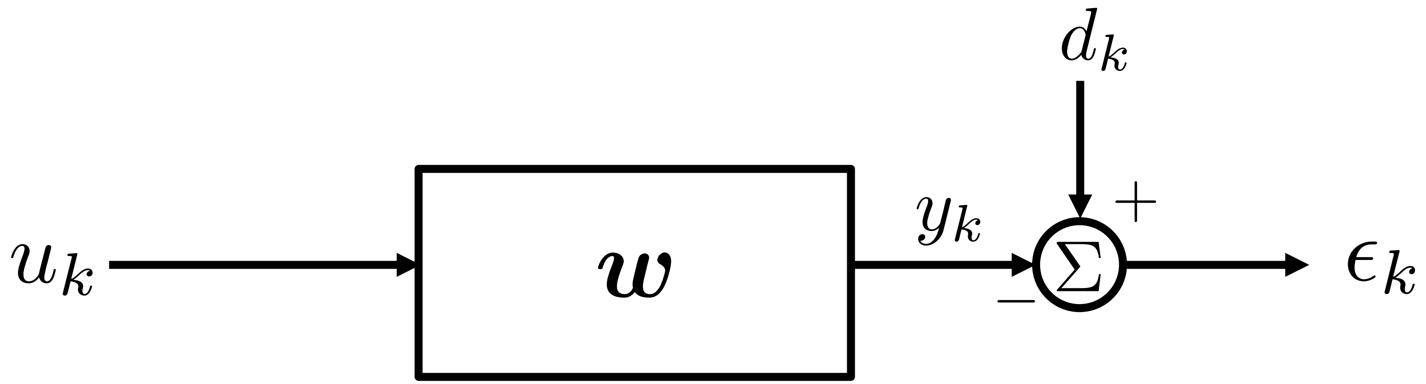
Residual error between original speech and linear prediction



Widely used for speech coding for data compression
(cf. code excited linear prediction coder: CELP)

Adaptive filter

- Objective of adaptive filter is to extract the desired signal by statistical learning of observed signal
- Obtaining filter $\boldsymbol{w} = [w_1, \ldots, w_K]^\mathsf{T}$ so that output signal y_k of input u_k corresponds to the desired signal d_k



Wiener filter

- Wiener filter: Applying MMSE estimation to adaptive filtering under the constraint of linear estimator
- Filter coefficients and input signal vectors are defined as

$$\boldsymbol{w} = [w_1, \ldots, w_K]^{\mathsf{T}}$$
$$\boldsymbol{u}_k = [u_k, u_{k-1}, \ldots, u_{k-K+1}]^{\mathsf{T}}$$

- Then, filter convolution is described by inner product

$$y_k = \boldsymbol{w}^{\mathsf{T}} \boldsymbol{u}_k$$
$$= \sum_{i=1}^{K} w_i u_{k-i+1}$$

Wiener filter

- Estimation error is represented by using known desired signal d_k as

$$\epsilon_k = d_k - y_k$$

- Cost function for Wiener filter is defined as mean square error

$$\begin{aligned}\mathcal{J} &= E[|\epsilon_k|^2] \\ &= E\left[(d_k - \boldsymbol{w}^\top \boldsymbol{u}_k)(d_k - \boldsymbol{w}^\top \boldsymbol{u}_k)^\top\right] \\ &= E[d_k^2] - \boldsymbol{w}^\top E[\boldsymbol{u}_k d_k] - E[d_k \boldsymbol{u}_k^\top]\boldsymbol{w} + \boldsymbol{w}^\top E[\boldsymbol{u}_k \boldsymbol{u}_k^\top]\boldsymbol{w} \\ &= \sigma_d^2 - \boldsymbol{w}^\top \boldsymbol{r}_{ud} - \boldsymbol{r}_{du}\boldsymbol{w} + \boldsymbol{w}^\top \boldsymbol{R}_u \boldsymbol{w}\end{aligned}$$

where $\begin{cases} \sigma_d^2 := E[d_k^2]\,, \quad \boldsymbol{R}_u := E[\boldsymbol{u}_k \boldsymbol{u}_k^\top] \\ \boldsymbol{r}_{ud} := E[\boldsymbol{u}_k d_k]\,, \ \boldsymbol{r}_{du} := E[d_k \boldsymbol{u}_k] \end{cases}$

Wiener filter

- By differentiating cost function by $\boldsymbol{w}$

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}} = 2(-\boldsymbol{r}_{ud} + \boldsymbol{R}_u \boldsymbol{w})$$

- By setting the above derivative to 0, the following normal equation (or Wiener—Hopf equation) is obtained

$$\boldsymbol{R}_u \boldsymbol{w} = \boldsymbol{r}_{ud}$$

- Then, optimal filter is obtained as

$$\hat{\boldsymbol{w}} = \boldsymbol{R}_u^{-1} \boldsymbol{r}_{ud}$$

Adaptive algorithm is practically used for real-time adaptation

Speech enhancement

- Speech enhancement from noisy observation
- (Non-causal) Wiener filtering in frequency domain

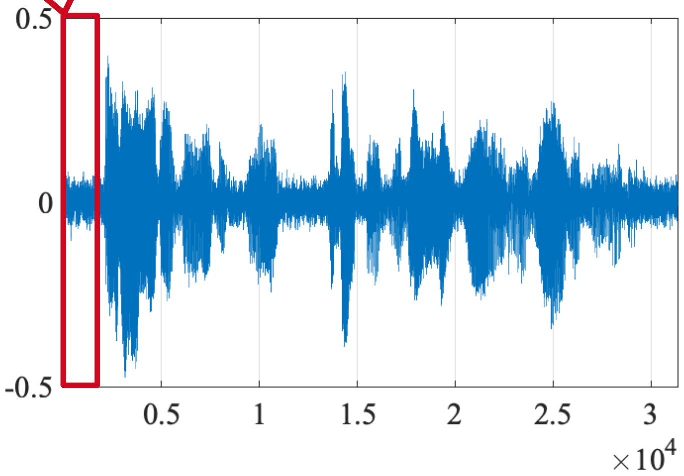
$$W(\omega) = \frac{S(\omega)}{S(\omega) + N(\omega)}$$

Power spectral density of original signal

Power spectral density of noise

SNR is estimated in non-speech section

Noisy observation



Enhanced speech