National Institute of Standards and Technology
U.S. Department of Commerce

Gradient-Based Solution of Maximum Likelihood Angle Estimation for Virtual Array Measurements

Peter Vouras

Wireless Networks Division (673)
Communications Technology Lab

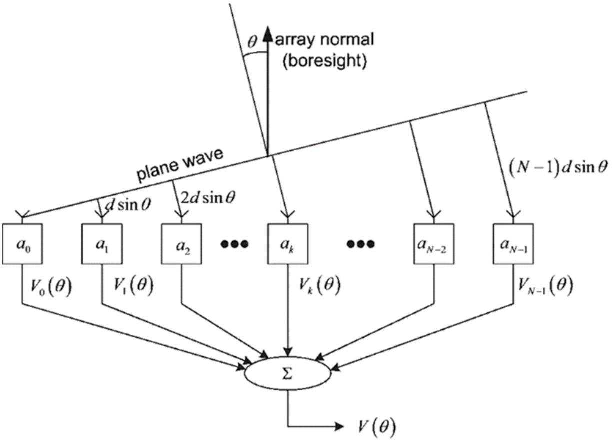
29 November 2018

Overview

- Synthetic Aperture or Virtual Array Concept
- Advantages of Synthetic Apertures
- Overview of Maximum Likelihood Angle Estimation
- Alternating Projections Algorithm
- Initial Conditions
- Measured Results
- Summary

Synthetic Aperture Concept

Phased Array Operation



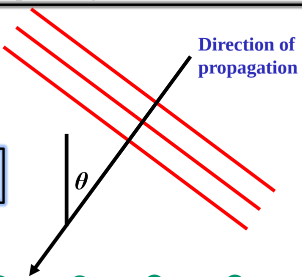
As a plane wave propagates across the aperture, the incremental delay between elements corresponds to a phase shift of $\frac{2\pi}{\lambda} n d \sin\theta$

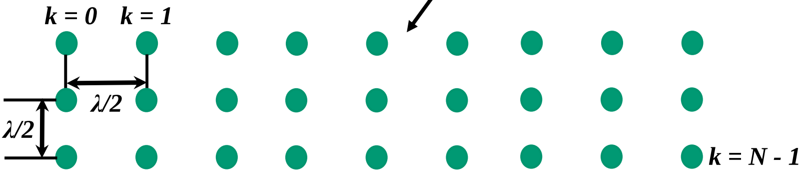
Synthetic Aperture

Want to capture a propagating plane wave by uniformly sampling space along a planar grid

Assume signal arrives from a source that is far enough away that the signal wavefronts are well approximated by straight lines (far field assumption)

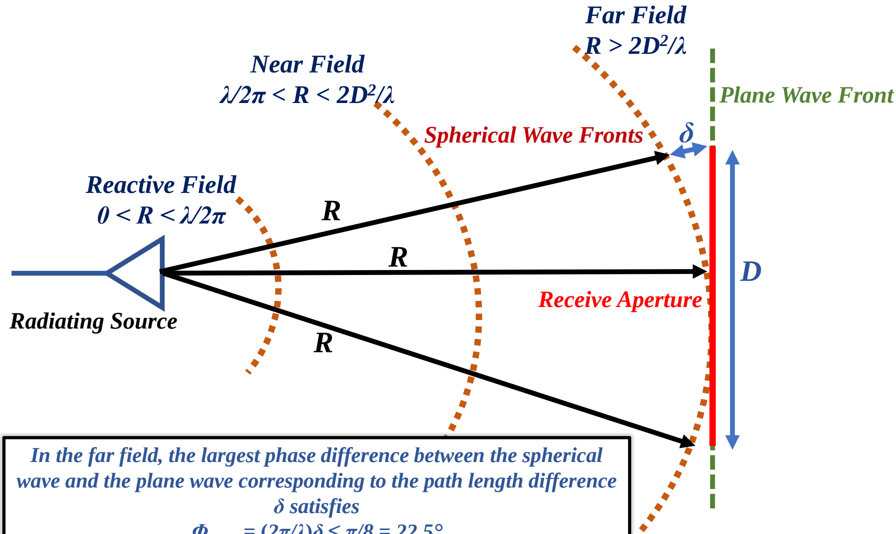
Far field distance = $2D^2/\lambda$ where D is the largest dimension of the antenna aperture





Sample space at $N=N_xN_y$ locations spaced $\lambda/2$ apart. This samples the virtual antenna aperture at twice the spatial frequency of the carrier to prevent spatial aliasing (grating lobes) from occurring

Plane Wave Assumption

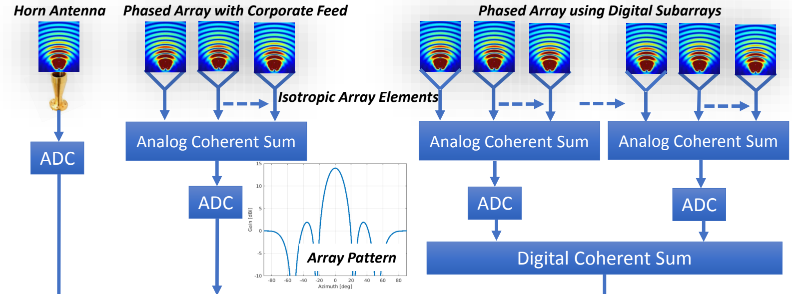


In the far field, the largest phase difference between the spherical wave and the plane wave corresponding to the path length difference δ satisfies

$$\Phi_{error} = (2\pi/\lambda)\delta \leq \pi/8 = 22.5°$$

Advantages of Synthetic Apertures

Evolution of Digital Arrays

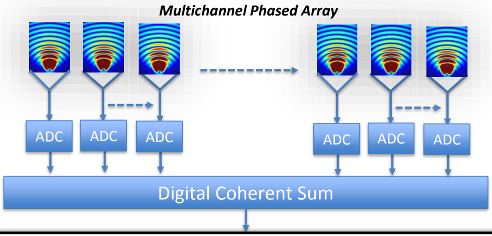


1. Dynamic range limited: (ADC sees full signal amplitude including horn gain)
2. Binary AOA estimation: (signal in main beam or not)

1. Dynamic range limited: (ADC sees full integration gain of array beamformer)
2. Binary AOA estimation: (signal in main beam or not, no AOA performance in sidelobes)

1. Dynamic range improvement: (ADC sees lower integration gain of subarray)
2. Improved AOA estimation possible if digitized subarray outputs are available. Otherwise binary AOA performance (signal in main beam or not)
3. Scan angles are limited to subarray beamwidth

Every Element Digitization



1. **Maximum dynamic range: ADC sees signal output of single element**
2. **Maximum AOA performance: Full resolution of array aperture available**
3. **Output of every element digitized: Maximum degrees of freedom for AOA**
4. **Arbitrary scan angles**
5. **Hugely expensive**
6. **Lots of data**

Synthetic aperture attains AOA estimation performance equivalent to a large multichannel phased array at a tiny fraction of the cost!

Experiment Configuration

Laboratory Environment

Synthetic aperture created using precise mechanical positioner

Overview of Maximum Likelihood Estimation

Array Signal Model

The complex signals received by the virtual array are,

$$\mathbf{x}(k) = \sum_{j=1}^{D} \mathbf{a}(u_j, v_j) s_k(k) + \mathbf{n}(k)$$

where $\mathbf{a}(u, v)$ is a steering vector, $s_j(k)$ is a baseband sample of the jth signal, $\mathbf{n}(k)$ is additive noise vector, D is the number of signal sources

Array Signal Model

The complex signals received by the virtual array are,

$$\mathbf{x}(k) = \sum_{j=1}^{D} \mathbf{a}(u_j, v_j) s_k(k) + \mathbf{n}(k)$$

where $\mathbf{a}(u, v)$ is a steering vector, $s_j(k)$ is a baseband sample of the jth signal, $\mathbf{n}(k)$ is additive noise vector, D is the number of signal sources

For data collected over N array snapshots,

$$\mathbf{X} = \mathbf{A}(\mathbf{u}, \mathbf{v})\mathbf{S} + \mathbf{N}$$

where $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$, $\mathbf{A}(\mathbf{u}, \mathbf{v}) = [\mathbf{a}(u_1, v_1), \ldots, \mathbf{a}(u_N, v_N)]$,

$\mathbf{N} = [\mathbf{n}(1), \ldots, \mathbf{n}(N)]$, $\mathbf{S} = [\mathbf{s}(1), \ldots, \mathbf{s}(N)]$

Steps to Compute ML Estimator

- Compute joint probability density function of the sampled data
- Compute the likelihood function
- Maximize the likelihood function with respect to the unknown parameters

Joint PDF and Likelihood Function

The joint PDF of $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$ is,

$$f(\mathbf{X}) = \prod_{k=1}^{N} \frac{1}{\pi \det[\sigma^2 \mathbf{I}]} \exp\left[\frac{1}{\sigma^2} \left|\mathbf{x}(k) - \mathbf{A}(\mathbf{u}, \mathbf{v})\mathbf{s}(k)\right|^2\right]$$

Joint PDF and Likelihood Function

The joint PDF of $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$ is,

$$f(\mathbf{X}) = \prod_{k=1}^{N} \frac{1}{\pi \det[\sigma^2 \mathbf{I}]} \exp\left[\frac{1}{\sigma^2} \left|\mathbf{x}(k) - \mathbf{A}(\mathbf{u}, \mathbf{v})\mathbf{s}(k)\right|^2\right]$$

The log-likelihood function is,

$$J(u, v) = -ND \log \sigma^2 - \frac{1}{\sigma^2} \sum_{k=1}^{N} \left|\mathbf{x}(k) - \mathbf{A}(\mathbf{u}, \mathbf{v})\mathbf{s}(k)\right|^2$$

Optimization Function

The optimization program to be solved is,

$$\max_{\mathbf{u},\mathbf{v}} \quad J(\mathbf{u},\mathbf{v}) = \max_{\mathbf{u},\mathbf{v}} \quad \mathrm{tr}[P_{\mathbf{A}(\mathbf{u},\mathbf{v})}\hat{\mathbf{R}}_{xx}]$$

Optimization Function

The optimization program to be solved is,

$$\max_{\mathbf{u},\mathbf{v}} \quad J(\mathbf{u},\mathbf{v}) = \max_{\mathbf{u},\mathbf{v}} \quad \mathrm{tr}[P_{\mathbf{A}(\mathbf{u},\mathbf{v})}\hat{\mathbf{R}}_{xx}]$$

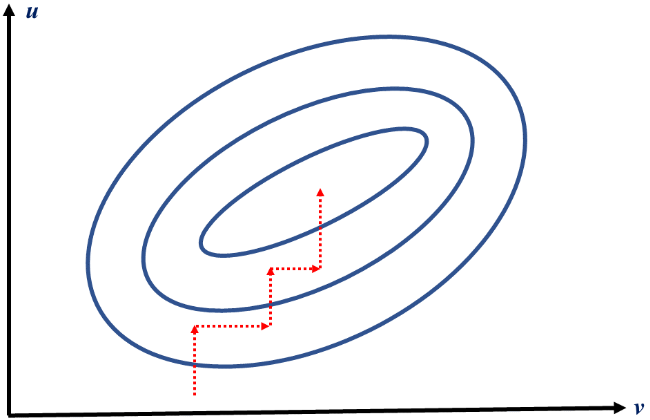
$$P_{\mathbf{A}(\mathbf{u},\mathbf{v})} = \mathbf{A}(\mathbf{u},\mathbf{v})[\mathbf{A}(\mathbf{u},\mathbf{v})^H\mathbf{A}(\mathbf{u},\mathbf{v})]^{-1}\mathbf{A}(\mathbf{u},\mathbf{v})^H = \mathbf{A}(\mathbf{u},\mathbf{v})\mathbf{A}(\mathbf{u},\mathbf{v})^{\dagger}$$

and $\hat{\mathbf{R}}_{xx} = \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{x}(k)\mathbf{x}(k)^H$

Solving the Optimization Program

- The ML angle estimates are obtained by searching over the array manifold for those D steering vectors that form a D-dimensional signal subspace closest to the measured data vectors $\mathbf{x}(1), \ldots, \mathbf{x}(N)$
- "Closeness" is measured by the norm of the projection of the $\mathbf{x}(k)$ vectors onto the signal subspace
- The Alternating Projections (AP) algorithm maximizes $J(\mathbf{u}, \mathbf{v})$ with respect to one parameter while holding the other parameters fixed
- Since $J(u_k, v_k)$ will have multiple local maximas, proper initialization is critical for global convergence

Alternating Projections Algorithm



Algorithm searches for peak along lines parallel to the coordinate axes

Gradient Vector and Hessian Matrix

The gradient vector is determined by computing the directional derivative of the cost function $J(\mathbf{p};\mathbf{d})$,

$$J'(\mathbf{p};\mathbf{d}) = \nabla J(\mathbf{p})^T \mathbf{d},$$

$$\text{for } \mathbf{p} = [u_k \quad v_k]^T, \mathbf{d} = [\delta_{u_k} \quad \delta_{v_k}]^T$$

Gradient Vector and Hessian Matrix

The gradient vector is determined by computing the directional derivative of the cost function $J(\mathbf{p};\mathbf{d})$,

$$J'(\mathbf{p};\mathbf{d}) = \nabla J(\mathbf{p})^T\mathbf{d},$$

$$\text{for } \mathbf{p} = [u_k \quad v_k]^T, \mathbf{d} = [\delta_{u_k} \quad \delta_{v_k}]^T$$

The Hessian matrix $\mathbf{H}(\mathbf{p})$ is determined by computing the Taylor series expansion of $J(\mathbf{p}+\mathbf{d})$ and retaining the second order terms,

$$J(\mathbf{p}+\mathbf{d}) \approx \mathbf{d}^T\mathbf{H}(\mathbf{p})\mathbf{d} + \text{ other terms}$$

Gradient Vector and Hessian Matrix

The gradient vector is determined by computing the directional derivative of the cost function $J(\mathbf{p};\mathbf{d})$,

$$J'(\mathbf{p};\mathbf{d}) = \nabla J(\mathbf{p})^T\mathbf{d},$$

$$\text{for } \mathbf{p} = [u_k \quad v_k]^T, \mathbf{d} = [\delta_{u_k} \quad \delta_{v_k}]^T$$

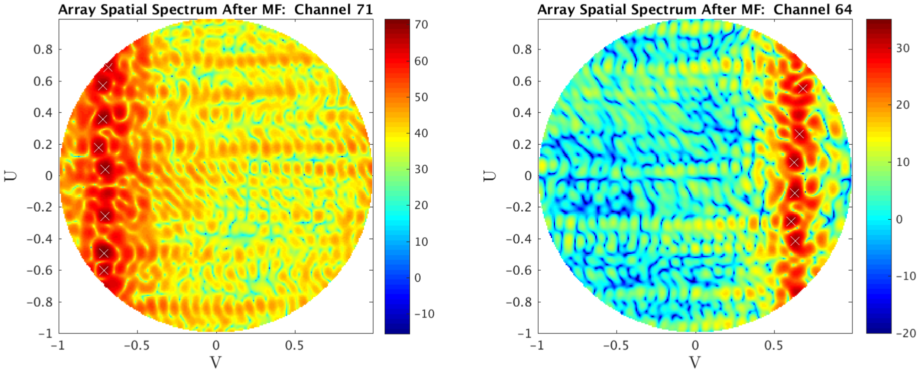
The Hessian matrix $\mathbf{H}(\mathbf{p})$ is determined by computing the Taylor series expansion of $J(\mathbf{p}+\mathbf{d})$ and retaining the second order terms,

$$J(\mathbf{p}+\mathbf{d}) \approx \mathbf{d}^T\mathbf{H}(\mathbf{p})\mathbf{d} + \text{ other terms}$$

The gradient vector and the Hessian matrix can be used to implement the conjugate gradient algorithm and Newton's method respectively

Alternating Projections Algorithm Initialization

Spatial Spectrum of Raw Data

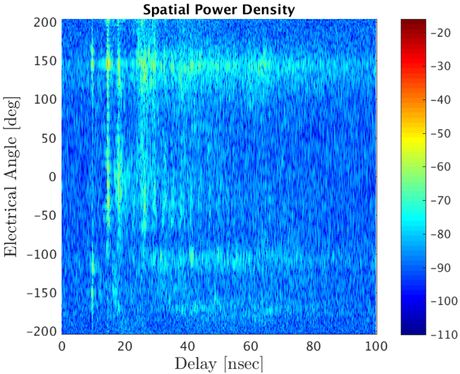


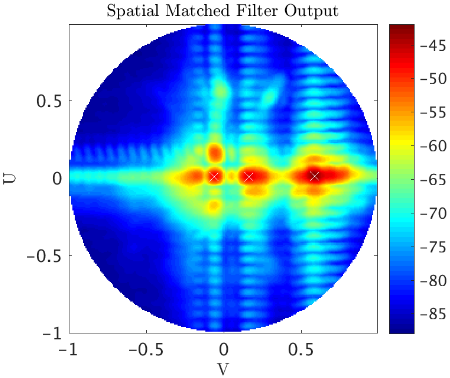
Output of 2-D FFT Across Array Elements for Fixed Delay Index

$$X(\frac{u}{\lambda}, \frac{v}{\lambda}) = \sum_n \sum_m s(k) e^{[\frac{-j2\pi}{\lambda}(md_x u + nd_y v)]}$$

Measured Results

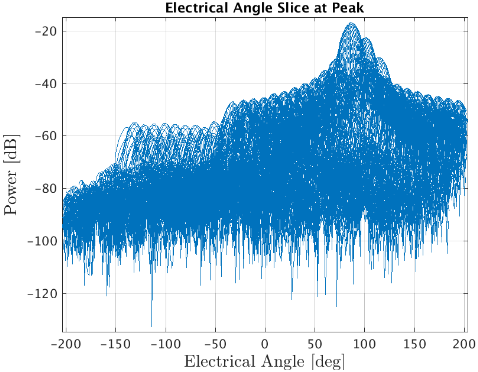
Multipath Environment in Lab

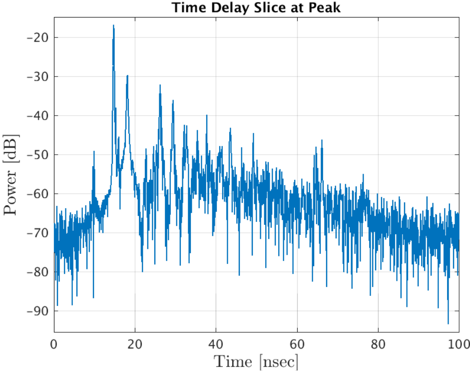




Measured data shows scattering from cabinet

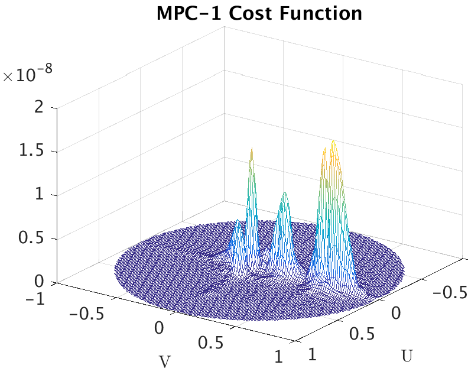
Angle and Delay Plots

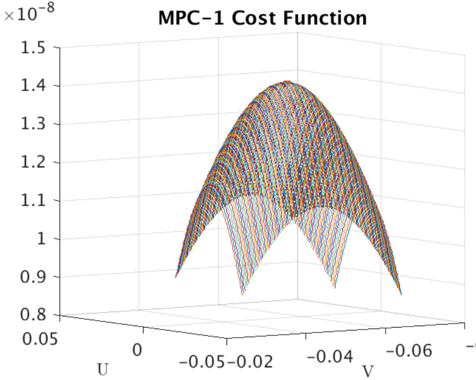




High sidelobes in angular domain due to FFT processing are visible

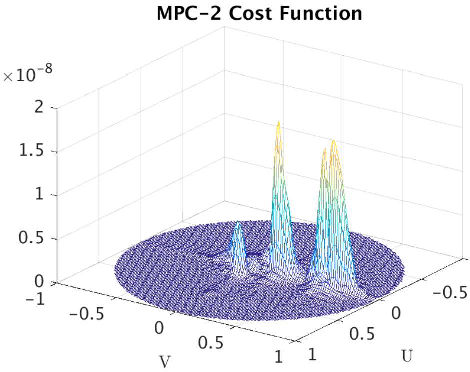
Cost Function: MPC-1

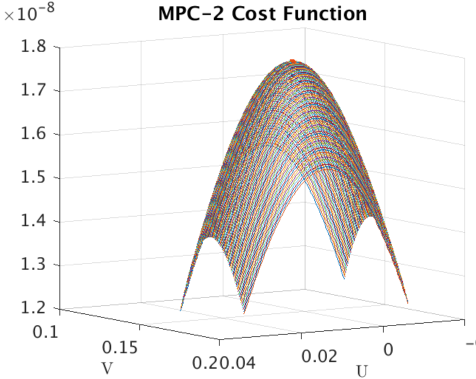




Data quality helps optimization program converge quickly

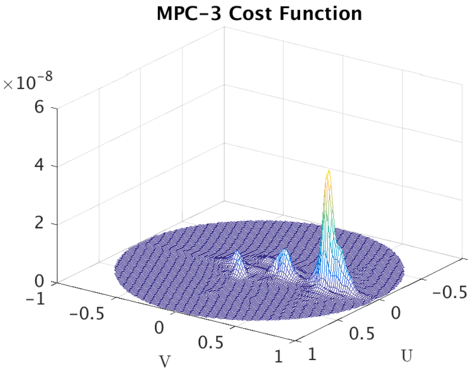
Cost Function: MPC-2

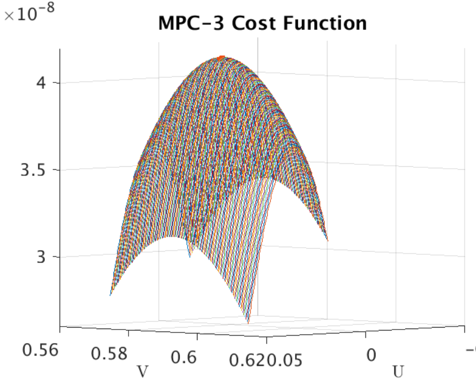




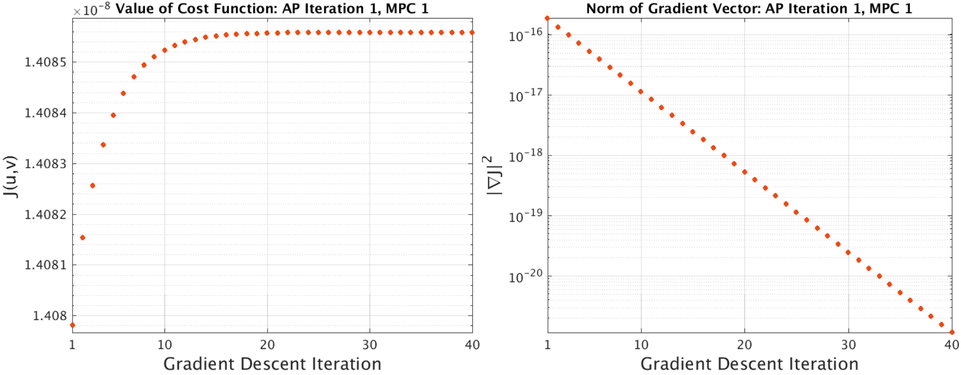
The global cost function has several peaks and valleys

Cost Function: MPC-3



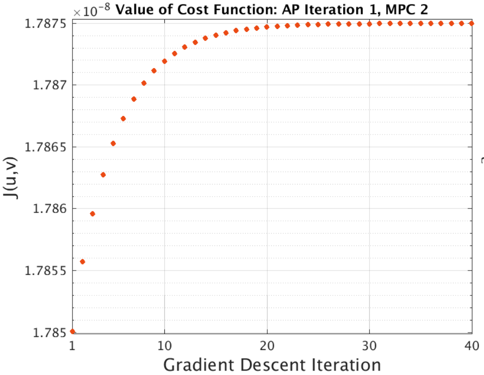


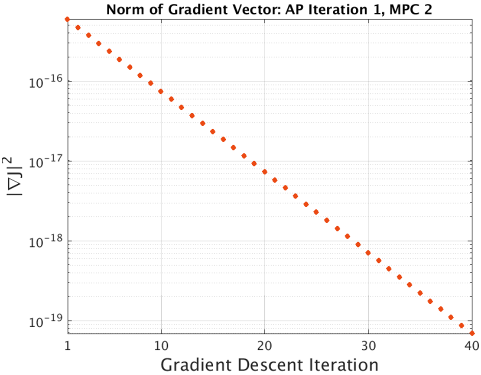
Convergence: MPC-1



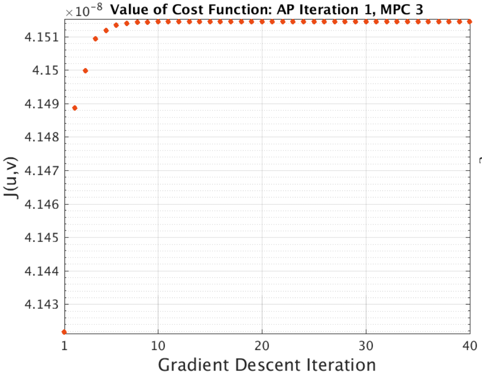
Algorithm converges in a few iterations

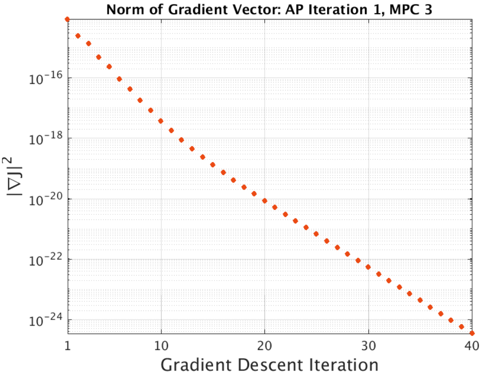
Convergence: MPC-2



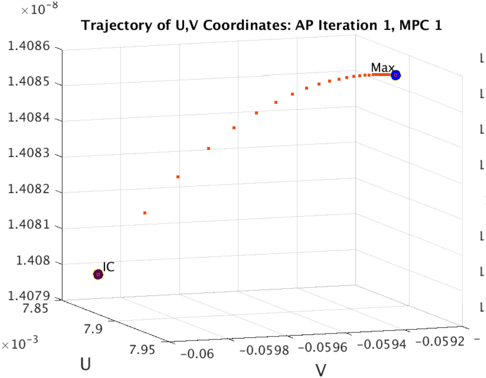


Convergence: MPC-3

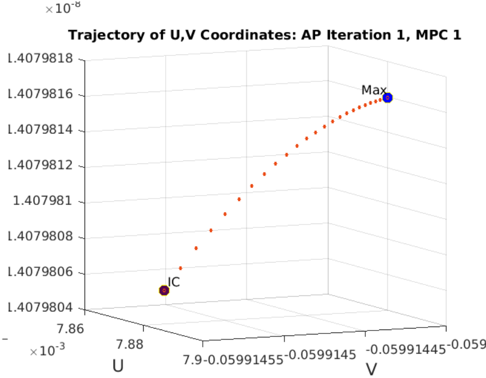




CG vs Newton: MPC-1

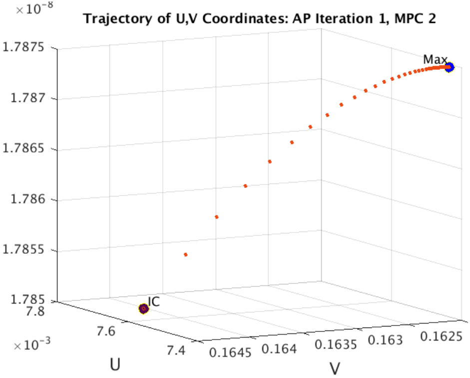


Conjugate Gradient

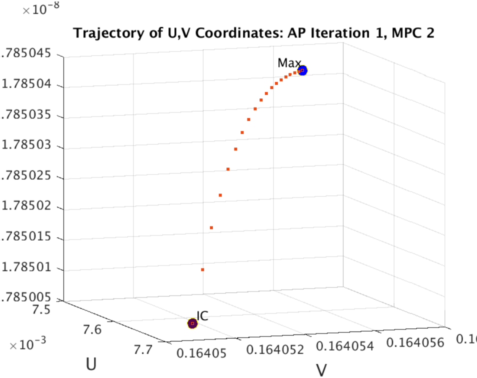


Newton's Method

CG vs Newton: MPC-2

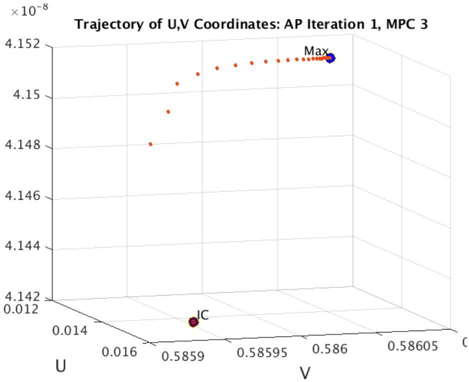


Conjugate Gradient

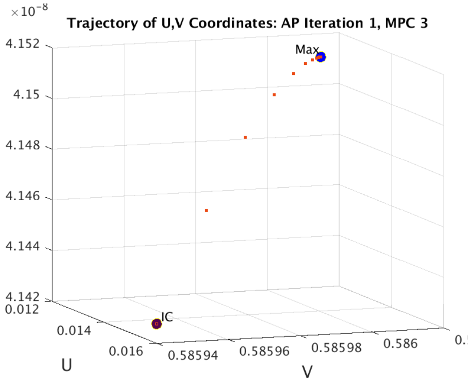


Newton's Method

CG vs Newton: MPC-3



Conjugate Gradient

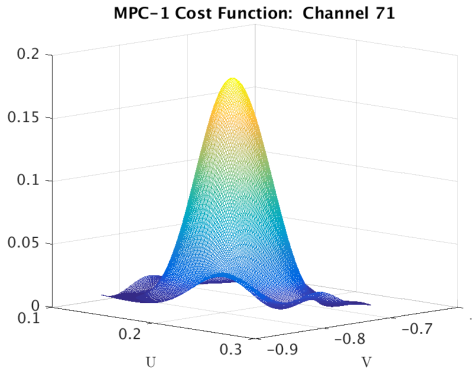


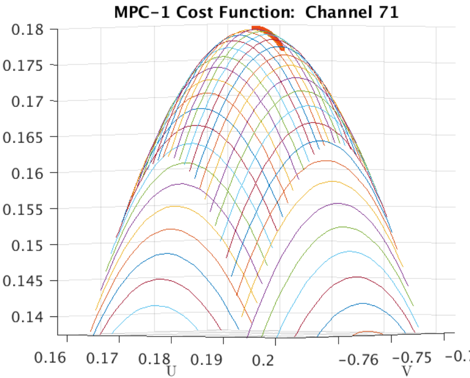
Newton's Method

Algorithms took slightly different paths to same solution

Effect of Step-Size

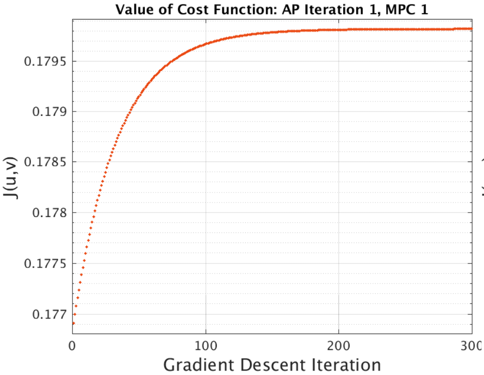
MPC-1 Cost Function



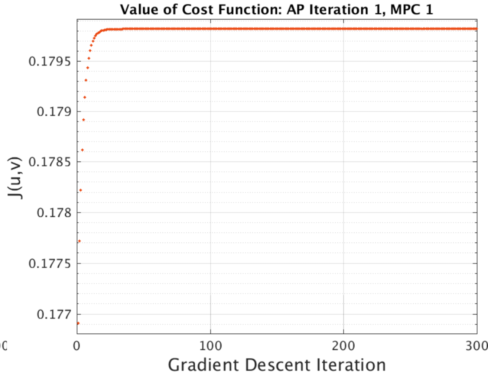


Cost function is concave in neighborhood of desired solution

Rate of Convergence



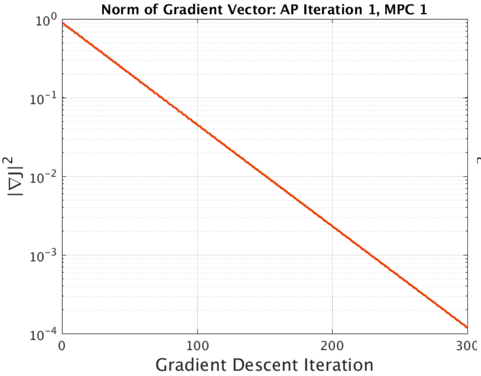
$\mu = 10^{-4}$



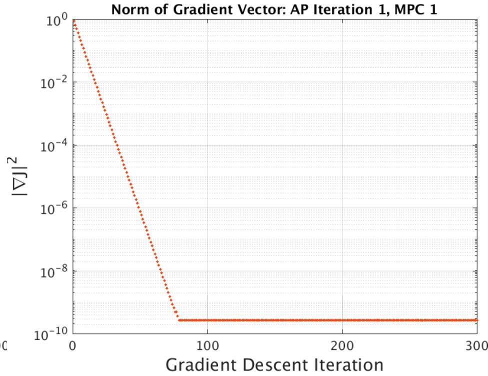
$\mu = 10^{-3}$

Algorithm reaches the peak sooner for $\mu = 10^{-3}$ vs $\mu = 10^{-4}$

Norm of Gradient Vector



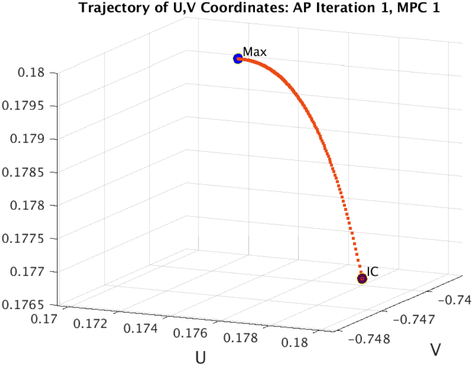
$\mu = 10^{-4}$



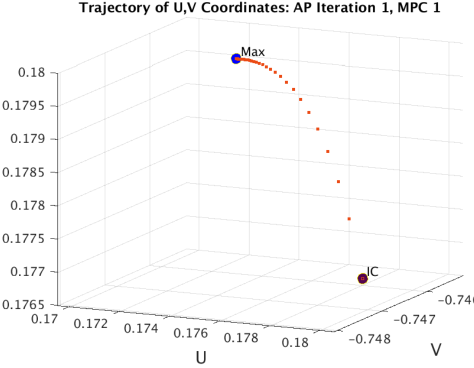
$\mu = 10^{-3}$

Algorithm convergences in less than 100 iterations for $\mu = 10^{-3}$

Convergence Trajectory



$\mu = 10^{-4}$



$\mu = 10^{-3}$

Larger step sizes converge faster but risk overshooting the peak

Summary

- A synthetic aperture is a powerful architecture for sounding stationary channels
- This paper derived an analytical expression for the gradient vector and Hessian matrix of the MLE cost function for AOA estimation
- The MLE AOA algorithm assumes a superposition of plane waves at the receive array. Without this assumption the phase at each array element must be modeled as range-dependent
- The MLE algorithm requires good initial conditions to converge
- Measured results using a synthetic aperture show the algorithm converges in tens of iterations

Bibliography I

[1] D. Munoz, F. Bouchereau, C. Vargas, R. Enriquez-Caldera, Position Location Techniques and Applications, by Elsevier Press, 2009.

[2] I. Ziskind and M. Wax, "Maximum Likelihood Estimation Via The Alternating Projection Maximization Algorithm," in *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing*, 1987.

[3] I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 10, October 1988.

[4] S. D. Blunt, T. Chan, and K. Gerlach, "Robust DOA Estimation: The Reiterative Superresolution (RISR) Algorithm," in *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 47, Issue 1, January 2011.