CMSC 660 Project Solutions
Optimization methods for Sound Source Localization using Microphone arrays

Vikas C. Raykar

December 19, 2002

Microphone arrays are widely employed for applications like teleconferencing, high quality sound capture, speaker recognition/identification, acoustic surveillance, head aid devices, speech acquisition in automobile environments etc. For all these applications the benefits that a microphone array provides over a single microphone are two fold. First using a microphone array we can localize a sound source and track its position accurately. The second benefit is that once the source location is known the microphone array can be electronically steered to the source providing spatial filtering. So the essential requirement for all these applications is the ability of the microphone array to locate a speech or sound source accurately. Broadly three types of methods exist for localizing a sound source: Focalization using a steered beamformer, High resolution spectral-estimation methods and Time Difference of Arrival (TDOA) based methods. Most commonly used method in practice is the TDOA based method and we will explore the optimization methods and other scientific computing techniques that can be applied to this method.

The TDOA method can be summarized as follows. In this assignment we assume that the sound source is a point source and the microphones we use have omni-directional receiving pattern(although realistic modelling can be done). So there is a delay between the sound wave received by a pair of microphones. So Sound source localization is a two step problem.

- First the signal received by several microphones is processed to obtain information about the time-delay between pairs of microphones. Various methods exist for estimating the time-delay which are based on the cross correlation between the signals.

- The estimated time-delays for pairs of microphones can be used for getting the location of the speaker.

In this project we won't go in detail regarding how the delay is estimated. Once the time delays are estimated the source localization problem can be formulated as follows:

Let there be M pairs of microphones. Let $\mathbf{m_i^1}$ and $\mathbf{m_i^2}$ for $i \in [1, M]$ be the vectors representing the spatial coordinates(x,y and z coordinates) of the two microphones in the i^{th} pair of microphones. Let the source be located at $\mathbf{s}$. The actual delay associated with a source at $\mathbf{s}$ and the i^{th} pair of microphones is given by,

$$T_i(\mathbf{s}) = \frac{|\mathbf{s} - \mathbf{m_i^1}| - |\mathbf{s} - \mathbf{m_i^2}|}{c} \tag{1}$$

where, c is the speed of propagation of sound in the acoustical medium(Use $c = 342m/s$). Let τ_i be the estimated time-delay. In practice, for a given microphone pair, the estimated delay τ_i and the actual delay $T_i(\mathbf{s})$ will never be equal because the estimated delay is corrupted by noise and also due to room reverberation. Given M pairs of sensors, their spatial coordinates and the estimated delays we can get an estimate $\hat{\mathbf{s}}$ of the source location.

Problem 1[5 points]

Consider one pair of microphones whose spatial coordinates are m_1 and m_2 and let τ be the time delay estimated for this pair of microphones. Using just one microphone pair is it possible to get a unique source location? Describe the region of ambiguity present?What is the minimum number of microphones required to get the source source coordinates?

Given m_1 and m_2 are the (x, y, z) coordinates of the two microphones and say let s be the location of the source. If τ the estimated delay is equal to the actual delay then

$$\tau = \frac{|s - m_1| - |s - m_2|}{c} \tag{2}$$

$$|s - m_1| - |s - m_2| = \tau c \tag{3}$$

where c is the speed of sound in air. Equation 3 represents one half(since we have to take into account the sign of τ) of a hyperboloid of two sheets with $\frac{m_1+m_2}{2}$ as the center with m_1 and m_2 being the two focal points and the line joining the two microphones as the axis of symmetry. Figure 1 shows one half of the hyperboloid of two sheets for a given microphone pair. Hence two microphones cannot uniquely determine the source location in 3D space.

The minimum number of microphones required to get the 3D location is 3. Since given 3 microphones there are 3 possible pairs of microphones. For each pair the source should lie on one half of a hyperboloid. 3 such hyperboloids intersect to specify a unique point in 3D space. Note that two pairs are not sufficient since they intersect to give a curve and not a unique point.

Problem 2[10 points]

In general we never have an perfect time delay estimation procedure. Explain how by using a large number of microphones larger than the minimum

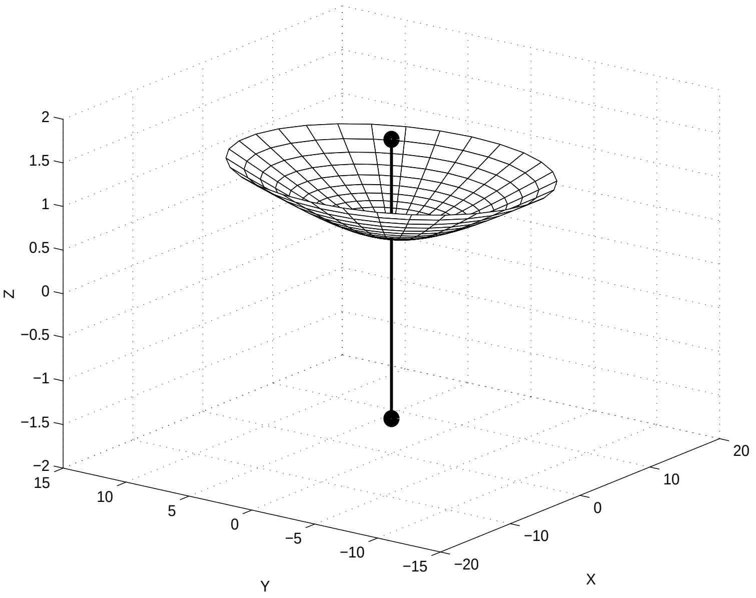

Figure 1: The region in space which corresponds to a given microphone pair for a given delay τ. This region corresponds one sheet of a hyperboloid of two sheets with $\frac{m_1+m_2}{2}$ as the center and the line joining the two microphones as the axis of symmetry.

required we can formulate the problem as one which minimizes an error criterion? Specifically us the fact that for a given microphone pair, the estimated delay τ_i and the actual delay $T_i(\mathbf{s})$ will never be equal and the true location would be one which minimizes this difference among all microphone pairs?

Let there be M pairs of microphones. Let $\mathbf{m_i^1}$ and $\mathbf{m_i^2}$ for $i \in [1, M]$ be the vectors representing the spatial coordinates(x,y and z coordinates) of the two microphones in the i^{th} pair of microphones. Let the source be located at $\mathbf{s}$. The actual delay associated with a source at $\mathbf{s}$ and the i^{th} pair of microphones is given by,

$$T_i(\mathbf{s}) = \frac{|\mathbf{s} - \mathbf{m_i^1}| - |\mathbf{s} - \mathbf{m_i^2}|}{c} \tag{4}$$

where, c is the speed of propagation of sound in the acoustical medium($c = 342m/s$ depends on the room temperature). Let τ_i be the estimated time-delay. In practice, for a given microphone pair, the estimated delay τ_i and the actual delay $T_i(\mathbf{s})$ will never be equal because the estimated delay is corrupted by noise and also due to room reverberation. Given M pairs of sensors, their spatial coordinates and the estimated delays the source location s is one which minimizes the error between the actual and the estimated time delay over all microphone pairs.

$$\hat{\mathbf{s}} = \arg_s(min(J(\mathbf{s}))) \tag{5}$$

where,

$$J(\mathbf{s}) = \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})]^2 \tag{6}$$

This does not have a closed-form solution since it is a non-linear function of s. We will have to use different optimization methods here.

The geometrical interpretation is as follows. Ideally if there were $M \geq 3$ mics then the intersection of all the hyperboloids corresponding to each microphone pair would be a unique point. Due to the errors in the estimation of time delays we do not have a unique intersection point. So by using the redundant mics we try to find the source location which best belongs to all the the hyperboloids in the least square sense.

Problem 3[10 points]

In the previous problem you formulated the problem as one which minimized a certain error criterion. The same can be formulated in a prbabilistic frame work. Assuming that the time delays estimated at each microphone pair are independently corrupted by zero mean white additive gaussian noise derive an Maximum likelihood estimator(ML) for the source location?(Remember that the ML estimator is found by maximizing the likelihood function)Show that the result is the same as the previous formulation?

Let τ_i the estimated time delay be corrupted by zero-mean additive white Gaussian noise with known variance $var(\tau_i)$.[This variance is usually the result of the particular time delay estimation method]. So τ_i is normally distributed with mean $T_i(\mathbf{s})$ and variance $var(\tau_i)$. $T_i(\mathbf{s})$ is the actual delay associated with a source at $\mathbf{s}$ and the i^{th} pair of microphones is given by Equation 1.

$$\tau_i \sim \mathrm{N}(T_i(\mathbf{s}), var(\tau_i)) \tag{7}$$

Assuming that each of the time delays are independently corrupted by zero-mean additive white Gaussian noise the likelihood function can be written as:

$$p(\tau_1, \tau_1, ..., \tau_M; s) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi var(\tau_i)}} exp[\frac{-(\tau_i - T_i(\mathbf{s})^2}{2var(\tau_i)}] \tag{8}$$

The log-likelihood ratio is:

$$ln(p(\tau_1, \tau_1, ..., \tau_M; s)) = -\sum_{i=1}^{M} ln(\sqrt{2\pi var(\tau_i)}) + [\frac{(\tau_i - T_i(\mathbf{s})^2}{2var(\tau_i)}] \tag{9}$$

The Maximum Likelihood(ML) location estimate, $\hat{\mathbf{s}}_{ML}$is the position which maximizes the log likelihood ratio or equivalenlty one which minimizes:

$$J_{ML}(\mathbf{s}) = \sum_{i=1}^{M} \frac{[\tau_i - T_i(\mathbf{s})]^2}{var(\tau_i} \tag{10}$$

This is same as the previous case except that the variance term comes into picture. Therefore

$$\hat{\mathbf{s}}_{ML} = \arg_s(min(J_{ML}(\mathbf{s}))) \tag{11}$$

Problem 4[5 points]
Try to get a feel for this function. Assume a certain room size, using say 16 microphones(four fixed on each wall) plot the function? Since it is a function of 3 variables fix one of them and plot the function.

The function $J_{ML}(\mathbf{s})$ was plotted for the room setup as shown in Figure 2. There where a total of 16 microphones. Four mics are made to lie at the corners of a square of side $50cm$. Each such square was fixed on the center of one of the four walls of the room. The coordinate system is also shown. For each square which contains 4 mics there are 6 possible pairs of microphones. In our case we used a total of 24 pairs of microphones with 6 corresponding to each square. Note that the total number of pairs possible is $(16 \times 15)/2 = 120$ of which we are using only 24 pairs.

Figure 3 shows the function for the source present at the center of the room. The function is shown as slices through the room. Each slice is parallel to the x-y plane. Also the function is plotted in the decibel scale to emphasize the minima. The minima now corresponds to the maxima which the darkest red part of plots. Note that at the center of the room the function has the maximum value. In this

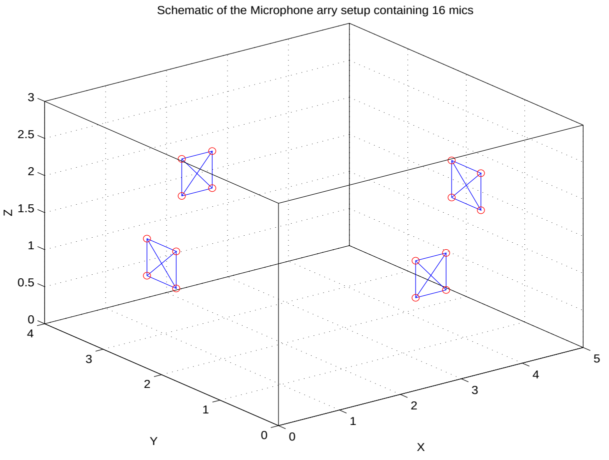

Figure 2: Schematic of the room setup and the microphone array

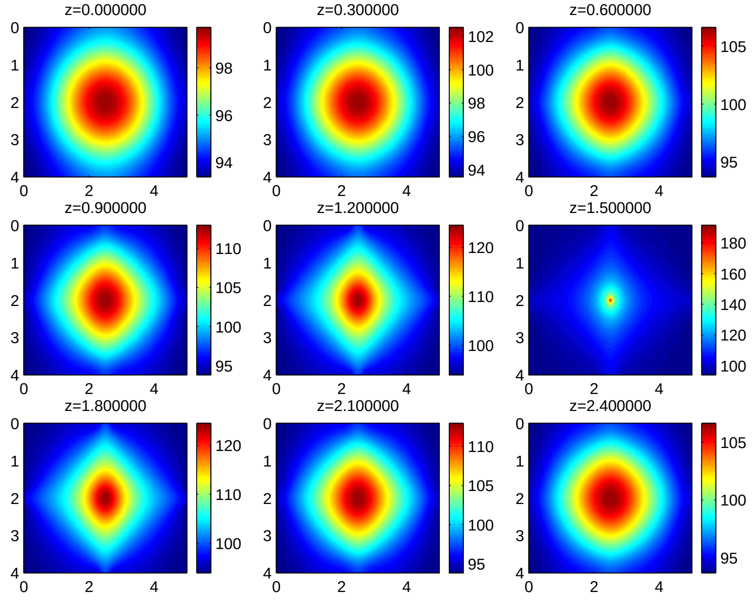

Figure 3: Error function for the source present at the center of the room. The function is shown as slices through the room. Each slice is parallel to the x-y plane.

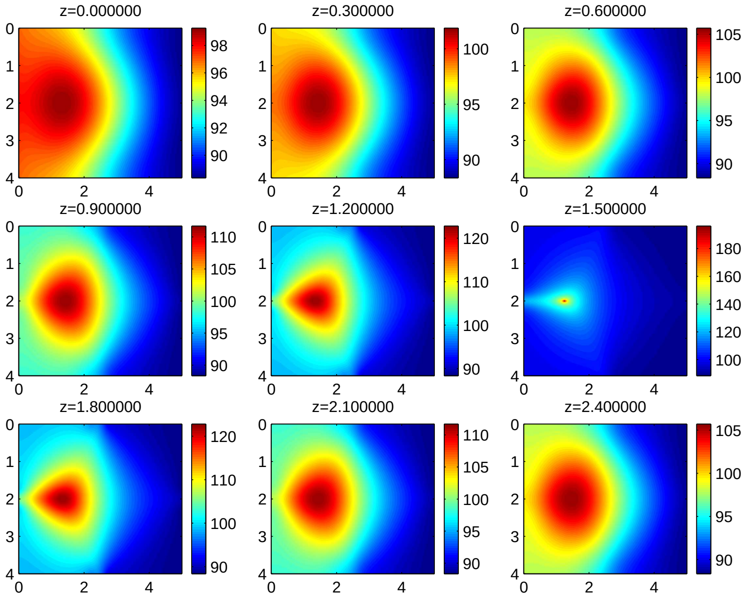

Figure 4: Error function for one more source position. The function is shown as slices through the room. Each slice is parallel to the x-y plane.

case the delays were perfect that is they were not corrupted by noise and hence we have not introduced the variance term in the error formulation. Figure 4 shows the same for one more position of the source.

Note that the function is very much dependent on the position of the microphones. For example Figure 6 shows the function for the source present at the center of the room for the microphone array configuration as in Figure 5.

Problem 5[20 points]
The function to be minimized is non-linear.Explore different methods to minimize this function. Compare them by using a fixed source location and generating the time-delays. Evaluate by using a large number of trials and report the error and the number of iteration required and also the number of function evaluations?

In this problem we evaluate different nonlinear optimization techniques to minimize the function. Once again the problem can be stated as:

$$\hat{\mathbf{s}} = \arg_s(min(J(\mathbf{s}))) \tag{12}$$

where,

$$J(\mathbf{s}) = \sum_{i=1}^{M}[\tau_i - T_i(\mathbf{s})]^2 \tag{13}$$

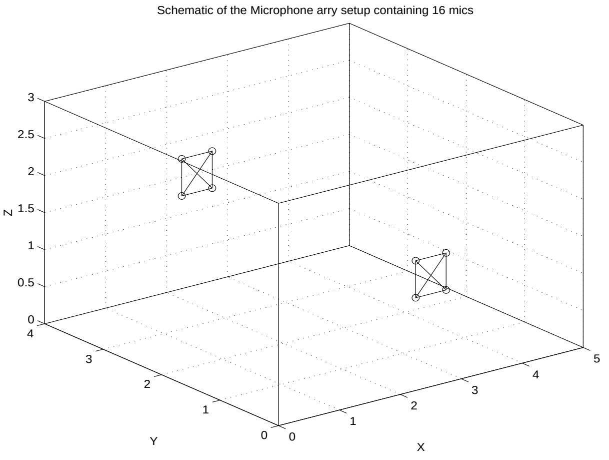

Figure 5: Schematic of the room setup and the microphone array

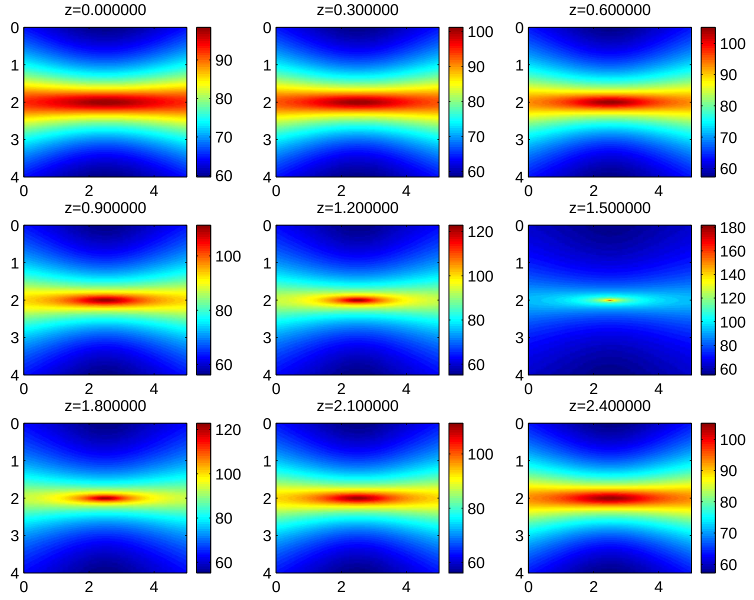

Figure 6: Error function for the source present at the center of the room. The function is shown as slices through the room. Each slice is parallel to the x-y plane.

We have not included the error variance since in this problem we will be evaluating different optimization methods for the noise free case. In the next problem we will consider the effect of noise on the TDOA estimates. We evaluate 4 different methods as described below. The simulations were done using MATLAB. Much of the details given here are taken from *Optimization Toolbox User's Guide*.

1.Nelder Mead Simplex Method This is a direct search method that does not use numerical or analytic gradients A simplex in n-dimensional space is characterized by the $n+1$ distinct vectors that are its vertices. In two-space, a simplex is a triangle. In three-space, it is a pyramid. At each step of the search, a new point in or near the current simplex is generated. The function value at the new point is compared with the function's values at the vertices of the simplex and, usually, one of the vertices is replaced by the new point, giving a new simplex. This step is repeated until the diameter of the simplex is less than the specified tolerance. Can handle discontinuities. Suitable for nonlinear problems with a large number of discontinuities. Does not require the evaluation of gradient or Hessian. But this method has a slow convergence
MATLAB function **fminsearch**

2.Quasi Newton Methods(BFGS,DFP) Newton-type methods use the Hessian H to get the descent direction. Calculating H numerically involves a large amount of computation. Quasi-Newton methods avoid this by using appropriate Hessian update schemes. The two commonly used update schemes are the BFGS(Broyden , Fletcher, Goldfarb and Shanno) method and the DFP(Davidon, Fletcher, and Powell) method. The BFGS method uses rank 2 update. DFP method avoids the inversion of the Hessian H, by using a formula that makes an approximation of the inverse Hessian at each update.For both we use the mixed quadratic and cubic polynomial line-search procedure.
MATLAB functions
fminunc with options.LargeScale set to 'off' uses the BFGS Quasi-Newton method with a mixed quadratic and cubic line search procedure.
fminunc with options.LargeScale set to 'off' and options. HessUpdate to 'dfp' uses the DFP Quasi-Newton method with a mixed quadratic and cubic line search procedure.

3.Nonlinear Least Square Methods(Gauss Newton,Levenberg Marquardt) The problem we have is a nonlinear least square problem. Although the function can be minimized using a general unconstrained minimization certain characteristics of the problem can often be exploited to improve the iterative efficiency of the solution procedure.The gradient and Hessian matrix of LS problem have a special structure. Levenberg Marquardt and Gauss Newton methods are the two methods widely used. These methods

have very high accuracy and also low iteration count.
MATLAB functions
lsqnonlin with options.LargeScale set to 'off' uses the Levenberg-Marquardt method
lsqnonlin with options.LargeScale set to 'off' and options.LevenbergMarquardt to 'off' uses the Gauss-Newton method.

Simulation Parameters

- The simulations were done for the room as shown in Figure 2. There where a total of 16 microphones. Four mics are made to lie at the corners of a square of side $50cm$. Each such square was fixed on the center of one of the four walls of the room. For each square which contains 4 mics there are 6 possible pairs of microphones. In our case we used a total of 24 pairs of microphones with 6 corresponding to each square.
- The results presented are averaged over 200 randomn trials were the actual source position was assigned randomnly to lie in the room.
- options = optimset('Display','off','TolFun',1e-12,'TolX',1e-12,'LargeScale','off').
- Initial guess was in the center of the room.
- All these methods are compared based on the localization error(The localization error is the Euclidean distance between the actual source position and as found by the optimization method), the number of iterations and the number of function evaluations required.
- For each of these the mean,median and the maximum value are noted.

The following table summarizes the results:

200 Trials	Localization Error (in m)			Iterations			Function Evaluations		
METHOD	Median	Mean	Max	Median	Mean	Max	Median	Mean	Max
Nelder Mead Simplex	3.95e-13	4.15e-13	7.59e-13	176	177	209	320	324	377
Quasi Newton (BFGS)	3.71e-03	1.50e-04	2.73e-01	15	16	59	113	120	305
Quasi Newton (DFP)	0.65	1.63	19.19	41	37	43	302	274	305
Gauss Newton	5.72e-08	8.89e-07	9.76e-06	5	5	6	33	31	41
Levenberg Marquardt	4.82e-06	1.44e-06	1.76e-05	7	7	9	44	44	63

Observations

- The first observation that can be made is that the Quasi Newton method with DFP update performs the worst. Hence we eliminate this method from our further discussion.
- The best method so far in terms of error is the Simplex method however it is the worst in terms of number of iterations and functional evaluations.
- Quasi Newton with BFGS update has higher error than Simplex while having significantly less number of iterations.
- However the best methods are the Gauss Newton and the Levenberg-Marquardt algorithms which give very low error and also very low iteration count.
- Among these two Gauss Newton method outperforms Levenberg-Marquardt.

Problem 6[10 points]

As mentioned earlier the the estimated delay τ_i and the actual delay $T_i(\mathbf{s})$ will never be equal because of noise and reverberation. For a given time delay τ assume and error of $\Delta\tau$ and explain the how the nature of the function to be *minimized changes as $\Delta\tau$ is increased?*

The function to be minimized is

$$J(\mathbf{s}) = \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})]^2 \tag{14}$$

The estimated delay τ_i and the actual delay $T_i(\mathbf{s})$ will never be equal because of noise and reverberation. In order to study the effect of the error in time delay estimation in the source location we perturb each of the time delays τ_i by a small factor $\Delta\tau$ and see how the nature of the function to be minimized changes.

$$\begin{aligned} J(\mathbf{s}) &= \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})]^2 \\ \hat{J}(\mathbf{s}) &= \sum_{i=1}^{M} [\tau_i + \Delta\tau - T_i(\mathbf{s})]^2 \\ \hat{J}(\mathbf{s}) &= \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})]^2 + \Delta\tau^2 + 2\Delta\tau[\tau_i - T_i(\mathbf{s})] \\ \hat{J}(\mathbf{s}) &= \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})]^2 + M\Delta\tau^2 + 2\Delta\tau \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})] \\ \hat{J}(\mathbf{s}) &= J(\mathbf{s}) + M\Delta\tau^2 + 2\Delta\tau \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})] \\ &\mathit{Ignoring\ the\ } M\Delta\tau^2 \mathit{term} \\ \hat{J}(\mathbf{s}) &\simeq J(\mathbf{s}) + 2\Delta\tau \sum_{i=1}^{M} [\tau_i - T_i(\mathbf{s})] \\ \hat{J}(\mathbf{s}) &\simeq J(\mathbf{s}) + 2\Delta\tau E(\mathbf{s}) \end{aligned} \tag{15}$$

$J(\mathbf{s})$ is the function which we want to minimize. But due to error in the estimation of time delays we are minimizing $\hat{J}(\mathbf{s})$ which is the sum of $J(\mathbf{s})$ and $E(\mathbf{s})$.As $\Delta\tau$ increases $E(\mathbf{s})$ dominates.

Figure 7 shows this effect for the source present at the center of the room. The function is shown as one slice parallel to the x-y plane at a height corresponding to the actual source location. Also the function is plotted in the decibel scale to emphasize the minima. The minima now corresponds to the maxima which the darkest red part of plots. Note that at the center of the room the function has the maximum value. Also shown are the functions by adding $\Delta\tau$ to the timedelays along with the corresponding $E(\mathbf{s})$. In each case the actual source location is marked. As can be seen that as $\Delta\tau$ increases the function maxima no longer corresponds to the actual source location.

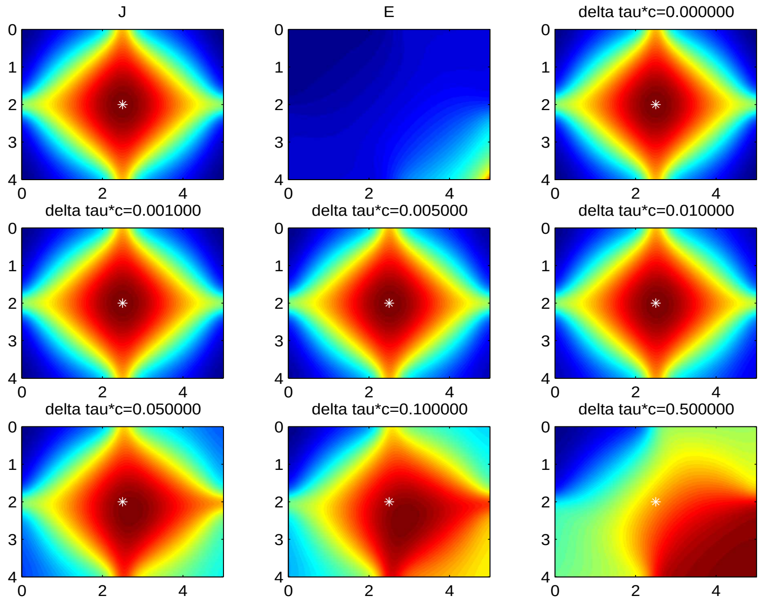

Figure 7: The first plot is $J(\mathbf{s})$,the second plot is $E(\mathbf{s})$ the rest of the plots are $\hat{J}(\mathbf{s})$ for different values of $\Delta\tau$

Problem 7[20 points]
Evaluate the different methods by varying adding a gaussian noise to the estimated time delays and compare different methods?

Figure 8 plots the mean,median and the maximum localization error,number of iterations and number of function evaluations for different optimization methods for varying noise conditions. The results are summed over 200 trials. The x axis in each represents $var(\tau)c$. Each of the time delay esimated was corrupted by adding zero mean white Gaussian noise with variance $var(\tau)$. Following four methods were compared:

- Nead Melder Simplex Method
- Quasi Newton with BFGS Hessian update and quadcubic line search.
- Gauss Newton .
- Levenberg Marquardt.

Observations

- The median localization error is same for all the methods.

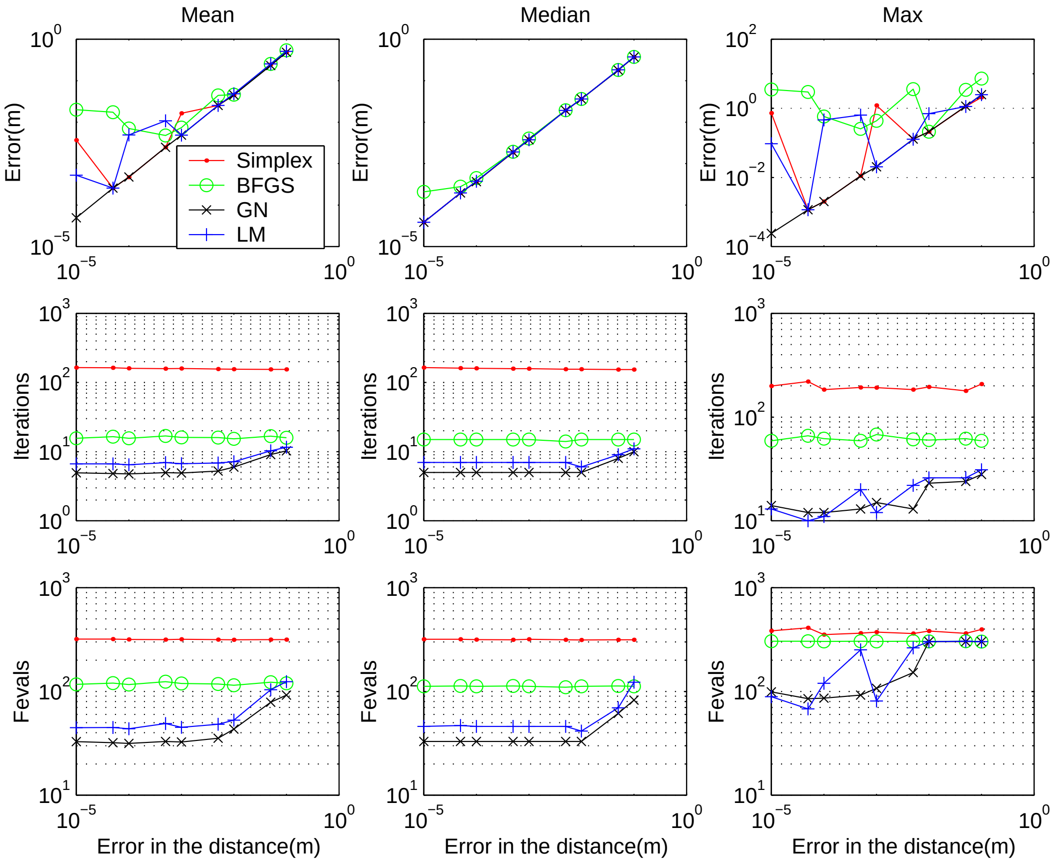

Figure 8: The plots show the mean,median and the maximum localization error,number of iterations and number of function evaluations for different optimization methods for varying noise conditions. The results are summed over 200 trials. The x axis in each represents $var(\tau)c$. Each of the time delay estimated was corrupted by adding zero mean white Gaussian noise with variance $var(\tau)$.

- With respect to the mean and max localization error Gauss Newton method has the least error followed by Levenberg Marquardt,Simplex with BFGS having the largest error.
- Also as the noise is increased the the error increases.
- With respect to the number of iterations and function evaluation the Simplex method requires the highest and the Gauss Newton requires the least.

Conclusion
Gauss Newton method is the best choice.

References

[1] Microphone Arrays Signal Processing Techniques and Applications Brandstein, M., Harvard University, Cambridge, MA, USA; Ward, D., Imperial College, London, UK (Eds.)

[2] Optimization toolbox Users guide