Introduction

The Problem:

Radio interferometry is an imaging technique used to study astronomical radio sources such as galaxies and supernovae. Radio interferometers like the **Very Large Array (VLA)**, one of the most prominent interferometers today, collect massive amounts of interferometric data from radio signals to reconstruct images, a process that comes at an unnecessary cost to time and data storage space.

The Inspiration:

Compressed sensing, a very new signal processing approach, proposes the efficient and accurate reconstruction of images with far fewer data. Traditional methods requires one **sample** of the signal per point source of radio emission, whereas compressed sensing allows image reconstruction with only a small fraction of those samples. (Mathematically, this appears to be impossible - it forces us to solve an underdetermined system of equations!) However, compressed sensing has never been applied to interferometry in practice. How can we ensure successful application of compressed sensing to radio interferometry?

The Research:

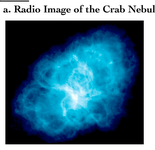
As in all realms of signal processing, **sample distribution**, or *where* samples of a radio signal are taken, is a key ingredient to image reconstruction success. What is the optimal sample distribution for the application of compressed sensing to radio interferometry? The relationship between the interferometric data and the object of reconstruction is provided by a system of linear equations, and compressed sensing performs best when the matrix of the system has a mathematical quality known as high **incoherence**. Through computer simulations, I studied how different sample distributions affect the matrix incoherence, and therefore the ability of a compressed sensing algorithm, **orthogonal matching pursuit**, to reconstruct images from simulated data. The results point to specific modifications to optimize the **Very Large Array (VLA) - one of the most prominent interferometers used today - and make it much more conducive to compressed sensing**. The results also indicate a general principle for optimizing any interferometric sample distribution, allowing other radio interferometers, and related instruments such as microwave interferometers, to make sampling processes much more efficient.

What is a Radio Image?

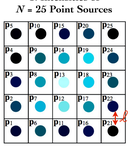
A radio image is a representation of the **intensity**, or optical power, of a radio signal. The total signal is composed of N individual radio waves, each emitted from a different point source $\mathbf{p}_l$ in the sky, and the intensity of one point source corresponds to one pixel in a radio image. The angular resolution is given by λ. For mathematical convenience, the total signal intensity, called the object $\mathbf{I}$, is vectorized as an N-vector.

Figure 1.

a. Radio Image of the Crab Nebula



b. Intensities of N = 25 Point Sources



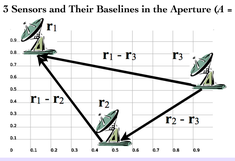
Radio Sensors and Baselines

A radio interferometer consists of n radio sensors, labeled $\mathbf{r}_j$ where j = 1, 2, ... n, located in an A x A square region called the **aperture** of size A. The displacement vectors between all possible pairs of sensors are called **baselines**, and are central to the sampling mechanism in radio interferometry (see "Radio Interferometry"). If there are n sensors, then there are $n(n - 1)/2$ possible combinations of two sensors, and therefore $n(n - 1)/2$ baselines. The VLA has 27 radio sensors, giving 351 baselines. Figure 2.a shows the VLA sensor distribution, and 2.b provides a simple example of the relationship between 3 sensors, the baselines, and the aperture.

Figure 2.

a. The VLA Radio Sensors

b. 3 Sensors and Their Baselines in the Aperture (A = 1)

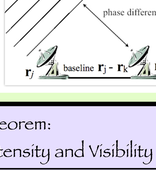


Radio Interferometry

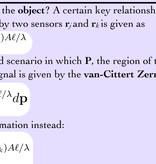
Radio interferometry uses the **interference patterns** of radio waves to reconstruct the intensity of the radio signal. Interference patterns result when **mutually coherent** waves, which have a stable phase relationship, are **superimposed**, or added together. Interference patterns are determined by the **phase difference** between the superimposed waves, or how much the waves are aligned (Figure 3.a).

Like all electromagnetic waves from natural sources, radio signals are mutually incoherent; the N waves have no stable phase relationship with each other. However, each wave is certainly coherent to itself. Therefore, when a wave from $\mathbf{p}_l$ is **sampled** with two sensors $\mathbf{r}_j$ and $\mathbf{r}_k$ at different locations, interferometers like the VLA can measure a phase difference (Figure 3.b). This phase difference is determined by the displacement vector between the two sensors, the baseline $\mathbf{r}_j - \mathbf{r}_k$, and is used to construct interference patterns. This interferometric data is called the **visibility**, and is essentially a result of superimposing a wave on itself. A system of equations, given by the van-Cittert Zernike Theorem, is then solved to reconstruct the signal intensity from the visibility.

Figure 3. **a. Phase Difference Between Two Waves**



b. Measuring Phase Difference for a Single Wave



The van-Cittert Zernike Theorem: A Mathematical Relationship between Intensity and Visibility

How does radio interferometry use the visibility data to reconstruct the intensity $\mathbf{I}$, the **object**? A certain key relationship between visibility and intensity is used. The visibility of one wave in the signal from $\mathbf{p}_l$ as sampled by two sensors $\mathbf{r}_j$ and $\mathbf{r}_k$ is given as

$$\mathbf{V}(\mathbf{r}_j - \mathbf{r}_k, \mathbf{p}_l) = \mathbf{I}(\mathbf{p}_l)e^{i2\pi/\lambda\,\mathbf{p}_l\cdot(\mathbf{r}_j-\mathbf{r}_k)A\ell/\lambda}$$

where $\mathbf{I}(\mathbf{p}_l)$ is the intensity of the point source $\mathbf{p}_l$, and $\mathbf{r}_j - \mathbf{r}_k$ is the baseline. In the idealized scenario in which $\mathbf{P}$, the region of the sky being imaged, is a continuous plane of infinitely many point sources, the visibility of the total signal is given by the **van-Cittert Zernike theorem:**

$$\mathbf{V}(\mathbf{r}_j - \mathbf{r}_k) = \int_{\mathbf{P}} \mathbf{I}(\mathbf{p})e^{i2\pi\,\mathbf{p}\cdot(\mathbf{r}_j-\mathbf{r}_k)A\ell/\lambda}d\mathbf{p}$$

As we are limited to imaging a finite number of point sources, we use the discretized summation instead:

$$\mathbf{V}(\mathbf{r}_j - \mathbf{r}_k) = \sum_{l=1}^{N} \mathbf{I}(\mathbf{p}_l)e^{i2\pi\,\mathbf{p}_l\cdot(\mathbf{r}_j-\mathbf{r}_k)A\ell/\lambda}$$

The visibility of the total signal is simply a sum of the visibilities of the N individual waves. Additionally, in order for images to be **well-resolved**, the criterion $A\ell/\lambda$ = 1 must be met. (This gives the VLA's current resolution capability). Such **diffraction-limited** interferometers have the ability to reconstruct images with resolutions as high as the theoretical resolution maximum.

Interferometers take one visibility sample of the signal with every baseline $\mathbf{r}_j - \mathbf{r}_k$. If there are n sensors, then there are $n(n - 1)/2$ baselines - for example, the VLA, with 27 sensors, has 351 baselines. The visibility $\mathbf{V}$ is then a vector with $n(n - 1)/2$ components. The goal of radio interferometry is to solve for the object $\mathbf{I}$, the total signal intensity, given $\mathbf{V}$, the collected visibility data.

The Linear Algebra Context

Figure 4.

$\mathbf{V}$ = $\mathbf{\Phi}\mathbf{I}$ with 3 Baselines and a Signal from 9 Point Sources

$$\begin{bmatrix}\text{baseline 1}\\ \text{baseline 2}\\ \text{baseline 3}\end{bmatrix} = \begin{bmatrix}\text{baseline 1}\\ \text{baseline 2}\\ \text{baseline 3}\end{bmatrix} \times \begin{bmatrix}\mathbf{p}_1\\ \mathbf{p}_2\\ \mathbf{p}_3\\ \mathbf{p}_4\\ \mathbf{p}_5\\ \mathbf{p}_6\\ \mathbf{p}_7\\ \mathbf{p}_8\\ \mathbf{p}_9\end{bmatrix}$$

(visibility measured by: = rows corresponding to: × intensity of:)

The discrete approximation of the van-Cittert Zernike theorem,

$$\mathbf{V}(\mathbf{r}_j - \mathbf{r}_k) = \sum_{l=1}^{N} \mathbf{I}(\mathbf{p}_l)e^{i2\pi\,\mathbf{p}_l\cdot(\mathbf{r}_j-\mathbf{r}_k)A\ell/\lambda}$$

provides us with a system of equations, corresponding to a linear algebra problem of the form $A\mathbf{x} = \mathbf{b}$. After the visibility data $\mathbf{V}$ has been collected, the goal is to solve for the object $\mathbf{I}$ from the system of equations $\mathbf{V} = \mathbf{\Phi}\mathbf{I}$, where $\mathbf{\Phi}$ is called the **sensing matrix** and its entry in the ith row and lth column, Φ_{il}, is defined as:

$$\Phi_{il} = e^{i2\pi\,\mathbf{p}_l\cdot(\mathbf{r}_j-\mathbf{r}_k)A\ell/\lambda} \quad \text{where } i = (j-1)n + k$$

where i = 1, 2, ... $n(n - 1)/2$ and l = 1, 2, ... N. This type of matrix is known as a **Fourier matrix**, and $\mathbf{V}$ is a **Fourier transform** of $\mathbf{I}$. Each ith row of $\mathbf{\Phi}$ corresponds to a different baseline, and each lth column corresponds to a different point source. When the ith row of $\mathbf{\Phi}$ is multiplied by $\mathbf{I}$, the product is the ith entry of $\mathbf{V}$, or the visibility of the signal measured by that specific baseline. In other words, any equation in the system corresponds to the visibility of the signal as measured by one specific baseline, as shown in Figure 4.

Doing it the Slow Way: Traditional Radio Image Reconstruction

A key problem arises in solving the system $\mathbf{V} = \mathbf{\Phi}\mathbf{I}$. It is an underdetermined system, meaning there are more variables than equations, because there are far more point sources than baselines. High quality images are often of millions of point sources, while the VLA only has 351 baselines. Underdetermined systems have infinitely many solutions, and the correct one cannot be found through the traditional method of using a **matrix inverse**. The VLA currently solves this problem by sampling the signal over an extended period of time, allowing the rotation of the earth to naturally create "new" baselines. This technique, known as **earth rotational aperture synthesis**, allows the VLA to eventually accumulate one sample per point source, and an adequately determined system

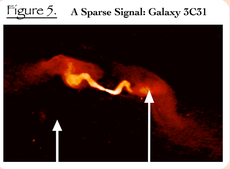
However, there are inherent physical constraints that prevent the VLA from completely sampling the Fourier space. Unknown components of $\mathbf{V}$ are set to 0, or interpolated from surrounding values, and can introduce significant error into the inverse reconstruction. Furthermore, the **deconvolution algorithm** CLEAN, which the VLA uses to eliminate this error, often requires that a technician have prior knowledge of the celestial region being imaged. Experience with where sources "should" be in the sky can be crucial to using CLEAN, and therefore deconvolution can require a considerable amount of manual, subjective work [4]. Furthermore, this method is tedious, as a colossal amount of interferometric data must be acquired - one sample must be taken for every pixel in a desired image.

Fast and Efficient: Radio Interferometry by Compressed Sensing

Can we skip the expensive, meticulous step of capturing the entire data set? Is it necessary to obtain one sample for each and every pixel of a desired image? Compressed sensing is a very new signal processing approach that got its research jumpstart in about 2006, and its answer is *No*. Compressed sensing operates off of the idea that we only need to capture the most significant coefficients of the signal to reconstruct it. One fundamental premise is that the signal is a **sparse** signal, one with mostly coefficients of zero and relatively few non-zero coefficients [3]. Most signals occurring in nature, including radio signals, are sparse. Figure 5 shows intuitively that only the relatively few point sources in the galaxy emit radio waves, while all the point sources in the "blank" background have an intensity of zero. One can toss out these zeros, and the compressed sensing idea is that sampling only the relatively few large coefficients of the signal gives enough information for a faithful reconstruction [3].

Again, we now face an underdetermined system; we have collected far fewer samples than the number of point sources. However, if the object has s non-zero components (an **s-sparse** signal), under certain conditions it has been theoretically proven that only slightly more than s samples, rather than all N samples, are needed [2] for an exact reconstruction.

Figure 5. **A Sparse Signal: Galaxy 3C31**



wave intensities of 0 (point sources in the black background)

non-zero wave intensities (point sources in the galaxy)

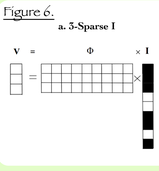
Orthogonal Matching Pursuit (OMP): An Intuitive Explanation

How is it possible to solve the underdetermined system? Why is signal sparsity central to compressed sensing? The key is that if the number of non-zero components, s, is small enough, we can create an adequately determined or overdetermined system, instead of the original underdetermined system. As a simple example, Figure 6.b shows that if $\mathbf{I}$ is 3-sparse, $\mathbf{V}$ is a linear combination of only the 3 color-coded columns of $\mathbf{\Phi}$, rather than a linear combination all N columns of $\mathbf{\Phi}$.

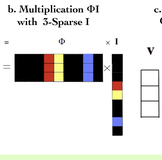
The compressed sensing algorithm that I used, called **orthogonal matching pursuit** (**OMP**), is an **iterative algorithm** that first locates, one by one, where the s non-zero components of the object are. OMP then extracts the columns of $\mathbf{\Phi}$ that correspond to the locations of the non-zero components, in order to create an adequately determined system (Figure 6.c). (If s is small enough, this step could also result in an overdetermined system.) This system can be solved with matrix inversion or pseudo-inversion, giving the non-zero components of the object. These are placed back into their original locations in a zero vector (Figure 6.d), giving the reconstruction.

Figure 6.

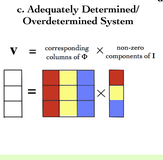
a. 3-Sparse I



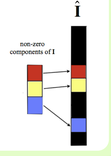
b. Multiplication ΦI with 3-Sparse I



c. Adequately Determined/ Overdetermined System



d. Final Reconstruction



OMP: A Rigorous Step-by-Step Explanation

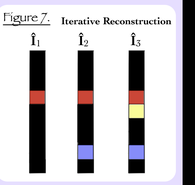
The Set-Up

OMP reconstructs $\mathbf{I}$ by finding the **sparsest solution** $\hat{\mathbf{I}}$ from the set of infinitely many solutions to $\mathbf{V} = \mathbf{\Phi}\mathbf{I}$. Numerous empirical and theoretical studies have demonstrated that the sparsest solution is, with overwhelming probability, the original object [2]. A prerequisite to reconstruction success is that s is less than or equal to the the number of samples, or the number of rows in $\mathbf{\Phi}$.

Finding the solution set to $\mathbf{V} = \mathbf{\Phi}\mathbf{I}$ is synonymous to finding solutions $\hat{\mathbf{I}}$ such that the **l$_2$-norm** of the residual $\mathbf{r} = \mathbf{\Phi}\hat{\mathbf{I}} - \mathbf{V}$, denoted $\|\mathbf{r}\|_2$, is minimized (ideally, 0). The l$_2$-norm, or Euclidean length, of a vector $x = [x_1, x_2, \ldots x_n]$ is defined as $\sqrt{(x_1^2 + x_2^2 + \ldots + x_n^2)}$. When $\|\mathbf{r}\|_2 = \|\mathbf{\Phi}\hat{\mathbf{I}} - \mathbf{V}\|_2 = 0$, then $\mathbf{V} = \mathbf{\Phi}\hat{\mathbf{I}}$ and $\hat{\mathbf{I}}$ is part of the solution set to $\mathbf{V} = \mathbf{\Phi}\mathbf{I}$.

The square root expression of the l$_2$-norm can be clumsy to manipulate algebraically, so OMP approaches the problem by minimizing the square of the l$_2$-norm of the residual, or $\|\mathbf{r}\|_2^2$. OMP is an **iterative algorithm**, meaning it starts off with $\hat{\mathbf{I}}_0$ as a zero vector and repeats (iterates) its procedure, each time producing a more accurate $\hat{\mathbf{I}}_k$ by finding the value and location of one more non-zero component of the $\mathbf{I}$ (Figure 7).

Figure 7. **Iterative Reconstruction**



Step 1: Finding the Location of $\Delta\hat{\jmath}_j$

At the k+1 iteration, OMP tries to minimize $\|\mathbf{r}_{k+1}\|_2^2$, which is equivalent, after algebraic manipulation, to

$$\|\mathbf{r}_k\|_2^2 + \mathbf{\Phi}_j^T\mathbf{\Phi}_j(\Delta\hat{\jmath}_j)^2 + 2\mathbf{\Phi}_j^T\mathbf{r}_k(\Delta\hat{\jmath}_j).$$

where $\Delta\hat{\jmath}_j$ is the one new non-zero component of $\mathbf{I}$ to be found in this iteration. By making the last two terms as negative as possible, we can aim to "cancel out" $\|\mathbf{r}_k\|_2^2$, the residual left over from the previous iteration, and thus minimize $\|\mathbf{r}_{k+1}\|_2^2$. To do this, OMP essentially finds the minimum of the quadratic $\mathbf{\Phi}_j^T\mathbf{\Phi}_j(\Delta\hat{\jmath}_j)^2 + 2\mathbf{\Phi}_j^T\mathbf{r}_k(\Delta\hat{\jmath}_j)$. The more negative the minimum is, the closer $\|\mathbf{r}_{k+1}\|_2^2$ will be to 0, so OMP first searches for the j that creates the quadratic with the most negative minimum possible. This index j will be location of $\Delta\hat{\jmath}_j$ (OMP must find both the location and value of this one new non-zero component). Solving for the minimum of this quadratic gives $-(\mathbf{\Phi}_j^T\mathbf{r}_k)^2/\|\mathbf{\Phi}_j\|_2^2$, where $\|\mathbf{\Phi}_j\|_2^2$ is a scalar constant for all j because the columns of a Fourier matrix have the same l$_2$-norm. To make this expression as negative as possible, OMP's first step is then to calculate

$$\hat{\jmath}_{k+1} = \arg\max_{1\le j\le N} \left|\mathbf{\Phi}_j^T\mathbf{r}_k\right|$$

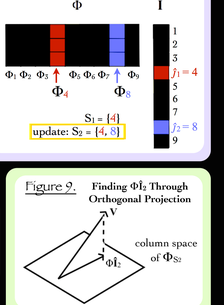
This finds the index $\hat{\jmath}_{k+1}$ of the column $\mathbf{\Phi}_{\hat{\jmath}_{k+1}}$ that maximizes the absolute value of $\mathbf{\Phi}_j^T\mathbf{r}_k$, the term that determines the magnitude of the minimum. OMP has now found the location of $\Delta\hat{\mathbf{I}}_k$, the one new non-zero component of $\mathbf{I}$.

Step 2: Updating the Set S_k

OMP now adds $\hat{\jmath}_{k+1}$ to $S_k = \{j_1, j_2, \ldots j_k\}$, the set of the indices of all the non-zero components of $\mathbf{I}$ found by Step 1 in the previous 1st through k^{th} iterations. The updated S_{k+1} now has the elements $\{j_1, j_2, \ldots j_k, j_{k+1}\}$. At the end of the reconstruction, this set S will ideally have s components, one locating each non-zero component of $\mathbf{I}$. For example, in Figure 8, this 2nd iteration would yield $S_2 = \{4, 8\}$, giving the locations of the red component (4), and the blue component (8).

The locations of the k+1 non-zero components correspond to the locations of k+1 columns of $\mathbf{\Phi}$. $\Phi\hat{\mathbf{I}}_{k+1}$ is a linear combination of k+1 columns of $\mathbf{\Phi}$ - precisely the k+1 columns whose indices are in S_{k+1}. $\Phi\hat{\mathbf{I}}_{k+1}$ lies in the **column space** of $\Phi_{S_{k+1}}$, or the vector space defined by all possible linear combinations of the columns of $\Phi_{S_{k+1}}$.

Figure 8. **Updating S_2**



Step 3: Finding $\mathbf{\Phi}\hat{\mathbf{I}}_{k+1}$

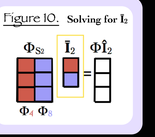
OMP now performs an **orthogonal projection** of $\mathbf{V}$ onto the column space of $\Phi_{S_{k+1}}$ (Figure 9), a transformation of $\mathbf{V}$ that is a standard method of finding $\mathbf{\Phi}\hat{\mathbf{I}}_{k+1}$ such that a) $\mathbf{\Phi}\hat{\mathbf{I}}_{k+1}$ is in the column space of $\Phi_{S_{k+1}}$ and b) $\|\mathbf{r}_{k+1}\|_2$ is minimized.

Figure 9. **Finding $\Phi\hat{\mathbf{I}}_2$ Through Orthogonal Projection**

Step 4: Solving for the Non-Zero Components of $\mathbf{I}$

The k+1 non-zero components of $\hat{\mathbf{I}}_{k+1}$ can then be found through matrix inversion or pseudo-inversion of the system $\Phi_{S_{k+1}}\hat{\mathbf{I}}_{k+1} = \mathbf{\Phi}\hat{\mathbf{I}}_{k+1}$, (Figure 10). We can use inversion as this is an adequately determined or overdetermined system, instead of the original underdetermined system, because $\mathbf{I}$ is sufficiently sparse. A prerequisite to reconstruction success is that s is less than or equal to the number of samples. Therefore, the number of indices in S_{k+1} will always be less than or equal to the number of samples; $\mathbf{\Phi}\hat{\mathbf{I}}_{k+1}$ is a linear combination of at most only s columns of $\mathbf{\Phi}$. The number of equations in the system $\Phi_{S_{k+1}}\hat{\mathbf{I}}_{k+1} = \mathbf{\Phi}\hat{\mathbf{I}}_{k+1}$ is now greater than or equal to the number of variables (non-zero components).

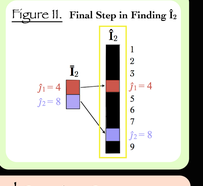
Figure 10. **Solving for $\hat{\mathbf{I}}_2$**



Step 5: "Filling in the Zeros"

The k+1 non-zero components found in Step 4 are then placed into a zero N-vector, precisely at the entries given by the components' indices $j_1, j_2 \ldots j_{k+1}$ (Figure 11). This produces the reconstruction $\hat{\mathbf{I}}_{k+1}$. OMP then starts the k+2th iteration and repeats the whole process, finding more non-zero components and updating $\hat{\mathbf{I}}$ until the residual $\|\mathbf{r}\|_2$ falls below some stopping criterion. In my experiments, I set this criterion as 1% of $\|\mathbf{I}\|_2$.

Figure 11. **Final Step in Finding $\hat{\mathbf{I}}_2$**



Sensing Matrix Incoherence: A Key to Compressed Sensing Success

Compressed sensing algorithms such as OMP perform best when the sensing matrix is highly **incoherent** [3], meaning its column vectors are nearly orthogonal to each other (the angle between them is close to π/2 radians). Because each column vector corresponds to the visibility of one point source, high incoherence is synonymous to having highly unique measurements of the signal.

As a measure of coherence, the **mutual coherence** [3] of a sensing matrix is defined as

$$\mu(\mathbf{\Phi}) = \max_{1\le i\ne i'\le N} \frac{|\mathbf{\Phi}_i \cdot \mathbf{\Phi}_{i'}|}{\|\mathbf{\Phi}_i\|_2\|\mathbf{\Phi}_{i'}\|_2}$$

where $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_{i'}$ denote the i^{th} and $i^{'th}$ columns of $\mathbf{\Phi}$. Considering the expression for the cosine of the angle between two vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\cos\theta_{\mathbf{x},\mathbf{y}} = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}$$

mutual coherence simply gives the maximum cosine value of the angles between all possible pairs of column vectors (excluding whenever $\mathbf{\Phi}_i = \mathbf{\Phi}_{i'}$, as a vector is always perfectly coherent to itself). We want the mutual coherence to be as low as possible. The closer it is to 0, the closer the angle between the most coherent column vectors is to π/2 radians, because cos(π/2) = 0.

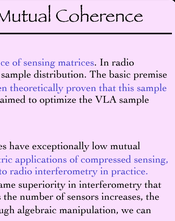
Though the ideal sensing matrix would be perfectly incoherent, in which case all of its columns would be orthogonal to each other, in compressed sensing this is inherently impossible because an underdetermined system is used. There are far more columns than the **dimension** of the columns, or the number of rows, and a basic theorem of linear algebra states that, for a vector space of dimension d, there can only be a maximum of d orthogonal vectors. However, the columns can still be nearly orthogonal when the mutual coherence is very low. A central theorem to my research [1] provides that if the sparsity s of the object satisfies

$$s \le \frac{1 + \frac{1}{\mu}}{2}$$

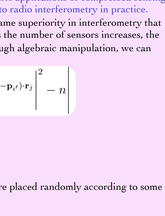
then OMP will reconstruct the object exactly (in the absence of **noise**, or measurement error in the data). The smaller μ is, the greater the sparsity s of the object that OMP is guaranteed to exactly reconstruct. Objects of greater sparsity carry more information and complexity, because there are more non-zero point sources (Figure 12); we want to be able to reconstruct as sparse images as possible. Based on this theorem, in my experiments I used mutual coherence as a theoretical indicator of how successful OMP reconstructions would be.

Figure 12.

a. Less Sparsity: Galaxy 3C75



b. Greater Sparsity: Supernova W30



The Central Connection: Sample Distribution and Mutual Coherence

Why does Sample Distribution Matter?

Sample distribution, or *where* samples of a signal are taken, directly affects the mutual coherence of sensing matrices. In radio interferometry, we take one sample of the signal at each baseline, so the **baseline distribution** is the sample distribution. The basic premise of my research: if I can design a sample distribution that gives very low mutual coherence, it has been theoretically proven that this sample distribution will also provide very high probability of OMP reconstruction success [1]. I ultimately aimed to optimize the VLA sample distribution - can I modify it to make it more conducive to compressed sensing?

Previous Studies of Sample Distribution

[3] summarizes the findings of many papers and states that randomly sampled sensing matrices have exceptionally low mutual coherence. However, previous studies of sample distribution have been focused on non-interferometric applications of compressed sensing, where the mechanisms of sampling are fairly different. Compressed sensing has never been applied to radio interferometry in practice.

One question addressed in my research was whether random sample distributions show the same superiority in interferometry that they have in non-interferometric applications. It can be proven in the interferometric context that as the number of sensors increases, the mutual coherence given by a uniform random sample distribution will keep approaching zero. Through algebraic manipulation, we can derive the identity for mutual coherence

$$\mu(\mathbf{\Phi}) = \max_{i\le i'\ne i'\le N} \frac{|\mathbf{\Phi}_i \cdot \mathbf{\Phi}_{i'}|}{\|\mathbf{\Phi}_i\|_2\|\mathbf{\Phi}_{i'}\|_2} = \max_{1\le i'\ne i'\le N} \frac{1}{n(n-1)} \left|\left|\sum_{j=1}^{n} e^{i2\pi(\mathbf{p}_i-\mathbf{p}_{i'})\cdot\mathbf{r}_j}\right|^2 - n\right|$$

Because we aim to minimize mutual coherence, minimizing

$$\max_{i\ne i'} \left|\sum_{j=1}^{n} e^{i2\pi(\mathbf{p}_i-\mathbf{p}_{i'})\cdot\mathbf{r}_j}\right|$$

appears to be a good standard for designing an optimal sample distribution. If the radio sensors $\mathbf{r}_j$ are placed randomly according to some distribution $f(\mathbf{r})$ on $[0,1]^2$ (the aperture of size 1), then

$$\frac{1}{n}\sum_{j=1}^{n} e^{i2\pi(\mathbf{p}_i-\mathbf{p}_{i'})\cdot\mathbf{r}_j}$$

is the **Monte-Carlo approximation** (a type of Riemann sum) of the integral

$$\int_{[0,1]^2} e^{i2\pi(\mathbf{p}_i-\mathbf{p}_{i'})\cdot\mathbf{r}} f(\mathbf{r})d\mathbf{r}$$

This integral equals 0 (indicating a minimization of the approximation, and therefore the mutual coherence) when the probability density function f = 1, meaning a random distribution that is uniform. The Monte-Carlo expression becomes a more and more accurate approximation of the integral - that is, it gets closer and closer to 0 - as the number of sensors n increases. Though this does not prove that the uniform random sample distribution is the best one, it does mathematically prove that the mutual coherence provided by this distribution will continue to approach 0. In contrast, there is **absolutely no** mathematical theory predicting how the VLA sample distribution will affect mutual coherence in the compressed sensing context.

Image Credits

Title Image: Image courtesy of National Radio Astronomy Observatory/Associated Universities, Inc., A. C. Boley and L. van Zee (Indiana University), D. Schade and S. Côté (Herzberg Institute for Astrophysics), J. Hibbard, and J.M. Uson.

Figures 1.b, 2.b, 3.b, 4, 6-11, 13-23, all "Sensor Distributions" and "Sample Distributions": Images created by Finalist.

Figure 1.a: Image courtesy of NRAO/AUI and M. Bietenholz.

Figures 2.a, 5, 24: Image courtesy of NRAO/AUI.

Figure 3.a: Image taken from CambridgeScienceFestival.org.

Figures 12.a, 19: Image courtesy of NRAO/AUI and F.N. Owen, C.P. O'Dea, M. Inoue, and J. Eilek.

Figure 12.b: Image courtesy of NRAO/AUI and Brogan et al.

References

[1] D. Donoho, M. Elad and V. Temlyakov. "Stable recovery of sparse overcomplete representations in the presence of noise." *IEEE Trans. Infor. Theory.* 52.1 (2006): 6-18.

[2] Donoho, David. "Compressed sensing." *IEEE Trans. Inform. Theory.* 52.4 (2006): 1289 - 1306.

[3] E. J. Candès and M. Wakin. "An introduction to compressive sampling." *IEEE Signal Proc. Mag.* 21.30 (2008).

[4] Myers, S.T. "Image reconstruction in radio interferometry." *National Radio Astronomy Observatory.* Jan. 2008. Web. Jan. 2011.

[5] "VLA Green Book." *National Radio Astronomy Observatory.* 1992. Web. Nov. 2010.

[6] *The Location of the GMRT Antennas. The Three Arms of the 'Y' Shaped Array Are Each about 14 Km Long.* Digital image. *National Radio Astronomy Observatory.* Web. Apr. 2011.

[7] Kritzer, Peter. "An Exact Formula for the ell-2 Discrepancy of the Shifted Hammersley Point Set". *Uniform Distribution Theory.* 1.1 (2006): 1-13.