Estimation

Nonlinear Estimation

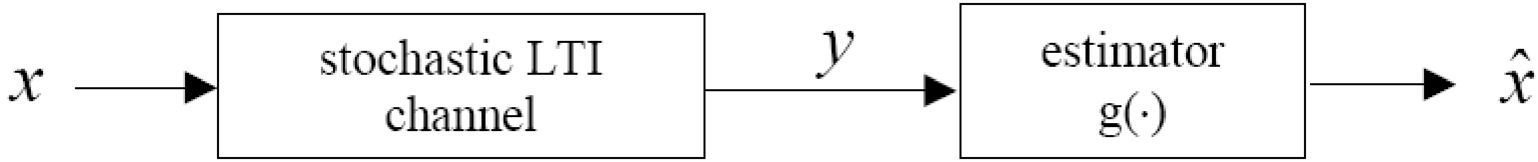
Linear & Polynomial Estimation

Minimum Mean Square Error Estimation:

Wiener Filtering

Estimation Problem

- Given:



  - Source generating nonobservable stochastic signals X
    - e.g., signal + noise, bits, characters, ...
  - Stochastic LTI channel (extendable to LTV)
    - e.g., $1/f$ noise from resistor, ionospheric propagation, ...
  - Observable outputs of channel, Y
    - constitutes a <u>new</u> random variable, because <u>stochastic</u> channel
- Find:
  - "best" estimate $\hat{X}$ of X, based on Y

Optimal Estimation

- Assume:
  - If we know the distribution of *X*, $f_X(x)$, then we do not know *Y* exactly, but we know its conditional distribution (=forward probability), $f_{Y|X}\left(y \mid X = x\right)$
    - NB: $f_{X|Y}(x \mid Y = y)$ (=backward probability) is unknown!
- Then:
  - We know the joint distribution of *X* and *Y*:

$$f_{X,Y}(x, y) = f_{Y|X}(y \mid X = x) \cdot f_X(x)$$

- Now: design optimal estimator for *X*
  - i.e., design system that out of all possible transfer functions *g*(.) it selects the one $g_{\mathrm{opt}}(.)$ for which $\hat{x} = g_{\mathbf{opt}}(y)$ is “closest” to *x*

Optimal MMSE Estimator

- Criterion: minimise the mean squared deviation

$$\mathrm{E}[(X-\hat{X})^2] = \iint [x - g(y)]^2 \, f_{X,Y}(x,y)\, dx\, dy$$

$$= \int dy\, f_Y(y) \int [x-g(y)]^2 \, f_{X|Y}(x \,|\, Y=y)\, dx$$

$$\doteq \int f_Y(y) K(y) dy$$

- Since $K(y) \geq 0$, we should minimise it:

$$K(y) = \int [x-g(y)]^2 \, f_{X|Y}(x \,|\, Y=y)\, dx$$

$$= \mathrm{E}(X^2 \,|\, Y=y) - 2g(y)\mathrm{E}(X \,|\, Y=y) + g^2(y)$$

- Thus, $\frac{dK(y)}{dg} = 0 \quad \Leftrightarrow \boxed{g(y) = \mathrm{E}(X \,|\, Y=y)}$

yields best (MMSE) estimate

Optimal MMSE Estimation Error

- Associated mean squared deviation (error) for this choice of *g*(.):

$$\varepsilon_{\mathrm{opt}}^2 = E(X^2) - \int E^2(X \mid Y = y)\, f_Y(y)\, dy$$

- Using Schwartz inequality $\Rightarrow$ $0 \leq \varepsilon_{\mathrm{opt}}^2 \leq \sigma_X^2$
- Optimal estimator is in general nonlinear

Optimal MMSE Estimation

- Why MMSE?
  - With this choice: estimation error is <u>orthogonal</u> to $g(y)$
    - Proof: projection of MMSE error onto $g(y)$:

$$\iint [x - \mathrm{E}(X \mid y)] g(y)\, f_{X,Y}(x,y)\, dx\, dy = \iint x\, g(y)\, f_{X,Y}(x,y)\, dx\, dy - \iint \mathrm{E}(X \mid y)\, g(y)\, f_{X,Y}(x,y)\, dx\, dy$$

$$= \iint x\, g(y)\, f_{X,Y}(x,y)\, dx\, dy - \int \mathrm{E}(X \mid y) g(y)\, f_Y(y)\, dy$$

$$= \iint x\, g(y)\, f_{X,Y}(x,y)\, dx\, dy - \int x\, f_{X|Y}(x \mid y) dx \int g(y)\, f_Y(y)\, dy$$

$$= \iint x\, g(y)\, f_{X,Y}(x,y)\, dx\, dy - \iint x\, g(y)\, f_{X,Y}(x,y)\, dx\, dy$$

$$= 0$$

Optimal Estimation

- Special cases:
  - (i) X and Y are statistically independent
    - i.e., observation of Y does not teach us anything about X:

$$g(y) = \mathrm{E}(X \mid Y = y) = \mathrm{E}(X), \qquad \varepsilon_{\mathrm{opt}}^2 = \sigma_X^2$$

    - Corollary: if X and Y are dependent, then optimal estimate of X after observation of Y yields MSE that is never larger than if we choose $\hat{x} = \mathrm{E}(x)$ without observing Y
  - (ii) Deterministic lossless channel: $Y = \phi(X), \quad X = \phi^{-1}(Y)$

$$\hat{X} = \phi^{-1}(Y) = X, \qquad \varepsilon_{\mathrm{opt}}^2 = 0$$

Linear Estimator: Wiener Filter

- Optimal estimator often difficult to realise
- Suboptimal: approximate $\mathrm{E}(X \mid Y = y)$ by best linear estimator: $\boxed{\hat{x} = \alpha y + \beta}$
  - Then: $\begin{cases} \dfrac{\partial \mathrm{E}[(X - \alpha Y - \beta)^2]}{\partial \alpha} = 0 \\ \dfrac{\partial \mathrm{E}[(X - \alpha Y - \beta)^2]}{\partial \beta} = 0 \end{cases}$ yields $\boxed{\begin{cases} \alpha = r \dfrac{\sigma_X}{\sigma_Y}, \\ \beta = \mathrm{E}(X) - \alpha \mathrm{E}(Y) \end{cases}}$ $r = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$

$$\boxed{\varepsilon^2_{\min} = \sigma^2_X (1 - r^2)}$$

- Using Schwartz inequality (do!) $\Rightarrow$ $\varepsilon^2_{\mathrm{opt}} \le \varepsilon^2_{\min} \le \sigma^2_X$

Optimal Linear Estimation

- For Gaussian *X* and *Y* (do!):

$$\mathrm{E}(X \mid Y = y) = r\frac{\sigma_X}{\sigma_Y}\left[y - \mathrm{E}(Y)\right] + \mathrm{E}(X)$$

i.e., linear!

Thus, linear estimate for <u>Gaussian</u> random variable is its <u>optimal</u> estimate

Polynomial Estimation

- Extension of suboptimal linear estimator to quadratic:

$$\hat{x} = \alpha y^2 + \beta y + \gamma$$

i.e., a linear problem of solving for α, β, γ

$$\begin{cases} \alpha\langle Y^4\rangle + \beta\langle Y^3\rangle + \gamma\langle Y^2\rangle & = & \langle X\,Y^2\rangle \\ \alpha\langle Y^3\rangle + \beta\langle Y^3\rangle + \gamma\langle Y\rangle & = & \langle X\,Y\rangle \\ \alpha\langle Y^2\rangle + \beta\langle Y\rangle + \gamma & = & \langle X\rangle \end{cases}$$

Now, 3rd- and 4th-order moments are needed

For Gaussian *X*, *Y*: 3rd-order moments =0; 4th-order moments calculable from 2nd-order moments via Isserlis's theorem

Optimum LSE Filters

Minimum Mean Square Error Estimation:

IIR Wiener Filtering

Optimum LSE Filters

Innovations Process & Whitening Filter

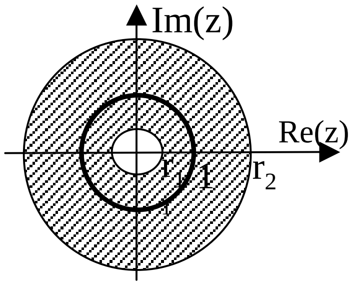
Innovations Process

- Problem statement:
  - Find:
    - (I) linear causal filter such that a given WSS random process $x(t)$ as input yields a white noise output signal: <u>whitening filter</u>
    - (II) (*inverse problem*) linear causal filter such that white noise input process produces a given WSS random process $x(t)$ as its output ("correlator")
  - We shall show: transfer functions of both filters are each other's inverse
  - Practical significance: non-ideal WSS process (non-vanishing correlation) can be represented as a <u>linearly filtered</u> ideal (white) WSS process

Innovations Process

- Define:
  - $\{x(n)\}$: wide-sense stationary (WSS) random process
  - $\{\gamma_{xx}(m)\}$: autocorrelation sequence of $\{x(n)\}$
  - $\Gamma_x(f)$: power spectrum of $\{x(n)\}$

- Assume $\ln \Gamma_x(z)$ is analytic (holomorph) in an annular region $r_1 < 1 < r_2$ in the z-plane comprising the unit circle $|z| = 1$



Innovations Process

- Parenthetically, Wiener-Khinchine theorem (for z-transforms):

$$\Gamma_x(z) = \sum_{m=-\infty}^{+\infty} \gamma_{xx}(m)\, z^{-m}$$

- Expansion of $\ln\Gamma_x(z)$ in Laurent series w.r.t. z=0 (generalization of Taylor series in z-plane): $\ln\Gamma_x(z) = \sum_{m=-\infty}^{+\infty} v(m)\, z^{-m}$
  - Alternative interpretation of $\ln\Gamma_x(z) = \sum_{m=-\infty}^{+\infty} v(m)\, z^{-m}$: the function $\ln\Gamma_x(z)$ is the z-transform of $\{v(m)\}$

Innovations Process

- For $\Gamma_x(f)$ real & even: Fourier coefficients are even:

$$v(m) = \int_{-0.5}^{0.5} \left(\ln \Gamma_x(f)\right) \exp\left(j2\pi f m\right) df = v(-m)$$

- Factorisation:

$$\ln \Gamma_x(z) = \sum_{m=-\infty}^{+\infty} v(m)\, z^{-m}$$

$$= v(0) + \sum_{m'=1}^{+\infty} v(-m') \left(z^{-1}\right)^{-m'} + \sum_{m=1}^{+\infty} v(m)\, z^{-m}$$

- Thus,

$$= \ln \sigma_w^2 + \sum_{m'=1}^{+\infty} v(m') \left(z^{-1}\right)^{-m'} + \sum_{m=1}^{+\infty} v(m)\, z^{-m}$$

$$\boxed{\Gamma_x(z) = \exp\left(\sum_{m=-\infty}^{+\infty} v(m)\, z^{-m} \right) = \sigma_w^2\, H(z^{-1})\, H(z)}$$

where

$$H(z) = \exp\left(\sum_{m=1}^{+\infty} v(m)\, z^{-m} \right)$$

- $\{v(m)\}$ = cepstrum of $\{\gamma_{xx}(m)\}$ and $\Gamma_x(f)$

Innovations Process

- Interpretation:
  - $H(z) = \exp\left(\sum_{m=1}^{+\infty} v(m)\, z^{-m}\right)$ for $r_1 \le |z| < 1$: causal response
  - $H(z^{-1}) = \exp\left(\sum_{m=1}^{+\infty} v(m) \left(z^{-1}\right)^{-m}\right)$ for $1 < |z| \le r_2$: non-causal
  - In $r_1 \le |z| < 1$, $H(z)$ is analytic (no poles) with Laurent series reducing to a Taylor series:

$$H(z) = \sum_{n=0}^{+\infty} h(n)\, z^{-n}$$

  - If $H(z)$ is analytic in $r_1 \le |z| < 1$, then $H(z^{-1})$ is analytic in $1 < |z| \le 1/r_1$

Innovations Process

- Further interpretation:
  - From $\Gamma_x(z) = \sigma_w^2 H(z) H(z^{-1})$, i.e., $\Gamma_x(f) = \sigma_w^2 H(f) H^*(f) = \sigma_w^2 |H(f)|^2$ this represents a linear filter $H(z)$ having input $\{w(n)\}$ with power spectral density $\Gamma_w(f) = \sigma_w^2$, output $\{x(n)\}$ with power spectral density $\Gamma_x(f)$
  - Since $H(z)$ is causal, $\sigma_w^2 = \Gamma_x(f)/|H(f)|^2$ represents a linear filter $1/H(z)$ with input $\{x(n)\}$ and as output the innovation process $\{w(n)\}$: whitening filter $1/H(z)$

$$w(n) \longrightarrow \boxed{H(z)} \longrightarrow x(n) \qquad\qquad x(n) \longrightarrow \boxed{1/H(z)} \longrightarrow w(n)$$

Optimum LSE Filters

IIR Wiener Filter

IIR Wiener Filter

- MMSE problem:

$$E\left(\left|e(n)\right|^2\right) = E\left(\left|d(n) - \sum_{k=0}^{+\infty} h(k)\,x(n-k)\right|^2\right)$$

  - Filter cffs. $h(k)$ to be chosen as solutions of infinite linear system (Wiener-Hopf equation)

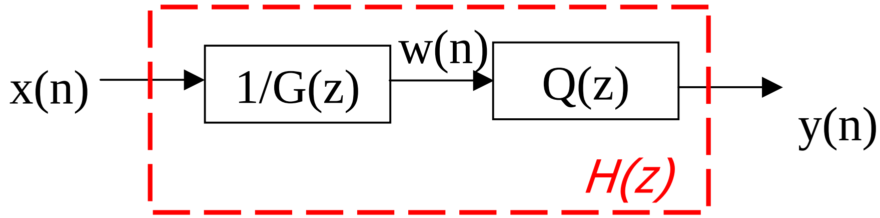
$$\sum_{k=0}^{+\infty} h(k)\,\gamma_{xx}(m-k) = \gamma_{dx}(m), \qquad m \geq 0$$

  - Problem: system cannot be solved with z-transform for $m \geq 0$ only (Wiener-Khinchine N/A)

  - Solution: use auxiliary equivalent representation of input $x(n)$, which transforms it into sequence defined over $-\infty < m < +\infty$ to which z-transform can be applied: *innovations representation*

IIR Wiener Filter

- Innovations representation of $x(n)$:
  - Idea: Cascading whitening filter $1/G(z)$ with second filter $Q(z)$ such that cascade is optimum Wiener filter $H(z)$ for $x(n)$:



$$\Gamma_x(z) = \sigma_w^2 G(z) G(z^{-1})$$

$$y(n) = \sum_{k=0}^{+\infty} q(k) w(n-k) = \sum_{k=0}^{+\infty} h(k) x(n-k), \qquad H(z) = Q(z)/G(z)$$

$$\sum_{k=0}^{+\infty} q(k) \gamma_{ww}(m-k) = \gamma_{dw}(m), \qquad m \geq 0$$ (Wiener-Hopf)

  - Since $\gamma_{ww}(m-k) = \sigma_w^2 \delta(m-k)$, filter cffs. are

$$\boxed{q(m) = \frac{\gamma_{dw}(m)}{\sigma_w^2}, \qquad m \geq 0}$$

IIR Wiener Filter

- Representation of $\gamma_{dw}(m)$ in terms of $\gamma_{dx}(.)$:

$$x(n) \longrightarrow \boxed{V(z)=1/G(z)} \longrightarrow w(n)$$

$$w(n) = \sum_{k=0}^{+\infty} v(k)\, x(n-k), \qquad V(z) = \sum_{k=0}^{+\infty} v(k)\, z^{-k} = \frac{1}{G(z)}$$

hence

$$\begin{aligned} \gamma_{dw}(k) &= E[d(n)\, w^*(n-k)] \\ &= \sum_{m=0}^{+\infty} v(m)\, E[d(n)\, x^*(n-k-m)] \\ &= \sum_{m=0}^{+\infty} v(m)\, \gamma_{dx}(k+m) \end{aligned}$$

IIR Wiener Filter

- z-transformation:

$$\Gamma_{dw}(z) = \sum_{k=-\infty}^{+\infty} \gamma_{dw}(k)\, z^{-k} = \sum_{k=-\infty}^{+\infty} \left[\sum_{m=0}^{+\infty} v(m)\, \gamma_{dx}(k+m) \right] z^{-k}$$

$$= \sum_{m=0}^{+\infty} v(m) \sum_{k=-\infty}^{+\infty} \gamma_{dx}(k+m)\, z^{-k}$$

$$= \sum_{m=0}^{+\infty} v(m)\, z^{+m} \sum_{k+m=-\infty}^{+\infty} \gamma_{dx}(k+m)\, z^{-(k+m)}$$

$$= V\left(z^{-1}\right) \Gamma_{dx}(z) \equiv \frac{\Gamma_{dx}(z)}{G(z^{-1})}$$

- Extracting causal part:

$$\Gamma_{dw}^{(+)}(z) = \sum_{k=0}^{+\infty} \gamma_{dw}(k)\, z^{-k} \equiv \left[\frac{\Gamma_{dx}(z)}{G(z^{-1})} \right]^{(+)}$$

⇒ Optimum IIR Wiener filter: (causal)

$$\boxed{H(z) = \frac{1}{\sigma_w^2 G(z)} \left[\frac{\Gamma_{dx}(z)}{G(z^{-1})} \right]^{(+)}}$$