# Optimal Filtering
# III: Signal Manifolds

ECE416 Adaptive Algorithms

Prof. Fred L. Fontaine

Department of Electrical Engineering
The Cooper Union

Spring 2022

# Signal Manifolds

Sometimes the signal space of interest is not a linear space.

Consider for example signals embedded in a possibly high dimensional linear space, but represented by a small number of parameters.

A simple example is a sinewave:

$$\mathbf{s}\left(\omega\right) = \begin{bmatrix} 1 \\ e^{j\omega} \\ \vdots \\ e^{j(M-1)\omega} \end{bmatrix} \in \mathbb{C}^M$$

An expression of the form $\sum_i \alpha_i \mathbf{s}\left(\omega_i\right)$ is not itself a sinewave, and hence the collection $\{\mathbf{s}\left(\omega\right)\}$ do not form a linear space.

# Signal Manifolds

If a class of signal vectors $\mathbf{s}(\Theta)$ depend on an $K$-dimensional parameter vector $\Theta$, then the signals can be thought of as lying on a $K$-dimensional surface in $M$-dimensional space. These surfaces are, in general curved, and are called *signal manifolds.*

The mathematical definition of a manifold is a smooth (differentiable) surface, although this condition will not always apply to what are called signal manifolds.

Before discussing the problem from a more general perspective, we will present one more specific example.

# Sensor Arrays

We consider a set of sensors at spatial locations $\vec{r}_m$, $1 \leq m \leq M$, that measure a wave or field. Examples include acoustic sensors (microphones), or electromagnetic sensors (antennas).

We first briefly review the concept of waves and especially plane waves, viewed at baseband (i.e., phasor domain).

Modern arrays can have anywhere from a handful to hundreds of sensors!

# Waves

A space-time function $g\left(\vec{r}, t\right)$ is a *wave* if it satisfies the wave equation:

$$\nabla^2 g - \frac{1}{v^2}\frac{\partial^2 g}{\partial t^2} = 0$$

where $v$ is a fixed parameter (that will turn out to be *velocity)*. An example is:

$$g\left(z, t\right) = A\cos\left(\omega t - \beta z + \phi_0\right)$$

which can be shown to be a wave if:

$$v = \frac{\omega}{\beta}$$

# Travelling Waves

This is a *travelling wave:* it propagates in the $+z$-direction with velocity $v$ in the sense that:

$$g\left(z+\Delta z, t+\Delta t\right) = g\left(z, t\right) \text{ if } \Delta z = v\Delta t$$

In other words, time $\Delta t$ later, the wave has moved through a distance $\Delta z = v\Delta t$.

# Phasor Form of a Wave

With $\omega = 2\pi f$, and observe the spatial periodicity $\lambda$, called wavelength, is related to $\beta$ via:

$$\beta = \frac{2\pi}{\lambda}$$

we get the familiar equation:

$$v = f\lambda$$

It will be convenient to view the function in the phasor domain:

$$g\left(z\right) = \left(Ae^{j\phi}\right) e^{-j\beta z}$$

# Phasor Form of a Wave

Since $\frac{\partial}{\partial t} \longrightarrow j\omega$ in phasor form, the wave equation becomes the *Helmholtz equation:*

$$\nabla^2 g + k^2 g = 0$$

where:

$$k = \frac{\omega}{v} = \frac{2\pi}{\lambda}$$

Observe in this case we must have $\beta = k$.

# Plane Waves

More generally, consider a function of the form:

$$g\left(\vec{r}\right) = e^{-j\vec{k}\cdot\vec{r}}$$

It is called a *plane wave* because the surface of constant phase is $\vec{k}\cdot\vec{r}$=constant, which is a plane.

The surface of constant phase is called a *wavefront*, which here is a plane.

Note that $\vec{k} \perp$ the plane.

# Plane Waves

It can be shown a plane wave satisfies the wave equation if:

$$\left|\vec{k}\right| = \frac{2\pi}{\lambda}$$

Back in the time domain this is:

$$g\left(\vec{r}, t\right) = \cos\left(\omega t - \vec{k} \cdot \vec{r}\right)$$

Writing:

$$\vec{k} = \frac{2\pi}{\lambda}\hat{a}_k$$

where $\hat{a}_k$ is a unit vector in the direction of $k$, we see this a travelling wave that propagates in the $\hat{a}_k$-direction.

That is, if $\Delta\vec{r} = \Delta r \hat{a}_k$, then $g\left(\vec{r} + \Delta\vec{r}, t + \Delta t\right) = g\left(\vec{r}, t\right)$ when $\Delta r = v\Delta t$.

It can be shown that any wave can be expressed as the superposition of plane waves (via Fourier analysis).

A general space-time Fourier representation is:

$$g\left(\vec{r},t\right)=\frac{1}{\left(2\pi\right)^4}\int_{\vec{k}\in\mathbb{R}^3}\int_{\omega=-\infty}^{\infty}G\left(\vec{k},\omega\right)e^{j\left(\omega t-\vec{k}\cdot\vec{r}\right)}d\vec{k}d\omega$$

For $g$ to satisfy the wave equation with velocity parameter $v$, then the support of $G\left(\vec{k},\omega\right)$ must lie on the cone:

$$\left|\vec{k}\right|=\frac{\omega}{v}$$

$\vec{k}$ is called the *wavenumber vector*, and $\left|\vec{k}\right|$ is called the *wavenumber*. It has units of $rad/m$.

# Narrowband Sensor Arrays

Assuming the velocity $v$ is fixed, if the frequency $f$ is also fixed then $\lambda = v/f$ is a constant and all waves are superpositions of plane waves with:

$$\vec{k} = \frac{2\pi}{\lambda}\hat{a}_k$$

where the only freedom of choice is in the direction of $\hat{a}_k$.

This is actually a *narrowband model:* we treat $f$ and hence $\lambda$ as constant. We will discuss *narrowband techniques only!*

# Wideband Sensor Arrays

A more complicated situation is where the frequency band is wide enough that $\lambda$ may be variable. Then we should write:

$$\vec{k} = \frac{2\pi}{v} f \hat{a}_k$$

where the variable parameters are not only the unit vector $\hat{a}_k$ but also $f$.

In this case, it is called a *wideband model* and more sophisticated techniques would be required.

In many cases, using a narrowband model is reasonable if the bandwidth is below 10% of the nominal (carrier) frequency.

We will use spherical coordinates: $\theta$ is the *polar angle* off the $z$-axis, $0 \le \theta \le \pi$, with 0 the "north pole", $\pi/2$ the "equator", and $\pi$ the "south pole." In other words, $\theta$ is like latitude. The *azimuth* $\phi$ is the angle off the $x$-axis in the $xy$-plane. In other words, $\phi$ is like longitude, $0 \le \phi \le 2\pi$.

With $\Theta = (\theta, \phi)$, called the *angle of arrival (AOA):*

$$\hat{a}\left(\Theta\right) = \cos\phi sin\theta\hat{a}_x + sin\phi sin\theta\hat{a}_y + \cos\theta\hat{a}_z$$

# Steering Vector for Sensor Arrays

The output of the sensor array, assuming a unit amplitude, zero reference phase plane wave with AOA $\Theta$ is incident on the array, is:

$$\mathbf{s}\left(\Theta\right) = \begin{bmatrix} e^{-j\frac{2\pi}{\lambda}\hat{a}_k(\Theta)\cdot\vec{r}_1} \\ \vdots \\ e^{-j\frac{2\pi}{\lambda}\hat{a}_k(\Theta)\cdot\vec{r}_M} \end{bmatrix}$$

This is called a *steering vector*, and the set of all steering vectors, parametrized by $\Theta = (\theta, \phi)$, comprise the *array manifold*.

# Sensor Arrays

Normally the array manifold is two-dimensional– it characterized by two parameters, but is *not* a linear space!

In some cases, however, it may actually be only one-dimensional. This depends on the locations of the sensors (as we shall see later).

Also note a pair of steering vectors at different angles are not necessarily linearly independent! In fact, we shall see, it may even be possible that $\mathbf{s}\left(\Theta_1\right) = \mathbf{s}\left(\Theta_2\right)$ for two different AOA's!

They are certainly not necessarily orthogonal, although in many cases we may expect because of the phasing relations that the inner product $\mathbf{s}^H\left(\Theta_1\right) s\left(\Theta_2\right)$ will be small.

# Uniform Linear Array

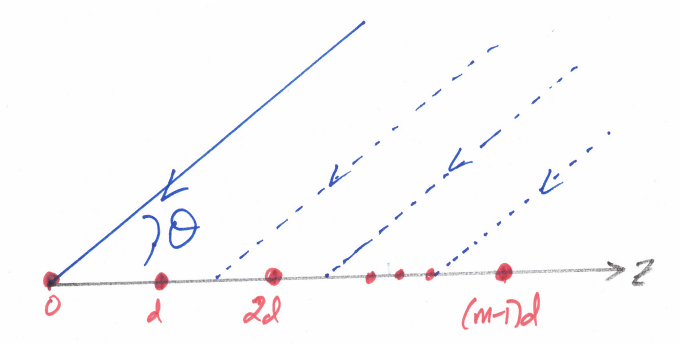
An important special case is a *uniform linear array* in which the sensors are at locations:

$$\vec{r}_m = (m-1)\, d\hat{a}_z, 1 \leq m \leq M$$

They are equally spaced along the $z$-axis. In this case:

$$\hat{a}_k(\Theta) \cdot \vec{r}_m = (m-1)\frac{2\pi d}{\lambda}\cos\theta$$

# Uniform Linear Array



# Uniform Linear Array

Note the azimuth $\phi$ drops out. The sensor array cannot distinguish among waves with AOA rotated around the axis of the array.

Only the angle $\theta$ relative to the axis of the array can be discerned.

Also note that in some cases, instead of the polar angle $\theta$ (the angle off the array axis), people work with $\theta'$ the angle relative to the normal, and then $\cos\theta$ gets replaced by $sin\theta'$, but we will continue to work with the cosine form.

# Uniform Linear Array

Define the *electrical angle* $\hat{\omega}$ as:

$$\hat{\omega} = \frac{2\pi d}{\lambda} \cos\theta$$

Then the steering vector can be parametrized by $\hat{\omega}$ (as $\theta : 0 \longrightarrow \pi$, $\cos\theta : -1 \longrightarrow 1$ one-to-one):

$$\mathbf{s}\left(\hat{\omega}\right) = \begin{bmatrix} 1 \\ e^{-j\hat{\omega}} \\ \vdots \\ e^{-j(M-1)\hat{\omega}} \end{bmatrix}$$

# Uniform Linear Array

Thus the steering vector is equivalent to a *discrete-time sinewave at normalized radian digital frequency* $\hat{\omega}$ (rad), of length $M$.

Note that the sign of $\hat{\omega}$ matters, and $\pm\hat{\omega}$ correspond to different AOAs.

Go back to:

$$\hat{\omega} = \frac{2\pi d}{\lambda} \cos\theta$$

If two electrical angles differ by multiples of $2\pi$, the steering vectors are *identical* and hence corresponding AOA's would be indistinguishable!

The range of values are:

$$-\frac{2\pi d}{\lambda} \leq \hat{\omega} \leq \frac{2\pi d}{\lambda}$$

There are two possibilities, depending on $\frac{2\pi d}{\lambda} \gtrless \pi$, in other words, $d \gtrless \lambda/2$.

# Uniform Linear Array

If $d \leq \lambda/2$, then:

$$-\pi \leq -\frac{2\pi d}{\lambda} \leq \hat{\omega} \leq \frac{2\pi d}{\lambda} \leq \pi$$

so every AOA $\theta$ yields a distinguishable electrical angle $\hat{\omega}$, and all the steering vectors are distinct. Note that there is a range of digital frequencies, however, i.e., :

$$\frac{2\pi d}{\lambda} < |\hat{\omega}| < \pi$$

which do not correspond to a physical AOA. This is called the *invisible region*, while $|\hat{\omega}| < \frac{2\pi d}{\lambda}$ is the *visible region*.

# Uniform Linear Array

If $d > \lambda/2$, then there will exist two different AOA's $\theta$ that lead to $\hat{\omega}$ values separated by $2\pi$, and hence the steering vectors will be *identical*. This is called *spatial aliasing*.

In practice, can we allow $d > \lambda/2$? YES!

# Isotropic Versus Directional Sensors

We have been assuming *isotropic sensors:* they sense the field the same from all directions.

Real sensors will produce different output powers for equal intensity incident plane waves depending on the AOA, and are characterized by a *directional gain* function.

# Isotropic Versus Directional Sensors

In the simplest case, suppose a sensor measures the incident wave with unit amplitude scaling if $\{\theta_1 < \theta < \theta_2\}$, and 0 if $\theta$ is outside this range. This range is called the *main beam* of the sensors.

Using such *directional sensors* can avoid the spatial aliasing problem, and we can have sensors spread further apart (with the disadvantage being that a more limited spatial range can be measured).

From this point on we will assume isotropic sensors.

# General Sensor Array Output

Assume there are $L$ sources with AOA's $\{\Theta_\ell\}_{1\leq\ell\leq L}$. A general output would have the form:

$$\mathbf{x} = \sum_{\ell=1}^{L} \beta_\ell \mathbf{s}\left(\Theta_\ell\right) + \mathbf{v}$$

where $\boldsymbol{\beta} = \left[\beta_1 \cdots \beta_L\right]^T$ represent the amplitude and phase of the signals "riding" on each of the plane waves, and $\mathbf{v}$ is 0-mean "noise" uncorrelated (and orthogonal, in a statistical sense) to $\mathbf{x}$.

# General Sensor Array Output

Before proceeding, it will be convenient to use steering vectors of unit length. So from now on let us define:

$$\mathbf{s}\left(\Theta_\ell\right) = \frac{1}{\sqrt{M}} \left[e^{-j\vec{k}\cdot\vec{r}_m}\right]_{1\leq m\leq M}$$

In principle, $\beta_\ell$'s could be deterministic but it will be more convenient to model them as random.

# Snapshots

In reality, we don't measure the output only once. We collect several *snapshots* across time: $x[n], 1 \leq n \leq N$:

$$\mathbf{x}[n] = \sum_{\ell=1}^{L} \beta_\ell [n] \, \mathbf{s}(\Theta_\ell) + \mathbf{v}[n]$$

Variables: $\ell$ is the source index, $n$ is the time index. Not shown is the index $m$, $1 \leq m \leq M$, is the sensor or *spatial* index. We can stack the snapshots as columns in a matrix.

# Data Model: Final Form

Define our data matrix as:

$$X = [\mathbf{x}[1] \ \ \mathbf{x}[2] \ \ \cdots \ \ \mathbf{x}[N]]$$

$X$ is $M \times N$ (space $\times$ time), and:

$$X = SB + v$$

where $B = \left[ \ \boldsymbol{\beta}[1] \quad \cdots \quad \boldsymbol{\beta}[N] \ \right]$ is the information riding on the sources, is $L \times N$ (source $\times$ time), $S = [\mathbf{s}(\Theta_1) \cdots \mathbf{s}(\Theta_L)]$ is the matrix whose columns are the source steering vectors (space $\times$ source), and $v = [\mathbf{v}[1] \cdots \mathbf{v}[N]]$ is the noise matrix (space $\times$ time).

# Data Model: Final Form

Note that we need to keep track of each dimension here: spatial (sensor index), time (snapshot) index, and source index. This allows us to interpret operations, e.g.:

$$
\begin{aligned}
SB \quad &\sim \quad [\text{space} \times \text{source}] \cdot [\text{source} \times \text{time}] \\
&\sim \quad \text{sum over the sources for each } (\text{space,time}) \\
&\longrightarrow \quad [\text{space} \times \text{time}]
\end{aligned}
$$

We also specify our statistical model based on these "axes". For example we usually assume the noise is 0-mean white across space and time, meaning it is uncorrelated with itself across space and time, and is also uncorrelated with $S$ (in case $\Theta_\ell$ are random) and $B$.

# Data Model: Final Form

We assume that $\boldsymbol{\beta}[n]$ is stationary with correlation matrix, $R_\beta$.

They are usually 0-mean and uncorrelated, which means $R_\beta$ is diagonal, but this is not always the case.

For example, there may be a multipath environment, where the waves undergo reflections off multiple surfaces so the same source may lead to multiple incident AOA's. This leads to correlated sources.

# Data Correlation Matrix

$$R_X = SR_\beta S^H + \sigma_v^2 I$$

Usually $L \ll M$, but definitely $L < M$ (if we are to properly discern the sources). We also prefer $N > M$, usually $N \approx 2M$ suffices.

We assume the noise power $\sigma_v^2$ is small compared to the signal power, i.e., $\|R_\beta\| = \lambda_{\max}\left(R_\beta\right)$. Thus we have a strong SNR at least with the dominant sources.

# Data Correlation Matrix

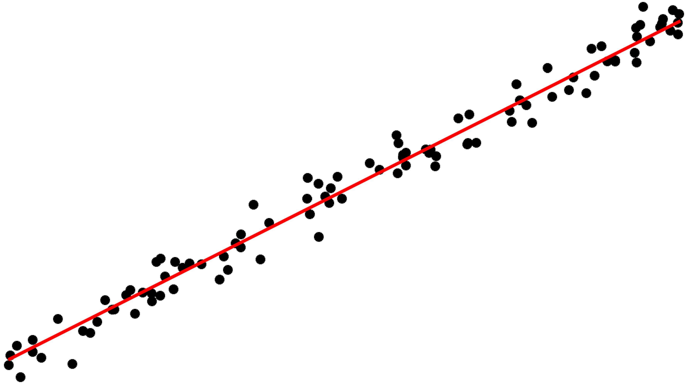
It is possible to discern sources with negative SNR, but of course that is more difficult.

Although $R_X$ is $M \times M$, it only has approximate rank $L$. The presence of noise is what gives it full rank $M$. A more precise statement is that $R_X$ can be well approximated by a low rank matrix.

Viewed another way, at each snapshot, the output of the array is *approximately* a linear combination of a (relatively) small number of steering vectors.

# Low Dimensional Signal Subspace

The signal vectors (at different snapshots) are clustered near a low-dimensional linear subspace, that is spanned by canonical vectors chosen from a low-dimensional manifold. Can we recover those canonical vectors?



# Identifying the Signal Subspace

Define $\mathcal{S} = span\left\{\mathbf{s}\left(\Theta_\ell\right)\right\}_{1\leq\ell\leq L}$, the "ideal" signal subspace. Then *in a loose sense:*

$$span\left\{\mathbf{x}\left[n\right]\right\} \approx \mathcal{S}$$

This suggests the following approach.

Perform an SVD on $X$, and we should identify $L$ "dominant" singular values, which tells us there are $L$ sources. Then take the first $L$ left singular vectors:

$$U_L = \left[\mathbf{u}_1 \cdots \mathbf{u}_L\right]$$

# Identifying the Signal Subspace

We could also do an eigendecomposition of $R_x$:

$$R_x \approx \frac{1}{N} X X^H \approx \frac{1}{N} U_L diag\left\{\sigma_1^2 \cdots \sigma_L^2\right\} U_L^H = \frac{1}{N} \sum_{\ell=1}^{L} \sigma_\ell^2 \mathbf{u}_\ell \mathbf{u}_\ell^H$$

*The range space of $U_L$ is an approximation to the signal subspace $\mathcal{S}$.*

# Identifying the Steering Vectors Spanning the Signal Subspace

Note that the $\mathbf{u}_\ell$'s are orthonormal, while the underlying steering vectors are not.

So we cannot expect the singular vectors to be actual steering vectors (they will not have the particular mathematical form of steering vectors).

But they will collectively (hopefully) span the same subspace.

# Identifying the Steering Vectors Spanning the Signal Subspace

It is still not clear how to identify the steering vectors from this information.

It turns out the techniques we will cover (based on MVDR spectrum and MUSIC spectrum) are closely related to the concept of *beamforming*, so we will discuss that first.

# Steering the Array

Given a sensor array output $\mathbf{x}$, say we want to recover the information riding on a particular steering vector $\mathbf{s}\left(\Theta_0\right)$. This would suggest:

$$y = \mathbf{s}^H\left(\Theta_0\right)\mathbf{x}$$

Since $\mathbf{s}\left(\Theta_0\right)$ is a unit vector, this is a projection onto the $\mathbf{s}\left(\Theta_0\right)$ direction.

# Steering the Array

Viewed another way:

$$y = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} e^{j\vec{k}(\Theta_0)\cdot\vec{r}_m} x_m$$

By combining the individual sensor outputs with certain phase factors, we can cancel out the relative phases for the prescribed AOA.

This yields a maximal combining effect for the case of an incident signal at the given AOA.

# Steering the Array

Indeed, originally beamforming was accomplished with analog circuitry, which was constrained (for physical reasons) to usually implement phase shifts only (i.e., not gain or attenuation), hence the term *phased arrays*.

Even today, limitations of the speed, precision and cost of A/D converters means that it is not uncommon that at least some beamforming is performed in the analog domain, with additional computations performed after digitization.

For example, subsets of sensors (subarrays) can be linearly combined separately, then digitized in parallel for further processing.

# Steering the Array

In general, $\mathbf{x}$ may have other sources present (from other AOAs), as well as noise.

Because steering vectors are not necessarily orthogonal, the simple scheme proposed above, i.e., $y = \mathbf{s}^H(\Theta_0)\,\mathbf{x}$, may produce a strong interference from the other sources.

It may be advantageous to try to cancel out some of the other sources, not simply to maximize the output due to the desired input signal.

**OPTIMAL FILTERING**
**III: SIGNAL MANIFOLDS**

**ECE416 ADAPTIVE ALGORITHMS**