Speaker Tracking and Beamforming

Dr. John McDonough

Spoken Language Systems
Saarland University

January 13, 2010

Introduction

- Many problems in science and engineering can be formulated in terms of estimating a *state* based on observations.
- In this tutorial, we will discuss the *Kalman filter*, which is one possible solution for this problem.
- We will briefly mention both the *probabilistic* and *joint probabilistic data association* filters, which represent other possible solutions.
- We will also present the fundamentsl of *beamforming*, whereby the signals from all sensors in a microphone array are combined to for optimal speech enhancement.

Bayesian Filtering

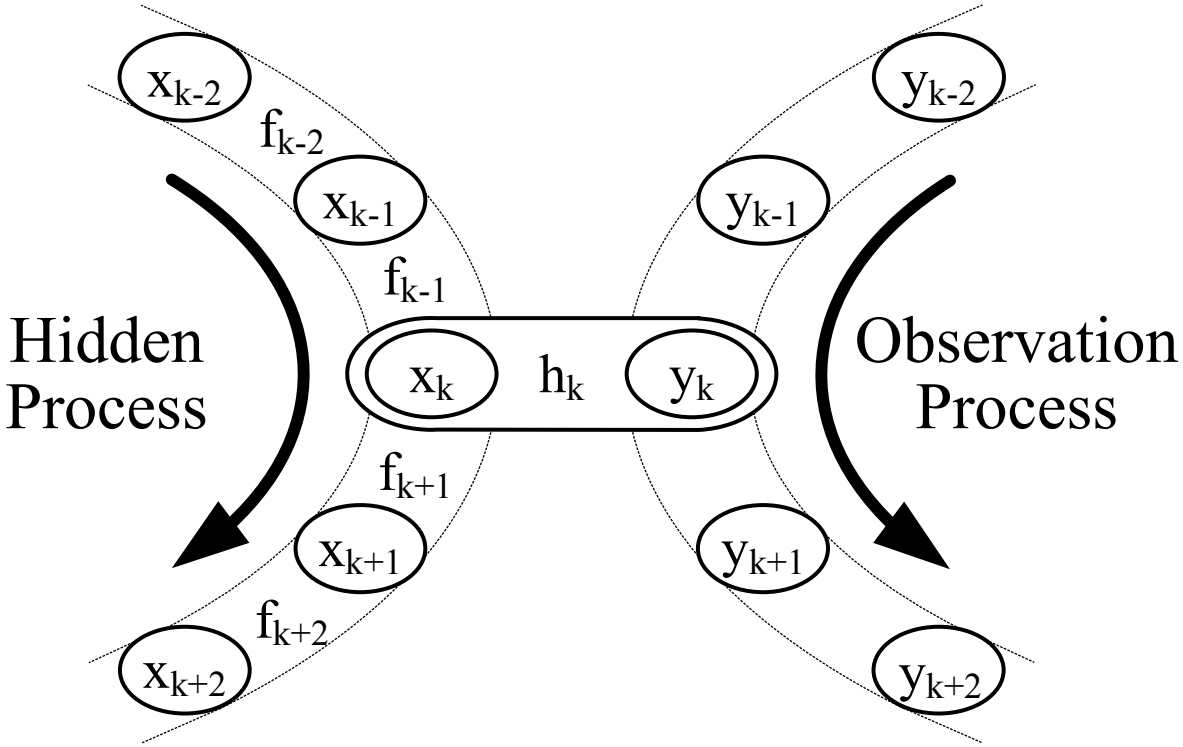


Figure: State-space model

Kalman Filter System and Observation Equations

- The state model of the Kalman filter is given by

$$\mathbf{x}_k = \mathbf{F}_{k|k-1}\,\mathbf{x}_{k-1} + \mathbf{u}_{k-1}, \tag{1}$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, \tag{2}$$

  where $\mathbf{F}_{k|k-1}$ and $\mathbf{H}_k$ are the known *transition* and *observation* matrices.
- The noise terms $\mathbf{u}_k$ and $\mathbf{v}_k$ in (1–2) are by assumption zero mean, white Gaussian random vector processes with covariance matrices

$$\mathbf{U}_k = \mathcal{E}\{\mathbf{u}_k\mathbf{u}_k^T\}, \qquad \mathbf{V}_k = \mathcal{E}\{\mathbf{v}_k\mathbf{v}_k^T\}.$$

- By assumption $\mathbf{u}_k$ and $\mathbf{v}_k$ are statistically independent.

First Order Markov Process

- The system model (1) represents a *first-order Markov process*, which implies that the current state of the system depends solely on the state immediately prior, such that

$$p(\mathbf{x}_k|\mathbf{x}_{0:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}) \; \forall \; k \; \in \; \mathbb{N}. \tag{3}$$

- The observation model (2) implies that the observations are dependent only on the current system state.

$$p(\mathbf{y}_k|\mathbf{x}_{0:k}, \mathbf{y}_{1:k-1}) = p(\mathbf{y}_k|\mathbf{x}_k) \; \forall \; k \; \in \; \mathbb{N}. \tag{4}$$

- Our goal is to track the *filtering density* $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ in time.
- From Bayes' rule, we can write

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_{1:k}|\mathbf{x}_k)\, p(\mathbf{x}_k)}{p(\mathbf{y}_{1:k})}. \tag{5}$$

Tracking the Filtering Density

- This filtering density can be tracked sequentially as:
  - *prediction*: The prior pdf of $\mathbf{x}_k$ can be obtained from

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\, d\mathbf{x}_{k-1}, \qquad (6)$$

    where the evolution of the state $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is defined by the state equation (1).
  - *correction* or *update*: The current observation is "folded" into the estimate of the filtering density through an invocation of Bayes' rule:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} = \frac{p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})}, \qquad (7)$$

    where the normalization constant is given by

$$p(\mathbf{y}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{y}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{y}_{1:k-1})\, d\mathbf{x}_k.$$

Innovations Sequence

- Let $\mathbf{y}_{1:k-1}$ denote all past observations up to time $k-1$, and let $\hat{\mathbf{y}}_{k|k-1}$ denote the MMSE estimate,

$$\hat{\mathbf{y}}_{k|k-1} \triangleq \mathcal{E}\{\mathbf{y}_k|\mathbf{y}_{1:k-1}\}.$$

  By definition, the *innovation* is the difference

$$\mathbf{s}_k \triangleq \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1} \tag{8}$$

  between the actual and the predicted observations.
- The innovation contains the "new information" required for sequentially updating the filtering density $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$.
- This information cannot be predicted from the state space model.

Properties of the Innovations Sequence

The innovations process has three important properties:

- *orthogonality*
  The innovation process $\mathbf{s}_K$ at time K is *orthogonal* to all past observations $\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}$, such that

$$\mathcal{E}\{\mathbf{s}_K \mathbf{y}_k^T\} = \mathbf{0} \; \forall \; 1 \; \leq k \leq K-1.$$

- *whiteness*
  The innovations are orthogonal to each other, such that

$$\mathcal{E}\{\mathbf{s}_K \mathbf{s}_k^T\} = \mathbf{0} \; \forall \; 1 \leq k \leq K-1.$$

- *reversibility*
  There is a one-to-one correspondence between the observed data $\mathbf{y}_{1:k} = \{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$ and the sequence of innovations $\mathbf{s}_{1:k} = \{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$, such that one can always be uniquely recover from the other.

Outline of Kalman Filter Development

- We will now present the principal quantities and relations in the operation of the KF.
- This presentation is intended to convey intuition rather than provide a rigorous derivation.
- The development of the KF will proceed in four phases:
  1. Provide an expression for calculating the covariance matrix of the innovations process.
  2. Obtain an expression for the sequential update of the MMSE state estimate.
  3. Define the Kalman gain, which plays the pivotal role in the sequential state update considered in phase two.
  4. State the Riccati equation, which provides the means to update the state estimation error covariance matrices required to calculate the Kalman gain.

Predicted State Estimation Error

- We begin by stating how the predicted observation may be calculated based on the current state estimate, according to,

$$\hat{\mathbf{y}}_{k|k-1} = \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}. \tag{9}$$

- In light of (8) and (9), we may write

$$\mathbf{s}_k = \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}. \tag{10}$$

- Substituting (2) into (10), we find

$$\mathbf{s}_k = \mathbf{H}_k \epsilon_{k|k-1} + \mathbf{v}_k, \tag{11}$$

where

$$\epsilon_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \tag{12}$$

is the *predicted state estimation error* at time k, using data up to time $k-1$.

Predicted State Estimation Error Covariance Matrix

- It holds that $\epsilon_{k|k-1}$ is orthogonal to $\mathbf{u}_k$ and $\mathbf{v}_k$.
- The correlation matrix of the innovations sequence can be expressed as

$$\mathbf{S}_k \triangleq \mathcal{E}\left\{\mathbf{s}_k\mathbf{s}_k^T\right\} = \mathbf{H}_k\mathbf{K}_{k|k-1}\mathbf{H}_k^T + \mathbf{V}_k, \tag{13}$$

where the *predicted state estimation error covariance matrix* is defined as

$$\mathbf{K}_{k|k-1} \triangleq \mathcal{E}\left\{\epsilon_{k|k-1}\epsilon_{k|k-1}^T\right\}. \tag{14}$$

Kalman Gain

The sequential update of the filtering density has two steps:

1. First, there is a *prediction*, which can be expressed as

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k|k-1}\hat{\mathbf{x}}_{k-1|k-1}, \tag{15}$$

   The prediction does *not* use the current observation $\mathbf{y}_k$.

2. Second, there is an *update* or *correction*,

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k\mathbf{s}_k, \tag{16}$$

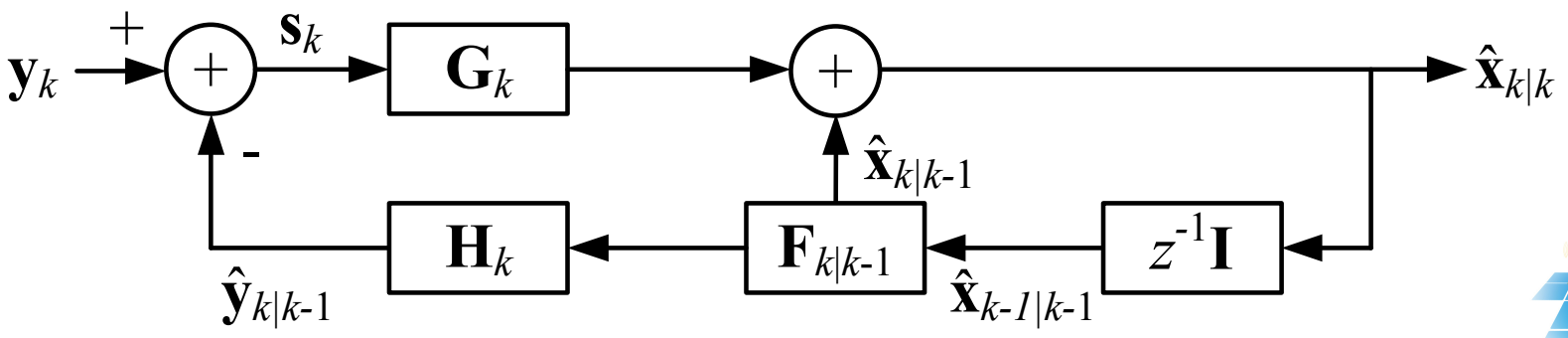
   where the *Kalman gain* is defined as

$$\mathbf{G}_k \triangleq \mathcal{E}\{\mathbf{x}_k\mathbf{s}_k^T\}\mathbf{S}_k^{-1}, \tag{17}$$

   for $\mathbf{x}_k$, $\mathbf{s}_k$, and $\mathbf{S}_k$ given by (1), (10), and (13).

Predictor-Corrector Structure

- To perform the Kalman update it is necessary to:
  - Premultiply the prior estimate $\hat{\mathbf{x}}_{k|k-1}$ by $\mathbf{F}_{k|k-1}$.
  - Add a correction factor consisting of the Kalman gain $\mathbf{G}_k$ multiplied by the innovation $\mathbf{s}_k$.
- The predictor-corrector structure of the Kalman filter is shown below.
- The problem of Kalman filtering reduces to calculating the Kalman gain.



Riccati Equation

- The Kalman gain (17) can be calculated as

$$\mathbf{G}_k = \mathbf{K}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}, \tag{18}$$

  where the correlation matrix $\mathbf{S}_k$ of the innovations sequence is defined in (13).
- The *Riccati equation*, then specifies how $\mathbf{K}_{k|k-1}$ can be sequentially updated, namely as,

$$\mathbf{K}_{k|k-1} = \mathbf{F}_{k|k-1}\,\mathbf{K}_{k-1}\,\mathbf{F}_{k|k-1}^T + \mathbf{U}_{k-1}. \tag{19}$$

Filtered State Estimation Error

- The matrix $\mathbf{K}_k$ in (19) is, in turn, obtained through the recursion,

$$\mathbf{K}_k = \mathbf{K}_{k|k-1} - \mathbf{G}_k \mathbf{H}_k \mathbf{K}_{k|k-1} = (\mathbf{I} - \mathbf{G}_k \mathbf{H}_k)\mathbf{K}_{k|k-1}. \qquad (20)$$

- This matrix $\mathbf{K}_k$ can be interpreted as the covariance matrix of the *filtered state estimation error*,

$$\mathbf{K}_k \triangleq \left\{ \epsilon_k \epsilon_k^T \right\},$$

where

$$\epsilon_k \triangleq \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}.$$

Filtered vs. Predicted State Estimation Error

- $\epsilon_{k|k-1}$ is the error in the state estimate made *without* knowledge of the current observation $\mathbf{y}_k$,
- ϵ_k is the error in the state estimate *with* knowledge of $\mathbf{y}_k$.

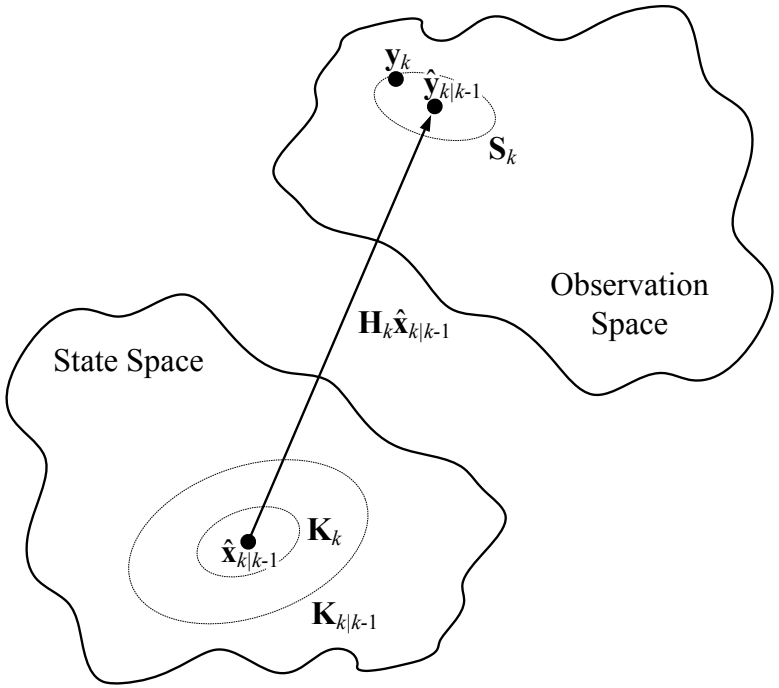


Figure: The operation of the Kalman filter.

Kalman Filtering Algorithm

The steps leading to the sequential update of $\hat{\mathbf{x}}_{k|k}$ are:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k|k-1}\,\hat{\mathbf{x}}_{k-1|k-1} \text{ (prediction)} \tag{21}$$

$$\mathbf{K}_{k|k-1} = \mathbf{F}_{k|k-1}\,\mathbf{K}_{k-1}\,\mathbf{F}_{k|k-1}^{T} + \mathbf{U}_{k-1} \tag{22}$$

$$\mathbf{S}_k = \mathbf{H}_k\,\mathbf{K}_{k|k-1}\,\mathbf{H}_k^{T} + \mathbf{V}_k \tag{23}$$

$$\mathbf{G}_k = \mathbf{K}_{k|k-1}\,\mathbf{H}_k^{T}\mathbf{S}_k^{-1} \tag{24}$$

$$\mathbf{s}_k = \mathbf{y}_k - \mathbf{H}_k\,\hat{\mathbf{x}}_{k|k-1} \tag{25}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k\,\mathbf{s}_k \text{ (correction)} \tag{26}$$

$$\mathbf{K}_k = (\mathbf{I} - \mathbf{G}_k\,\mathbf{H}_k)\,\mathbf{K}_{k|k-1} \tag{27}$$

Prediction and Correction (Again)

- The connection between the KF and sequential Bayesian filtering can be stated explicitly as writing
  - the *prediction* as

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{K}_{k|k-1}) \tag{28}$$

  - and the *correction* as

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{K}_k), \tag{29}$$

where $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multidimensional Gaussian pdf with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

PDAF Illustration

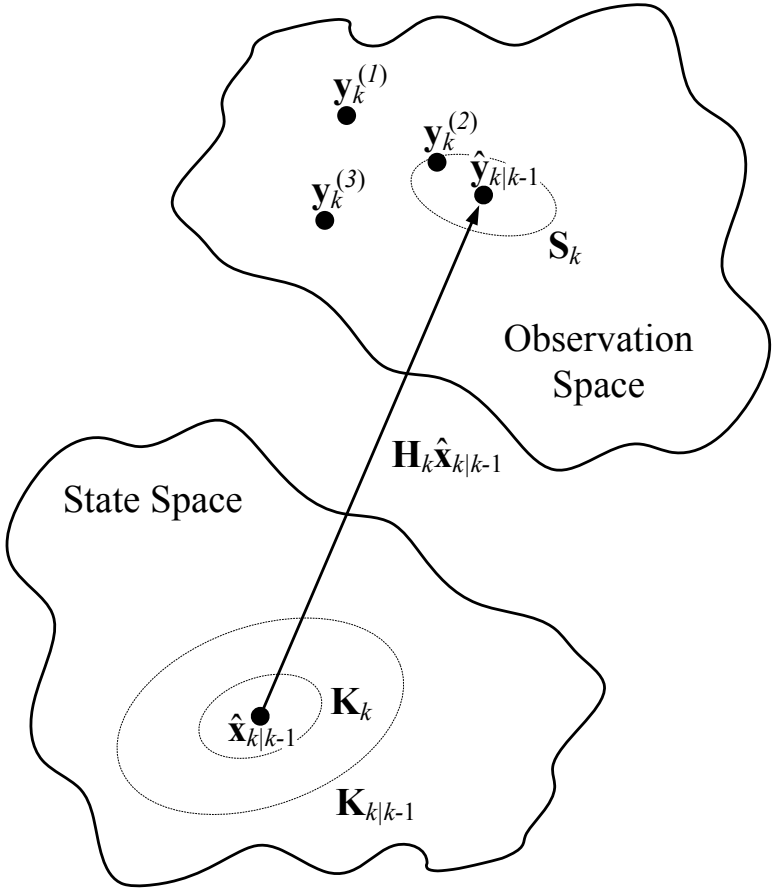


Figure: Schematic illustrating the operation of the probabilistic data association filter.

JPDAF Illustration

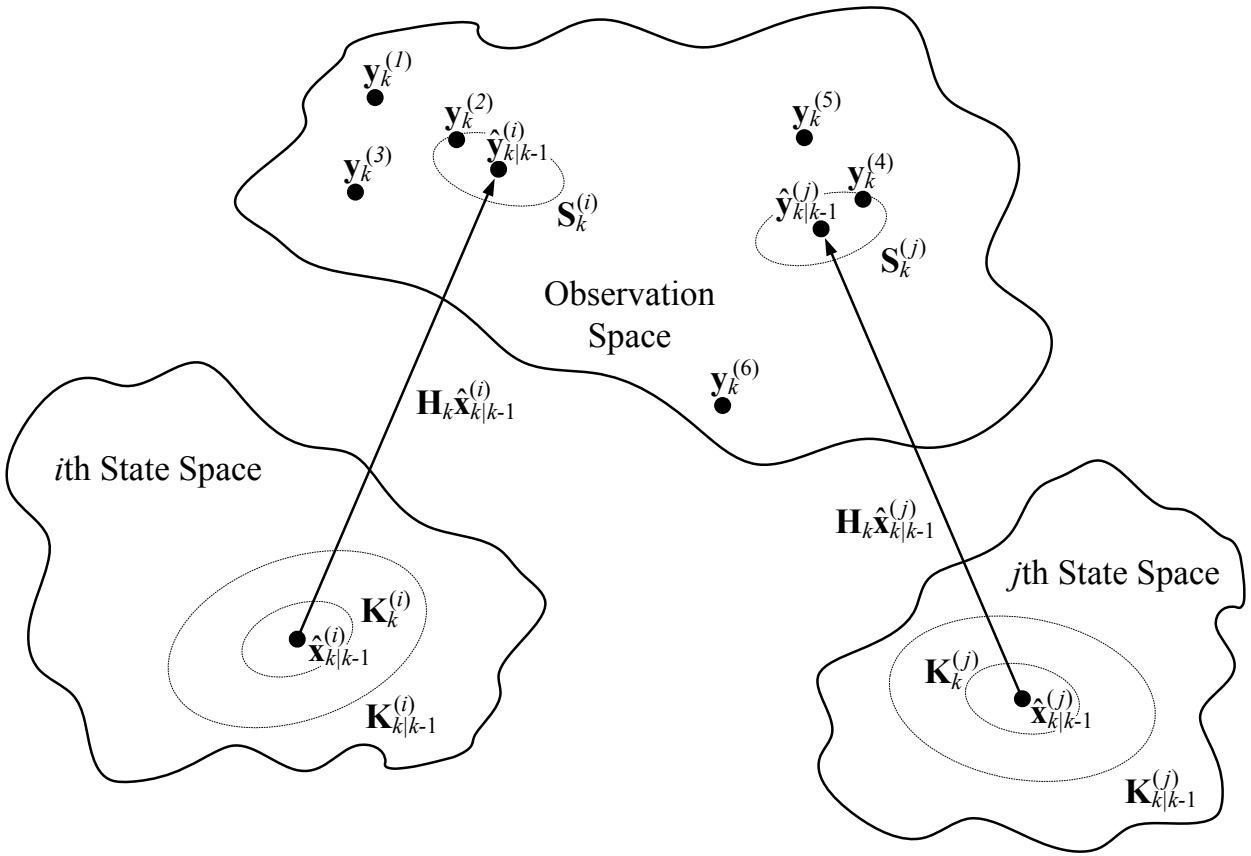


Figure: Schematic illustrating the operation of the joint probabilistic data association filter.

Speaker Tracking with the Kalman Filter

- Solving for that $\mathbf{x}$ minimizing (43) would be eminently straightforward were it not for the fact that the objective function is nonlinear in $\mathbf{x}$.
- In the coming development, we will find it useful to have a linear approximation.
- Hence, we take a partial derivative with respect to $\mathbf{x}$ and write

$$\nabla_{\mathbf{x}} T_n(\mathbf{x}) = \frac{1}{c} \cdot \left[\frac{\mathbf{x} - \mathbf{m}_{n1}}{D_{n1}} - \frac{\mathbf{x} - \mathbf{m}_{n2}}{D_{n2}} \right],$$

where

$$D_{nm} = |\mathbf{x} - \mathbf{m}_{nm}| \;\; \forall \;\; n = 1, \ldots, M; m = 1, 2,$$

is the distance between the source and the microphone located at $\mathbf{m}_{nm}$.

A Taylor Series Approximation

- The speaker's position cannot change instantaneously.
- Thus both the present $\hat{\tau}_i(k)$ and past TDOA estimates $\{\hat{\tau}_i(n)\}_{n=1}^{k-1}$ are potentially useful in estimating a speaker's current position $\mathbf{x}(k)$.
- Let us approximate $T_n(\mathbf{x})$ with a first order Taylor series expansion about the last position estimate $\hat{\mathbf{x}}(k-1)$ by writing

$$T_n(\mathbf{x}) \approx T_n(\hat{\mathbf{x}}(k-1)) + \mathbf{c}_n^T(k)\left[\mathbf{x} - \hat{\mathbf{x}}(k-1)\right], \tag{30}$$

where the row vector $\mathbf{c}_n^T(k)$ is given by

$$\mathbf{c}_n^T(k) = \left[\nabla_{\mathbf{x}} T_n(\mathbf{x})\right]^T_{\mathbf{x}=\hat{\mathbf{x}}(k-1)} = \frac{1}{c}\cdot\left[\frac{\mathbf{x}-\mathbf{m}_{n1}}{D_{n1}} - \frac{\mathbf{x}-\mathbf{m}_{n2}}{D_{n2}}\right]^T_{\mathbf{x}=\hat{\mathbf{x}}(k-1)} \tag{31}$$

Linearization

- Substituting the linearization (30) into (43) provides

$$\epsilon(\mathbf{x}; k) = \sum_{n=1}^{M} \frac{1}{\sigma_n^2} \left[\bar{\tau}_n(k) - \mathbf{c}_n^T(k)\mathbf{x} \right]^2 \tag{32}$$

where

$$\bar{\tau}_n(k) = \hat{\tau}_n(k) - \left[T_n(\hat{\mathbf{x}}(k-1)) - \mathbf{c}_n^T(k)\hat{\mathbf{x}}(k-1) \right] \ \forall\, n = 1, \ldots, M. \tag{33}$$

Let us define

$$\bar{\bar{\tau}}(k) = \begin{bmatrix} \bar{\tau}_1(k) \\ \bar{\tau}_2(k) \\ \vdots \\ \bar{\tau}_M(k) \end{bmatrix}, \hat{\tau}(k) = \begin{bmatrix} \hat{\tau}_1(k) \\ \hat{\tau}_2(k) \\ \vdots \\ \hat{\tau}_M(k) \end{bmatrix}, \tag{34}$$

Linearization (cont'd.)

- Also define,

$$\mathbf{T}(\hat{\mathbf{x}}(k)) = \begin{bmatrix} T_1(\hat{\mathbf{x}}(k)) \\ T_2(\hat{\mathbf{x}}(k)) \\ \vdots \\ T_M(\hat{\mathbf{x}}(k)) \end{bmatrix}, \mathbf{C}(k) = \begin{bmatrix} \mathbf{c}_1^T(k) \\ \mathbf{c}_2^T(k) \\ \vdots \\ \mathbf{c}_M^T(k) \end{bmatrix}. \tag{35}$$

Linearization (cont'd.)

- Then (33) can be expressed in matrix form as

$$\bar{\bar{\tau}}(k) = \hat{\tau}(k) - [\mathbf{T}(\hat{\mathbf{x}}(k-1)) - \mathbf{C}(k)\hat{\mathbf{x}}(k-1)]\,. \tag{36}$$

- Similarly, defining

$$\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2) \tag{37}$$

enables (32) to be expressed as

$$\epsilon(\mathbf{x}; t) = [\bar{\bar{\tau}}(k) - \mathbf{C}(k)\mathbf{x}]^T \mathbf{\Sigma}^{-1} [\bar{\bar{\tau}}(k) - \mathbf{C}(k)\mathbf{x}]\,. \tag{38}$$

Definition: Time Delay of Arrival

- Consider then a sensor array consisting of $N+1$ microphones located at positions $\mathbf{m}_i \; \forall \; i = 0, \ldots, N$.
- Let $\mathbf{x} \in \mathbf{R}^3$ denote the position of the speaker in a three-dimensional space, such that,

$$\mathbf{x} \triangleq \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \text{ and } \mathbf{m}_n \triangleq \begin{bmatrix} m_{n,x} \\ m_{n,y} \\ m_{n,z} \end{bmatrix}.$$

- The *time delay of arrival* (TDOA) between the microphones at positions $\mathbf{m}_1$ and $\mathbf{m}_2$ can be expressed as

$$T(\mathbf{m}_1, \mathbf{m}_2, \mathbf{x}) = \frac{|\mathbf{x} - \mathbf{m}_1| - |\mathbf{x} - \mathbf{m}_2|}{c} \tag{39}$$

where c is the speed of sound.

Phase Transform

- Let $\hat{\tau}_{mn}$ denote the observed TDOA for the m-th and nth microphones.
- The TDOAs can be observed or estimated with a variety of well-known techniques.
- The most common method is based on the *phase transform*,

$$\rho_{mn}(\tau) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{Y_m(e^{j\omega\tau}) Y_n^*(e^{j\omega\tau})}{\left| Y_m(e^{j\omega\tau}) Y_n^*(e^{j\omega\tau}) \right|} e^{j\omega\tau} \, d\omega, \qquad (40)$$

where $Y_n(e^{j\omega\tau})$ denotes the short-time Fourier transform of the signal arriving at the nth sensor in the array.

Time Delay of Arrival

- The definition of the GCC in (40) follows directly from the frequency domain calculation of the cross-correlation of two sequences.
- The normalization term $\left| Y_m(e^{j\omega\tau}) Y_n^*(e^{j\omega\tau}) \right|$ in the denominator of the integrand in (40) is intended to weight all frequencies equally.
- Such a weighting leads to more robust TDOA estimates in noisy and reverberant environments.
- The TDOA estimate is obtained from

$$\hat{\tau}_{mn} = \max_{\tau} \ \rho_{mn}(\tau). \tag{41}$$

Details of Calculating TDOAs

- In other words, the "true" TDOA is taken as that which maximizes the GCC $\rho_{mn}(\tau)$.
- For reasons of computational efficiency, $\rho_{mn}(\tau)$ is typically calculated with an inverse FFT.
- Thereafter, an interpolation is performed to overcome the granularity in the estimate corresponding to the sampling interval.
- Usually, $Y_n(e^{j\omega_k})$ appearing in (40) are calculated with a Hamming analysis window of 15 to 25 ms in duration.

Maximum Likelihood Speaker Tracking

- Consider two microphones located at $\mathbf{m}_{n1}$ and $\mathbf{m}_{n2}$ comprising the nth microphone pair.
- Define the TDOA as

$$T(\mathbf{m}_1, \mathbf{m}_2, \mathbf{x}) = \frac{|\mathbf{x} - \mathbf{m}_1| - |\mathbf{x} - \mathbf{m}_2|}{c} \tag{42}$$

  where c is the speed of sound, $\mathbf{x}$ denotes the position of an active speaker.
- Source localization based on the maximum likelihood criterion proceeds by minimizing the error function

$$\epsilon(\mathbf{x}) = \sum_{i=n}^{M} \frac{1}{\sigma_n^2} \left[\hat{\tau}_n - T_n(\mathbf{x})\right]^2, \tag{43}$$

  where σ_n^2 is the error covariance and $\hat{\tau}_n$ is the observed TDOA from (40) and (41).

Sound Propagation

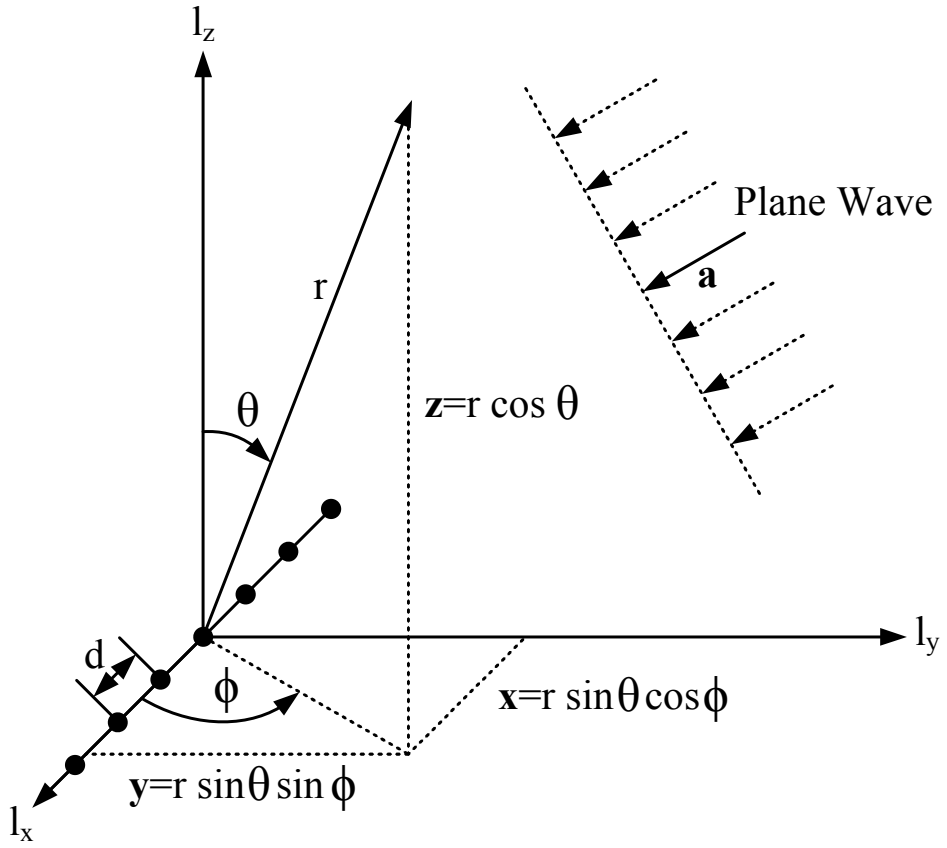


Figure: Relation between the spherical coordinates (r, θ, ϕ) and Cartesian coordinates (x, y, z).

Array Processing

- We process the output of each sensor with a LTI filter with impulse response $h_n(\tau)$ and sum the outputs:

$$y(t) = \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} h_n(t-\tau)\, f_n(\tau, \mathbf{m}_n)\, d\tau$$

- In matrix notation

$$y(t) = \int_{-\infty}^{\infty} \mathbf{h}^T(t-\tau)\, \mathbf{f}(\tau, \mathbf{m})\, d\tau$$

where

$$\mathbf{h}(t) = \begin{bmatrix} h_0(t) \\ h_1(t) \\ \vdots \\ h_{N-1}(t) \end{bmatrix}$$

Frequency Domain Array Processing

- In the frequency domain, this becomes

$$
\begin{aligned}
Y(\omega) &= \int_{-\infty}^{\infty} y(t)\, e^{-j\omega t}\, dt \\
&= \mathbf{H}^T(\omega)\, \mathbf{F}(\omega)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{H}(\omega) &= \int_{-\infty}^{\infty} \mathbf{h}(t) e^{-j\omega t}\, dt \\
\mathbf{F}(\omega, \mathbf{m}) &= \int_{-\infty}^{\infty} \mathbf{f}(t, \mathbf{m}) e^{-j\omega t}\, dt
\end{aligned}
$$

Array Manifold Vector

- In acoustic beamforming it may happen, however, that the distance from the source to any one of the sensors of an array is of the same order as the aperture of the array itself.
- The array manifold vector must be defined as

$$\mathbf{v}_{\mathbf{x}}(\mathbf{x}) \triangleq \begin{bmatrix} e^{-j\omega|\mathbf{x}-\mathbf{m}_0|/c} \\ e^{-j\omega|\mathbf{x}-\mathbf{m}_1|/c} \\ \vdots \\ e^{-j\omega|\mathbf{x}-\mathbf{m}_{N-1}|/c} \end{bmatrix},$$

where $\mathbf{x}$ is the position of the desired source in Cartesian coordinates.

Delay-and-Sum Beamformer Implementations

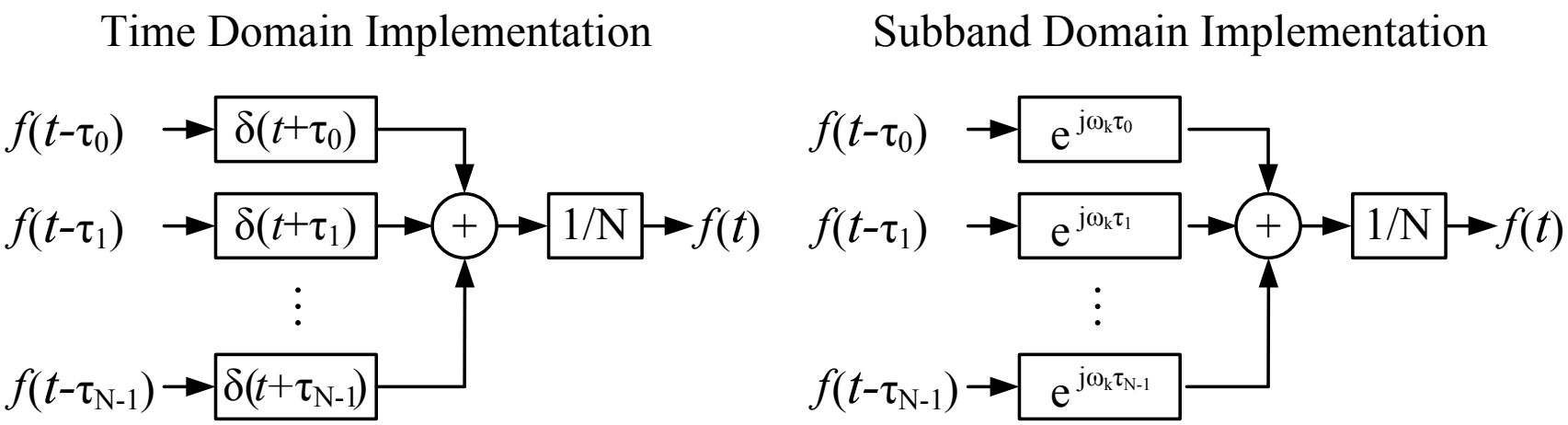


Figure: Time and subband domain implementations of the delay-and-sum beamformer.

Beam Pattern in Various Spaces

- The beam pattern is given by

$$B_\phi(\phi) = \frac{1}{N} \frac{\sin\left(\frac{N}{2} \cdot \frac{2\pi}{\lambda} \cos\phi \cdot d\right)}{\sin\left(\frac{1}{2} \cdot \frac{2\pi}{\lambda} \cos\phi \cdot d\right)} \text{ for } 0 \leq \phi \leq \pi$$

- In u-space this becomes

$$B_u(u) = \frac{1}{N} \frac{\sin\left(\frac{\pi N d}{\lambda} u\right)}{\sin\left(\frac{\pi d}{\lambda} u\right)} \text{ for } -1 \leq u \leq 1$$

- And in ψ-space,

$$B_\psi(\psi) = \frac{1}{N} \frac{\sin\left(N \frac{\psi}{2}\right)}{\sin\left(\frac{\psi}{2}\right)} \text{ for } -\frac{2\pi d}{\lambda} \leq \psi \leq \frac{2\pi d}{\lambda}$$

Effect of Element Spacing

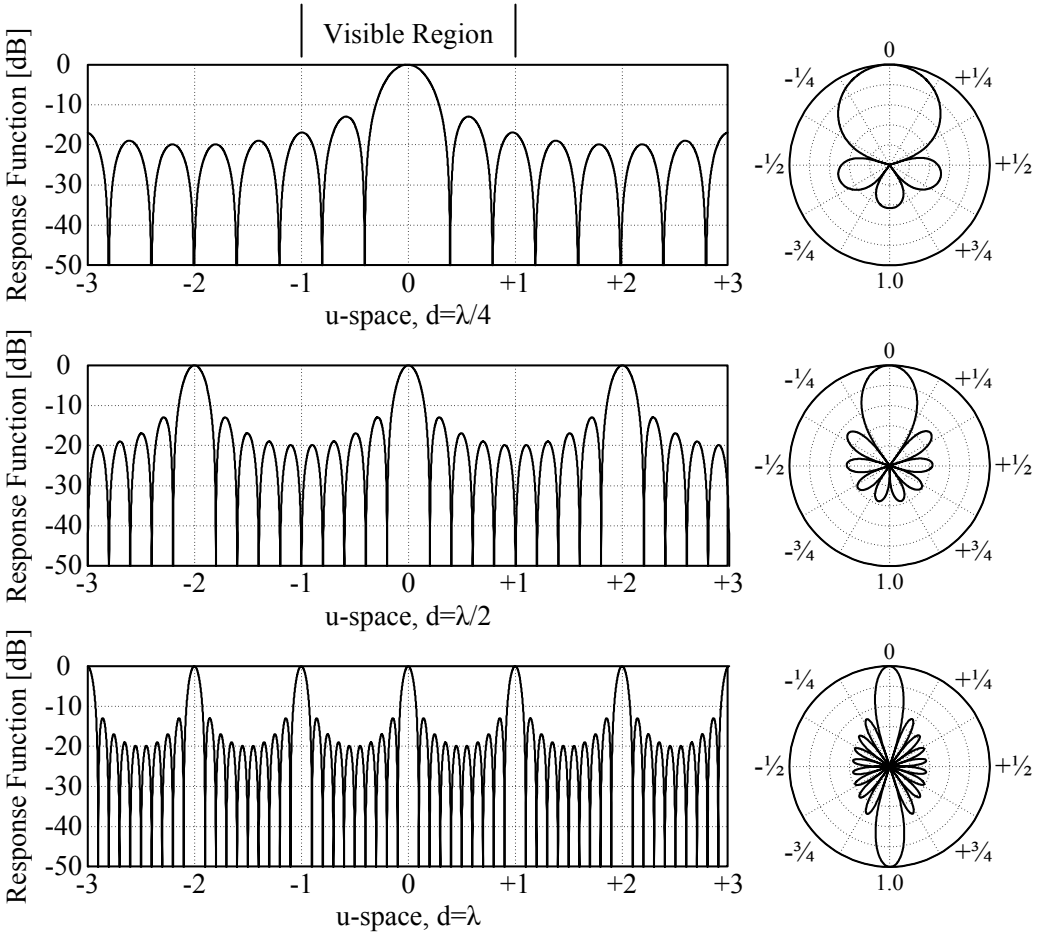


Figure: Effect of element spacing on beam patterns in linear and polar coordinates for $N = 10$.

Effect of Steering

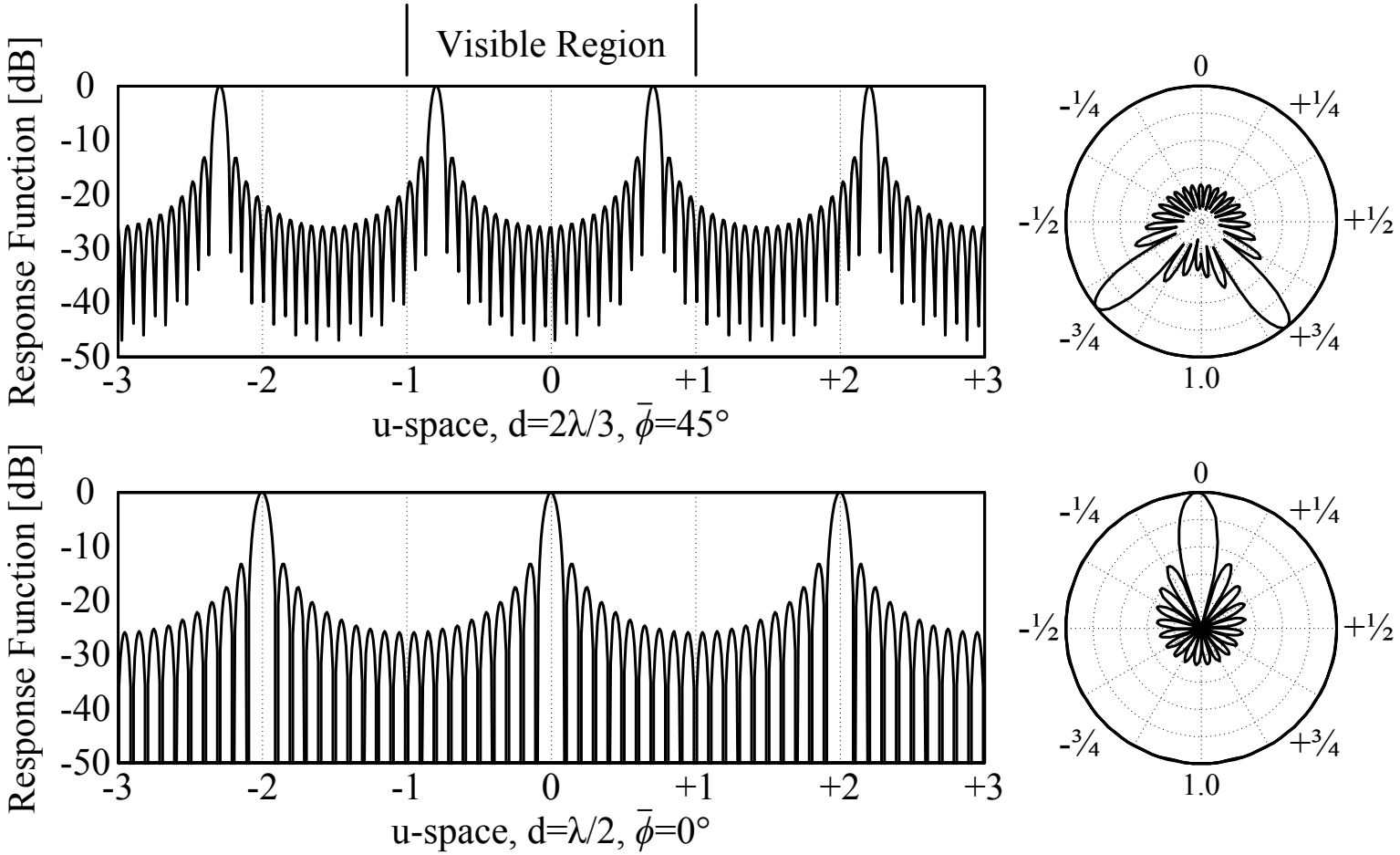


Figure: Effect of steering on the grating lobes for $N = 20$ plotted in linear and polar coordinates.

Inter-Sensor Spacing

- The position of the first grating lobe is $u - u_{\mathrm{T}} = \lambda/d$.
- Hence, keeping grating lobes out of the visible region requires

$$|u - u_{\mathrm{T}}| \leq \frac{\lambda}{d}.$$

- This inequality can be further overbounded as

$$|u - u_{\mathrm{T}}| \leq |u| + |u_{\mathrm{T}}| \leq 1 + |\sin \bar{\phi}_{\mathrm{max}}| \leq \frac{\lambda}{d},$$

  where $\bar{\phi}_{\mathrm{max}}$ is the maximum broadside angle to which the pattern is to be steered.
- Excluding grating lobes from the visible region requires

$$\frac{d}{\lambda} \leq \frac{1}{1 + |\sin \bar{\phi}_{\mathrm{max}}|}.$$

- If the array is to be steered over the entire half plane, it must hold that $d \leq \lambda/2$.

Array Gain

- We will in all cases take *word error rate* (WER) as the most important measure of system performance.
- It is useful, however, to have other performance measures intended a single component.
- *Signal-to-noise ratio* (SNR), a very common metric for signal quality, is the ratio of signal power to noise power.
- *Array gain* is a measure of how much improvement in SNR is achieved by a sensor array.
- Array gain is the ratio of SNR at the output of the array to that at input of any given sensor.

Snapshot Model

- Let $\mathbf{X}(\omega) \in \mathbb{C}^N$ denote a subband domain snapshot:

$$\mathbf{X}(\omega) = \mathbf{F}(\omega) + \mathbf{N}(\omega), \tag{44}$$

  where $\mathbf{F}(\omega)$ denotes the subband-domain snapshot of the desired signal and $\mathbf{N}(\omega)$ denotes that of the noise or other interference impinging on the sensors of the array.
- By assumption, $\mathbf{F}(\omega)$ and $\mathbf{N}(\omega)$ are uncorrelated and the signal vector $\mathbf{F}(\omega)$ is

$$\mathbf{F}(\omega) = F(\omega)\,\mathbf{v_k}(\mathbf{k}), \tag{45}$$

  where $\mathbf{v_k}(\mathbf{k})$ is the array manifold vector.

Second Order Statistics

- We now introduce the notation necessary for specifying the second order statistics of random variables and vectors.
- In general, for some complex scalar random variable $Y(\omega)$, we will define
$$\Sigma_Y(\omega) \triangleq \mathcal{E}\{|Y(\omega)|^2\}.$$
- Similarly, for a complex random vector $\mathbf{X}(\omega)$, let us define the *spatial spectral matrix* as
$$\mathbf{\Sigma}_{\mathbf{X}}(\omega) \triangleq \mathcal{E}\{\mathbf{X}(\omega)\mathbf{X}^H(\omega)\}.$$

Signal and Noise Components

- Let us begin by assuming that the component of the desired signal reaching each component of a sensor array is $F(\omega)$ and the component of the noise or interference reaching each sensor is $N(\omega)$.
- This implies that the SNR at the input of the array can be expressed as

$$\mathrm{SNR}_{\mathrm{in}}(\omega) \triangleq \frac{\Sigma_F(\omega)}{\Sigma_N(\omega)}. \tag{46}$$

- In the frequency domain, the output of the beamformer is

$$Y(\omega) = \int_{-\infty}^{\infty} y(t)\, e^{-j\omega t}\, dt = \mathbf{H}^T(\omega)\, \mathbf{X}(\omega). \tag{47}$$

Signal and Noise Components (cont'd.)

- Let us now define $\mathbf{w}^H(\omega) = \mathbf{H}^T(\omega)$.
- Then (47) can be rewritten as

$$Y(\omega) = \mathbf{w}^H(\omega)\,\mathbf{X}(\omega) = Y_F(\omega) + Y_N(\omega), \tag{48}$$

where $Y_F(\omega) = \mathbf{w}^H(\omega)\,\mathbf{F}(\omega)$ and $Y_N(\omega) = \mathbf{w}^H(\omega)\,\mathbf{N}(\omega)$ are, respectively, the signal and noise components in the output of the beamformer.

Beamformer Output

- When the delay-and-sum beamformer (DSB) is steered to wavenumber $\mathbf{k} = \mathbf{k}_T$, the sensor weights become

$$\mathbf{w}^H = \frac{1}{N}\,\mathbf{v_k}^H(\mathbf{k}_T). \tag{49}$$

- The variance of the output of the beamformer can be calculated according to

$$\Sigma_Y(\omega) = \mathcal{E}\{|Y(\omega)|^2\} = \Sigma_{Y_F}(\omega) + \Sigma_{Y_N}(\omega), \tag{50}$$

where

$$\Sigma_{Y_F}(\omega) = \mathbf{w}^H(\omega)\,\mathbf{\Sigma_F}(\omega)\,\mathbf{w}(\omega), \tag{51}$$

is the signal component of the beamformer output, and

$$\Sigma_{Y_N}(\omega) = \mathbf{w}^H(\omega)\,\mathbf{\Sigma_N}(\omega)\,\mathbf{w}(\omega), \tag{52}$$

is the noise component.

- Equations (50–52) follow directly from the assumption that $\mathbf{F}(\omega)$ and $\mathbf{N}(\omega)$ are uncorrelated.

Signal Component in Beamformer Output

- Expressing the snapshot of the desired signal once more as in (45), we find that the spatial spectral matrix $\mathbf{F}(\omega)$ of the desired signal can be written as

$$\mathbf{\Sigma_F}(\omega) = \Sigma_F(\omega)\,\mathbf{v_k}(\mathbf{k}_\mathrm{s})\,\mathbf{v_k}^H(\mathbf{k}_\mathrm{s}), \tag{53}$$

  where $\Sigma_F(\omega) = \{|F(\omega)|^2\}$.
- Substituting (53) into (51), we can calculate the output signal spectrum as

$$\Sigma_{Y_F}(\omega) = \mathbf{w}^H(\omega)\,\mathbf{v_k}(\mathbf{k}_\mathrm{s})\,\Sigma_F(\omega)\,\mathbf{v_k}^H(\mathbf{k}_\mathrm{s})\,\mathbf{w}(\omega) = \Sigma_F(\omega), \tag{54}$$

  where the final equality follows from the definition (49) of the DSB.

Array Gain of Delay-and-Sum Beamformer

- Substituting (49) into (52) it follows that the noise component present at the output of the DSB is given by

$$\begin{aligned}\Sigma_{Y_N}(\omega) &= \frac{1}{N^2}\mathbf{v}^H(\mathbf{k}_s)\,\boldsymbol{\Sigma}_{\mathbf{N}}(\omega)\,\mathbf{v}(\mathbf{k}_s)\\ &= \frac{1}{N^2}\mathbf{v}^H(\mathbf{k}_s)\boldsymbol{\rho}_{\mathbf{N}}(\omega)\mathbf{v}(\mathbf{k}_s)\Sigma_N(\omega),\end{aligned} \tag{55}$$

where the *normalized spatial spectral matrix* $\boldsymbol{\rho}_N(\omega)$ is

$$\boldsymbol{\Sigma}_{\mathbf{N}}(\omega) \triangleq \Sigma_N(\omega)\,\boldsymbol{\rho}_{\mathbf{N}}(\omega). \tag{56}$$

- Hence, the SNR at the output of the beamformer is

$$\mathrm{SNR}_{\mathrm{out}}(\omega) \triangleq \frac{\Sigma_{Y_F}(\omega)}{\Sigma_{Y_N}(\omega)} = \frac{\Sigma_F(\omega)}{\mathbf{w}^H(\omega)\,\boldsymbol{\Sigma}_{\mathbf{N}}(\omega)\mathbf{w}(\omega)}. \tag{57}$$

Array Gain of DSB (cont'd.)

- Then based on (46) and (57), we can calculate the array gain of the DSB as

$$A_{\text{dsb}}(\omega, \mathbf{k}_s) = \frac{\Sigma_{Y_F}(\omega)}{\Sigma_{Y_N}(\omega)} \Big/ \frac{\Sigma_F(\omega)}{\Sigma_N(\omega)} = \frac{N^2}{\mathbf{v}^H(\mathbf{k}_s)\, \rho_{\mathbf{N}}(\omega)\, \mathbf{v}(\mathbf{k}_s)}. \tag{58}$$

Distortionless Constraint

- Many adaptive beamforming algorithms impose a *distortionless constraint*.
- In particular, for a plane wave arriving along the main response axis $\mathbf{k}_s$

$$Y(\omega) = F(\omega), \tag{59}$$

  where $Y(\omega)$ is the beamformer output, and $F(\omega)$ is the Fourier transform of the source signal.
- It follows that

$$Y(\omega) = F(\omega)\,\mathbf{w}^H(\omega)\,\mathbf{v}(\mathbf{k}_s) = F(\omega).$$

- Hence, the distortionless constraint can be expressed as

$$\mathbf{w}^H(\omega)\,\mathbf{v}(\mathbf{k}_s) = 1. \tag{60}$$

Distortionless Constraint and the DSB

- Clearly setting
$$\mathbf{w}^H(\omega) = \frac{1}{N}\mathbf{v}^H(\mathbf{k}_s),$$
as is the case for the DSB, will satisfy (60).
- Thus, the DSB satisfies the distortionless constraint.

Snapshot Model

- Now let us characterize the noise snapshot model as a zero-mean process with spatial spectral matrix

$$\mathbf{\Sigma_N}(\omega) = E\{\mathbf{N}(\omega)\mathbf{N}^H(\omega)\} = \mathbf{\Sigma_c}(\omega) + \sigma_w^2\mathbf{I},$$

  where $\mathbf{\Sigma_c}$ and $\sigma_w^2\mathbf{I}$ are the spatially correlated and uncorrelated portions, respectively, of the noise covariance matrix.
- Spatially correlated interference is due to the propagation of some interfering signal through space.
- Uncorrelated noise is typically due to the self-noise of the sensors.

Snapshot Model (cont'd.)

- The beamformer output will be specified as

$$Y(\omega) = \mathbf{w}^H(\omega)\,\mathbf{X}(\omega) = Y_F(\omega) + Y_N(\omega).$$

- When noise is present, we can write

$$Y(\omega) = F(\omega) + Y_N(\omega),$$

where $Y_N(\omega) = \mathbf{w}^H(\omega)\mathbf{N}(\omega)$ is the noise component remaining in the beamformer output.

Optimization Criterion

- In addition to satisfying the distortionless constraint, we wish also to minimize the variance of the output.
- To solve the constrained optimization problem, we can apply the method of Lagrange multipliers.
- To wit, we first define the "symmetric" objective function

$$F \triangleq \mathbf{w}^H(\omega)\,\mathbf{\Sigma_N}(\omega)\,\mathbf{w}(\omega)+\lambda[\mathbf{w}^H(\omega)\mathbf{v}(\mathbf{k}_s)-1]+\lambda^*[\mathbf{v}^H(\mathbf{k}_s)\mathbf{w}-1],$$

  where λ is a complex Lagrange multiplier, to incorporate the constraint into the objective function.
- Taking the complex gradient with respect to $\mathbf{w}^H$, equating this gradient to zero, and solving yields

$$\mathbf{w}^H_{\text{mvdr}}(\omega) = -\lambda\,\mathbf{v}^H(\mathbf{k}_s)\,\mathbf{\Sigma}_\mathbf{N}^{-1}(\omega). \tag{61}$$

Minimum Variance Distortionless Response Beamformer

- Applying now the distortionless constraint (60), we find

$$\lambda = -\left[\mathbf{v}^H(\mathbf{k}_s)\,\boldsymbol{\Sigma}_{\mathbf{N}}^{-1}(\omega)\,\mathbf{v}(\mathbf{k}_s)\right]^{-1}.$$

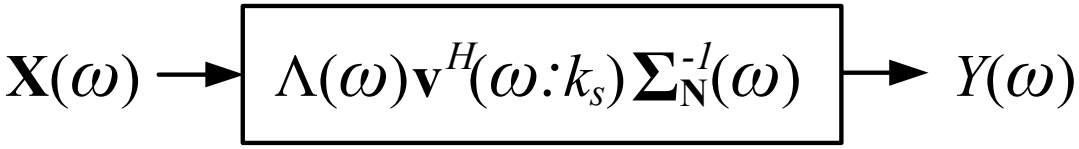
- Thus, the optimal sensor weights are given by

$$\mathbf{w}_{\mathrm{o}}^H(\omega) = \Lambda(\omega)\,\mathbf{v}^H(\mathbf{k}_s)\,\boldsymbol{\Sigma}_{\mathbf{N}}^{-1}(\omega) = \mathbf{w}_{\mathrm{mvdr}}^H(\omega), \qquad (62)$$

where

$$\Lambda(\omega) \triangleq \left[\mathbf{v}^H(\mathbf{k}_s)\,\boldsymbol{\Sigma}_{\mathbf{N}}^{-1}(\omega)\,\mathbf{v}(\mathbf{k}_s)\right]^{-1}. \qquad (63)$$

MVDR Schematic



- The figure shows a schematic of the MVDR beamformer.
- $\Lambda(\omega)$ is the spectral power of the noise component in $Y(\omega)$, as is apparent from

$$\begin{aligned}\Sigma_{Y_N}(\omega) &= \mathbf{w}_{\text{mvdr}}^H(\omega)\,\boldsymbol{\Sigma}_{\mathbf{N}}(\omega)\,\mathbf{w}_{\text{mvdr}}(\omega)\\ &= \mathbf{v}^H(\mathbf{k}_s)\,\boldsymbol{\Sigma}_{\mathbf{N}}^{-1}(\omega)\,\boldsymbol{\Sigma}_{\mathbf{N}}(\omega)\,\boldsymbol{\Sigma}_{\mathbf{N}}^{-1}(\omega)\,\mathbf{v}(\mathbf{k}_s)\cdot\Lambda^2(\omega)\\ &= \Lambda(\omega). \end{aligned} \tag{64}$$

Advantages of Subband Processing

- The foregoing implies that the sensor weights for each subband are designed independently.
- This is one of the chief advantages of subband domain adaptive beamforming.
- In particular, the transformation into the subband domain has the effect of a divide and conquer optimization scheme.
- A single optimization problem over MN free parameters, where M is the number of subbands and N is the number of sensors, is converted into M optimization problems, each with N free parameters.
- Each of the M optimization problems can be solved independently, which is a direct result of the statistical independence of the subband samples.

Array Gain of MVDR Beamformer

- As $\mathbf{w}_{\text{mvdr}}^{H}(\omega)$ satisfies the distortionless constraint, the power spectrum of the desired signal at the output of the beamformer can be expressed as

$$\Sigma_{Y_F}(\omega) = \Sigma_F(\omega),$$

  where $\Sigma_F(\omega)$ is the power spectrum of the desired signal $F(\omega)$ at the input of each sensor.
- Hence, based on (64), the output SNR can be written as

$$\frac{\Sigma_F(\omega)}{\Sigma_{Y_N}(\omega)} = \frac{\Sigma_F(\omega)}{\Lambda(\omega)}.$$

MVDR Beamformer with Plane Wave Interference

- Consider a desired signal with array manifold vector $\mathbf{v}(\mathbf{k}_s)$ and a single plane-wave interfering signal with manifold vector $\mathbf{v}(\mathbf{k}_1)$.
- In addition, there is uncorrelated sensor noise with power σ_w^2.
- In this case, the spatial spectral matrix $\mathbf{\Sigma_N}(\omega)$ of the noise can be expressed as

$$\mathbf{\Sigma_N}(\omega) = \sigma_w^2 \mathbf{I} + M_1(\omega)\, \mathbf{v}(\mathbf{k}_1)\, \mathbf{v}^H(\mathbf{k}_1), \tag{65}$$

where $M_1(\omega)$ is the spectrum of the interfering signal.

Plane Wave Interference (cont'd.)

- Applying the matrix inversion lemma to (65) provides

$$\mathbf{\Sigma_N}^{-1} = \frac{1}{\sigma_{\mathrm{w}}^2}\left[\mathbf{I} - \frac{M_1}{\sigma_{\mathrm{w}}^2 + N M_1}\mathbf{v}_1\mathbf{v}_1^H\right], \tag{66}$$

  where we have suppressed ω and $\mathbf{k}$ for convenience, and defined $\mathbf{v}_1 \triangleq \mathbf{v}(\mathbf{k}_1)$.
- The noise spectrum at each element of the array can be expressed as

$$\Sigma_N = \sigma_{\mathrm{w}}^2 + M_1. \tag{67}$$

MVDR Beamformer with Plane Wave Interference (cont'd.)

- Substituting (66) into (62), we find

$$\mathbf{w}_{\text{mvdr}}^{H} = \frac{\Lambda}{\sigma_{\text{w}}^{2}} \mathbf{v}_{\text{s}}^{H} \left[\mathbf{I} - \frac{M_1}{\sigma_{\text{w}}^{2} + N M_1} \mathbf{v}_1 \mathbf{v}_1^{H} \right]. \tag{68}$$

- The *spatial correlation coefficient* between the desired signal and the interference is defined as

$$\rho_{\text{s1}} \triangleq \frac{\mathbf{v}_{\text{s}}^{H} \mathbf{v}_1}{N}, \tag{69}$$

Plane Wave Interference (cont'd.)

- Note that

$$\rho_{s1} = B_{dsb}(\mathbf{k}_1 : \mathbf{k}_s),$$

where $B_{dsb}(\mathbf{k}_1 : \mathbf{k}_s)$ is the delay-and-sum beam pattern aimed at $\mathbf{k}_s$, the wavenumber of the desired signal, and evaluated at $\mathbf{k}_1$, the wavenumber of the interference.

- With this definition (68) can be rewritten as

$$\mathbf{w}_{mvdr}^H = \frac{\Lambda}{\sigma_w^2}\left[\mathbf{v}_s^H - \rho_{s1}\frac{NM_1}{\sigma_w^2 + NM_1}\mathbf{v}_1^H\right]. \tag{70}$$

Schematic: MVDR Beamformer with Plane Wave Interference

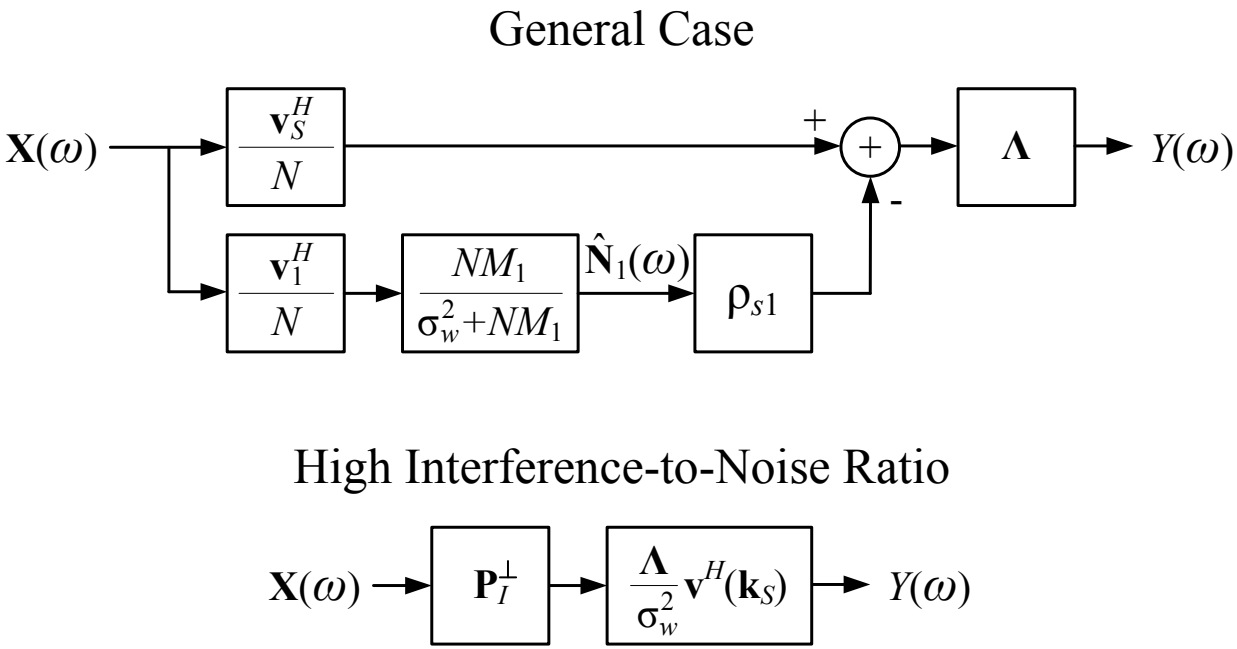


Figure: Optimum MVDR beamformer in the presence of a single interferer.

Analysis of Plane Wave Interference

- The normalizing coefficient (63) then reduces to

$$\Lambda = \left\{ \frac{1}{\sigma_{\mathrm{w}}^2} N \left[1 - \frac{NM_1}{\sigma_{\mathrm{w}}^2 + NM_1} |\rho_{\mathrm{s}1}|^2 \right] \right\}^{-1} . \tag{71}$$

- It is clear that the upper and lower branches of this MVDR beamformer correspond to conventional beamformers pointing at the desired signal and the interference, respectively.
- The necessity of the bottom branch is readily apparent if we reason as follows:
  - The path labeled $\hat{\mathbf{N}}_1(\omega)$ is the minimum mean-square estimate of the interference plus noise.
  - This noise estimate is scaled by $\rho_{\mathrm{s}1}$ and subtracted from the output of the DSB in the upper path, in order to remove that portion of the noise and interference captured by the upper path.

Analysis (cont'd.)

- Observe that in the case where $NM_1 \gg \sigma_{\mathrm{w}}^2$, we may rewrite (70) as

$$\mathbf{w}_{\mathrm{mvdr}}^H = \frac{\Lambda}{\sigma_{\mathrm{w}}^2} \mathbf{v}_{\mathrm{s}}^H \mathbf{P}_I^{\perp},$$

  where $\mathbf{P}_I^{\perp} = \mathbf{I} - \mathbf{v}_1 \mathbf{v}_1^H$ is the projection matrix onto the space orthogonal to the interference.
- This case is shown schematically in Figure 9, which indicates that the beamformer is placing a perfect null on the interference.

Limiting Case: Plane Wave Interference

- Substituting (67) and (71) and provides

$$A_{\text{mvdr}} = N(1+\sigma_{\text{I}}^2)\left[\frac{1+N\sigma_{\text{I}}^2(1-|\rho_{\text{s1}}|^2)}{1+N\sigma_{\text{I}}^2}\right],$$

  where the *interference-to-noise ratio* (INR), defined as

$$\sigma_{\text{I}}^2 \triangleq \frac{M_1}{\sigma_{\text{w}}^2},$$

  is the ratio of spatially correlated to uncorrelated noise.
- Observe that the suppression of the interference is not perfect when either σ_{I}^2 is verly low, or u_{I} is very small such that the interference moves within the main lobe region of the delay-and-sum beam pattern.

MVDR Beam Patterns

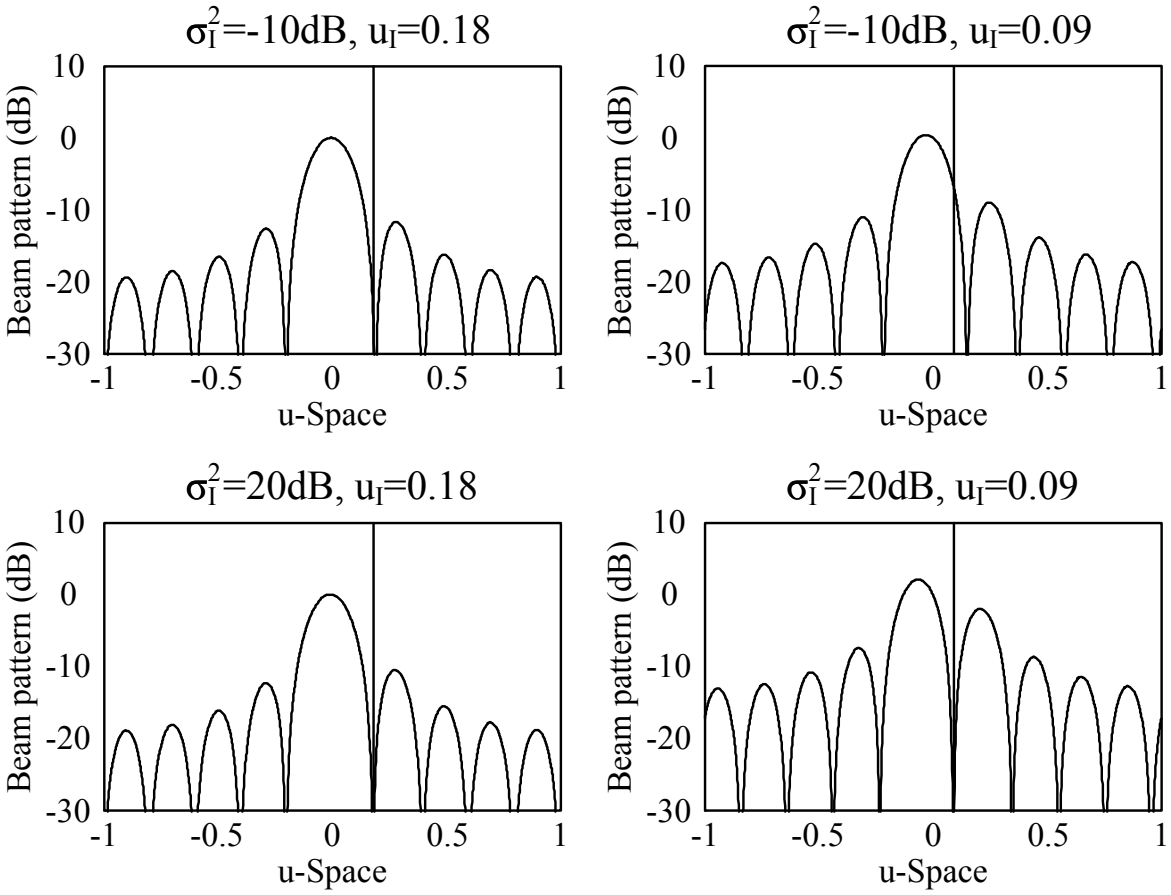


Figure: MVDR beam patterns for single plane wave interference

Performance Curves

- The array gain of the DSB in the presence of a single interferer is readily obtained by substituting (65) and (67) into (58), whereupon we find

$$A_{\text{dsb}}(\omega, \mathbf{k}_{\text{s}}) = \frac{N^2(\sigma_{\text{w}}^2 + M_1)}{\mathbf{v}_{\text{s}}^H(\sigma_{\text{w}}^2\mathbf{I} + M_1\mathbf{v}_1\mathbf{v}_1^H)\mathbf{v}_{\text{s}}} = \frac{N(1+\sigma_{\text{I}}^2)}{1+\sigma_{\text{I}}^2 N|\rho_{\text{s}1}|^2}.$$

- The array gains for both DSB and optimal MVDR beamformer at various INR levels are plotted in the figure.

Summary

- In this lecture, we presented the principal components and steps in the Kalman filter.
- We also discussed how these components can be sequentially updated.
- In addition, we presented the interpretation of the Kalman filter as either the:
  - optimal minimum mean square error estimator, or
  - optimal Bayesian filter.
- We also presented two variations of the classic Kalman filter:
  - the probabilistic data association filter;
  - the joint probabilistic data association filter.
- In addition, we covered the basics of beamforming.