Q1. Discuss the evaluation of covariance matrices in relation to Kalman and extended Kalman filtering

Problem Setup

- The KF assumes the following dynamic state-space model:

$$\begin{aligned}\text{Process equation:} \quad x_k &= F_k x_{k-1} + v_{k-1} \\ \text{Measurement equation:} \quad z_k &= H_k x_k + w_k\end{aligned}$$

- Assumptions:
  - Additive uncorrelated Gaussian noise sequences with known statistics. i.e., $v_k \sim \mathcal{N}(0, Q_{k-1})$ and $w_k \sim \mathcal{N}(0, R_k)$
  - Known initial estimate. i.e., $x_0 \sim \mathcal{N}(\hat{x}_{0|0}, P_{0|0})$
- Objective: Estimate the state at time k recursively given $\{z_1, z_2, \ldots z_k\}$.
- Performance criteria: Minimum mean-square error.

Two Basic Operations

- Predict:

$$
\begin{aligned}
\hat{x}_{k|k-1} &= F_k \hat{x}_{k-1|k-1} \\
P_{k|k-1} &= F_k P_{k-1|k-1} F_k^T + Q_{k-1}
\end{aligned}
$$

- Correct:

$$
\begin{aligned}
\hat{z}_{k|k-1} &= H_k \hat{x}_{k|k-1} \\
S_{k|k-1} &= H_k P_{k|k-1} H_k^T + R_k \\
W_k &= P_{k|k-1} H_k^T S_{k|k-1}^{-1} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + W_k (z_k - \hat{z}_{k|k-1}) \\
P_{k|k} &= P_{k|k-1} - W_k H_k P_{k|k-1}
\end{aligned}
$$

Error Covariances: Analysis

- As a self-assessment of its own errors, the KF yields the error covariance matrices:
  - state $\Rightarrow$predicted and posterior error covariances
  - measurement $\Rightarrow$ innovation covariance
- Why we evaluate covariances?
  - To verify the credibility of the filter: if the actual error is consistent with the filter-computed error?
  - To compare various filter performances
  - To probe into modeling errors

Error Covariances (Cont'd)

- Tools for evaluation:
  - Mean-Squared Error (MSE)
  - Posterior Cramer-Rao Lower Bound (PCRLB)
  - Normalized Innovation-Squared (NIS)
  - Normalized Estimation Error-squared(NEES)
- The PCRLB and the NIS can be used in real-time applications.

Mean-Squared Error

- Given the true state x_k, the MSE (matrix) of the filter estimate is defined by

$$\text{MSE}(k) \;=\; \mathbb{E}((x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T)$$

- The gain-posterior covariance relationship: $W_k = P_{k|k} H^T R^{-1}$.
- 3 practical cases:
  - $P_{k|k} = \text{MSE}(k) \Rightarrow$ optimal
  - $P_{k|k} > \text{MSE}(k) \Rightarrow$ pessimistic
  - $P_{k|k} < \text{MSE}(k) \Rightarrow$ optimistic
- MSE = variance + bias-squared (in a scalar case).

PCRLB

- The covariance matrix $P_{k|k}$ of an unbiased state estimator $\widehat{\mathbf{x}}_{k|k}$ has a lower bound

$$P_{k|k} \quad \succeq \quad J_k^{-1}$$

where the Fisher information matrix

$$J_k = D_{k-1}^{22} - D_{k-1}^{21}\left(J_{k-1} + D_{k-1}^{11}\right)^{-1} D_{k-1}^{12} \qquad (k > 0)$$

where

$$\begin{aligned}
D_{k-1}^{11} &= -E\left\{\nabla_{\mathbf{x}_{k-1}}\left[\nabla_{\mathbf{x}_{k-1}} \ln p(\mathbf{x}_k|\mathbf{x}_{k-1})\right]^T\right\} \\
D_{k-1}^{21} &= -E\left\{\nabla_{\mathbf{x}_{k-1}}\left[\nabla_{\mathbf{x}_k} \ln p(\mathbf{x}_k|\mathbf{x}_{k-1})\right]^T\right\} \\
D_{k-1}^{12} &= -E\left\{\nabla_{\mathbf{x}_k}\left[\nabla_{\mathbf{x}_{k-1}} \ln p(\mathbf{x}_k|\mathbf{x}_{k-1})\right]^T\right\} = \left[D_{k-1}^{21}\right]^T \\
D_{k-1}^{22} &= -E\left\{\nabla_{\mathbf{x}_k}\left[\nabla_{\mathbf{x}_k} \ln p(\mathbf{x}_k|\mathbf{x}_{k-1})\right]^T\right\} - E\left\{\nabla_{\mathbf{x}_k}\left[\nabla_{\mathbf{x}_k} \ln p(\mathbf{z}_k|\mathbf{x}_k)\right]^T\right\}
\end{aligned}$$

PCRLB for the KF

- For the LG case, information matrices

$$\begin{aligned} D^{11}_{k-1} &= F^T_{k-1} Q^{-1}_{k-1} F_{k-1} \\ D^{12}_{k-1} &= [D^{21}_{k-1}]^T = -F^T_{k-1} Q^{-1}_{k-1} \\ D^{22}_{k-1} &= Q^{-1}_{k-1} + H^T_k R^{-1}_k H_k \end{aligned}$$

- Hence we get the Fisher information matrix at time k as

$$\begin{aligned} J_k &= Q^{-1}_{k-1} - Q^{-1}_{k-1} F_{k-1} \left(J_{k-1} + F^T_{k-1} Q^{-1}_{k-1} F_{k-1} \right)^{-1} F^T_{k-1} Q^{-1}_{k-1} + J^z_k \\ &= \left(Q_{k-1} + F_{k-1} J^{-1}_{k-1} F^T_{k-1} \right)^{-1} + H^T_k R^{-1}_k H_k \end{aligned}$$

- Obtain the same when replacing J_k with $P^{-1}_{k|k}$ $\Rightarrow$ KF is an efficient estimator for a linear-Gaussian system.

Credibility Check on Innovation Covariance

- Properties of innovation sequences
  - zero-mean Gaussian
  - uncorrelated.
- Define normalized innovation-squared (NIS):

$$\epsilon(k) = \nu_k^T S_{k|k}^{-1} \nu_k$$

where the innovation and its covariance

$$\nu_k = (z_k - \hat{z}_{k|k-1})$$

$$S_{k|k-1} = \mathrm{cov}(\nu_k)$$

- Distribution of $\epsilon(k)$:

$$\epsilon(k) \sim \chi_m^2$$

where m is the measurement-vector dimension or the dof.

Credibility Check (Cont'd)

- Postulate the null hypothesis

$$H_0 : \quad \mathbb{E}[\epsilon(k)] \quad = \quad m.$$

- Accept both the innovation and its covariance commensurate with theoretical results if

$$\epsilon(k) \quad \in \quad [r_1, \quad r_2]$$

- The acceptance interval $[r_1 \quad r_2]$ is determined such that

$$P(\epsilon(k) \in [r_1 \quad r_2] | H_0) \quad = \quad 1 - \alpha$$

where α is the level of significance.

- For $\alpha = 0.05$, the limits

$$r_{1,2} \quad \approx \quad \frac{1}{2m}\big(\pm 1.96 + \sqrt{2m-1}\big)^2$$

Remarks

- For an unbiased estimator, the NIS check is directly comparable to a credibility check on the innovation covariance.
- In an N Monte Carlo runs, we use the averaged NIS statistics.

Credibility Check on Posterior Error Covariance

- Define normalized estimation error-squared (NEES):

$$\epsilon(k) = \tilde{x}_k^T P_{k|k}^{-1} \tilde{x}_k$$

where the estimation error

$$\tilde{x}_k = (x_k - \hat{x}_{k|k})$$

- Distribution of $\epsilon(k)$:

$$\epsilon(k) \sim \chi_n^2$$

where n is the state-vector dimension.

- Following a similar procedure as in the NIS case, we may check the credibility of filter-estimated covariance at an α level.

Extended Kalman Filters

- The EKF assumes the following dynamic state-space model:

$$\text{Process equation:} \quad x_k = f(x_{k-1}) + v_{k-1} \quad (1)$$

$$\text{Measurement equation:} \quad z_k = h(x_k) + w_k \quad (2)$$

- Idea: Linearize nonlinear functions using the first-order Taylor series expansion.

- Let $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$. Then we write y, where $y = f(x)$ as

$$y = f(x) \approx f(\bar{x}) + \underbrace{F(x - \bar{x})}_{\text{linear}},$$

where the Jacobian

$$F = [\nabla_x f(x)^T]^T_{x=\bar{x}}.$$

EKF (Cont'd)

- Approximates y to be Gaussian with the following mean and covariance:

$$\begin{aligned}\bar{y} &= \mathbb{E}(f(x)) \\ &\approx f(\bar{x}) \\ \Sigma_y &\approx F\Sigma_x F^T.\end{aligned}$$

Final Remarks

- For a nonlinear system, all conditional densities are non-Gaussian. However all the above methods assume Gaussianity suggesting that they wont fully characterize a nonlinear filter accuracy
- EKF estimate is biased $\Rightarrow P_{k|k} <$ MSE
- To meet the condition of zero-mean error (unbiased mean error), we subtract off mean errors before applying the NIS or NEES.
- In the EKF case, the information matrices of the CRLB are given by:

$$\begin{aligned} D_{k-1}^{11} &= \mathbb{E}(F_{k-1}^T Q_{k-1}^{-1} F_{k-1}) \\ D_{k-1}^{12} &= [D_{k-1}^{21}]^T = -\mathbb{E}(F_{k-1}^T) Q_{k-1}^{-1} \\ D_{k-1}^{22} &= Q_{k-1}^{-1} + \mathbb{E}(H_k^T R_k^{-1} H_k) \end{aligned}$$

  where F and H are Jacobians of the state and measurement functions.

- The expectation operators are replaced by Monte Carlo averages

(useful in simulations!)

References

- Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with applications to target tracking* , Wiley, 2001.
- D. Simon *Optimal state estimation,* Wiley 2006.

Thank you!

Q2. Discuss solution of ill-posed systems of equations with regard to methods

Well (ill)-Posed Problems

- The quality of solution depends on
  - the problem itself and
  - the computer
- According to Hadamard, a problem is well-posed if the solution
  - exists
  - is unique and
  - depends continuously on the data (stable).
- Typically practical inverse problems are all ill-posed.
- Even well-posed problems may be unstable or ill-conditioned when implemented in digital computers.

Solutions for Linear Systems

- A linear system of equations in a matrix form:

$$Ax = b$$

where the coefficient matrix $A \in \mathbb{R}^{m \times n}$, the constant vector $b \in \mathbb{R}^m$ and the variable vector $x \in \mathbb{R}^n$.

- Suppose $m = n$. The solution

$$\hat{x} = A^{-1}b$$

may be disastrous especially when n is large.

- Types of the solution:
  - No solution
  - Multiple solutions
- Solving methods: decompositions and regularization

Cholesky Decomposition

- Suppose the matrix A is
  - Symmetric and
  - Positive definite
- Decompose A into a unique lower and upper triangular matrices:

$$A = LL^T$$

- On the LHS, we have

$$A\hat{x} = LL^T\hat{x} = L(L^T)\hat{x}$$

- Solve by the forward and backward substitutions:

$$\begin{aligned} Ly &= b \\ L^T\hat{x} &= y \end{aligned}$$

- fast, stable and requires less space!

Truncated SVD

- Decomposes any matrix A as

$$A = UDV^T,$$

where U and V are orthogonal matrices such that $U^TU = V^TV = I$; D is diagonal with singular values of A.

- If A is non-singular, we write

$$A^{-1} = V\Sigma^{-1}U^T,$$

where $\Sigma^{-1} = [\text{diag}(\frac{1}{\sigma_i})]$.

Overdetermined Systems

- More equations than unknowns
- b does not lie in $\mathscr{R}(A) \Rightarrow$ No solution
- Yields the unique solution by minimizing the residual $||Ax - b||_2$.
- Using the SVD, we write

$$\begin{aligned} \min ||Ax - b|| &= \min ||U\Sigma V^T - b|| \\ &= \min ||\Sigma V^T x - U^T b|| \\ &= \min ||\Sigma v - \tilde{b}||, \end{aligned}$$

where $v = V^T x$ and $\tilde{b} = U^T b$.

- The min. length solution for v is

$$v = \Sigma^+ \tilde{b}.$$

- Hence

$$\hat{x} \;=\; V\Sigma^{+}\tilde{b} = V\Sigma^{+}U^{T}b.$$

- Summary:
  - Compute the SVD of A : $\quad A = U\Sigma V^{T}$
  - Zero-out 'small' σ_i's of Σ.
  - Obtain

$$\hat{x} \;=\; V\Sigma^{+}(U^{T}b),$$

where $\Sigma^{+} = [\text{diag}(\frac{1}{\sigma_i})]$.

Underdetermined Systems

- Effectively, fewer equations than unknowns
- $b \in \mathscr{R}(A) \Rightarrow$ Multiple solutions
- We may choose the smallest norm solution similarly to the overdetermined case.
- Pros:
  - Robust when A is singular or near singular
  - Treats both the underdetermined and overdetermined systems identically
- Cons:
  - Computational more demanding.

Regularized LS method

- To improves the stability, add regularization in the minimization:

$$\hat{x} = \arg\min ||Ax - b||^2 + ||\Gamma x||^2$$

  where Γ is the regularization matrix or *Tikhonov* matrix

- Regularized solution:

$$\hat{x} = (A^T A + \Gamma^T \Gamma)^{-1} A^T b$$

- $\Gamma = 0 \Rightarrow$ Conventional LS solution.

Choice of Γ Using the KF Theory

- Perceive the following to be the measurement equation:

$$b_k = Ax_k + w_k$$

- Suppose $\hat{x}_{k|k-1} \sim \mathcal{N}(0, \sigma_x^2 I)$, and $w_k \sim \mathcal{N}(0, \sigma_b^2 I)$.
- Then the updated state:

$$\begin{aligned}
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + W(b_k - \hat{b}_{k|k-1}) \\
\hat{x} &= Wb \\
&= \left[A^T(\sigma_b^2 I)^{-1}A + (\sigma_x^2 I)^{-1}\right]^{-1} A^T(\sigma_b^2 I)^{-1} b \\
&= \frac{1}{\sigma_b^2}\left[\frac{1}{\sigma_b^2}A^T A + \frac{1}{\sigma_x^2} I\right]^{-1} A^T b \\
&= \left[A^T A + \left(\frac{\sigma_b}{\sigma_x}\right)^2 I\right]^{-1} A^T b
\end{aligned}$$

- The above expression suggests to choose Γ to be $\Gamma = \alpha I$, where $\alpha = \frac{\sigma_b}{\sigma_x}$.

Pseudo-inverses

- works well for full-rank matrix A.
- Case (i): Overdetermined systems
  - Yields the unique solution in the minimum residual, $||Ax - b||$ sense.
  - Write

$$A^T A\hat{x} = A^T b$$

  - As $A^T A$ is non-singular, we get

$$\hat{x} = A^+ b,$$

    where the pseudo-inverse matrix

$$A^+ = (A^T A)^{-1} A^T.$$

Pseudo-inverse (Cont'd)

- Case (ii): Underdetermined systems
  - Yields the unique solution in the smallest length, $||x||$ sense:

$$\hat{x} \;=\; A^{+}b,$$

where the pseudo-inverse matrix

$$A^{+} \;=\; A^{T}(AA^{T})^{-1}.$$

- Limitations:
  - $A^{T}A$ may be singular or near-singular
  - matrix-squared form may amplify roundoff errors !
  - Remedy: Use the SVD on $A^{T}A$ or the QR on A directly.

QR Decomposition in Pseudo-inverses

- Decompose A into

$$A = QR,$$

where R is upper triangular; Q is orthogonal such that $QQ^T = I$.

- For an overdetermined case,

$$\begin{aligned}
\hat{x} &= (A^T A)^{-1} A^T b \\
&= (R^T Q^T Q R)^{-1} R^T Q^T b \\
&= (R^T R)^{-1} R^T Q^T b = R^{-1} Q^T b \\
\Rightarrow R\hat{x} &= \underbrace{Q^T b}_{\text{rotate}}
\end{aligned}$$

- Use back substitution to get the stable solution.

References

- J. Reilly, *ECE 712:Matrix Computations for Signal Processing*, Course notes.
- G. Golub and C. Van Loan, *Matrix Computations*, John Hopkins, 1996.
- T. Moon and W. Stirling, *Mathematical Methods and Algorithms*, Prentice-Hall, 2001.

Thank you!

Q3. Derive the Wiener filter (non-causal) for a stationary process with given spectral characteristics

Background

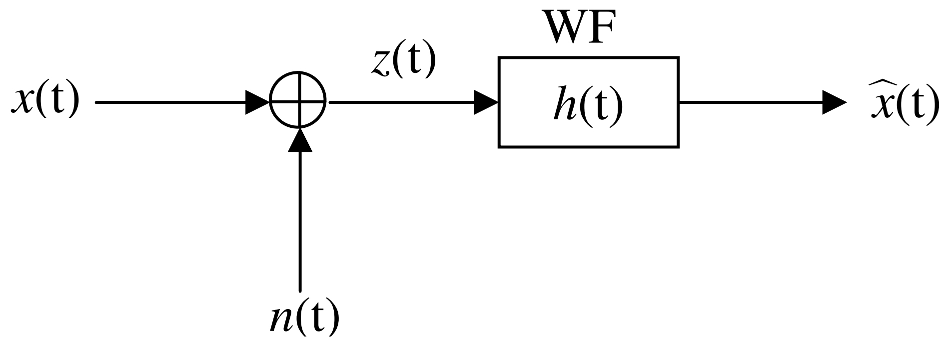

Figure 1: Signal Flow Diagram

- Goal: filter out noise that has corrupted a signal
- Assumptions:
  - Additive noise
  - signal and noise are wide-sense stationary processes
  - Spectral characteristics are known *a priori*.
- Performance criteria: Minimum mean-square error (MMSE).

Model setup

- The input-output relationship of the linear time-invariant system:

$$\hat{x}(t) = \int_{-\infty}^{t} h(\tau) z(t-\tau) d\tau = h(t) * z(t)$$

where

- $h(t)$ is the impulse response of the filter
- $z(t)$ is the observed process and related to the unknown signal process $x(t)$ via

$$z(t) = x(t) + n(t)$$

where $n(t)$ is the additive noise.

- $\hat{x}(t)$ is the output process.

Non-causal Wiener Filtering

- Under non-causality, past and future observations are known; hence the Wiener filter(WF) acts as a smoother.
- Redefine the goal to estimate

$$\hat{x}(t) = \mathbb{E}[x(t)|z(\xi), -\infty < \xi < \infty] \tag{1}$$

- The WF minimizes the MSE function:

$$\begin{aligned} J &= \mathbb{E}[\big(x(t) - \hat{x}(t)\big)^2] \\ &= \mathbb{E}[(x(t) - \textstyle\int_{-\infty}^{\infty} h(\alpha) z(t-\alpha) d\alpha)^2] \end{aligned}$$

- Solve the above minimization problem using the orthogonality principle.

Orthogonality Principle

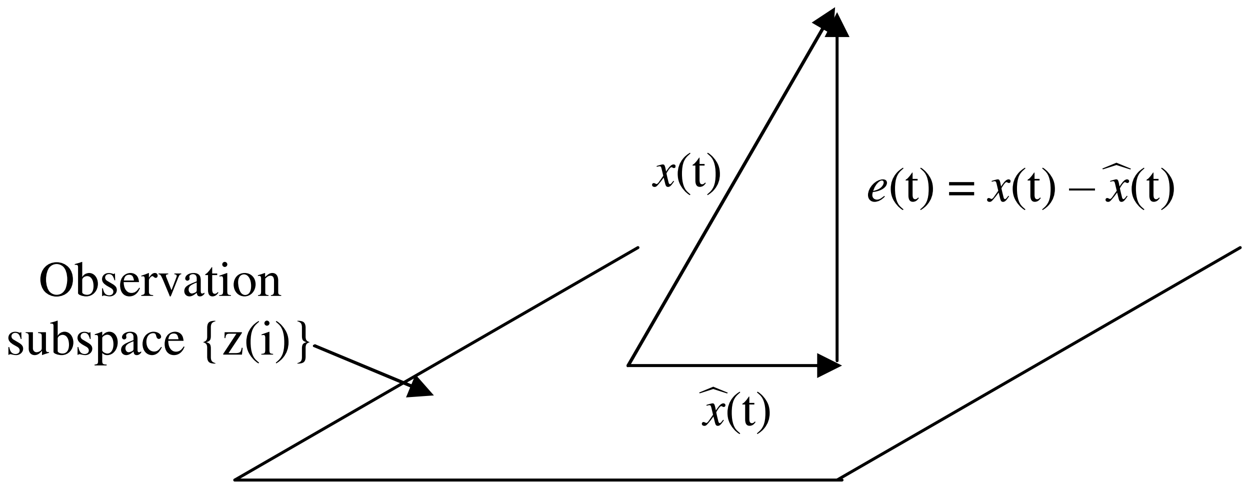

Figure 2: Projection of signal onto the observation subspace

- Idea: To get the minimum error in the MSE sense, the observation subspace has to be orthogonal to the estimation-error subspace.

Orthogonality Principle (Cont'd)

- Using the orthogonality principle we have

$$\begin{aligned}
\mathbb{E}\big[(x(t)-\hat{x}(t))z(t-\tau)\big] &= 0 \\
\mathbb{E}\big[\{x(t)-\textstyle\int_{-\infty}^{\infty} h(\alpha)z(t-\alpha)d\alpha\}z(t-\tau)\big] &= 0 \\
\mathbb{E}[x(t)z(t-\tau)] - \textstyle\int_{-\infty}^{\infty} h(\alpha)E[z(t-\alpha)z(t-\tau)]d\alpha &= 0 \\
\Rightarrow R_{xz}(\tau) = \int_{-\infty}^{\infty} h(\alpha)\mathbb{E}[z(t-\alpha)z(t-\tau)]d\alpha.
\end{aligned}$$

- Substituting $t = \xi + \tau$ yields

$$\begin{aligned}
\text{RHS} &= \int_{-\infty}^{\infty} h(\alpha)\mathbb{E}[z(\xi+\tau-\alpha)z(\xi)]d\alpha \\
&= \int_{-\infty}^{\infty} h(\alpha)R_z(\tau-\alpha)d\alpha \\
&= h(\tau) * R_z(\tau)
\end{aligned}$$

- Hence, we get the cross-correlation function:

$$R_{xz}(\tau) = h(\tau) * R_z(\tau)$$

Orthogonality (Cont'd)

- Taking the FT yields the cross-power spectral density:

$$S_{xz}(\omega) = H(\omega)S_z(\omega) \tag{2}$$

- The transfer function of the WF

$$H(\omega) = \frac{S_{xz}(\omega)}{S_z(\omega)} \tag{3}$$

- Further assumptions:
  - $x(t)$ and $n(t)$ are independent stochastic processes $\Rightarrow$ zero cross-correlation.
  - $n(t)$ has zero-mean.
- Under this assumption, we have
  - $R_{xz}(\tau) = R_x(\tau)$
  - $R_z(\tau) = R_x(\tau) + R_n(\tau)$

Orthogonality Principle (Cont'd)

- Equivalently, taking the FT yields
  - $S_{xz}(\omega) = S_x(\omega)$
  - $S_z(\omega) = S_x(\omega) + S_n(\omega)$
- A simplified transfer function of the WF

$$H(\omega) = \frac{S_x(\omega)}{S_x(\omega) + S_n(\omega)} \tag{4}$$

- Interesting Observations:
  - The transfer function is non-zero only where the signal has power content.
  - $H(\omega) = \frac{1}{1+\frac{1}{\frac{S_x(\omega)}{S_n(\omega)}}} \Rightarrow$ emphasizes frequencies where the SNR is large.

Wiener Filter: Block Diagram

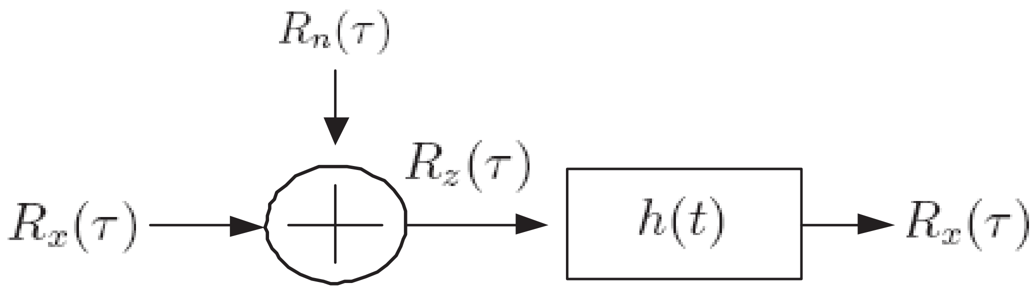

Figure 3: Time Domain

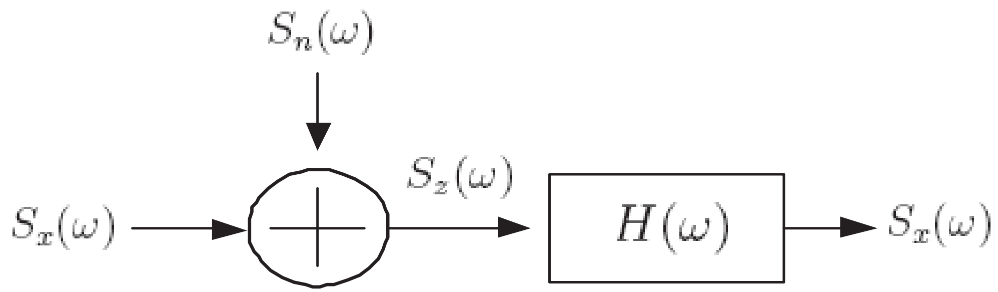

Figure 4: Frequency Domain

Toy Example

- Known. $R_x(\tau) = \frac{\sin(W\tau)}{W\tau}$ with $W = 5 \times 10^3$; $S_n(\omega) = 10^{-5}$
- Find $S_x(\omega)$:

$$S_x(\omega) = \frac{1}{2W} \prod\left(\frac{2\pi\omega}{W}\right)$$

- Find $H(\omega)$:

$$H(\omega) = \begin{cases} \frac{1}{1.1}, & |\omega| \leq W; \\ 0, & \text{otherwise.} \end{cases}$$

- WF acts as an ideal LPF.

Final Remarks

- WF is applied in image restoration.
- WF has the following limitations:
  - Not amenable to state-vector estimation problems
  - Not applicable to non-stationary signals
  - Non-causal WFs are not suitable for real-time applications

References

- A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed., McGraw Hill, 2002.
- K. M. Wong, *ECE 762: Detection and Estimation Theory*, Course Notes.
- R. Yates and D. Goodman *Probability and Stochastic Processes*, Wiley, 2004.

Thank you!